

The Health Benefits, Resource Use and Cost-Effectiveness of Current and Future Cervical Cancer Screening Policies in Norway

Kine Pedersen



Thesis submitted for the Degree of Philosophiae Doctor (PhD)

Department of Health Management and Health Economics
Institute of Health and Society
Faculty of Medicine
University of Oslo
Norway
2017

© **Kine Pedersen, 2018**

*Series of dissertations submitted to the
Faculty of Medicine, University of Oslo*

ISBN 978-82-8377-174-9

All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

Cover: Hanne Baadsgaard Utigard.
Print production: Reprintsentralen, University of Oslo.

CONTENTS

Acknowledgements 5

List of papers7

List of abbreviations 9

Summary..... 11

1 INTRODUCTION 15

2 BACKGROUND 17

 2.1 HPV and cervical cancer 17

 2.1.1 Human papillomavirus and related diseases..... 17

 2.1.2 HPV and cervical carcinogenesis 18

 2.2 Cervical cancer prevention 20

 2.2.1 Cervical cancer screening..... 20

 2.2.2 HPV vaccination 30

3 THEORETICAL FRAMEWORK..... 32

 3.1 Priority setting and economic evaluation in healthcare 32

 3.1.1 Priority setting in healthcare..... 32

 3.1.2 Economic evaluation 33

 3.1.3 Cost-efficiency and cost-effectiveness 38

 3.1.4 Measuring and valuing health outcomes..... 40

 3.1.5 Costs 42

 3.2 Decision-analytic modeling 44

 3.2.1 Types of decision-analytic models 45

 3.2.2 Components of decision-analytic modeling..... 50

4 THESIS OBJECTIVES 55

5 MATERIALS AND METHODS 56

5.1	Analytic overview	56
5.2	Decision-analytic models.....	60
5.2.1	Decision-tree model for novel biomarkers (<i>Paper I</i>)	60
5.2.2	Microsimulation model of cervical carcinogenesis (<i>Papers II-IV</i>)	62
5.3	Health-related quality of life	67
5.4	Costing.....	68
5.5	Comparator screening strategies.....	70
5.6	Assumptions and analyses	71
5.6.1	Paper I.....	71
5.6.2	Papers II-IV	72
6	SUMMARY OF RESULTS.....	74
6.1	Paper I.....	74
6.2	Paper II.....	75
6.3	Paper III	77
6.4	Paper IV	78
7	DISCUSSION	81
7.1	Discussion of results	81
7.2	Methodological considerations	85
7.2.1	Model input parameters and assumptions.....	85
7.2.2	Analytic approach	89
7.3	Policy implications.....	91
7.4	Future research.....	93
8	CONCLUSIONS.....	96
9	REFERENCES.....	97
10	PAPERS I-IV	113

Acknowledgements

I am grateful to all who made this thesis possible and for all the experiences it brought with it – both personally and academically.

First, I would like to thank my supervisor Emily A. Burger, PhD (University of Oslo and Harvard University) and co-supervisors Ivar Sønbo Kristiansen, MD, PhD, MPH (University of Oslo), Eline Aas, PhD (University of Oslo) and Henrik Støvring, PhD (University of Aarhus). Emily, your enthusiasm for research is contagious and I could not have asked for a better mentor. Thank you for guiding and encouraging me through this thesis, and for generously spending time to share your expertise in HPV and cervical cancer as well as simulation modeling and economic evaluation. A special thank you for making it possible for me to come as a visiting scholar to the Center for Health Decision Science (CHDS) at the Harvard T.H. Chan School of Public Health in Boston (for almost seven months in total), and for your efforts in making my Boston visits such a great experience! Ivar, thank you for introducing me to economic evaluation, to HPV and cervical cancer, and to research in general, already five years ago. Your guidance through my master thesis and the opportunity to present the work at the Society for Medical Decision Making conference in Singapore in January 2014 (my very first presentation at an international congress), inspired me to pursue this research fellowship. Our discussions are always interesting and motivational; thank you for taking me on as one of your last PhD students! Finally, thank you, Eline and Henrik, for offering your methodological expertise and for many interesting discussions. I am privileged to have had the four of you as my supervisors, thank you!

In addition to my supervisors, I would like to thank my co-authors Jane J. Kim, PhD (Harvard University), Stephen Sy, MPH (Harvard University), Mari Nygård, MD, PhD (Cancer Registry of Norway), and Sveinung W. Sørbye, MD, PhD (University Hospital of North Norway). Thank you, Jane, for inviting me to the CHDS and for the opportunity to work with you and your wonderful team. My time at the center has been invaluable and I am grateful that you always found time to discuss my projects. To Stephen, thank you for helping me to navigate both the Harvard model and the streets of Boston – it was great fun! To the rest of the HPV team and the people at

CHDS, thank you for sharing your knowledge and always making me feel welcome. Thank you, Mari, for welcoming me to the Cancer Registry of Norway (and your journal club) and offering your expertise in HPV and cervical cancer. Lastly, thank you, Sveinung, for enthusiastically sharing your knowledge of HPV and cervical cancer and providing helpful information along the way.

I would also like to thank the Research Council of Norway for funding this project and research visits to Boston.

I owe many thanks to the University of Oslo and my colleagues at the Department of Health Management and Health Economics, where I spent most of my time working on this thesis. A special thanks to Hege, Camilla and Gudrun for many great time-outs (either running or at the coffee shop), to Camilla, for reading through this thesis, and to my office-buddy Ge Ge, for fabulous (and patient!) company all day long. To all my colleagues at HELED, thank you for all the engaging discussions, coffee breaks and fun times – it would not have been the same without you!

Thank you to my family and friends who have encouraged and supported me throughout this fellowship. A special thanks to my parents, Inga-Lise and Steinar, and my brother Stian, who have inspired me to have a passion for learning and encouraged me to pursue my interests and aspirations.

Finally, to Espen, for your incredible encouragement, support and patience, and for enthusiastically supporting my Boston visits and worldwide travels, even if it meant spending so much time apart – thank you!

List of papers

- I. **Pedersen K**, Sørbye SW, Kristiansen IS, Burger EA. Using novel biomarkers to triage young adult women with minor cervical lesions: A cost-effectiveness analysis. *BJOG: An International Journal of Obstetrics and Gynaecology*. 2017 Feb; 124(3):474-484. doi: 10.1111/1471-0528.14135. Epub 2016 Jun 6.
- II. **Pedersen K**, Burger EA, Sy S, Kristiansen IS, Kim JJ. Cost-effective management of women with minor cervical lesions: Revisiting the application of HPV DNA testing. *Gynecologic Oncology*. 2016 Nov; 143(2):326-333. doi: 10.1016/j.ygyno.2016.08.231. Epub 2016 Aug 17.
- III. Burger EA*, **Pedersen K***, Sy S, Kristiansen IS, Kim JJ. Choosing wisely: A model-based analysis evaluating the trade-offs in cancer benefit and diagnostic referrals among alternative HPV testing strategies in Norway. *British Journal of Cancer*. 2017 Aug 3. doi: 10.1038/bjc.2017.248. Epub ahead of print.
* Denotes equal contribution.
- IV. **Pedersen K**, Burger EA, Nygård M, Kristiansen IS, Kim JJ. Adapting cervical cancer screening for women vaccinated against human papillomavirus infections: The value of stratifying guidelines. Submitted.

List of abbreviations

ASC-H	Atypical squamous cells, cannot exclude high-grade lesion
ASC-US	Atypical cells of undetermined significance
CBA	Cost-benefit analysis
CC	Cervical cancer
CIN	Cervical intraepithelial neoplasia
CEA	Cost-effectiveness analysis
EQ-5D	EuroQol-5 Dimensions
HPV	Human papillomavirus
HRQoL	Health-related quality of life
HSIL	High-grade squamous intraepithelial neoplasia
ICER	Incremental cost-effectiveness ratio
INMB	Incremental net monetary benefit
LEEP	Loop electrosurgical excision procedure
LYG	Life years gained
NICE	National Institute for Health and Care Excellence
NOK	Norwegian Kroner
PSA	Probabilistic sensitivity analysis
QALY	Quality-adjusted life year

Summary

In Norway, cervical cancer (CC) remains the third most common cancer among mid-adult women, despite substantial reductions in CC incidence since the widespread introduction of cytology-based CC screening. The Norwegian CC Screening Program currently involves triennial cytology-based screening for women aged 25-69 years, yet novel CC prevention technologies such as biomarkers for CC screening and prophylactic HPV vaccination are changing the landscape of CC prevention. For example, novel biomarkers (e.g., HPV DNA or mRNA testing, identifying the most carcinogenic HPV types -16/-18 using genotyping, p16/Ki67 dual staining) may help improve the effectiveness and efficiency of the current cytology-based screening program by revisiting management guidelines for women with minor cervical cytological lesions. These women are at an elevated risk of progressing to more severe lesions within the next screening round thus active surveillance (often referred to as triage testing) is suggested for appropriate clinical management, yet there is lack of consensus in optimal management guidelines. Furthermore, primary HPV DNA testing starting at age 34 years (with five-year intervals) is under consideration in Norway; however, implementation is challenged by concerns for capacity constraints (e.g., number of gynecologists to perform colposcopies) and uncertainty around health benefit and resource use trade-offs. Lastly, the first cohort of Norwegian girls vaccinated against human papillomavirus (HPV) infections at age 12 years (in 2009) will become eligible for CC screening in 2022. With increased heterogeneity of CC risk in the population, stratifying CC screening guidelines according to HPV vaccination status may help maintain high-value prevention approaches.

The general aim of this thesis is to inform decision-makers about the health benefits, resource use and cost-effectiveness of current and future CC screening policies in Norway. Specifically, the four papers in this thesis evaluate: (i) the short-term health and economic outcomes of using novel biomarkers to triage younger (i.e., aged 25-33 years) unvaccinated women with minor cervical cytological lesions, (ii) the most cost-effective strategy to triage unvaccinated women with minor cervical cytological lesions, (iii) the trade-offs in health benefits and resource use associated with adopting primary HPV DNA testing strategies for unvaccinated women, and (iv) the most cost-effective

CC screening strategies for women vaccinated against HPV-infections in adolescence. By quantifying the health benefits and resource use associated with candidate screening strategies, this thesis has the potential to aid decision-makers in their continuous, complex work in refining CC prevention policies in Norway.

The analyses performed in *Papers I-IV* utilized a decision-analytic approach to evaluate the health and economic consequences associated with candidate screening strategies for vaccinated and unvaccinated women. *Paper I* used a decision-tree model to estimate the short-term consequences of candidate strategies for younger adult women with minor cervical cytological lesions, while *Papers II-IV* employed a microsimulation state-transition model of HPV and cervical carcinogenesis to quantify the health and economic outcomes associated with candidate screening strategies for a hypothetical cohort of individual women over their lifetime. All papers quantified the health and economic outcomes associated with candidate screening strategies, such as the number of precancers detected (*Paper I*), the quality-adjusted life expectancy (*Papers II and IV*), CC incidence (*Papers II-IV*), the total average cost per woman and the number of colposcopy referrals (*Papers I-IV*).

A cost-effectiveness framework was used to identify cost-efficient and cost-effective strategies in *Papers I, II and IV*, while *Paper III* focused on the trade-offs in health benefits and resource use associated with candidate strategies. In *Papers I, II and IV*, cost-efficient algorithms were identified using the incremental cost-effectiveness ratio (ICER). In *Paper I*, the ICER was defined as the additional cost per additional precancer detected, and the average cost per detected precancer associated with the current Norwegian strategy was used as a proxy for the willingness-to-pay threshold (to detect one additional precancer). In *Papers II and IV*, the ICER was defined as the additional cost per additional quality-adjusted life year (QALY) gained, and a commonly cited Norwegian willingness-to-pay threshold of a \$100,000 per QALY was used as a benchmark to identify the most cost-effective strategy.

Paper I indicates that, in the short-term, the efficiency and effectiveness (in terms of precancer detection) of the current cytology-based screening program can be improved using reflex HPV mRNA testing to triage women with minor cervical

cytological lesions. *Paper II* suggests that the long-term efficiency and effectiveness (in terms of CC incidence) of the current triage algorithm for women with minor cervical cytological lesions can be improved using reflex HPV DNA testing with direct colposcopy referral for women positive for HPV-16/-18 infections. *Paper III* indicate that in order to maximize the CC preventive benefits of the future primary HPV-based screening program, while controlling colposcopy referral rates, HPV-based screening should start at an earlier age and rather utilize a less intensive triage algorithm for HPV-positive/cytology-negative women. Finally, in order for screening to remain cost-effective for women who received the HPV vaccine in adolescence, *Paper IV* suggests that a de-intensified HPV-based screening strategy (e.g., screening once or twice over a lifetime) is required.

In conclusion, this thesis highlights opportunities to improve the effectiveness and efficiency of current and future CC screening policies for HPV-vaccinated and unvaccinated women. However, model-based consequence- and cost-effectiveness analyses can only inform one aspect of the decision-making process, and the optimal screening strategy depends on multiple factors such as available resources and the preferences of both decision-makers and individual women for the trade-off between health benefits (e.g., reduced risk of developing CC) versus potential costs and harms associated with participating in CC screening.

1 INTRODUCTION

In Norway, cervical cancer (CC) remains the third most common cancer among women aged 25-49 years, who are in their productive social and working years [1]. This is despite substantial reductions in CC incidence following more than two decades of organized cytology-based screening [2]. A persistent infection with high-risk human papillomavirus (HPV) is the necessary cause of CC [3-5]; this discovery led to the development of groundbreaking CC prevention technologies that are changing the landscape of CC prevention, including biomarkers (e.g., HPV tests) for CC screening and prophylactic HPV vaccination. Worldwide, decision-makers are considering the application of these technologies to improve the effectiveness and efficiency of CC prevention policies. In addition, the introduction of HPV immunization programs is expected to reduce the risk of developing CC among vaccinated individuals, prompting decision-makers to consider the impact of HPV vaccination on optimal CC screening approaches. Within a healthcare sector with pressing demands, there are critical challenges to designing screening policies that continue to reduce the burden of CC, while providing efficient use of resources and ‘good value for money’. Importantly, screening programs seek to maximize benefits and minimize the harms of screening; the aim is to prevent CC from developing by detecting and treating precancers (before they have an opportunity to progress to cancer), while simultaneously ensuring that screening algorithms are efficient and feasible in both the short- and long-term, and limit the burden to women.

No single empirical study can capture all the health and economic consequences of alternative interventions (e.g., screening strategies), which is required to inform decisions about whether and how to adopt emerging technologies in clinical practice. Another approach is to use decision-analytic modeling, which involves synthesizing best available evidence from multiple sources of data (e.g., clinical trials, population-based registries, meta-analyses) and explicitly comparing alternative strategies while scrutinizing uncertainty. These models can project the health benefits and resource use associated with candidate interventions, which can inform economic evaluation and cost-effectiveness analyses, and in turn, complex priority setting questions.

Using decision-analytic modeling and an economic evaluation framework, the general aim of this thesis is to inform policy makers about the health benefits, resource use and cost-effectiveness of current and future CC screening policies in Norway. In particular, this thesis addresses knowledge gaps related to CC screening for women who are not vaccinated against HPV infections and who are: (i) detected with minor cervical lesions within the current cytology-based screening program (i.e., triage using candidate biomarkers), and (ii) offered primary HPV testing (i.e., informing the impact of screening algorithm 'levers' on health benefits and resource use). Lastly, this thesis addresses the knowledge gap related to whether and how CC screening should be carried out for women who were vaccinated against HPV infections in adolescence. Addressing these knowledge gaps has the potential to aid decision-makers in choosing between candidate screening approaches when refining and designing current and future prevention policies.

This thesis is structured as follows: **Chapter 2** provides a general background to HPV and CC epidemiology, CC prevention strategies (i.e., screening and HPV vaccination) and current CC prevention policies in Norway. The theoretical framework is presented in **Chapter 3**, including priority setting and economic evaluation in healthcare, as well as the types and components of decision-analytic modeling. The thesis objectives are presented in **Chapter 4**, followed by an overview of the materials and methods in **Chapter 5**, including a description of the analytic framework employed in each paper, the decision-analytic models, the model inputs, the comparator screening strategies, and the analyses and assumptions. **Chapter 6** summarizes the results of *Papers I-IV*, followed by **Chapter 7**, which includes a discussion of the results, methodological considerations, policy implications and areas of future research. The thesis conclusions are summarized in **Chapter 8** and references are listed in **Chapter 9**. Finally, the full-text manuscripts and accompanying supplements for *Papers I-IV* are provided in **Chapter 10**.

2 BACKGROUND

2.1 HPV and cervical cancer

2.1.1 Human papillomavirus and related diseases

HPV is a common sexually transmitted infection which may cause genital warts, recurrent respiratory papillomatosis, and occasionally, cancer [5]. More than 200 HPV genotypes ('types') have been identified; however, only about 12 types are considered oncogenic (i.e., high-risk types) [5-7]. A persistent infection with high-risk HPV is the cause of virtually all CCs [3-5], as well as a proportion of other genital cancers (vaginal, vulvar, anal, penile) and oropharyngeal cancers. CC is predominantly caused by HPV types -16 and -18, which attributes to ~60% and ~15% of all CCs, respectively [8]. An additional ~15% of all CCs are cumulatively attributed to HPV-31, -33, -45, -52, and -58 infections. Infection with HPV-6 and -11 low-risk types is associated with most genital warts and recurrent respiratory papillomatosis [7].

A study using data from the GLOBOCAN 2012 database estimated that 4.5% of all incident cancer cases worldwide are attributable to HPV [9]. The study reported a considerable variation in the HPV-attributable fraction (for all cancers among both men and women) across geographical regions, which was generally higher in less developed countries than more developed countries, and ranged from 1.3% in Australia and New Zealand to 15.8% in Sub-Saharan Africa. In addition, the global HPV-attributable burden was higher for women than men (i.e., 8.6% vs. 0.8%, respectively), and CC constituted the majority (i.e., 83%) of all HPV-attributable cancers.

In Norway, an average of 619 cancers occurred annually in organs affected by HPV infections during 2010-2014; of these, CC contributed to nearly half the cases (i.e., 288 cases) [10]. Furthermore, in Norway, CC is the third most common cancer in young adult women (aged 25-49 years), while it is the 13th most common cancer among women of all ages [1]. The cumulative risk of developing CC by age 75 years is 0.9%. The overall 5-year relative survival of CC in Norway is 81%, but ranges from 25% to 93% depending on cancer stage at diagnosis [1].

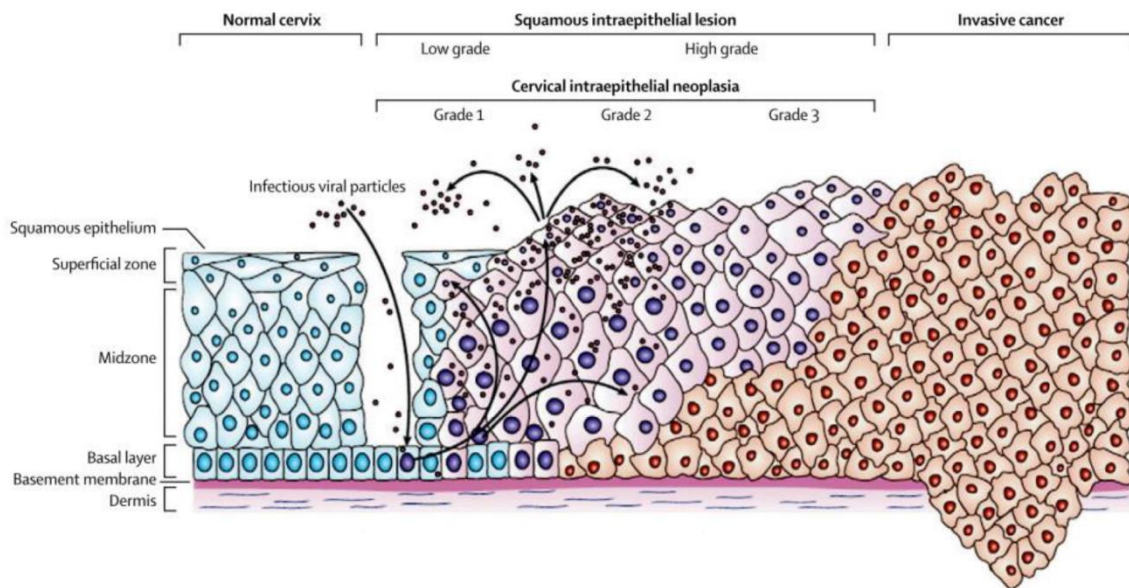


Figure 1. Stages of cervical carcinogenesis. A persistent infection with HPV in the cervical epithelium can cause dysplasia or precancer (histologically classified as cervical intraepithelial neoplasia grades 1, 2, or 3, of which grades 2 and 3 are often denoted as precancer), and invasive CC. Adapted from Crosbie et al. [11].

2.1.2 HPV and cervical carcinogenesis

The natural history of HPV and cervical carcinogenesis can be characterized as a stepwise process, involving: (i) acquisition of HPV, (ii) persistence (rather than clearance) of the infection, (iii) progression to precancer (i.e., a histology result of cervical intraepithelial neoplasia (CIN) grade 2 or 3), and (iv) invasive cancer (**Figure 1 and 2**) [6]. HPV is usually asymptomatic and is highly transmissible via skin-to-skin or skin-to-mucosa contact, thus most sexually active individuals acquire an HPV infection over their lifetime. The prevalence is highest in younger individuals and thereafter decreases by age [6]. There are several different HPV types, and individuals may acquire multiple new infections simultaneously or consecutively; concurrent infections are considered to be independent of one another [6, 12]. Results from a systematic review and meta-analysis of 86 studies (which together included more than 100,000 women) suggested that about half of HPV infections clear within 6-12 months after acquiring an infection [13], and ~90% clear within 2 years [14]. Moreover, an HPV infection may provide type-specific natural immunity against subsequent cervical HPV infections [15]. There are also studies suggesting that some HPV infections may become latent (inactive) or undetectable [6].

A persistent HPV infection may progress to cervical precancer and invasive cancer; a process that usually takes 10-20 years. The potential for persistence and progression of an HPV infection is influenced by HPV genotype, an individual's immune response, and behavioral cofactors (**Figure 2**) [6]. For example, the 3-year cumulative risk of developing cervical precancer given presence of an infection with HPV-16 or -18 is nearly 10-fold higher compared to the other high-risk HPV types [16]. Furthermore, individuals with a weakened immune system due to an HIV-infection have been reported to be at an increased risk of developing cervical precancer and cancer [17, 18]. Behavioral cofactors that may impact the acquisition, persistence and progression of an HPV infection include smoking [19], multi-parity [20] and long-term use of hormonal contraceptives [21]. Due to ethical reasons, there is a scarcity of studies evaluating the progression from cervical precancer to CC; however, in an unethical natural history study from New Zealand, women detected with precancer (i.e., CIN3) during 1965 to 1974 did not receive treatment, of which approximately one-third developed CC within 30 years [22]. Because of limited available empirical data, the progression potential of cervical precancer remains uncertain.

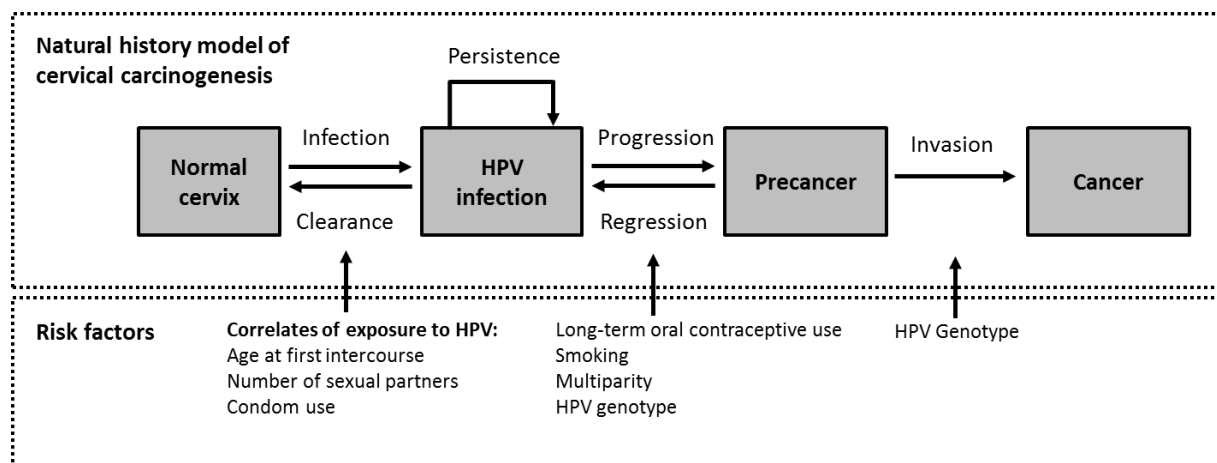


Figure 2. Example conceptual model of the natural history of cervical carcinogenesis, including correlates of HPV exposure and risk factors for progression to precancer and cancer. Adapted from Schiffman and Wentzensen [23].

Finally, invasive CC (International Classification of Diseases and Related Health Problems 10th Revision (ICD-10) code C53) can be classified by stage at diagnosis (e.g., local, regional and distant stages) and histological sub-type. The two main histological

sub-types of CC are squamous cell carcinoma and adenocarcinoma. Squamous cell carcinoma is most the common (accounting for 70-80% of all CCs) and is the type that is most readily prevented by screening [2]. Although a persistent infection with HPV has been established as the necessary cause of CC, recent evidence suggest that a small proportion of CCs (predominantly adenocarcinomas) are not HPV-positive, even in studies that applied the most sensitive detection methods [24].

In sum, HPV has been established as the causal agent of CC [3], and reported [25] to meet all of Hill's criteria for causation (i.e., strength, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment, and analogy) [26]. Although a persistent infection with HPV is the necessary cause of CC, it is not a sufficient cause, and the complete causal web of factors that affect progression and regression of cervical precancers remains uncertain. Nevertheless, improved understanding of cervical carcinogenesis has led to the development of novel technologies including biomarkers such as HPV tests for CC screening and secondary prevention as well as prophylactic HPV vaccination for primary prevention. These technologies provide opportunities to improve CC prevention efforts and reduce the burden of CC.

2.2 Cervical cancer prevention

This section provides an overview of CC prevention approaches, including the principles and aspects of CC screening, organized CC screening programs and HPV vaccination policies (in Norway and elsewhere), and outlines the key knowledge gaps for CC prevention this thesis aims to address.

2.2.1 Cervical cancer screening

Principles and aspects of cervical cancer screening

In medicine, screening involves the use of tests, examinations or other medical procedures to identify the likely presence of a specific disease or condition in asymptomatic individuals [27]. Individuals with a positive screening test are followed-

up for diagnosis and, if necessary, treatment. Population-based screening can target either an entire population or subgroup (i.e., mass screening) or a selected high-risk group (i.e., selective screening). The aim of screening is to prevent the specific disease from developing (e.g., by detecting its precursors) and/or to improve prognosis by detecting the disease at an early stage. In 1968, the World Health Organization suggested 10 principles that should be considered and satisfied prior to implementing a population-based screening program [27]:

1. The condition sought should be an important health problem.
2. There should be an accepted treatment for patients with recognized disease.
3. Facilities for diagnosis and treatment should be available.
4. There should be a recognizable latent or early symptomatic stage.
5. There should be a suitable test or examination.
6. The test should be acceptable to the population.
7. The natural history of the condition, including the development from latent to declared disease, should be adequately understood.
8. There should be an agreed policy on whom to treat as patients.
9. 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.
10. Case-finding should be a continuing process and not a “once and for all project”.

These principles are generally met for CC screening, which is targeted at adult women (exact recommendations for screening target ages vary across countries) with the aim to reduce morbidity and mortality from CC by detecting and removing cervical precancers before they have an opportunity to progress to cancer. Achieving this requires a sequential process, involving: (i) the primary screen, recommended to all women targeted by the program (who are perceived to be healthy), (ii) management of women with a positive screening test (often referred to as triage or secondary screening), (iii) diagnostic colposcopy with biopsy, and (iv) treatment.

Among those with a positive screening test, triage is necessary for deciding who should be referred for diagnostic colposcopy and who should receive intensified surveillance or return to routine screening. It has been suggested that risk thresholds (e.g., risk of developing cervical precancer or cancer within the next screening round) should guide follow-up management [16]. For example, follow-up testing is suggested for women with $\geq 2\%$ risk of developing precancer within the next 2-3 years, and a diagnostic colposcopy is recommended if the risk is $\geq 10\%$. However, these thresholds are based

on 'rules of thumb' and the optimal thresholds to guide clinical management have not been formally evaluated. As it is currently not possible to differentiate precancers destined to progress from those that will spontaneously regress in the absence of treatment, all women who are detected with CIN2 or CIN3 are recommended treatment. The most common treatment of precancers in developed countries involves removal of cone-shaped tissue from the cervix, which is usually performed with an electro-surgical loop (i.e., loop electro-surgical excision procedure (LEEP)) with local anesthesia. Women diagnosed with CC are treated depending on cancer stage.

The achieved health benefits of a CC screening program should be balanced against the potential harms and resource use of screening (**Figure 3**) [28]. Hence, prior to implementing a new screening policy, decision-makers should ensure that screening algorithms provide efficient use of resources, are feasible in both the short- and long-term, and keep the burden to women at an acceptable level. For example, physician consultations and colposcopy referrals require that a woman spend time and money (e.g., transportation costs and co-payments) to attend the procedure. Women may also experience anxiety from the procedure itself, or from awaiting and receiving test results. In addition, diagnostic colposcopy-directed biopsy may cause pain, bleeding, or discharge [29], and treatment of cervical precancer is associated with an increased risk of preterm birth and other adverse pregnancy outcomes [30, 31]. The majority of precancers would never have progressed to cancer in the absence of treatment [22] yet available technologies cannot distinguish between precancers that will progress to CC from those likely to regress; consequently, CC screening involves some over-treatment. Thus, when designing or refining a screening program, decision-makers must consider several factors in order to ensure an acceptable balance between the screening benefits, harms and resource use.

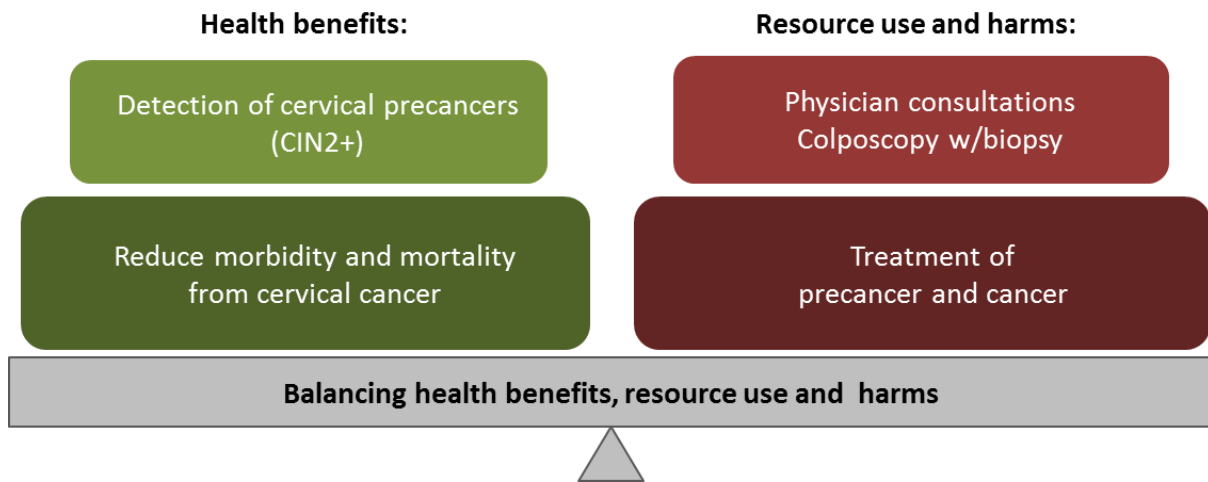


Figure 3. Example trade-offs between health benefits, resource use and harms of CC screening.

Balancing screening trade-offs requires the consideration of imperfect diagnostic tests and the distribution of test outcomes (i.e., true positive, false positive, true negative and false negative). This includes an assessment of the sensitivity, defined as the probability of testing positive given that the disease is present, and specificity, defined as the probability of testing negative given that the disease is absent, which together impacts the probability of having a positive screening test (i.e., positivity rates). Estimates of diagnostic accuracy are convoluted by different classification systems for reporting of cytology and histology results, which serves as a proxy for the underlying disease (**Table 1**). Moreover, the sensitivity and specificity estimates of a diagnostic test may be biased due to differing disease severity in different populations (i.e., spectrum bias) and if the gold standard test has not been used for the controls (i.e., verification bias). For example, one of the few studies that adjusted for verification bias when estimating the diagnostic accuracy of cytology and HPV testing showed that adjusted estimates yielded a lower sensitivity and a higher specificity than unadjusted estimates [32]. Using the formula of Bayes revision, one can further calculate the probability of having the disease given that the screening test was positive (i.e., positive predictive value), and the probability of not having the disease given that the screening test was negative (i.e., negative predictive value). These estimates depend on both the diagnostic accuracy and prevalence of the disease. Although the ideal screening test would have both high sensitivity and high specificity, it has been

suggested that sensitivity is most important in primary screening to ensure follow-up of individuals who are at an elevated risk of developing cancer, while the triage algorithm can be designed to increase the specificity of the overall screening program [33].

Table 1. Classification of cytology and histology results used in Norway.

Cytology		Histology	Underlying disease
NILM	Normal, no intraepithelial lesion or malignancy	Normal	Normal cervix
ASC-US	Atypical squamous cells of undetermined significance		
LSIL	Low-grade squamous intraepithelial lesion	CIN1	HPV infection
ASC-H	Atypical squamous cells – cannot exclude HSIL		
HSIL	High-grade squamous intraepithelial lesion	CIN2 CIN3	Precancer
		Cancer	Cancer

Adapted from Schiffman et al. [6].

Several diagnostic tests are available for primary CC screening; the two most established technologies include cervical cytology and HPV DNA testing. Cytology is a subjective test that requires morphological interpretation by a cytotechnician; consequently, reproducibility is low and diagnostic accuracy varies widely across studies [32, 34-36]. In a recent meta-analysis, the sensitivity (specificity) of cytology to detect (exclude) CIN2+ at a threshold of atypical squamous cells of undetermined significance or more severe (ASC-US+) was 72% (68%) [36]. Meta-analyses suggest that the diagnostic accuracy of cytology does not vary by collection method (i.e., conventional Pap smear versus liquid-based cytology), although liquid-based collection may result in fewer inadequate tests and is preferable because it allows re-using the sample for HPV testing (i.e., reflex testing) [35, 37]. However, recent studies suggest that the sensitivity for detecting CIN2+ depends on the type of liquid-based cytology test [38, 39]. In contrast to cytology, HPV DNA tests are automated and reproducible. The HPV DNA test is a biomarker (defined by the National Institute of

Cancer as ‘a biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease’ [40]), which detects presence or absence of HPV DNA. These tests have proven more sensitive, but less specific than cytology as most HPV infections are transient [32, 36, 41]; a recent meta-analysis reported a sensitivity (specificity) to detect (exclude) CIN2+ of 91% (61%) [36]. However, a wide range of HPV DNA tests are available that differ in test method as well as which HPV types are included, and consequently, the accuracy varies by test [42]. The improved sensitivity and negative predictive value of an HPV test makes it more favorable for primary CC screening; several large, randomized clinical trials have demonstrated that HPV DNA testing in primary screening provides greater protection against CC and allows extension of primary screening intervals [41]. However, the reduced specificity of HPV DNA testing necessitates appropriate triage management to avoid unnecessary follow-up, yet uncertainty remains about the health and resource use trade-offs associated with candidate triage algorithms for HPV-positive women [43].

While cytology and HPV DNA tests are widely adopted in CC screening, emerging biomarkers provide opportunities to improve the effectiveness and efficiency of CC screening. For example, HPV DNA tests that can provide genotype-specific information (e.g., isolating positivity for the two most carcinogenic genotypes, HPV-16 and -18, rather than positivity for an array of high-risk HPV types) are increasingly available. Other biomarkers include HPV viral messenger ribonucleic acid (mRNA) transcripts of E6/E7 proteins (‘HPV mRNA test’) and p16/Ki67 dual staining (which involves staining of p16 and Ki67 proteins from cytology/histology slides) [44]. These biomarkers may help improve follow-up of women with minor cervical cytological lesions, yet the health and economic trade-offs associated with using these biomarkers in the screening triage algorithm remains unknown.

Cervical cancer screening programs

The introduction of CC screening has contributed to reducing CC incidence [2, 45-47]. For example, a study using 50 years of data on CC incidence in the Scandinavian

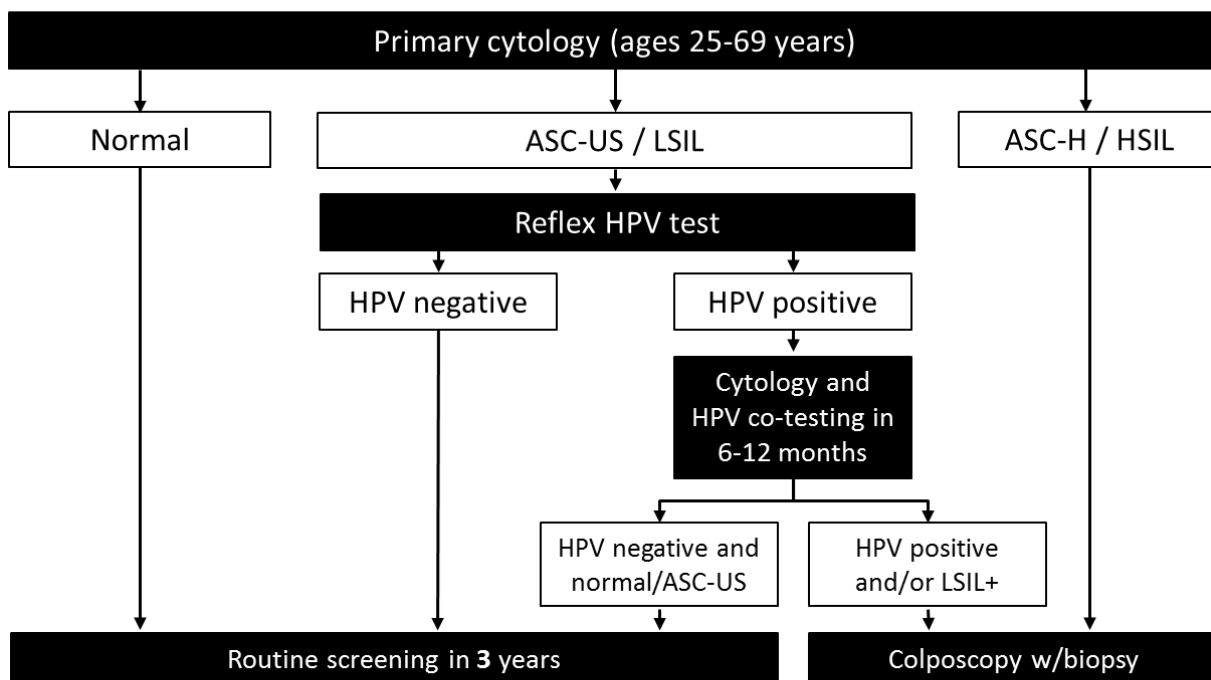
countries of Denmark, Norway and Sweden estimated that the introduction of cytology-based screening might have prevented almost half of the CC cases that would have been expected in the absence of screening [47]. In a Norwegian-specific study, the proportion of prevented CCs since the introduction of opportunistic screening with cytology in the 1970s was suggested to be as high as 68% [2]. Population-based screening programs are widely implemented in European countries and the high coverage achieved in many countries suggests that the program is perceived acceptable to screen-eligible women [48]. The performance of these programs (e.g., compliance rates) are continuously monitored and refined [49]. For example, audits of the Swedish CC screening program showed that women who attended CC screening according to guidelines had a reduced risk of developing CC [50], and improved CC prognosis [51]. Most countries recommend primary screening using cytology or cytology in combination with HPV testing (i.e., co-testing). Following the evidence from randomized controlled trials that HPV testing is more efficacious than cytology as the primary screening method [41], several countries have begun to switch to primary HPV testing, such as Norway [52], the Netherlands [53], Australia [54], Italy [55] and the United States [56].

The Norwegian CC Screening Program was implemented in 1995 and invites women aged 25 to 69 years to attend cytology-based screening every 3 years. The screening program is managed by the Cancer Registry of Norway, which collects and monitors data on screening and cancer data, such as cytology, HPV test and histology results, as well as CC diagnoses. In Norway, nearly 70% of women in screening target age have attended screening within the last 3.5 years [57]. However, a study using population-based data from the Cancer Registry of Norway to evaluate screening behavior over more than 20 years of organized screening found that less than half of screen-eligible women attended screening at the recommended repeated intervals [58]. This study also found that screening behavior was associated with CC outcomes; for example, women who attended screening every 3.5 years or more frequently had lower CC incidence and were diagnosed with CC at an earlier stage than women who attended screening less frequently.

The current screening guidelines in Norway (**Figure 4A**) recommends that women with a normal cytology can return to routine screening, while women who are detected with high-grade cervical lesions (i.e., ASC-H or HSIL) on their primary screen are advised direct colposcopy with biopsy. Reflex HPV DNA testing is used for women who are detected with minor cervical lesions (i.e., ASC-US or LSIL); reflex HPV-positive women are advised repeat cytology and HPV co-testing in 6-12 months while HPV-negative women can return to routine screening. For women with ASC-US/LSIL and reflex HPV-positive results, diagnostic colposcopy with biopsy is recommended for women who are persistently HPV-positive and/or have cytology results indicating LSIL or more severe, while the remaining women can return to routine screening. This triage algorithm for women with ASC-US/LSIL was implemented July 1st 2014. Between 2005 and 2014, triage of these women involved repeat delayed cytology and HPV co-testing in 6-12 months rather than reflex HPV testing (referred to as ‘former guidelines’). Moreover, during 2005-2011, both HPV DNA and mRNA tests were used, yet since 2011, only HPV DNA tests have been approved for reimbursement [59, 60].

Norwegian health authorities are currently evaluating switching from cytology to primary HPV testing starting at age 34 years (**Figure 4B**) [52], but will maintain triennial cytology for women aged between 25 and 33 years due to the high prevalence of transient HPV infections among younger women [61]. From February 2015, a controlled implementation pilot study was initiated in four Norwegian counties (covering ~25% of the population). For women aged ≥ 34 years, the screening algorithm currently under evaluation (i.e., ‘proposed guidelines’) involves primary HPV testing with return to routine screening in 5 years for HPV-negative women. Reflex cytology is used for HPV-positive women; women with an abnormal cytology (i.e., ASC-US+) are advised colposcopy with biopsy, while women with a normal cytology are recommended repeat HPV testing in 12 months. Women who are persistently HPV-positive will be referred for colposcopy with biopsy, while HPV-negative women can return to routine screening. As of July 2017, the Cancer Registry of Norway will send their recommendation for implementing primary HPV testing to the Norwegian Directorate of Health and the Ministry of Health and Care Services for consideration; if approved, national roll-out is expected within the next couple of years.

A. Current guidelines in Norway



B. Proposed guidelines in Norway

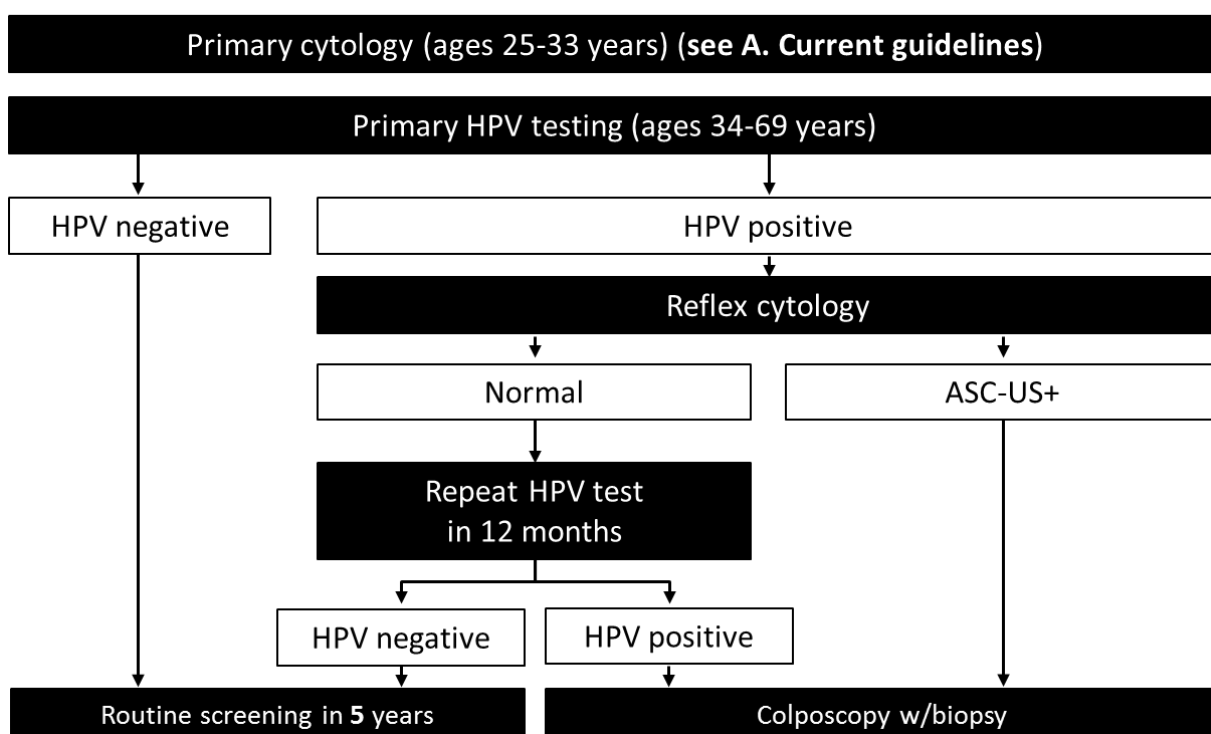


Figure 4. Cervical cancer screening guidelines in Norway: A) Current guidelines (i.e., as of August 2017) and B) Proposed guidelines. Black boxes indicate follow-up management; white boxes indicate a test outcome. LSIL+ (ASC-US+) indicate LSIL (ASCUS) or more severe.

The Norwegian CC Screening Program was estimated to require an annual societal cost of NOK730 million (2013-Kroner) [62], and these investments in the current (and proposed) guidelines have demonstrated ‘good value for money’ [63] according to current benchmarks for cost-effectiveness in Norway (see Section 3.1.3) [64]. In Norway, the cytology-based screening program will remain important for younger women (aged 25-33 years) unlikely to be recommended primary HPV-based screening, as well as for all women in screening target age until the HPV-based program is scaled up nationwide, at least for the next decade when the majority of Norwegian women have not received the HPV vaccine in adolescence. Within the current cytology-based screening program, follow-up of women with ASC-US/LSIL remains a challenge as these women are at an intermediate risk of developing CC within the next screening round, yet the risk is not considered high-enough to justify direct colposcopy referral [16, 65]. For these women, novel applications of biomarkers may help improve the follow-up algorithm, particularly for younger women not currently recommended primary HPV testing under the proposed HPV-based guidelines. The cost-effectiveness of primary HPV testing in Norway has been evaluated previously [63], suggesting that HPV-based screening starting at 34 years is cost-effective. However, the study also found that, if considering a wider range of strategies, switching at age 31 years would be preferred. Prior to national roll-out of primary HPV testing, evaluating the health benefits and resource use trade-offs associated with candidate algorithms may elucidate which algorithm ‘levers’ decision-makers may use to help maintain an acceptable balance between screening benefits and harms, such as the age of switching to primary HPV testing, the primary screening interval and follow-up of HPV-positive/cytology-negative women. Consequently, there is a need for studies that evaluate the use of novel biomarkers within the current cytology-based triage algorithm prior to nationwide roll-out of primary HPV DNA testing, as well as studies that can inform the design of future HPV-based screening algorithms.

2.2.2 HPV vaccination

The first generation of prophylactic HPV vaccines was approved by the European Medicines Agency and the US Food and Drug Administration in 2006-2007. These include the bivalent vaccine (Cervarix®, GlaxoSmithKline Biologicals, Belgium), protecting against HPV-16 and -18 high-risk infections, and the quadrivalent vaccine (Gardasil®, Merck & Co., Whitehouse Station, NJ USA), protecting against both HPV-16 and -18 infections and two low-risk types HPV-6 and -11 (associated with genital warts, see Section 2.1.1). The second-generation nonavalent HPV vaccine (Gardasil9®, Merck & Co., Whitehouse Station, NJ USA) was approved in 2009, which protects against HPV-6, -11, -16, -18, -31, -33, -45, -52, -58 infections. All vaccines have demonstrated safety (i.e., without major adverse events following immunization [66]) and high efficacy (i.e., achieve the intended beneficial effects in vaccinated individuals under ideal conditions of use [66]) against persistent HPV infection and cervical precancer [67-73]. There is also evidence suggesting that the bivalent and quadrivalent vaccines offer cross-protection against non-vaccine targeted HPV types [74], and that the vaccines provide herd immunity (i.e., protection against HPV infections and genital warts among unvaccinated individuals) [73, 75]. Vaccine administration was initially recommended to include a 3-dose schedule; however, a 2-dose schedule demonstrated non-inferiority compared with a 3-dose schedule [76, 77], which prompted the World Health Organization to recommend a 2-dose schedule for individuals receiving their first immunization at age ≤ 14 years [78]. With longer follow-up data from HPV immunization programs, the duration of the vaccine efficacy (especially duration of cross-protection [74]) and potential type-replacement can be evaluated.

Most developed countries, including Norway, have implemented national HPV immunization programs [79]. In Norway, school-based HPV vaccination for 12-year old girls was implemented in 2009. From the fall of 2017, the vaccination program will switch from the quadrivalent to the bivalent vaccine [80], and from a 3-dose to a 2-dose schedule [81]. A temporary (i.e., a two-year) 'catch-up' vaccination program targeting women born in 1991 or later (i.e., aged 26 years or younger) was implemented in November 2016. Following recommendations by several Norwegian guidance bodies

to expand the school-based vaccination program to include boys, the government decided to implement a gender-neutral HPV vaccination program starting from the fall of 2018 [82].

For women vaccinated against HPV infections in adolescence, the risk of developing CC is expected to decrease considerably; as a result, the heterogeneity of CC risk in the population will increase. The first birth cohort of women who received the HPV vaccine at age 12 years in 2009 in Norway will become eligible for the Norwegian CC Screening Program in 2022. However, the Norwegian screening guidelines have not yet been adapted for these lower-risk women, which may be important for CC screening to continue to provide ‘good value for money’ and balancing screening benefits and harms in the post-vaccination era. For example, previous model-based analyses have suggested that CC screening for HPV-vaccinated women may involve primary HPV testing, start at a later age, and occur less frequently than guidelines for unvaccinated women [63, 83-86]. Importantly, a study comparing the cost-effectiveness of candidate screening strategies in cohorts offered the nonavalent HPV vaccine in Australia, England, New Zealand and the US, suggested that optimal guidelines might differ between countries [86]. The cost-effectiveness of candidate screening guidelines for women vaccinated with the first generation HPV vaccines have been previously evaluated within the context of Norway [63]. However, no studies have evaluated a broad range of strategies (e.g., screening intervals >5 years) for women vaccinated with either the first or second generation vaccines in Norway, as well as evaluating the value of implementing a separate set of guidelines for HPV-vaccinated women (that differ from guidelines for unvaccinated women). Such analyses may aid decision-makers in designing CC screening guidelines in the post-vaccination era.

3 THEORETICAL FRAMEWORK

3.1 Priority setting and economic evaluation in healthcare

A fundamental problem in economic theory is how to allocate scarce resources in a society with unlimited wants. For example, we prefer more healthy years of life to less, which is reinforced with emerging technologies that provide opportunities for further health improvements. Increasing (or ‘unlimited’) demand for health services poses a challenge for society when resources are limited, such as health personnel, medical equipment, operating rooms and hospital beds. Moreover, there is an opportunity cost of utilizing resources for a specific purpose; spending more resources on one aspect of health care (e.g., preventative interventions) displaces resources that could alternatively be spent on other health measures (e.g., curative interventions). Similarly, spending more resources on health care overall displaces resources that could alternatively be spent within other sectors. Consequently, priority setting is unavoidable and considering the trade-offs of alternative courses of action is an essential part of decision-making. This chapter reviews the principles of priority setting and economic evaluation in healthcare, including an overview of the components of cost-effectiveness analysis and methods for measuring and valuing health and economic outcomes.

3.1.1 Priority setting in healthcare

In Norway, an over-arching health policy objective that has been cited in several official documents for more than two decades is to provide “*more healthy years of life for the population as a whole*” [87-89]. In addition, the Patients’ Rights Act (§2.1) states that all Norwegian citizens have the right to ‘necessary’ health care services [90]. Regulations further specify that this right should be based on (i) the severity of the disease, particularly the reduction in length and/or quality of life if the healthcare is delayed, (ii) the expected effectiveness of the healthcare, and (iii) that the expected costs are reasonable in relation to expected health benefit of the healthcare [91]. These criteria were based on two commissions on healthcare priority setting in Norway (referred to as Lønning I and II) in 1987 and 1997, respectively [92, 93]. Principles for

healthcare priority setting in Norway was recently evaluated by a commission in 2014 (referred to as the Norheim commission) [94, 95] and a working group in 2015 (referred to as the Magnussen group) [96]. Together with Lønning I and II, the reports of the Norheim commission and the Magnussen group formed the basis for the Norwegian white paper on priority setting in healthcare, which was published in 2016 [97]. In line with the previous Lønning-reports and the Patients' Rights Act, this white paper stated that priority setting in healthcare should be based on the expected benefits and costs of health interventions, as well as the severity of the disease. Importantly, the white paper emphasized that these criteria should be evaluated jointly such that the more severe the disease or the more benefit an intervention provides, the higher resource use can be accepted. In sum, this implies that economic arguments, including cost-effectiveness, should be an integrated part of priority setting in healthcare in Norway.

3.1.2 Economic evaluation

Economic evaluation is a decision support tool to inform different types of decision-makers about the efficient allocation of health care resources [98-100]. This methodological framework involves *'the comparative analysis of alternative courses of action in terms of both their costs and consequences'* [100, page 4]. An economic evaluation should reflect the existing evidence, link intermediate to final endpoints, extrapolate consequences over an appropriate time horizon of the evaluation, and make results applicable to the decision-making context [100]. Although various types of health care evaluations exist (e.g., consequence analysis and cost-minimization analysis), a complete economic evaluation requires the comparison of both costs and consequences of alternative interventions [100]. The two main types of economic evaluation are cost-benefit analysis (CBA) and cost-effectiveness analysis (CEA). While costs are measured in monetary units in both types of analysis, health consequences are measured differently. In CBA, health benefits are measured in monetary units, providing a useful framework to inform resource allocation decisions within and between sectors of the economy. In contrast, CEA measures health consequences in natural units, such as life years gained, precancers detected and cancers averted. Many

CEAs measure preference-based consequences such as quality-adjusted life years (QALYs), calculated as the life years associated with an intervention over the relevant time horizon weighted by the quality of those life years [101] (see details in Section 3.1.4), which represent a variant of CEA that is sometimes referred to as cost-utility analysis. Measuring health outcomes in non-monetary units is often preferred within health and medicine because it directly reflects the general health policy objective of maximizing health, as well as the difficulties associated with placing a monetary value on health consequences. CEA is therefore the most commonly-used approach to economic evaluation in healthcare, and is the recommended approach in Norway [64] and in other countries [98].

The theoretical foundation of economic evaluation lies in welfare economics, decision theory to inform individual preferences, and the mathematical theory of constrained optimization [98-100, 102]. CBA is grounded in welfare economics, which forms the basis for the two key Pareto principles of value judgements [99]. The first is referred to as ‘actual Pareto improvements’, which occurs when a policy improves the welfare of one or more persons without making anyone worse off. The second is referred to as ‘potential Pareto improvements’ (or the Kaldor-Hicks criterion), denoting a policy in which gainers in welfare could potentially compensate the losers in welfare while remaining better off after the policy change. Stemming from welfare economics, CBA has been referred to as the ‘welfarist approach’ [103]. Although the theoretical foundations of CEA have been subject to debate [98], it has been referred to as the ‘extra-welfarist approach’ [103]. The extra-welfarist approach differs from the welfarist approach in that outcomes other than individual utility can be considered in the analysis and that (healthy) individuals rather than the affected individual can be the source of valuation of the relevant outcomes. Furthermore, the extra-welfarist approach allows outcomes to be weighted according to other principles than preference-based utilities, and that comparisons between individuals in several dimensions is allowed [103]. A more detailed comparison and discussion of the welfarist and extra-welfarist approach is provided by Brouwer and colleagues [103]. CEA also relies on decision analysis, for which the core elements include: (i) the probability of outcomes, (ii) payoffs associated with the outcomes (e.g., costs, resource

use, health consequences), and (iii) expected values. Central to decision analysis is expected utility theory, a normative theory about individual decision-making under conditions of uncertainty [100]. This theory has been criticized for not reflecting how individuals make decisions in practice, and alternative theories such as prospect [104] and regret [105] theory have been suggested. Finally, a fundamental principle of CEA is constrained optimization, a process of maximizing desirable outcomes given constraints (e.g., budgetary or resource constraints) [98]. This process involves identifying: the possible strategies, the desired outcome(s), the constrained resource(s), the outcomes (e.g., health benefits and costs) associated with each possible intervention, dominated interventions (which are eliminated from further consideration), the trade-offs associated with the possible interventions and the optimal strategy given these trade-offs [98]. Alongside the process of constrained optimization, conducting an economic evaluation further requires a choice of the target population for receiving the intervention, the analytic perspective (i.e., the viewpoint of the analysis), the scope of the analysis (e.g., relevant outcomes, time horizon), and discounting of health and economic outcomes [98].

Several guidance documents for conducting health economic evaluation have been published in order to increase the quality and consistency of priority setting decisions. For example, a report on 'Cost-Effectiveness in Health and Medicine' was published in 1996 [102] following the work of an expert group referred to as the Panel on Cost-effectiveness in Health and Medicine (herein referred to as the 'First Panel'). The report synthesized available evidence and recommendations for conducting economic evaluation and cost-effectiveness analysis of healthcare interventions to guide priority setting in the US, and became an international reference book for conducting economic evaluation. Almost two decades later, in 2012, the Second Panel on Cost-effectiveness in Health and Medicine (herein referred to as the 'Second Panel') was formed with the goal of updating the recommendations from the First Panel. The results of their work was published in 2016 [98, 106]. While the recommendations of the Second Panel are intended for an international scope and impact, they are focused on the US context [98]. In Europe, Drummond and colleagues have published several editions of a reference book for health economic evaluation [99, 100], and the National

Institute for Health and Care Excellence (NICE) in England (a health agency widely known for evaluating the cost-effectiveness of technologies to inform the National Health Service has developed their own guide to methods for technology appraisal [107]. Economic evaluations should be specific to the relevant setting (e.g., a country); in Norway, the official document providing guidelines for economic evaluation in healthcare was published by the Norwegian Directorate of Health in 2012 [64]. In addition, the Norwegian Medicines Agency has published guidelines for pharmacoeconomic analyses, which inform their decisions of whether a drug should be accepted to the reimbursement scheme [108]. However, these guidelines are likely to be updated in the near future to be in accordance with the 2016 Norwegian white paper on priority setting in healthcare [97]. As the majority of evaluations conducted in this thesis were carried out prior to the fall of 2016, the thesis primarily reflects the guidelines outlined by the Norwegian Directorate of Health [64]. These guidelines and their differences from recommendations outlined in the 2016 Norwegian priority setting white paper and in international guidelines (e.g., Second Panel, NICE) are discussed throughout this thesis.

Components of economic evaluation

A first step to conducting an economic evaluation is to identify which strategies or interventions to evaluate. Ideally, all relevant strategies for a particular decision problem should be considered; identification of strategies thus requires in-depth understanding to the field and, often, consultations with experts. Omitting relevant strategies can lead to biased comparisons of strategies and incorrect identification of preferred strategies [109]. In addition, the analyst should define the target population of the analysis (i.e., to whom the intervention is intended).

Depending on the decision-making context, the analyst must decide which outcomes to evaluate and over what time horizon. The time horizon should be long enough to capture all relevant cost and health consequences of the relevant strategies, which often requires the evaluation of costs and consequences over a lifetime. The outcomes may be surrogate (e.g., precancers detected) or clinically relevant (e.g., cancers

prevented, life years gained), or both, which may help inform different aspects of the decision-making process such as the feasibility of implementation in the short- and long-term. Using QALYs as a measure of health outcome is often recommended in order to report health outcomes that are commensurable across diseases and patient groups, and reflecting the impact of health interventions on both the quantity and quality of life (see Section 3.1.4 for details about the QALY concept). Guidance bodies (e.g., the Second Panel [98]) have also encouraged reporting of other relevant outcomes such as cancer incidence reduction for cancer screening programs.

Which health and economic consequences are relevant to consider is usually defined by the viewpoint of the analysis. To improve the consistency in the reporting and enhance the comparability of CEAs, several guidance bodies recommend using ‘reference cases’ for reporting CEAs; that is, analytic perspectives that incorporate a certain set of costs and consequences as defined by a standard set of methods and assumptions [98, 107]. For example, the Second Panel recommend to report results using both a societal and a healthcare reference case perspective, alongside potential other perspectives that may be relevant for the decision-makers. In the healthcare sector perspective, it is recommended to include only the medical costs (within the formal healthcare sector) paid by third-party payers and out-of-pocket by patients [98], while the societal perspective should include *“all costs and health effects regardless of who incurs the costs and who obtains the effects”* [106]. In the UK, NICE recommends to use a healthcare reference case perspective [107], while in Norway, a societal analytic perspective is recommended for use in CEA [64]. QALYs is the recommended measure of health outcomes in the Norwegian guidelines for CEA [64], in the 2016 Norwegian priority setting white paper [97], and for use in the reference case analyses outlined by the Second Panel [98] and NICE [107]. However, recommendations for which costs to include in the societal perspective differ across guidelines (e.g., see Section 3.1.5 for discussion on how the Norwegian guidelines differ from the recommendations of the Second Panel). The Second Panel acknowledged that the societal perspective varied considerably across studies and recommended including an ‘impact inventory’ table to increase the transparency of the analytic

perspective [98]. An example of an impact inventory for *Papers II and IV* is displayed in **Table 3** in Section 5.1.

To allow comparison of alternative strategies with differential timing of costs and consequences, CEA guidelines recommend to discount both costs and health consequences (i.e., translating future flows of costs and consequences into their present values) [64, 98]. The rationale for discounting is the value of time; we value consumption more today than in the future because we are impatient and risk averse for the uncertain future (i.e., time preference), and there is a growth rate of consumption over time (i.e., the opportunity cost of investments; resources invested today could yield more resources tomorrow). The issue of whether costs and health outcomes should be discounted at the same rate and what the rate should be has been debated (a more in-depth discussion is provided elsewhere [98, 99]), and recommendations for discounting frequently vary across guidelines. For example, discounting costs and health outcomes at the same rate is recommended in countries such as Norway (4% per year [64, 97]), in the UK (3.5% per year [107]), and in the US (3% per year [98]). In contrast, countries such as the Netherlands recommend a differential rate for discounting costs and health outcomes (i.e., at 4% and 1.5%, respectively [110]).

3.1.3 Cost-efficiency and cost-effectiveness

Following the comparison of costs and consequences associated with alternative strategies, the analyst can provide a recommendation of which strategy provides ‘good value for money’ (i.e., which strategy is ‘cost-effective’, ‘optimal’ or ‘preferred’). When decision-makers are considering mutually exclusive strategies (i.e., only one strategy can be adopted, such as a screening algorithm), the predominant metric used to identify cost-efficient strategies is the incremental cost-effectiveness ratio (ICER). The ICER is defined as the difference in costs of a strategy (**A**) compared to the next least costly strategy (**B**) divided by the difference in health benefit (e.g., QALYs) of those strategies, defined by (1):

$$(1) \quad ICER_A = \frac{Cost_A - Cost_B}{Health\ benefit_A - Health\ benefit_B}$$

Strategies that are more costly and less effective than the next least costly strategy are considered strongly dominated, and strategies that are more effective but less cost-effective (i.e., have a higher ICER) than other strategies are considered weakly dominated. Both strongly and weakly dominated strategies are excluded from further consideration, while the remaining strategies are considered cost-efficient and often referred to as the ‘efficiency frontier’.

Determining which strategy is preferred (i.e., cost-effective or optimal) among the cost-efficient strategies requires a benchmark of how much decision-makers are willing-to-pay for an additional unit of health benefit. This benchmark is commonly referred to as the willingness-to-pay threshold (often expressed in terms of additional costs per additional QALY), and the preferred strategy is the one with an ICER just below the willingness-to-pay threshold. For example, in the US, threshold values of \$50,000-100,000 per QALY gained have been commonly-cited [111]. In Norway, prior to the 2016 priority setting white paper [97], a commonly-cited threshold value was 500,000 Norwegian Kroner (NOK) in 2005-values (~80,000 US Dollar (US \$)) [64, 112], which corresponds to ~\$100,000 in 2014-values (US \$1 = NOK6.30 [113]) when adjusted for changes in real income wage in Norway during years 2005-2014 [114]. However, the policy-makers emphasized that this should serve as a ‘reference value’ rather than a strict threshold as cost-effectiveness analysis can only inform one aspect of the decision-making process [112]. Rather than determining a value of what the health system should be willing-to-pay for additional benefits, another approach to deducing the threshold value is to reflect the opportunity cost, that is, the health benefits forgone by adopting a new intervention in clinical practice. In the UK, researchers recently estimated a threshold value reflecting the opportunity costs within the National Health Service (using expenditure data from 2008 and mortality data from 2008-10) of £12,936 per QALY (US \$20,212) [115]. Based on this threshold, an opportunity cost of NOK275,000 per QALY was suggested as a Norwegian benchmark in the 2016 Norwegian priority setting white paper [97]. The white paper further stated

that the ICER of an intervention should be evaluated together with the severity of the disease, such that the more severe a disease is, the higher an ICER can be accepted. However, the white paper did not suggest explicit criteria for weighting the ICER by severity of disease; rather, the white paper emphasized that priority setting decisions should be based on a holistic assessment of the documentation in light of the priority setting criteria.

An alternative approach to determining which strategy is optimal if the willingness-to-pay threshold is known is to use a metric referred to as the incremental net monetary benefit (INMB). The INMB is defined as the additional health benefits ($\Delta\text{Health benefit}$) of a strategy multiplied with the willingness-to-pay threshold (λ) minus the additional cost (ΔCost) of that strategy, defined by (2):

$$(2) \quad \text{INMB} = (\Delta\text{Health benefit} * \lambda) - \Delta\text{Cost}$$

By directly incorporating the willingness-to-pay threshold, a strategy with a positive INMB (i.e., $\text{INMB} > 0$) indicates that the additional costs required to achieve the additional health benefit is less than what decision-makers would at most be willing-to-pay for those health benefits.

Defining a single willingness-to-pay threshold remains a challenge worldwide, and few countries have explicitly stated a threshold value [98, 116]. Without an explicit threshold, which strategy is optimal will be uncertain; however, identifying strategies on the efficiency frontier provides a useful framework to narrow down the strategies decision-makers need to consider. For example, decision-makers may want to evaluate other aspects in addition to the ICER or INMB of a strategy, such as feasibility (i.e., given capacity constraints) and harm-benefit considerations.

3.1.4 Measuring and valuing health outcomes

CEAs require that the incremental health benefits of an intervention are included in the denominator of the ICER. Although other outcomes such as life years gained and

precancers detected can be used, a commonly used and recommended measure of health outcome in CEA is QALYs gained, which captures changes in both quantity and quality of life [98, 100]. The QALY concept is based on the idea that, over time, individuals transition between health states with an associated value that depends on the desirability or preference for that health state, referred to as the health-related quality of life (HRQoL) [101]. HRQoL is sometimes referred to as the weight, utility or preference score, and reflects the physical and mental well-being associated with a particular health state. The QALYs associated with an intervention is calculated by multiplying the number of life years (accumulated over the relevant time horizon) by the HRQoL for those life years. The weights used to adjust life years with the associated HRQoL should be usable across diseases and conditions, and are conventionally measured on a scale ranging from 0 (indicating death) to 1 (indicating perfect health) [98, 101]. Furthermore, the weights should have an interval property to allow aggregation of QALY gains such that an increase in HRQoL of 0.1 should reflect the same gain across the spectrum (e.g., an increase from 0.2 to 0.3 represents the same HRQoL gain as an increase from 0.7 to 0.8).

HRQoL can be measured directly or indirectly. Direct methods involve asking subjects (e.g., patients, medical experts, the general population) directly about their HRQoL using valuation methods such as the standard gamble, the time trade-off and the visual analogue scale [98, 100, 101]. The standard gamble is directly based on expected utility theory by involving both choices and uncertainty, while the time trade-off method involves choices under uncertainty [100, 101, 117]. Indirect methods involve first asking the patient to describe their current health state using a health status classification system, and subsequently score the health state using previously obtained scores based on surveys of other patients or individuals in the general population (i.e., a preference-based scoring system). The indirect approach is often carried out using multi-attribute utility instruments, a method based on multi-attribute utility theory (an extension of expected utility theory that allows expressing utilities of outcomes with multiple attributes) [117, 118]. In this approach, a multi-attribute utility is constructed based on preferences within and across health attributes as described in the health status classification system [101]. The health-status classification system can be generic across

diseases (e.g., the EuroQol (EQ)-5D, Health Utilities Index, Short Form 6D, 15D) or specific to a certain condition or disease [119]. For example, the commonly-used EQ-5D instrument with 3 levels (EQ-5D-3L) measures the health attributes mobility, self-care, usual activities, pain/discomfort and anxiety/depression with three levels of severity (i.e., no, moderate or extreme problems with any of the attributes). The scoring system can be based on the preferences of the community or patients. While patients have direct experience with the disease and associated interventions, using patient preferences may introduce potential bias such as adaptation to a condition or disease [120]. Using community preferences may be favorable as members of the community are potential patients as well as the payers of public healthcare (which will benefit virtually all members of the society). The issue of whose preferences should be used in CEA has been debated and a more detailed discussion can be found elsewhere [98, 101].

Finally, although the QALY metric is widely applied and recommended as a measure of health benefit in CEA, there are several issues related to the methods for valuation of HRQoL [101], as well as ethical issues related to its use for resource allocation decisions [121]. For example, the different methods for valuation of HRQoL have shown to yield different results (e.g., EQ-5D and SF-6D [122]), and the iterative methods (i.e., time trade-off and standard gamble) may depend on the starting point for valuation (referred to as ‘anchoring bias’) [123].

3.1.5 Costs

In CEA, the downstream resource use associated with an intervention should be identified, quantified and valued in monetary terms [99]. The costs associated with an intervention broadly include medical costs (e.g., physician consultations, inpatient and outpatient care, treatment), time costs of patients (i.e., time spent on travel to healthcare facilities, waiting time at the facility, and the time spent receiving the healthcare), time costs of informal (unpaid) caregivers, transportation costs, and productivity costs (i.e., the societal production value of time). Different classifications of costs have been used in the literature; for example, the First Panel distinguished

between direct healthcare costs associated with an intervention, direct non-healthcare costs (e.g., transportation costs), patient's time cost, and productivity costs. The Second Panel broadly categorized costs according to whether they accrue inside or outside the formal healthcare sector, of which the latter include patient and caregiver time costs, transportation costs, and productivity costs [98].

As mentioned in Section 3.1.2, the identification of which cost components to include in the analysis depends on the perspective; while all healthcare resources associated with an intervention should be included in a healthcare perspective, all costs both within and outside the healthcare sector should be included in a societal perspective. However, exact recommendations for which non-healthcare costs to include vary across guidelines, particularly with respect to productivity costs. For example, the Norwegian guidelines for economic evaluation [64] recommends that a societal analytic perspective is adopted (including patient time and travel costs), yet suggest that productivity costs can be included if they are available. The guidelines further suggest that the results are presented both with and without productivity costs due to methodological issues with estimating these costs, as well as the ethical issue of whether productivity costs should influence priority setting in healthcare [64]. In contrast, the 2016 Norwegian priority setting white paper [88] explicitly stated that productivity costs should not be included. Within the context of the US, the First Panel suggested that productivity costs was already captured in the QALYs and thus recommended against including these costs in the denominator of the ICER to avoid 'double-counting' [102]. Following years of research and debate, the Second Panel proposed that productivity costs are not included in the QALYs and recommended to include these costs in the societal reference case perspective [98].

Following the identification of the relevant cost components, quantification involves identifying units of measurement and specifying how these units should be quantified (e.g., using surveys, registries, reports, expert opinion). In order to value resources in monetary terms, the unit costs should reflect the opportunity costs of the resources. Ideally, the opportunity cost is reflected using marginal (rather than average) costs, for example using market prices (for competitive markets). However, as marginal costs are often unavailable, using the average cost is accepted because it often reflects the

marginal cost in the long term [98]. The two main approaches for valuing resources include micro-costing (e.g., primary collection of actual resource use associated with an intervention) and gross-costing (e.g., using Diagnostic Related Group reimbursement rate) [98]. The choice of method depends on feasibility of conducting micro-costing as well as the expected impact of using a more detailed costing approach.

3.2 Decision-analytic modeling

Economic evaluation may be conducted alongside clinical trials, which are essential in informing decisions about whether and how to adopt emerging technologies in clinical practice. However, no single trial can capture all the short- and long-term health and resource consequences of alternative courses of action (e.g., screening strategies) within multiple settings. In addition, delaying decisions to adopt new technologies while waiting for follow-up data is an active choice that may forgo health benefits. Mathematical simulation modeling (i.e., decision-analytic modeling) is an alternative approach for evidence acquisition, which involves synthesizing best available evidence from multiple sources of data (e.g., clinical trials, population-based registries, meta-analyses and independent studies) with explicit examination of uncertainty (e.g., diagnostic accuracy). Using available and (often) short-term data, mathematical models can extrapolate health and economic consequences over longer-term periods and quantify important trade-offs. These models have been broadly applied to inform decisions about health care, and its contribution to epidemiologic projections and policy guidance have received wide acceptance. For example, the US Preventive Services Task Force and NICE in the United Kingdom actively utilize mathematical models to inform the long-term consequences of an intervention [107, 124].

With roots in expected utility theory, decision analysis and decision-analytic modeling provide a systematic approach to decision-making under uncertainty by using mathematical relationships to project consequences associated with candidate strategies [125]. In addition to the general guidance literature for health economic evaluation (see Section 3.1.2), guidance for conceptualizing decision-analytic models

(‘models’) and dealing with model validation, uncertainty and transparency have been reported in a series of articles by the International Society for Pharmacoeconomics Outcomes Research—Society for Medical Decision Making (ISPOR-SMDM) Modeling Good Research Practices Task Force [126-132]. This chapter first provides an overview of different types of models that may be applied to quantify health and economic outcomes, followed by an overview of the components of decision-analytic modeling.

3.2.1 Types of decision-analytic models

A first step to conducting a decision analysis involves conceptualizing the decision problem and subsequently the decision-analytic model [127]. This includes a statement of the decision problem and analytic objective (e.g., informing clinical practice and resource allocation). Model conceptualization is governed by the decision problem and the characteristics of the disease/condition and should not depend on available data [98]. Multiple model types with different characteristics and properties are available, such as decision-tree models, state-transition models, dynamic transmission models, discrete event simulations, and agent-based models [98, 100, 127, 129-131, 133]. These models generally differ by their unit of representation (i.e., individuals versus members of a homogenous cohort), whether there is interaction between individuals, passage of time (e.g., time horizon and discrete versus continuous time), whether events are recurring and whether resource constraints should be explicitly modelled [127]. Ultimately, the choice of model scope and structure depends on the decision problem and policy contexts, and involves a trade-off between simplicity and transparency. For example, a more complex model may be warranted if the model is intended to inform multiple decision problems. Deciding the modeling framework further involves a consideration of how to account for passage of time as well as the unit of analysis (i.e., individuals, a homogenous cohort or a population) and whether interaction between individuals or other model components is required [98]. This section provides an overview of different model types with an emphasis on the two types used in this thesis: decision-tree and state-transition models.

Decision-tree models

A decision-tree model can be a useful tool to structure decision problems with less complex characteristics (e.g., a short and fixed time horizon) in a logical, sequential way [127]. The key elements of a decision-tree are square decision nodes (indicating a decision between alternative strategies), circular chance nodes (indicating two or more possible outcomes), triangular terminal nodes (indicating the end of a pathway), and branches forming pathways moving from left to right between the nodes (**Figure 5**) [100]. The pathways are mutually exclusive (i.e., an individual can only follow one pathway) and collectively exhaustive (i.e., an individual must follow a pathway completely until the terminal node). Each pathway is governed by probabilities, which indicate the likelihood of each possible outcome from the chance nodes and are conditional on earlier events. For example, the probability of surviving a disease may be conditioned on whether an individual received treatment for that disease or not, as indicated by probabilities p_{Alive_treat} and $p_{Alive_notreat}$ in **Figure 5**, respectively. Each pathway can be associated with health and economic outcomes; for example, in **Figure 5**, $Cost_treat_alive$ and $QALY_treat_alive$ indicate the costs and QALYs associated with the upper pathway. There are two main limitations of decision-tree models; first, the model becomes practically unmanageable when the number of pathways is large (i.e., the model becomes ‘bushy’), and second, passage of time is not accounted for unless explicitly built-in by the analyst. These limitations notwithstanding, decision-tree models have been widely applied to inform decision-making in multiple settings, such as treatment of rheumatoid arthritis [134], diagnosis of latent tuberculosis infection [135], and health interventions for drug and alcohol problems [136].

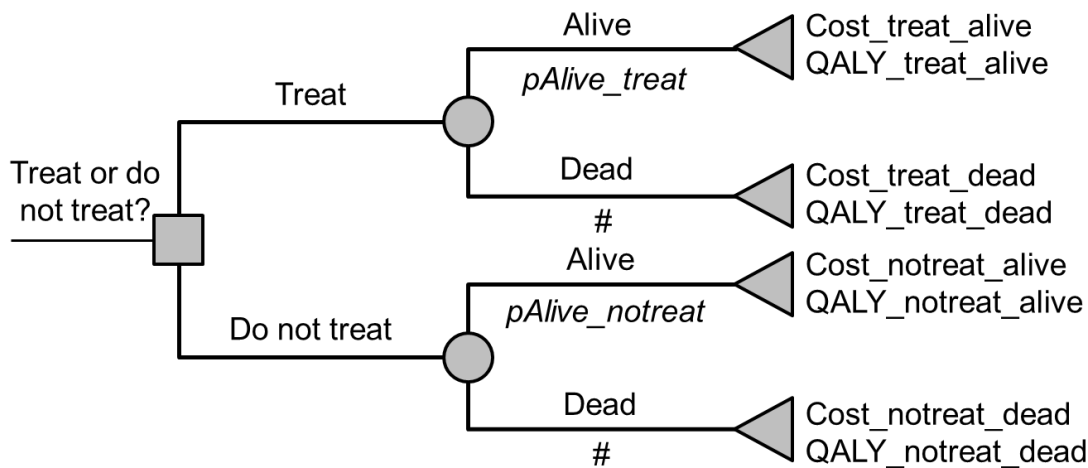


Figure 5. Simple decision-tree model schematic with options (i) treat and (ii) do not treat with outcomes alive or dead. # indicates the complement of the probabilities $p_{\text{Alive_treat}}$ and $p_{\text{Alive_notreat}}$. See description of decision-tree models in Section 3.2.1 for details.

State-transition models

To overcome the shortcomings of the decision-tree model, a state-transition model is useful when the decision problem can be conceptualized using health states to reflect the disease or treatment process [127, 131]. A state-transition model is particularly convenient when the decision problem is characterized by recurring events, changing health states over time and when explicit timing of events is required. The key elements of state-transition models are the health states and the associated values (e.g., the costs and HRQoL associated with a health state), cycle length, transitions and transition probabilities (**Figure 6**) [131]. The cycle length determines the length of time spent in a health state before a transition to another health state can occur. Transitions govern possible movements between health states and transition probabilities determine the likelihood of those transitions.

State-transition models differ by unit of analysis, which is either a hypothetical cohort (i.e., cohort simulation, often referred to as Markov models) or one individual at a time (i.e., microsimulation) [131]. Models using cohort simulation distribute an entire cohort across health states over time according to the transition probabilities. The main limitation of these models (the so-called ‘Markovian assumption’) is their inability to

allow transition probabilities to depend on event history, that is, the time spent in each of the health states. If dependency on event history is required, the Markovian assumption can be overcome by using ‘tunnel’ health states or microsimulation. Microsimulation models simulate a finitely large number of individuals (e.g., 1 million) one at a time between health states according to both transition probabilities and random numbers governing the outcome for a particular individual. As such, these models are stochastic and may require simulation of a large number of individuals in order to obtain stable estimates of the expected values, which may be computationally expensive.

Irrespective of using cohort simulation or microsimulation, the analysis can represent either a single cohort (e.g., a birth cohort or a single-age cohort) or a population [98]. For example, for evaluations of CC screening, a single cohort of women could be simulated over their lifetime starting at an early age (i.e., prior to acquisition of an HPV infection). Alternatively, a population approach could be used to simulate multiple cohorts reflecting women in screening target ages (e.g., women aged 25 to 69 years) over their remaining lifetime. The population approach enables projecting health and economic consequences for individuals currently eligible for screening, yet the consequences may then be influenced by the characteristics of each cohorts (e.g., changes in risk factors over time) and assumptions about the future (e.g., drug prices for future cohorts) [98, 137].

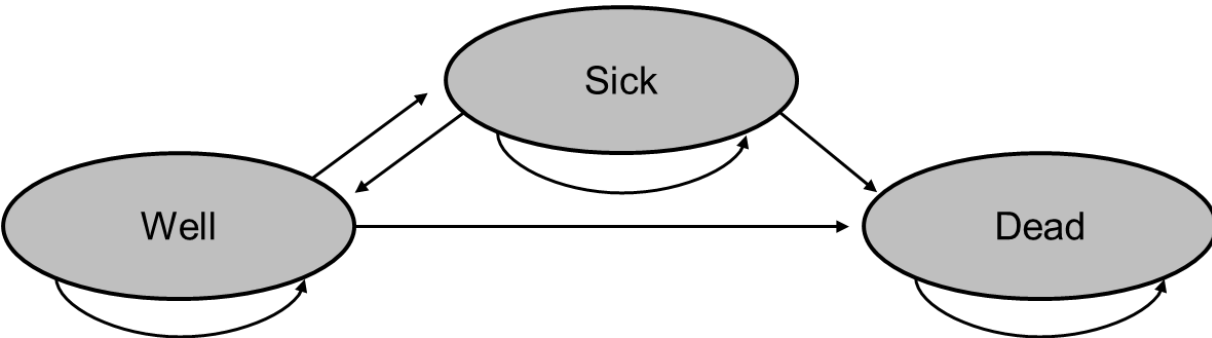


Figure 6. Schematic of a simple state-transition model with the health states well, sick and dead.

Other model types

When the conceptualization of the decision problem requires that individuals interact with each other (e.g., the spread of a disease) or other components of the healthcare system (e.g., resource constraints), model types such as dynamic transmission models (infectious disease models), discrete event simulation and agent-based models may be applied [127]. Dynamic transmission models are characterized by their ability to allow the risk of disease infection to depend on the number of infectious agents at a given point in time, which makes these models suitable for evaluating the direct and indirect (i.e., herd immunity) effects of interventions to prevent communicable diseases (e.g., vaccination programs) [129]. A common approach to dynamic transmission modelling is to represent infection status using compartments, such as the basic SIR (susceptible, infected, recovered) model, which is governed by differential equations. Other model types include discrete event simulation and agent-based models, which allow for interactions between individuals as well as between individuals and the environment or system (e.g., the healthcare system) and is recommended for decision problems characterized by constrained resources (e.g., organ transplants) [98, 130]. These models are characterized by entities (e.g., individuals), their attributes (e.g., age, health status), the events they experience (e.g., disease progression, hospital admission) and the resource they consume (e.g., an operating theatre), queues stemming from occupied resources (e.g., waiting for an available operating theatre), and discrete time intervals [130]. Agent-based models have been referred to as an extension of discrete event simulation to reflect a more complex interaction between entities or ‘agents’ [130]. While these other model types can capture interaction between individuals, they are usually complex in structure and programming, and consequently computationally expensive, which may limit the ability of these models to reflect complex interventions such as screening strategies.

3.2.2 Components of decision-analytic modeling

Following conceptualization of the decision problem and the decision-analytic model, a comprehensive synthesis of available data is required to inform the model structure and input parameters as well as to increase the likelihood of high-quality decisions [98]. The evidence synthesis involves identifying available evidence through literature searches in bibliographic databases and conducting qualitative analysis and/or quantitative synthesis of identified literature. Subsequently, the quality, transferability and risk of bias in the identified evidence should be assessed before summarizing the evidence and estimate model parameters. Some parameters may not be available from empirical data (i.e., unobserved or unobservable) or may only be available in some settings. In these cases, the model can be used to estimate the parameter values through calibration, an iterative process of adjusting input values and assessing results until the model outputs correspond with (i.e., ‘fits’) the observed data [98, 138, 139]. Lastly, decision-analytic modeling involves explicit evaluation of uncertainty, validation and transparency. This section provides an overview of model calibration, uncertainty, validation and transparency.

Model calibration

Model calibration is a stepwise process involving: (i) identifying and estimating calibration targets, (ii) defining a measure of goodness-of-fit, (iii) parameter search, and (iv) defining acceptance criteria and stopping rule for when the process is complete [138]. First, the observed data to which the model should fit (i.e., the calibration targets) must be identified. The calibration targets should be clinically meaningful and/or relevant for decision-making, and should be setting-specific. For example, a model evaluating CC screening policies in Norway could use age-specific estimates of HPV prevalence (for each high-risk HPV type) as a calibration target. Target values can then be estimated by synthesizing evidence similar to the approach for other model input parameters.

The next step involves defining a measure of goodness-of-fit in order to assess how well the model outputs correspond with the observed data [138]. One goodness-of-fit

measure that is easy to implement is ‘acceptable windows’, which involve comparing model outputs to a predefined value range for each target. However, this approach does not capture the degree of closeness to the target value. An alternative approach is to use ‘minimizing deviations’, which captures the degree of closeness to the target value by minimizing the mean percentage deviations (or the sum of squared errors) between the targets and model outputs. Finally, ‘likelihood functions’ can be used to estimate the likelihood of observing the model outputs given the empirical data, which requires information about sample sizes and distributions. For example, the likelihood function, L , for the binomial model is defined by

$$(3) \quad L = Pr(K = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Where:

n = sample size of target data

p = number of events observed in target data divided by n

k = number of events predicted by model for sample size n

L = likelihood of seeing model results in light of observed target data

If multiple calibration targets are used, a combined measure of goodness-of-fit across all targets is required. A mathematically convenient approach to combine likelihood results involves summing across the log-likelihoods (i.e., the logarithm of the likelihood), which can be weighted according to relative importance of each target, if desired.

Following identification of calibration targets and goodness-of-fit measure, the parameter search can be initiated to identify parameter values or parameter sets that produce model outputs that correspond with the calibration targets. This can either be done manually by adjusting the input values (i.e., a trial-and-error approach), or by using an automated search algorithm that select parameter values from pre-specified probability distributions. An automated search can use a ‘grid’ (i.e., grid search) with viable combination of parameters, either sampled randomly using random number generator or more efficiently such as the Latin Hypercube Sampling [138], which involves dividing the probability density function of a parameter into n intervals with equal probability and sampling randomly from each interval. Automated searches can also be directed, such as the Nelder-Mead approach [138].

In order to determine when a particular parameter set provides good-fit with the calibration targets, the acceptance criteria must be defined. For example, one may use the single best parameter set that minimizes or maximizes the goodness-of-fit (depending on which measure is used). Another approach is to estimate a confidence interval for the goodness-of-fit for the best-fitting parameter set; subsequently, all parameter sets with goodness-of-fit within that interval would be considered statistically indistinguishable. Eventually, a stopping rule is used to determine when the calibration process is complete, such as when an acceptable parameter space has been searched or based on convergence of a directed search algorithm. The final step of the calibration process involves integrating the calibration results within the model, either using the best-fitting set or using a number of best-fitting sets as a probabilistic sensitivity analysis (PSA) of the calibrated parameter values. The number of good-fitting sets used in PSA depends on practical considerations such as computation time.

Model uncertainty, validation and transparency

The conceptualization of a decision model (e.g., structure, unit of representation, data sources) requires several assumptions that represent sources of uncertainty in the analysis. In their report on model parameter estimation and uncertainty, the SMDM-ISPOR Modeling Task Force states that *'a model-based analysis' value lies not simply in its ability to generate a precise point estimate for a specific outcome but also in the systematic examination and responsible reporting of uncertainty surrounding this outcome and the ultimate decision being addressed'* [132]. The analyst should therefore characterize and report uncertainty of the analysis to assess the robustness of the results and the value of collecting additional information (e.g., conducting additional clinical trials) [132]. The relevant uncertainty assessment depends on the decision problem; for example, decisions may be based only on expected values when the decision-maker is unable to delay decisions. The ISPOR-SMDM Task Force characterizes uncertainty for decision modeling using four main concepts, including parameter uncertainty, stochastic uncertainty, heterogeneity, and structural uncertainty [132]. Parameter uncertainty stems from estimation of parameters that are

inherently uncertain, such as the probability of experiencing an event, and may be assessed deterministically or using PSA. In deterministic sensitivity analysis (SA), parameter values are varied manually one at a time (one-way SA) or two or more parameters simultaneously (e.g., two-way or multi-way SA) while holding the other parameters constant. PSA involves varying (preferably) all parameters simultaneously by assigning a pre-defined probability distribution (e.g., beta, gamma) to each parameter and sample multiple sets of parameter values. The results of PSA may inform confidence intervals, the probability of each strategy to be cost-effective (i.e., cost-effectiveness acceptability curves), and value of information analyses (e.g., expected value of perfect information). Stochastic uncertainty refers to the overall random variability between patients or patient subgroups, while heterogeneity refers to the variability between patients or patient subgroups that can be attributed to patients' characteristics (e.g., age, sex, vaccination status) [132]. The analyst may undertake separate CEAs for relevant subgroups to inform decisions about subgroup-specific interventions. Finally, structural uncertainty relates to the assumptions of the model structure, such as which health states are included in a state-transition model to represent the natural history of disease. Structural uncertainty is not usually formally quantified but may have a considerable impact on results; for example, a study evaluating the structural uncertainty of Markov models for evaluating treatment of advanced breast cancer found that the cost-effectiveness results differed considerably for different model structures [140].

In addition to a comprehensive assessment of uncertainty, the success of decision-analytic models relies on validation (i.e., how well the model reproduces reality) and transparency (i.e., allowing others insight to model structure and assumptions) [126]. To evaluate the validity of the model, the analyst may consult with experts to confirm the model structure, assumptions, inputs and outputs (i.e., face validity), or compare the model outputs with real-world data not used to inform the model (i.e., external validity). An example of external validation is provided in Section 5.2.2 (i.e., validation of the decision-analytic model used in *Papers II-IV*). The analyst may also compare the results with other models used to inform the same decision problem (i.e., cross validity). One example of comparative modeling is the National Cancer Institute-

sponsored Cancer Intervention and Surveillance Modeling Network (CISNET) consortium of decision-analytic modeling investigators, which aims to inform decisions about cancer control strategies by comparing multiple models for breast, cervical, colorectal, esophageal, lung and prostate cancer (<https://cisnet.cancer.gov/>). Finally, transparent reporting of the model structure and assumptions may involve providing additional technical documentation along with the manuscript. Reporting guidelines, such as the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) guidelines [141] and other checklists and recommendations (e.g., the inclusion of an impact inventory) suggested in economic evaluation textbooks [98, 100], may further help increase the transparency of models and CEAs. Another example is a recent initiative coined ‘HPV-FRAME’, which aims to develop a consensus statement and quality framework for the reporting of model-based analyses evaluating strategies to prevent HPV-related disease (<http://www.hpv-frame.org/>). Together, transparent reporting and explicit evaluation of validity and uncertainty may help improve the usefulness and application of model-based analyses evaluating healthcare interventions.

4 THESIS OBJECTIVES

Novel technologies are changing the landscape of CC screening; new diagnostic tests provide opportunities to improve the effectiveness and cost-effectiveness of the CC screening program. At the same time, CC screening is at a crossroads following implementation of prophylactic HPV vaccination, which gives rise to an increasing heterogeneity of CC risk in the population. These changes are prompting decision-makers to consider whether and how to adopt these new technologies in clinical practice, and to consider adaptations to the screening program in the era of HPV vaccination. In line with policy objectives for prioritizing healthcare in Norway, these decisions should be based on an assessment of the expected benefits and costs of health interventions. Using a decision-analytic framework, the **general aim** of this thesis is to inform decision-makers about the health benefits, resource use and cost-effectiveness associated with candidate CC screening strategies for HPV-unvaccinated and -vaccinated women. Specifically, four research objectives were outlined and assessed in separate papers as follows:

Paper I: To quantify the short-term health and economic outcomes associated with a set of candidate strategies involving novel biomarkers to triage younger unvaccinated women with minor cervical cytological lesions.

Paper II: To compare the long-term health and economic outcomes of alternative strategies involving reflex HPV DNA testing to triage unvaccinated women with minor cervical cytological lesions.

Paper III: To evaluate the trade-offs in health benefits and resource use associated with adopting primary HPV DNA testing strategies for unvaccinated women.

Paper IV: To identify the most cost-effective CC screening strategies for women vaccinated against HPV-infections in adolescence, and the value of stratifying screening guidelines according to HPV vaccination status.

5 MATERIALS AND METHODS

The papers included in this thesis employed a decision-analytic approach to evaluate the health benefits, resource use and cost-effectiveness associated with current and future CC screening policies in Norway, as outlined in the four thesis objectives in **Chapter 4**. This chapter first provides an analytic overview of *Papers I-IV*, followed by a description of the decision-tree model used in *Paper I* and the microsimulation model used in *Papers II-IV*. The chapter subsequently reviews the HRQoL weights used to estimate QALYs in *Papers II* and *IV*, and summarizes the costing approach and the cost estimates used in *Papers I, II* and *IV*. Finally, the chapter provides an overview of the comparator screening strategies, analyses and assumptions in *Papers I-IV*. Details of the model structure, calibration, validation, epidemiologic data and costing approach are further provided in the technical appendices of *Papers I-IV* (**Chapter 10**).

5.1 Analytic overview

The papers included in this thesis evaluated screening strategies considered relevant for both current (*Papers I-II*) and future (*Papers III-IV*) CC screening policies in Norway, and included both primary (*Papers III-IV*) and secondary (*Papers I-III*) screening algorithms with candidate screening tests (**Table 2**). The target populations included: women aged 25-33 years detected with minor cervical lesions on their primary cytology and not vaccinated against HPV infections (*Paper I*), unvaccinated women aged 25-69 years (*Papers II* and *III*), and women aged 25-69 years who were fully vaccinated against vaccine-targeted HPV infections in adolescence (*Paper IV*). *Papers I* and *II* evaluated the health benefits, resource use and cost-effectiveness associated with strategies involving novel biomarkers in secondary screening for unvaccinated women with minor cervical cytological lesions: (i) within a single screening round (*Paper I*) and (ii) over a lifetime (*Paper II*). *Paper III* evaluated the health benefit and resource use trade-offs associated with primary HPV testing (i.e., the proposed guidelines for future CC screening in Norway) for a cohort of unvaccinated women over their lifetime. Finally, *Paper IV* evaluated the health

benefits, resource use and cost-effectiveness associated with potential future screening guidelines for a cohort of HPV-vaccinated women over their lifetime.

Table 2. Main elements of *Papers I-IV*.

Element	Paper I	Paper II	Paper III	Paper IV
Target population	Women aged 25-33 years w/minor cervical lesions*	Women aged 25-69 years	Women aged 25-69 years	Women aged 25-69 years
HPV vaccination status of the target population	Unvaccinated	Unvaccinated	Unvaccinated	Vaccinated
Intervention: Current/future guidelines	Current	Current	Future	Future
Screening setting	Secondary	Secondary	Primary and secondary	Primary
Screening test(s)	Alternative biomarkers	HPV DNA test	HPV DNA test	Cytology and HPV DNA test
Time horizon	3 years (one screening round)	Lifetime	Lifetime	Lifetime
Discount rate (per year)	Base-case: 0% SA: 4%	Base-case: 4% SA: 0%	Base-case: 0%	Base-case: 4% SA: 0%
Costing year	2014	2014	NA	2014
Decision-analytic model	Decision-tree	Microsimulation state-transition	Microsimulation state-transition	Microsimulation state-transition
Type of analysis	CEA	CEA	Consequence analysis	CEA
Analytic outcomes	Cost per woman CIN2+ detected GP consultations Colpo referrals Preterm births	Cost per woman QALYs Life expectancy CC incidence Screening tests Colpo referrals Treatments	CC incidence Colpo referrals Screening tests Treatments	Cost per woman QALYs Life expectancy CC incidence Screening tests Colpo referrals

* In contrast to *Papers II-IV*, the analysis in *Paper I* only considered outcomes for women with an index result indicating minor cervical lesions, and not women with cytology results indicating no cervical abnormalities or high-grade cervical lesions.

Abbreviations: CC, Cervical cancer; CEA, cost-effectiveness analysis; CIN2+, Cervical intraepithelial neoplasia grade 2 or worse; Colpo, Colposcopy; HPV, human papillomavirus; GP, General practitioner; NA, Not applicable; QALY, quality-adjusted life years.

Papers I, II and IV adopted a societal analytic perspective and discounted costs and health benefits by 4% per year, as outlined in the Norwegian guidelines for economic

evaluation [64]. Of note, *Paper I* presented undiscounted results in the primary analysis and discounted results in uncertainty analysis due to the short time horizon. *Papers II* and *IV* evaluated undiscounted outcomes in uncertainty analysis. Costs were valued in 2014 Norwegian Kroners and converted to Euros (in *Paper I*, €1 = 8.35 [142]) or US \$ (in *Papers II* and *IV*, \$1 = 6.30 [113]) using the average annual exchange rates for 2014. Analytic outcomes reflecting resource use included the number of GP consultations (*Paper I*), screening (i.e., cytology and HPV) tests (*Papers II-IV*), colposcopy referrals (*Papers I-IV*), precancer treatments (*Papers II-III*), and the total cost per woman (*Papers I, II* and *IV*). Analytic health outcomes included the number of precancers detected and preterm birth rates (*Paper I*), CC incidence (*Papers II-IV*), life expectancy and QALYs (*Papers II* and *IV*). The Second Panel recommended presenting an impact inventory to clarify the inclusion of health and economic consequences in the analysis. To aid transparency, and as an example of how the impact inventory could look like for *Papers II* and *IV* under these new guidelines for reporting of CEAs, the relevant long-term consequences of CC screening within and outside the healthcare sector was retrospectively identified and displayed in **Table 3**.

A cost-effectiveness framework was used to identify cost-efficient and cost-effective strategies in *Papers I, II* and *IV*, while *Paper III* involved quantifying health and economic outcomes as a type of consequence analysis. In *Papers I, II* and *IV*, cost-efficient algorithms were identified using the ICER. In *Paper I*, the ICER was defined as the additional cost per additional precancer detected, and the average cost per detected precancer associated with the current Norwegian strategy was used as a proxy for the willingness-to-pay threshold (to detect one additional precancer). In *Papers II* and *IV*, the ICER was defined as the additional cost per additional QALY, and the commonly cited Norwegian willingness-to-pay threshold of a \$100,000 per QALY gained [64, 112] was used as a benchmark to identify the most cost-effective strategy. Finally, for *Paper IV*, in addition to identifying the optimal screening strategies for HPV-vaccinated women, the INMB was used to estimate the efficiency gains of stratifying CC screening according to HPV vaccination status (see Supplementary Appendix for *Paper IV* for further details).

Table 3. Impact inventory for *Papers II* and *IV*.*

Sector	Type of impact (within each sector with unit of measurement if relevant)	Included in this analysis from ...perspective?			Notes on sources of evidence
		Healthcare sector	'Restricted societal'†	Societal†	
Formal Health Care Sector					
Health	Health outcomes (effects)				
	Longevity effects	✓	✓	✓	See methods of <i>Papers I</i> and <i>IV</i>
	HRQoL effects	✓	✓	✓	
	Cancer incidence reduction	✓	✓	✓	
	Medical costs				
	Paid for by third-party payers	✓	✓	✓	See methods of <i>Papers I</i> and <i>IV</i>
	Paid for by patients out-of-pocket	✓	✓	✓	
	Future related medical costs	✓	✓	✓	
Future unrelated medical costs	✗	✗	✗		
Informal Health Care Sector					
Health	Patient-time costs	NA	✓	✓	See methods of <i>Papers I</i> and <i>IV</i>
	Unpaid caregiver-time costs	NA	✗	✗	See discussion
	Transportation costs	NA	✓	✓	See methods of <i>Papers I</i> and <i>IV</i>
Non-Health Care Sectors					
Productivity	Cost of lost production (paid and unpaid labor)	NA	✗	✓	See methods of <i>Papers I</i> and <i>IV</i>
	Cost of uncompensated household production	NA	✗	✗	
Consumption	Future consumption unrelated to health	NA	✗	✗	
Social services	Social services as part of intervention	NA	✗	✗	
Legal/criminal justice	None	NA			
Education	None	NA			
Housing	None	NA			
Environment	None	NA			

* Adapted from Neumann et al. [98].

† The perspectives labelled 'restricted societal' and 'societal' reflect the recommendations of the Norwegian guidelines for economic evaluation [64]. In *Papers I, II* and *IV*, these perspectives were referred to as 'societal' and 'expanded societal', respectively.

5.2 Decision-analytic models

The choice of model type (see Section 3.2.1) should be based on the characteristics of the decision problem and the disease in question. A decision-tree model was utilized to evaluate the short-term consequences of candidate biomarkers (*Paper I*) due to the interest from the decision-makers to evaluate consequences within a single screening round in order to assess short-term feasibility. As health and economic outcomes are extrapolated beyond a single screening round (e.g., over a lifetime), a greater accuracy of the underlying natural history of disease model is required. Therefore, to evaluate long-term consequences of candidate screening strategies, *Papers II-IV* utilized a microsimulation state-transition model reflecting health states of HPV and cervical carcinogenesis (i.e., from acquisition of an HPV infection and progression to precancer and cancer depending on HPV type and infection persistence), which also allowed tracking of individual expenditure and clinical events over a lifetime. This section provides an overview of the models used in *Papers I-IV*.

5.2.1 Decision-tree model for novel biomarkers (*Paper I*)

The decision-tree model was initially conceptualized together with Norwegian policy-makers to inform an ongoing process of adapting the triage algorithm for women with minor cervical lesions (i.e., ASC-US or LSIL) to incorporate reflex HPV DNA testing [143]. For *Paper I*, the model was extended to include screening triage strategies that involved novel biomarkers such as HPV DNA testing with HPV-16 and -18 genotyping ('genotyping'), HPV mRNA testing, and p16/Ki67 dual staining. The model was adapted to simulate a hypothetical cohort of women aged between 25 to 33 years and detected with minor cervical lesions (i.e., ASC-US or LSIL) at their primary cytology through one subsequent screening round (i.e., 3 years). A total of 13 strategies were included in the model, which broadly varied by the triage test (i.e., HPV DNA testing with or without genotyping, HPV mRNA testing, or p16/Ki67 dual staining), immediate or delayed follow-up, and criteria for prompting return to routine screening (see Section 5.5 for summary of comparative strategies). Starting with an index cytology result of ASC-US/LSIL, each pathway in the model reflects a single 3-year screening

history with consecutive test results, waiting time and test outcomes, and terminates in the event of loss-to-follow-up, detection of high-grade precancer, or the end of 3-years. To ensure consistent passage of time throughout the pathways, time units of a month was integrated to the model to capture time between screening tests and test outcomes.

The simulated cohort progress through the model according to empirical positivity rates, which were estimated from primary Norwegian epidemiologic data and published literature from European and North-American clinical trials. Primary data from Norway included observed age-specific positivity rates from the Norwegian CC Screening Program and nationwide opportunistic HPV mRNA testing during 2003-2004 (i.e., prior to implementation of HPV testing in the triage algorithm in 2005, thus not performed as part of the official screening guidelines). Literature review and evidence synthesis of data from CC screening clinical trials in Europe and North America informed parameters that could not be estimated from Norwegian-specific data.

The natural history of disease was reflected using the composite outcome of CIN2+. Given the short 3-year time horizon of the analysis, the model did not allow for progression from no lesion or CIN1 to CIN2+, yet allowed cases of CIN2+ to regress to CIN1 or no lesion at a monthly probability of 2% [144]. The baseline prevalence of CIN2+ was estimated using calibration. The calibration target was defined as the 3-year cumulative incidence of CIN2+ among women detected with ASC-US/LSIL on their primary cytology in Norway, which was estimated to be 19.7% using primary data from the Cancer Registry of Norway. The goodness-of-fit measure was defined as 'acceptable windows' with a range of plus-minus 15% around the target value as an acceptance criterion (i.e., 16.7-22.6% cumulative incidence of CIN2+). A trial-and-error search approach was used until an acceptable value was identified, which resulted in an estimated baseline prevalence of CIN2+ (among women with ASC-US/LSIL) of 29%.

The model was validated internally by cross-checking equations and inputs against their sources and using the software's (i.e., TreeAge Pro) built-in debugging tools. Face validity of clinical assumptions and model inputs was confirmed with Norwegian

experts. Finally, external validation was performed by comparing study results against published literature not used during calibration.

5.2.2 Microsimulation model of cervical carcinogenesis (*Papers II-IV*)

The microsimulation state-transition model (herein referred to as the ‘microsimulation model’) used in this thesis [145] has been previously developed and updated as new knowledge of HPV and cervical carcinogenesis have emerged [146-148]. The model has been adapted to evaluate CC prevention policies in several settings, such as the US [83, 85, 149-151], Norway [63, 152, 153] and other European countries [154], as well as developing countries [155-157]. The most recent version of the model [145] (used in *Papers II-IV*) was informed by leading HPV epidemiologists to update the previous model version [146] in accordance with current understanding of HPV and CC epidemiology. Specifically, the model was updated to reflect: (i) progression and clearance rates that depend on duration of HPV infection rather than age, (ii) a woman can be co-infected with multiple HPV types, and (iii) precancer was represented as two separate health states (i.e., CIN2 and CIN3) that are non-sequential with respect to cancer progression. The model comprises of health states reflecting HPV-induced cervical carcinogenesis, including no HPV infection, HPV infection status (stratified by HPV types -16, -18, -31, -33, -45, -52, -58, pooled other high-risk HPV types, and pooled low-risk types), cervical precancer (defined as CIN2 or CIN3), and CC (i.e., squamous cell carcinoma by stages local, regional and distant) (**Figure 7**). Using microsimulation, individual women enter the model at age 9 years with no HPV infection and progress through the model at monthly transitions. The model tracks individuals’ disease history (e.g., HPV infections, of which multiple can occur concurrently), clinical events (e.g., cancer incidence and life expectancy), and resource use from interventions, which overlay the natural history model (i.e., screening and/or vaccination). The model reflects age-specific acquisition of HPV infections, and regression/clearance and progression of an HPV infection or a precancerous lesion is governed by duration of the infection or lesion. Women with CC can have their cancer detected through screening or from symptoms, or progress to more advanced stages if the cancer is left undetected. Women are simulated until death, which can occur from

non-cervical causes from any health states based on life-tables from Norway [158], or from CC (i.e., stage-specific excess mortality).

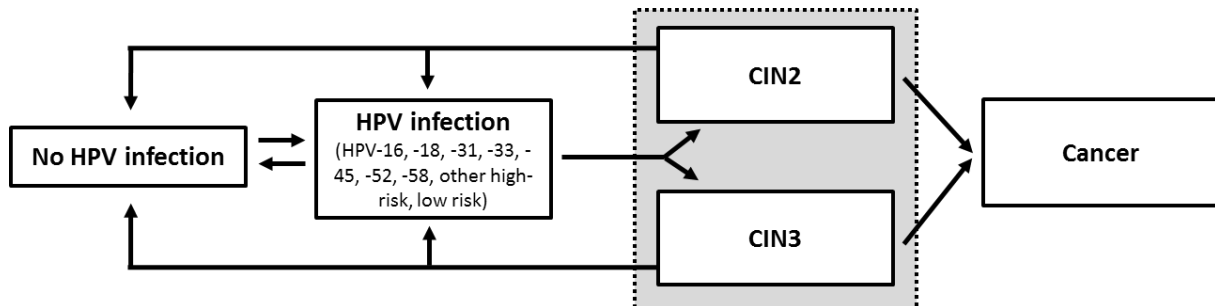


Figure 7. Schematic of the microsimulation model used in *Papers II-IV*.

Assuming that the underlying mechanism of cervical carcinogenesis does not vary across settings, the baseline parameter inputs for the natural history parameters were estimated from large epidemiologic studies and described in detail elsewhere [145]. However, as HPV and CC epidemiology (e.g., HPV prevalence, CC incidence) differ between countries, setting-specific data are needed to adjust baseline natural history inputs to ensure the model corresponds with observed epidemiologic data in the relevant setting. The initial likelihood-based calibration to the Norwegian setting for the previous model version [146] has been described elsewhere [63]. The model was recalibrated for *Papers II-IV* to reflect changes in the model structure since the initial calibration [145, 146]. As calibration is an iterative process following the accumulation of data and knowledge, the model was first calibrated for *Paper II* and subsequently recalibrated for *Papers III and IV*.

Specifically, for *Paper II*, we identified 63 calibration targets that included age-specific prevalence of high-risk HPV-16, -18, -33, -45, -52 and -58 (for ages 18-19 and ages 20-49 years with 5-year age-groups), the distribution of the respective high-risk HPV-types in precancers (i.e., CIN3) and squamous cell carcinoma. For *Papers III-IV*, seven additional targets for age-specific prevalence of pooled other high-risk HPV (i.e., HPV-35, -39, -51, -56 and -59) infections were included as these types are relatively more important when evaluating interventions among HPV-vaccinated women. Age-specific

HPV prevalence targets were estimated using preliminary data from a study affiliated with the Cancer Registry of Norway that included a random sample of 1,818 women aged 18-49 years with a normal cytology result, who attended screening in Trondheim in 2007 [63]. Estimates of HPV genotype-distribution in CIN3 and SCC were based on Norwegian data from the HERACLES and SCALE studies [63, 159], which included 178 women with CIN3 and 207 women with squamous cell carcinoma who were HPV DNA positive and aged 18 years or older.

Likelihood functions for the binomial model were used to measure the goodness-of-fit for each target, of which the sum of individual log-likelihood measures for all targets was used to calculate a composite goodness-of-fit score for each set of input parameters. The calibrated parameters included age- and type specific incidence of HPV infection, natural immunity following type-specific HPV infection, and disease progression. The search for parameter sets was based on random draws from a uniform distribution (with the minimum and maximum value reflecting a plausible range for the specific parameter based on empirical data). The random search process was used to create a repository of ~1 million parameter sets, which constituted the search space for calibration. Subsequently, the goodness-of-fit with the empirical data was calculated for each parameter set. To define the 'good-fitting' sets, a cut-off value was defined as the score of the last parameter set that was in the top 5% of a Chi-squared distribution with the number of degrees of freedom equal to the number of targets. **Figure 8** provides an example of model fit with empirical calibration targets (i.e., HPV type-distribution in CIN3 and cancer) using the top 10 good-fitting sets (model fit with all calibration targets have been presented in the technical appendices of the papers). The top 50 good-fitting parameter sets were selected to reflect the uncertainty in the natural history of disease.

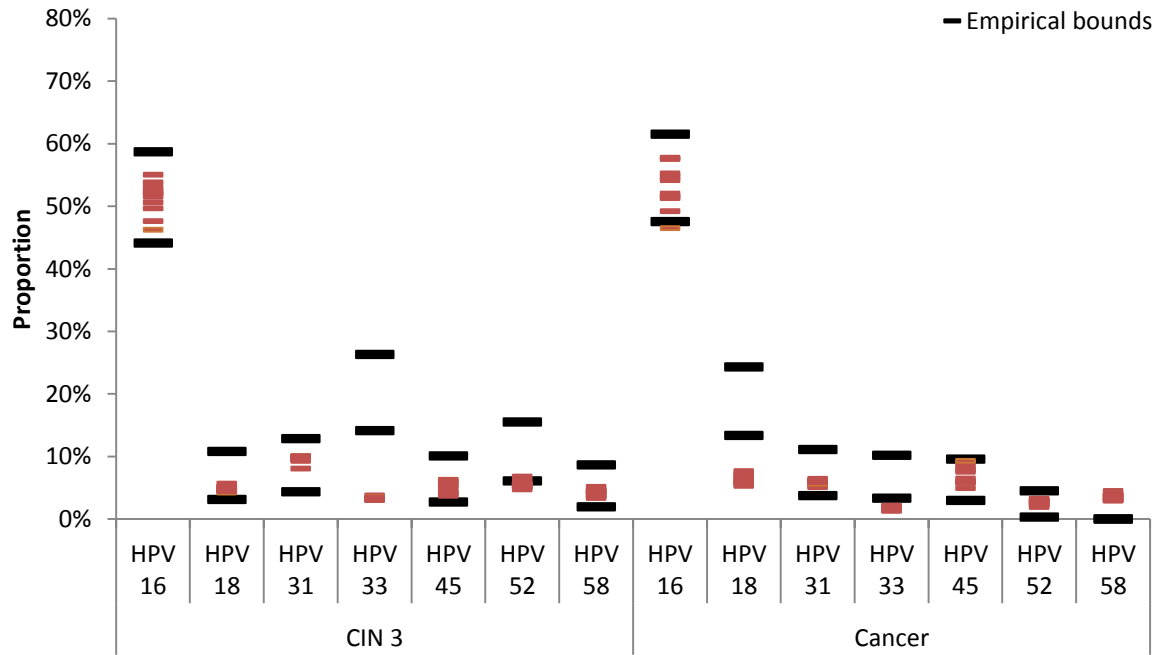


Figure 8. Calibration output for HPV type-distribution in CIN3 and CC: Model output from the 10 best-fitting sets (red lines) and the upper and lower bound (black bold lines) estimated from the empirical data.

Following model calibration in the absence of interventions, we assessed the external validity of the model by comparing model-projected CC incidence with observed age-specific incidence of squamous cell carcinoma in Norway during 2010-2014 [160]. In an effort to inform the screening compliance assumptions for model validation, population-based data from the Cancer Registry of Norway (including the screening histories of ~1.4 million women in Norway during years 1992-2013) was used to estimate longitudinal adherence to repeated screening intervals. The materials and methods for estimation of screening behavior has been described in detail elsewhere [58]. Briefly, women were categorized into five categories of longitudinal screening adherence based on average screening interval length over the time they were eligible for CC screening, which included never-screeners, severe under-screeners, moderate under-screeners, guidelines-based screeners and over-screeners. For model validation, we used the distribution of longitudinal adherence categories to determine the proportion of women with a specific screening frequency (i.e., every 1-10 years, 15-yearly, 20-yearly, or non-attender) (Table 4), as well as observed compliance to follow-up procedures following an abnormal screening result [57]. The model output was in

reasonable correspondence with observed Norwegian data (**Figure 9**). For example, following calibration for *Papers III and IV*, the model projected an annual incidence of squamous cell carcinoma incidence of 227 (range across the 50 good-fitting parameter sets: 146-271) cases for an average birth cohort of ~30,000 women in Norway, while an average of 223 cases were observed annually in Norway during 2010-2014.

Table 4. Derivation of screening frequency distribution used for model validation. Adapted from Pedersen et al [58].

Categories of longitudinal screening behavior		Over-screener	Guidelines-based	Moderate under	Severe under	Never	Screening frequency distribution in validation exercise:	
Estimated proportion of longitudinal screening behavior		27 %	19 %	29 %	17 %	8 %		
Screening frequency (years)	1	2 %					2 %	
	2	5 %					5 %	
	3	20 %	10 %				30 %	
	4		9 %	13 %			22 %	
	5			9 %			9 %	
	6			7 %			7 %	
	7				6 %		6 %	
	8				4 %		4 %	
	9				3 %		3 %	
	10				2 %		2 %	
	15				1 %		1 %	
	20				1 %		1 %	
	Never screen						8 %	8 %
	Sum		27 %	19 %	29 %	17 %	8 %	100 %

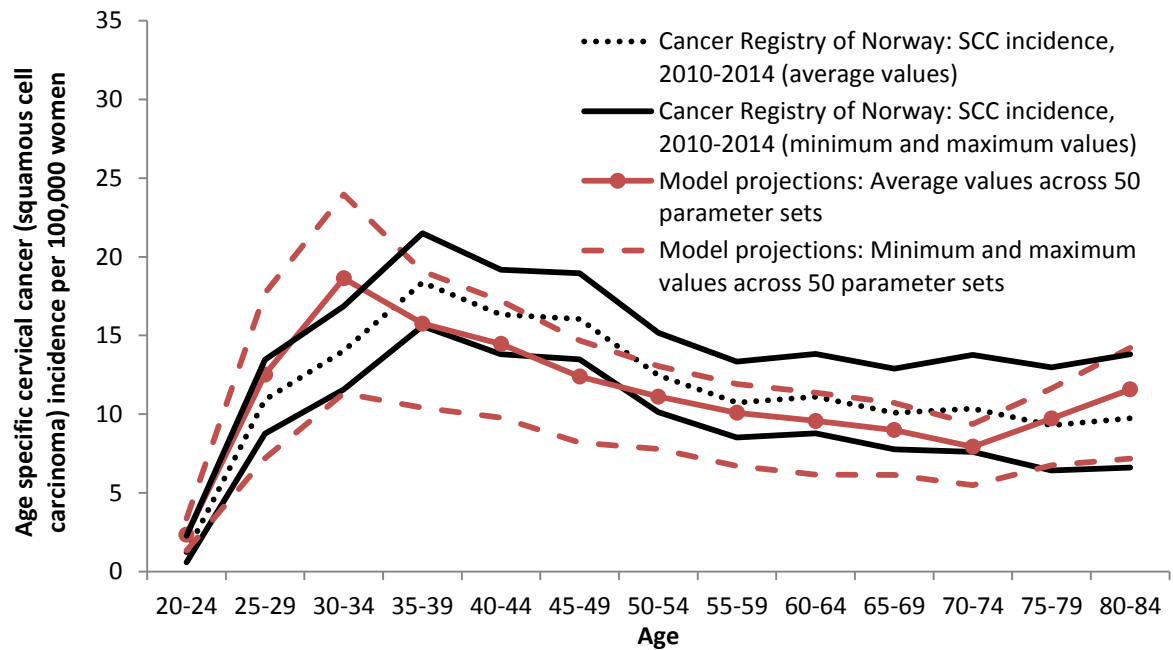


Figure 9. Age-specific cervical cancer incidence in Norway during 2005-2014 from the Cancer Registry of Norway (black line) and model output from the 50 good-fitting sets (red lines), with mean (solid lines) and minimum and maximum (dashed lines) values when assuming imperfect adherence to screening guidelines. Adapted from the technical appendix for *Papers III and IV*.

5.3 Health-related quality of life

As studies to inform the HRQoL of CC-related health states are scarce, the utility weights used to adjust for HRQoL (*Papers II and IV*) were based on estimates reported in a previous model-based study that evaluated the cost-effectiveness of expanding the HPV vaccination program to include adolescent boys within the Norwegian setting [161]. In this study, age-specific HRQoL weights for women with no CC were based on estimates for the general female population in Denmark (a neighboring Scandinavian country) [162] (**Table 5**). For women with CC, these HRQoL weights were adjusted using an adjustment factor (i.e., a multiplicative approach) that varied by cancer stage (**Table 5**). The utility decrements were informed by a study that elicited utilities using the time trade-off method [163]. Scandinavian studies found that the quality of life for survivors of gynecological cancer was similar to the general population [164, 165]; therefore, the utility decrement associated with CC was assumed to endure for 5 years.

Table 5. Age-specific HRQoL weights for the general female population in Denmark [162] and HRQoL adjustment factor for health states [161, 163].

Age group	HRQoL	Health state	HRQoL adjustment
<20	1	No cancer	1
20-29	0.9203	Local cancer	0.76
30-39	0.9118	Regional cancer	0.67
40-49	0.8763	Distant cancer	0.48
50-59	0.8499	Dead	0
60-69	0.8552		
70-79	0.8320		
80+	0.6919		

5.4 Costing

For *Papers I, II and IV*, cost estimates were based on a cost study of the Norwegian CC screening program [62] and a CEA of CC screening in Norway [63], which were updated for analyses performed in this thesis to reflect 2014 NOK. In all papers, the base-case costing assumptions reflected the cost components recommended for the societal analytic perspective as outlined in the Norwegian guidelines for economic evaluation [64]. Cost components for the societal perspective included medical costs (i.e., screening and diagnostic consultations, pathology analysis of test samples, and treatment of precancer and cancer), time costs of patients (i.e., time spent on travel to healthcare facilities, waiting time at the facility, and the time spent receiving the healthcare) and transportation costs (**Table 6**). Of note, this perspective does not include productivity costs and therefore represents a ‘restricted’ societal perspective (see Section 3.1.5). Therefore, in *Papers II and IV*, two additional sets of costing assumptions were explored in uncertainty analysis. First, an ‘expanded societal perspective’, which included the base-case cost components as well as productivity costs associated with sick leave after precancer and cancer treatment (i.e., more closely reflecting the societal reference case recommended by the Second Panel and NICE). Second, a healthcare perspective, which restricted the cost components to include medical costs only. Details about the costing are provided in the technical appendices of *Papers I, II and IV*.

Table 6. Cost estimates used in the different analytic perspectives for *Papers I, II and IV*.*

Cost component	Cost estimates used in the different analytic perspectives		
	Health Care	'Restricted societal'	Societal'
Formal Health Care Sector			
<i>Screening and diagnostic consultations</i>			
GP office visit	\$122	\$122 (61-244)	\$122
Colposcopy examination	\$258	\$258 (129-516)	\$258
<i>Analyzing test sample at pathology laboratory</i>			
Liquid-based cytology	\$45	\$45 (22-89)	\$45
HPV DNA test	\$39	\$39 (20-79), \$90†	\$39
HPV mRNA test (5 types)	--	\$83	--
HPV mRNA test (14 types)	--	\$93	--
PI6/Ki67 dual staining	--	\$126	--
Cervical biopsy	\$124	\$124 (62-247)	\$124
<i>Treatment of precancer and cancer (Papers II and IV only)</i>			
Precancer (CIN2 or CIN3)	\$1,209	\$1,209	\$1,209
Local cancer	\$24,283	\$24,283	\$24,283
Regional cancer	\$44,592	\$44,592	\$44,592
Distant cancer	\$29,005	\$29,005	\$29,005
Informal Health Care Sector			
<i>Patient-time costs</i>			
GP office visit	NA	\$88	\$88
Colposcopy examination	NA	\$118	\$118
Precancer treatment ‡	NA	\$441	\$441
Local cancer treatment §	NA	\$2,015	\$2,015
Regional cancer treatment ¶	NA	\$10,064	\$10,064
Distant cancer treatment ¶¶	NA	\$9,860	\$9,860
<i>Transportation costs</i>			
GP office visit/ colposcopy	NA	\$32	\$32
Precancer treatment ‡	NA	\$32	\$32
Local cancer treatment §	NA	\$643	\$643
Regional cancer treatment ¶	NA	\$1,945	\$1,945
Distant cancer treatment ¶¶	NA	\$2,502	\$2,502
Non-Health Care Sectors (productivity costs associated with sick leave after treatment)			
Precancer (CIN2 or CIN3) ‡	NA	NA	\$3,091
Local cancer §	NA	NA	\$12,915
Regional cancer ¶	NA	NA	\$105,129
Distant cancer ¶¶	NA	NA	\$105,204

* Costs are valued in 2014 Norwegian Kroner and converted to US \$ (\$1 = NOK6.30 [113]). The cost estimates in *Paper I* were expressed in Euros (€1 = 8.35 [142]) but are presented in US \$ in this table for consistency and comparability. *Paper I* assigned gamma distributions with 20% uncertainty around the mean estimates. *Paper I* adopted a 'restricted societal' perspective. In *Papers II and IV*, the base-case analysis reflected the restricted societal perspective (the values evaluated in one-way sensitivity analysis are indicated in parenthesis) while healthcare and societal perspectives were explored in uncertainty analysis. The 'restricted societal' and 'societal' perspectives reflect the recommendations of the Norwegian CEA guidelines [64]. In *Papers I, II and IV*, these perspectives were referred to as 'societal' and 'expanded societal', respectively. The table layout is adapted from Table 1 in Sanders et al [106].

† *Paper I* assumed a unit cost of HPV DNA testing of \$90 to reflect the cost of the test if used solely within a triage setting. *Papers II and IV* assumed a unit cost of \$45 to reflect application within both primary and triage testing (which was also evaluated in one-way uncertainty analysis for *Paper I*).

‡ The total cost of precancer treatment within the restricted societal perspective was \$1,682 (range in one-way uncertainty analysis: \$841-3,364), while the total cost was \$4,773 within the societal perspective.

§ The total cost of local cancer treatment within the restricted societal perspective was \$26,941 (range in one-way uncertainty analysis: \$13,471-53,882), while the total cost was \$39,856 within the societal perspective.

¶ The total cost of regional cancer treatment within the restricted societal perspective was \$56,601 (range in one-way uncertainty analysis: \$28,301-113,202), while the total cost was \$161,730 within the societal perspective.

¶¶ The total cost of regional cancer treatment within the restricted societal perspective was \$41,367 (range in one-way uncertainty analysis: \$20,684-82,735), while the total cost was \$146,571 within the societal perspective.

5.5 Comparator screening strategies

The screening strategies considered in *Papers I-IV* varied by primary screening test, screening start age and frequency, and triage screening approaches (**Table 7**). *Papers I and II* focused on improvements within the current cytology-based screening program, and hence only varied the triage algorithm for women with minor cervical lesions. *Papers III and IV* focused on future CC screening policies in Norway, and varied both primary and triage algorithm levers. For some novel biomarkers, there is limited information available about the interaction between the diagnostic test and the long-term underlying natural history of disease, thus evaluation of these biomarkers was restricted to *Paper I* (which used a decision-tree model with a short-term time horizon).

In *Paper I*, the target population included only women detected with ASC-US/LSIL, and the baseline comparator strategy included both the current (i.e., since July 2014) and former (i.e., years 2005-July 2014) triage algorithms in Norway. For *Papers II-IV*, the baseline comparator strategy included current guidelines in Norway and a no-intervention scenario (i.e., no screening or vaccination). *Paper IV* also included a strategy involving HPV vaccination only.

Table 7. Overview of algorithm levers considered and/or varied in screening strategies under evaluation in *Papers I-IV*.

Strategy lever	Paper I	Paper II	Paper III	Paper IV
Primary screening test	Cytology	Cytology	Cytology with switch to HPV, HPV only	Cytology w/wo switch to HPV, HPV only
Screening start age	25	25	25	25, 28, 31, 34
Primary cytology interval (years)	3	3	3	3, 5, 7, 10, 15, 20
Age to switch to HPV testing	Not considered	Not considered	28, 31, 34	34
Primary HPV interval (years)	Not considered	Not considered	3, 4, 5, 6, 8, 10	3, 5, 7, 10, 15, 20
Primary HPV testing 1-2 times per lifetime	Not considered	Not considered	Not considered	1 or 2 screens (15-years apart) starting at age 25, 30, 35 or 40 years
Triage population	ASC-US/LSIL	ASC-US/LSIL and reflex HPV-positive	HPV-positive/ Cytology-negative	Not considered
Triage approaches considered	-Reflex HPV DNA w/wo genotyping -Delayed co-test -Diff. management -Reflex HPV mRNA (5 and 14 gt) -HPV DNA & mRNA -p16/Ki67	Repeat testing in 6, 12 or 18 months: -Co-testing -HPV DNA for all -HPV DNA only for HPV-16/-18 neg (direct colpo for HPV-16/-18 pos)	Repeat testing in 6, 12 or 18 months with HPV DNA; Requiring 1, 2 or 3 repeat HPV-positive results to prompt colpo referral	Not considered
Number of screening strategies in primary analysis	13	10	216	74

Abbreviations: ASC-US, atypical squamous cells of undetermined significance; Colpo, Colposcopy; Diff, differential; gt, genotypes; HPV, human papillomavirus; LSIL, low-grade squamous intraepithelial lesion; neg, negative; pos, positive; Refl, reflex; w/wo, with/without.

5.6 Assumptions and analyses

5.6.1 Paper I

In the primary analysis, *Paper I* assumed imperfect compliance with recommended follow-up within the 3-year interval (i.e., 98.5% compliance to colposcopy examination and 90.0% compliance with other follow-up tests, reflecting observed compliance rates in Norway). Test characteristics for the diagnostic tests were based on meta-analyses and independent studies (reported in Table 1 in *Paper I*). The primary analysis was evaluated using probabilistic sampling of input parameters; analytic outcomes reflect the average value across 1,000 samples. Probabilities, costing parameters and time consumption parameters were assumed beta-, gamma- and Poisson-distributed, respectively. Beta distributions were informed by the number of individuals who

experienced an event among the total study population, or using the mean estimate with 95% confidence interval bounds if the study did not report total number of individuals in the study population. Empirical uncertainty around the mean cost estimates were not available, thus 20% was assumed.

The impact of key model parameters on the cost-effectiveness results were evaluated in deterministic uncertainty analyses. In one-way sensitivity analysis, the lower bounds of test characteristics reported in the studies used to inform the sensitivity parameters (Tables 1 and SI in *Paper I*) were evaluated for the diagnostic sensitivity (to detect CIN2+) of cytology, HPV DNA test, colposcopy w/biopsy, HPV mRNA testing with 5/14 genotypes, and dual staining. Multi-way sensitivity analysis explored the impact on results of assuming the lower bound value of diagnostic sensitivity for all diagnostic tests simultaneously. In addition to varying assumptions around screening test characteristics, the uncertainty analyses included one-way SA of screening compliance assumptions (i.e., 70% and 100% compliance to both screening tests and colposcopy exams), the baseline prevalence of CIN2+, the monthly regression probability of CIN2+, and the patient time and travel costs associated with attending a physician consultation.

5.6.2 Papers II-IV

In *Papers II-IV*, the primary and secondary analyses assumed 100% compliance to screening and follow-up procedures. In *Papers II* and *IV*, imperfect follow-up compliance was evaluated as a scenario uncertainty analysis assuming 72.3% compliance to follow-up testing, 82.8% compliance to colposcopy with biopsy, and 97% compliance to precancer treatment, based on observed compliance rates from the Cancer Registry of Norway [57]. In *Paper IV*, the imperfect screening behavior scenario also assumed that only 80% of screen-eligible women attended primary screening.

The primary analyses in *Papers II-IV* assumed a 70% sensitivity (91% specificity) of cytology to detect (exclude) ASC-US or more severe given presence (absence) of CIN2+ and a 86% sensitivity (89% specificity) of colposcopy with biopsy to detect (exclude) CIN2+ given presence (absence) of CIN2+. The sensitivity (specificity) of an HPV DNA

test is defined as the probability of HPV DNA-positive (-negative) given HPV DNA is present (absent) of 100%; thus, the sensitivity (and specificity) is modelled as the ability of the HPV test to detect the presence (or absence) of HPV infection. Clinical HPV sensitivity (and specificity) for detecting presence (and absence) of CIN2+ is a model output. While the model assumes that high-risk HPV is a necessary condition for progression to cancer, it also accounts for high-grade precancers attributable to low-risk HPV that may be detected by cytology but would not be detected by high-risk HPV testing. The model calculated sensitivity of HPV testing reflects that high-grade lesions will be missed due to low-risk HPV types; but in terms of progression to cancer, HPV testing will detect these clinically important high-risk infections. For example, in *Paper III*, the “implied” HPV test sensitivity and specificity for detecting lesions was 93% and 80%, respectively.

The analytic outcomes presented in the primary, secondary and uncertainty analyses reflect the average value across the top-fitting 50 natural history parameter sets (see Section 5.2.2), while the minimum and maximum value across the 50 sets represent uncertainty bounds. Model simulations included a cohort of 1 million individual women for each parameter set. As the probability of experiencing a disease event (e.g., HPV infection, CC) is lower for vaccinated women than for unvaccinated women, the cohort size was increased to 4 million women for analyses for *Paper IV* to ensure stability of results. Due to the computation time required to simulate 4 million women, the uncertainty analyses in *Paper IV* used a single parameter set that represents the average parameter input values across all 50 parameter sets.

6 SUMMARY OF RESULTS

6.1 Paper I

For women detected with ASC-US/LSIL on their primary cytology, both the current (i.e., since 2014) and former (i.e., years 2005-2014) triage guidelines in Norway detected fewer cases of CIN2+ and were more costly than candidate strategies. Among the 11 alternative triage strategies, five were identified as cost-efficient with ICERs ranging from €2,978-33,095 per additional CIN2+ detected. These strategies included reflex HPV mRNA testing with five or 14 genotypes, reflex HPV DNA testing with direct colposcopy for HPV-positive women (and either return to screening or repeat HPV test in 12 months for HPV-negative women), and a strategy involving differential management of ASC-US and LSIL (i.e., direct colposcopy referral for women with LSIL) (**Figure 10**). When the average cost per CIN2+ detected projected by the current Norwegian triage strategy was used as a benchmark for the willingness-to-pay threshold (i.e., €4,195 per CIN2+ detected), reflex HPV mRNA testing with 14 genotypes was the preferred strategy (€2,978 per CIN2+ detected). For a moderate increase in the willingness-to-pay threshold, reflex HPV DNA testing with direct colposcopy for HPV-positive women would be preferred (€5,522-5,674 per CIN2+ detected, depending on the follow-up of HPV-negative women). Compared to the current Norwegian guidelines, all of the cost-efficient strategies were projected to increase CIN2+ detection by 18-57% for women with ASC-US/LSIL on their primary cytology. The additional CIN2+ benefit was accompanied by a trade-off of increased colposcopy referral rates (i.e., 14-138% increase) and 1.8-10.3 additional preterm births per 10,000 women with an index result of ASC-US/LSIL, compared to current guidelines.

In uncertainty analyses, the strategy involving HPV mRNA genotyping with 14 genotypes consistently had ICERs below the benchmark willingness-to-pay threshold. In contrast, the strategy involving differential management of ASC-US and LSIL (a strategy employed in the US [166]) consistently required an ICER of ≥ 3 times the benchmark. Results were most sensitive to when we reduced the diagnostic sensitivity of either the HPV mRNA test with 14 genotypes or all biomarkers simultaneously; in

these scenarios, HPV mRNA testing with 14 genotypes was no longer cost-efficient, and was replaced by dual staining on the efficiency frontier. Finally, results were not influenced by discounting costs and cases of CIN2+ detected by 4% per year.

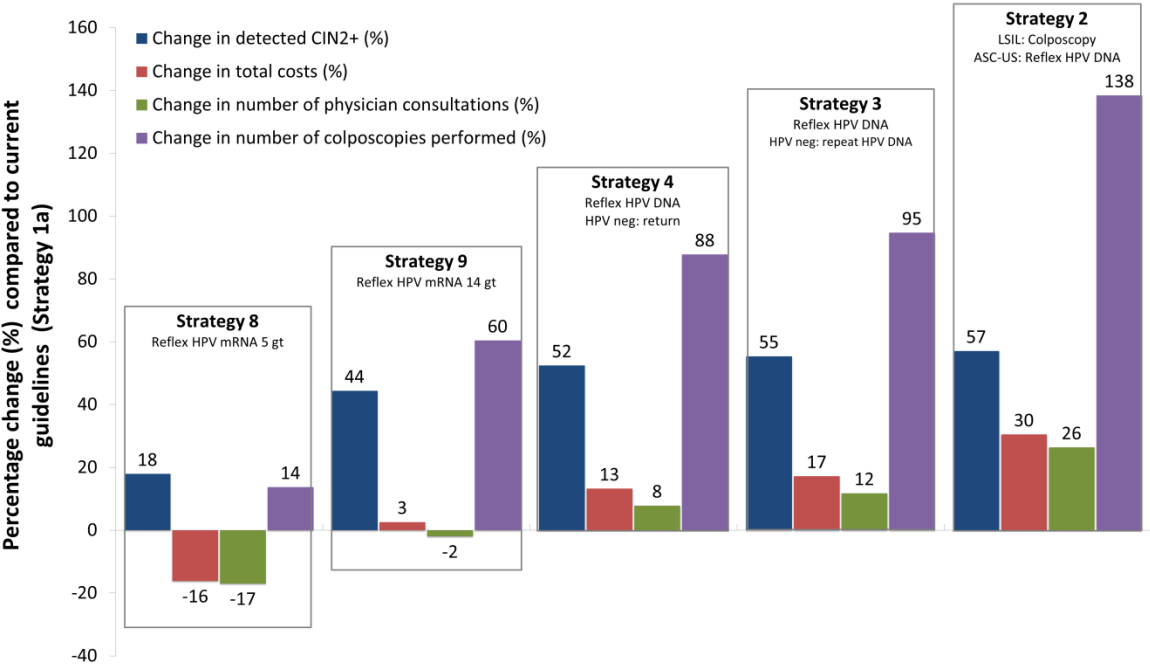


Figure 10. Health and economic trade-offs associated with cost-efficient triage strategies compared with the current Norwegian guidelines (Strategy 1a). Strategy 8; Reflex HPV mRNA testing (5 genotypes), Strategy 9; reflex HPV mRNA testing (14 genotypes), Strategy 4; reflex HPV DNA testing with direct colposcopy for HPV-positive women and return to screening for HPV-negative women; Strategy 3; reflex HPV DNA testing with direct colposcopy w/biopsy for HPV-positive women and return to screening for HPV-negative women with ASC-US only (repeat HPV DNA test in 12 months for women with LSIL), Strategy 2; differential management, involving reflex HPV DNA testing for women with ASC-US and direct colposcopy w/biopsy for women with LSIL. Adapted from *Paper I*.

6.2 Paper II

The ten candidate triage strategies for women with ASC-US or LSIL and who were reflex HPV-positive on their primary cytology were projected to reduce the lifetime risk of CC by 85.5-87.0%, compared to no screening. Although the strategies provided nominal differences in cancer benefit, resource use varied considerably. For example, the total percentage change in colposcopy referrals (for all women attending screening) ranged from a 7.3% reduction to a 21.4% increase, compared to current guidelines. Four strategies were identified as cost-efficient (i.e., using the cost per

QALY gained), including repeat HPV testing at 18 months without genotyping, HPV genotyping with direct colposcopy for HPV-16/-18-positive, and direct colposcopy for all women with ASC-US or LSIL and reflex HPV-positive (**Figure II**). For a willingness-to-pay threshold of a \$100,000 per QALY gained, the strategy involving genotype-specific management was preferred. All variations of the current guidelines (i.e., 6, 12, or 18 months follow-up wait time) were inefficient.

In secondary analysis, the additional comparator strategy involving differential management for women with ASC-US (i.e., reflex HPV DNA testing) and LSIL (i.e., direct colposcopy referral) was identified as cost-efficient but had an ICER that exceeded \$9 million per QALY, while the strategy involving genotype-specific management remained cost-effective, even if switching to primary HPV-based screening at ages 31 or 34 years. In uncertainty analysis, immediate colposcopy for all women with ASC-US or LSIL and reflex HPV-positive results was the preferred strategy under several analytic assumptions (e.g., imperfect screening adherence, lower HPV test sensitivity, 0% discount rate). When calculating the ICER using life years gained (LYG) rather than QALYs, the rank order of the strategies did not change except that delayed HPV DNA testing in 12 months, rather than in 18 months, was included on the efficiency frontier. Moreover, delayed HPV DNA testing in 12 months was identified as the preferred strategy (ICER: \$65,550 per LYG), while the strategy involving genotype-specific management had an ICER of \$118,980 per LYG.

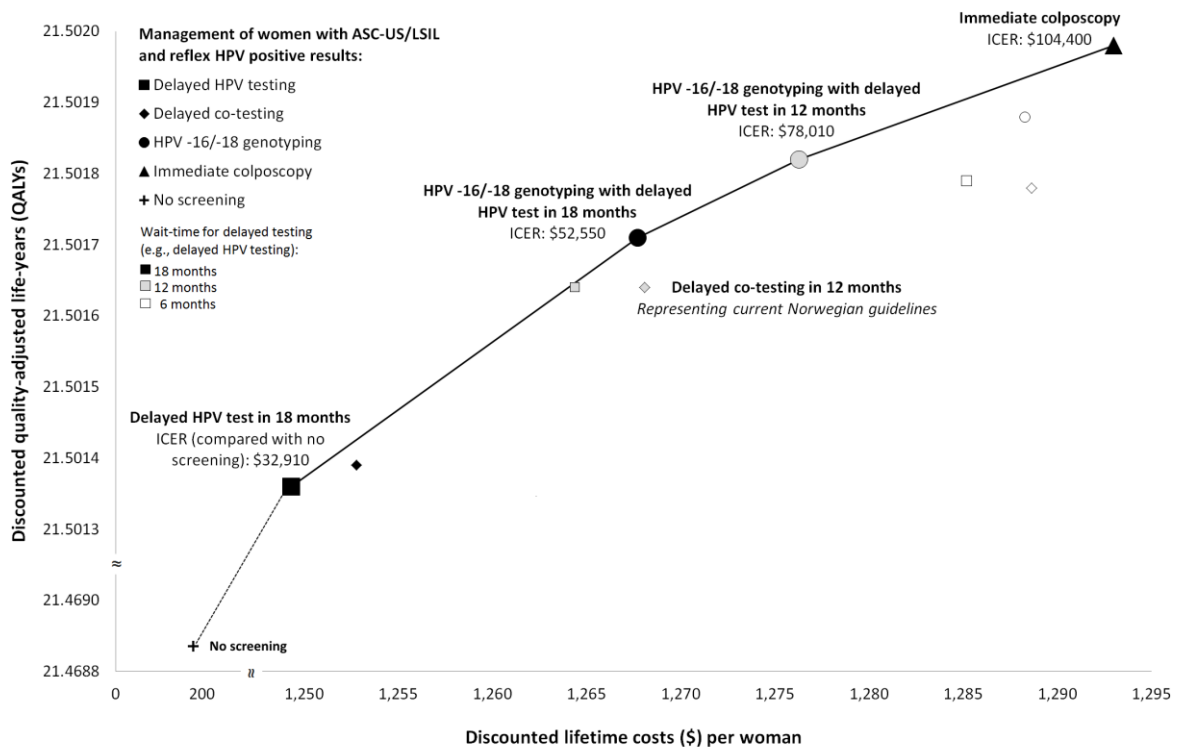


Figure II. Cost-effectiveness results for alternative triage algorithms for women with ASC-US or LSIL and high-risk HPV-positive results. Efficient strategies are accentuated with a larger symbol and connected by the solid line (i.e., efficiency frontier). All costs are expressed in 2014 US dollars (US\$ = NOK6.30). Adapted from *Paper II*.

6.3 Paper III

The candidate primary HPV-based strategies were projected to improve the cancer benefit compared to the current Norwegian guidelines; that is, the reduction in lifetime CC risk (compared to no screening) was 87.7% for the current cytology-based approach, while the reductions ranged from 90.9% to 96.3% for the HPV-based strategies (**Figure 12**). The cancer benefit was most influenced by the age of switching to HPV-based screening. In contrast, the number of screening tests (i.e., cytology and HPV tests) was most influenced by the primary screening frequency, while referral rates for colposcopy were most influenced by the triage approach for HPV-positive/cytology-negative women. Although all primary HPV-based strategies were expected to increase colposcopy referral rates compared to the expected level for the current Norwegian cytology-based guidelines, combining less intensive algorithm levers (e.g., screening frequency, number of persistent HPV-positive results to prompt

referral and the wait-time between repeat tests) could help temper the colposcopy referral rates. In uncertainty analysis, the cancer benefit was most influenced by a reduced HPV DNA test sensitivity; this assumption resulted in lower reductions in lifetime risk of CC (i.e., 84.2-94.6% reduction compared to no screening), which was mostly influenced by the primary screening frequency.

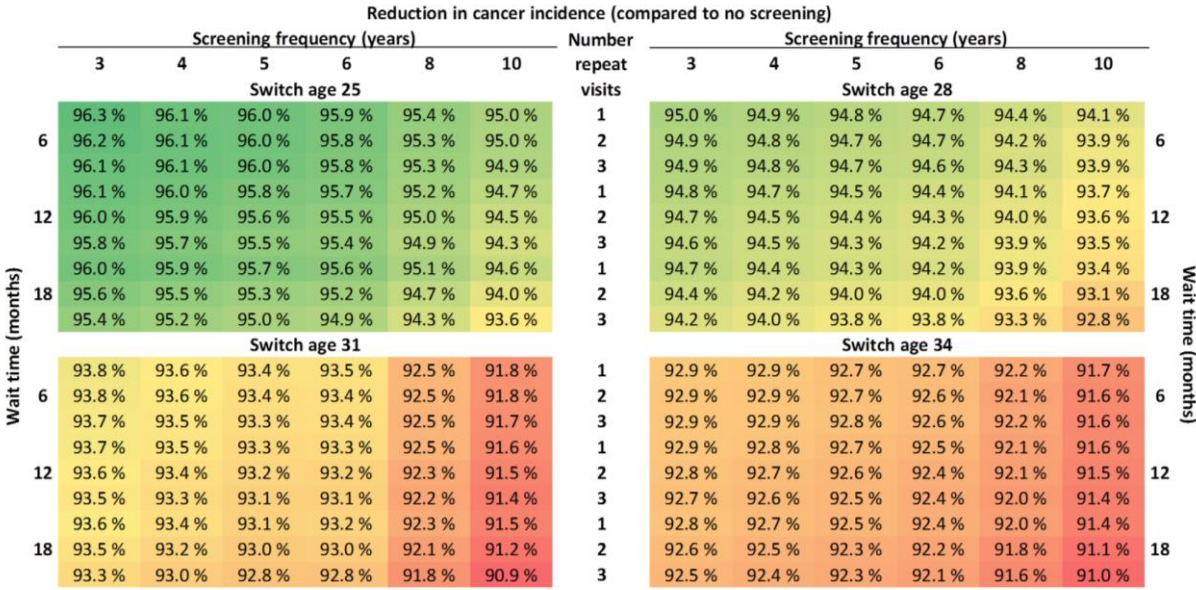


Figure 12. Reduction in cancer incidence compared to no screening associated with candidate primary HPV-based algorithms. Heat map formatting indicates low (dark red) to high (dark green) values. Adapted from *Paper III*.

6.4 Paper IV

For women fully vaccinated against HPV infections at age 12 years with any of the three available HPV vaccines, both the current cytology-based and proposed HPV-based Norwegian guidelines required higher costs and provided lower QALYs than candidate screening strategies. Importantly, all cost-efficient strategies involved HPV-based screening. Given a willingness-to-pay threshold of approximately \$100,000 per QALY, the optimal strategy for women vaccinated with the bivalent/quadrivalent vaccine involved two lifetime screens using HPV testing at ages 31 and 51 years (\$71,070 per QALY) or ages 30 and 45 years (\$102,910 per QALY) (**Figure 13**). For women vaccinated with the nonavalent HPV vaccine, screening once per lifetime using HPV testing at age 40 years (\$27,490 per QALY) was preferred. Strategies that provided

greater cancer benefits also required more colposcopy referrals. Importantly, a considerable amount could be allocated to identify the vaccination status of these women and screen them according to a separate set of guidelines. For example, for the first birth cohort of ~22,000 women who were fully vaccinated against HPV-16 and -18 infections in 2009 in Norway, between ~\$14-18 million could be allocated over their lifetime to identify individual vaccination status and stratify the screening program, while remaining cost-effective.

In uncertainty analysis, the results for the bivalent/quadrivalent vaccine was most sensitive to assuming lifelong cross-protection against non-vaccine targeted genotypes, imperfect screening compliance, including medical costs only (i.e., excluding patient time and travel costs), and assuming 0% discounting. For example, under assumptions of imperfect screening compliance, three lifetime screens using HPV testing and starting at age 28 years with 15 years apart was preferred (\$97,530 per QALY). Results for the nonavalent vaccine were most sensitive to assuming a lower vaccine efficacy of 90% (i.e., HPV testing at ages 31 and 51 years was preferred (\$77,700 per QALY)) and assuming 0% discounting (i.e., HPV testing at ages 25, 40 and 55 years was preferred (\$81,360 per QALY)). The rank order of the strategies and the ICERs for both the bivalent/quadrivalent and the nonavalent vaccine scenarios were reasonably robust to using LYG as measure of health benefit, in which case the preferred strategies did not change.

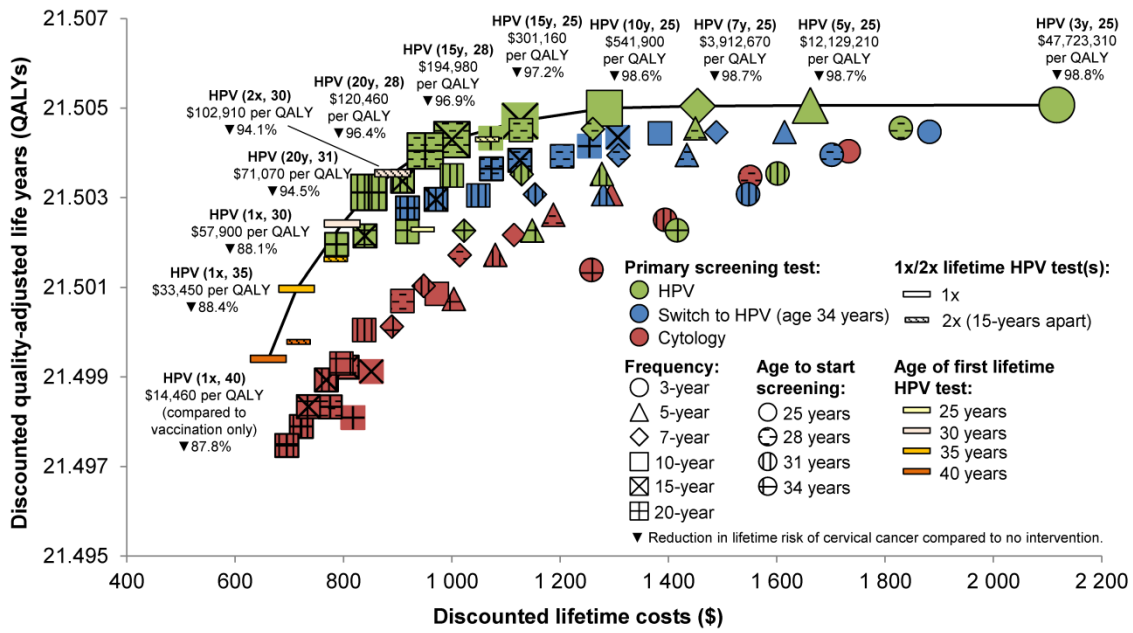


Figure B. Cost-effectiveness results for women vaccinated against HPV infections with the bivalent/quadrivalent vaccine. Efficient strategies are accentuated with a larger symbol and connected by the solid line (i.e. efficiency frontier). Parentheses for the efficient strategies indicate screening frequency (e.g. “20-yearly” or “1x/2x” indicate one or two lifetime screens) and age to start screening. All costs are expressed in 2014 US dollars (US\$ = NOK6.30). Abbreviations: HPV, human papillomavirus; QALY, quality-adjusted life years. Adapted from *Paper IV*.

7 DISCUSSION

The model-based analyses presented in this thesis addressed knowledge gaps related to the health benefits, resource use and cost-effectiveness associated with current and future CC screening policies in Norway. The findings highlight areas for improving the effectiveness and efficiency of the current and future screening algorithms that decision-makers can consider when designing future screening policies. This chapter addresses topics not discussed in-depth in each of the papers and starts with a broader discussion of the main findings of each paper, followed by a discussion of the methodological considerations, the policy implications, and areas of future research within CC prevention.

7.1 Discussion of results

The findings of *Papers I* and *II* indicate that there are opportunities to improve the effectiveness and efficiency of the current cytology-based screening program in Norway using biomarkers to triage women with minor cervical cytological lesions. *Paper I* showed that, in the short term and compared to current Norwegian guidelines, there is potential to detect additional precancers more efficiently using reflex HPV mRNA or DNA testing with direct colposcopy referral for women who are HPV-positive. In order to control total costs and colposcopy referrals, HPV mRNA testing may be preferred. However, the longer-term consequences of this strategy should be demonstrated prior to implementation. *Paper II* evaluated how strategies involving HPV DNA testing performs when also considering long-term outcomes, highlighting opportunities to improve the effectiveness and efficiency of the current guidelines using reflex HPV DNA testing (e.g., with genotype-specific management). These results are line with recent findings from a systematic review and meta-analysis evaluating HPV-16/-18 genotyping for women detected with ASC-US/LSIL, suggesting that direct colposcopy referral for HPV-16/-18 positive women can improve the efficiency of secondary screening yet at a cost of lower sensitivity compared to referring all high-risk HPV-positive women for colposcopy [167]. Both *Papers I* and *II* identified reflex HPV DNA testing with direct colposcopy for all HPV-positive women

as a cost-efficient strategy, which demonstrated ‘good value for money’ (according to the commonly cited Norwegian benchmark for cost-effectiveness, see Section 3.1.3) under several alternative analytic assumptions. However, this strategy was not preferred under base-case assumptions, and required increased resource use (e.g., colposcopy referrals) compared to current guidelines and other (less intensive) cost-efficient strategies. As the current Norwegian guidelines involves repeat cytology and HPV co-testing for all women detected with ASC-US/LSIL and who are reflex HPV-positive, direct colposcopy referral either restricted to HPV-16/-18 positive women or offered all HPV-positive women will increase colposcopy referrals in Norway. Implementation of these strategies thus requires available colposcopy resources and acceptance of increased colposcopy rates.

Within the next few years, Norwegian decision-makers will likely implement primary screening with HPV DNA testing for women aged 34 years and older (while maintaining cytology-based screening for women aged 25-33 years). As such, the insight gained from *Papers I and II* will remain relevant to optimize CC prevention among younger, unvaccinated women. As with the current cytology-based guidelines, the main concern with primary HPV-based screening is increased colposcopy referrals. To inform the trade-offs in health benefits and resource use associated with candidate primary HPV-based strategies, *Paper III* provides insight to which algorithmic levers are most influential on health benefits (i.e., the age to initiate primary HPV testing) and resource use (i.e., the triage algorithm for HPV-positive/cytology-negative women). These findings correspond with a recent evaluation of candidate triage strategies for HPV-positive women within the Canadian CC Screening Trial concluding that colposcopy referral rates could be tempered via triage [168]. Although Norwegian decision-makers have outlined a specific screening algorithm for primary HPV-based screening in Norway [52], decision-makers may use the consequence analysis provided in *Paper III* together with short-term follow-up data from the Norwegian implementation pilot study [52] when deciding the final guidelines for nationwide roll-out as well as in their continued refinement of the screening program.

The results of *Paper IV* underpin the advantage of transitioning CC screening guidelines to involve primary HPV testing, which was identified as the preferred

screening approach for HPV-vaccinated women. In Norway, the first cohort of women who were vaccinated against HPV infections in adolescence in 2009 will become eligible for screening in 2022. Consistent with the findings of other studies [63, 83-86], *Paper IV* demonstrated that for CC screening to remain cost-effective for these women, primary HPV-based screening coupled with a considerable de-intensified screening approach is required. Importantly, the results suggested that the optimal number of lifetime screens depended on vaccine type, which may influence future choice of vaccine type for the childhood immunization program. In a tender between the bivalent and nonavalent vaccine in May 2017, Norwegian decision-makers selected the bivalent vaccine for use in the immunization program. Although choosing the nonavalent vaccine would involve spending a higher upfront cost on one prevention program, *Paper IV* illustrate that these costs may be offset by scaling back the other prevention program (i.e., screening). While these decisions are often made in isolation, *Paper IV* highlights the importance of evaluating the interplay between primary and secondary prevention programs, for which decision-analytic modeling provides a useful tool.

Papers II and *IV* identified screening behavior as an important factor for the effectiveness and efficiency of a screening algorithm. While the short-term consequences of candidate strategies evaluated in *Paper I* were robust to assumptions concerning screening behavior, the cost-effectiveness results of *Papers II* and *IV* (which assumed 100% compliance to screening and follow-up procedures in the primary analysis) were influenced by introducing imperfect screening behavior assumptions. An argument in favor of assuming 100% screening compliance in the analysis is to allow comparisons of the 'best-case' benefits for each strategy. In addition, one may argue that the choice of a population-based prevention strategy should not be influenced by the behavior of some individuals, especially as compliance to a future algorithm (for which no empirical data exist) may not reflect compliance to the current algorithm. The uncertainty analyses of *Papers I, II* and *IV* primarily focused on compliance with follow-up procedures, which may depend on the follow-up approach. For example, in a Norwegian study including women with minor cervical cytological lesions, compliance was higher among women who were recommended direct

colposcopy referral compared to those recommended repeat testing in 12 months [169]. Consequently, when considering the results of *Papers I-IV* and when designing a screening strategy in general, decision-makers may trade-off the potential benefits of reduced loss-to-follow-up by choosing a more intensive strategy (e.g., direct colposcopy for all screen-positive) versus the accompanying cost of more unnecessary procedures.

All papers included in this thesis demonstrate that decision-makers can use several algorithm levers in order to design screening policies that balance the health benefits, resource use and potential harms of screening. Different diagnostic tests have strengths and weaknesses reflected by their diagnostic accuracy; in turn, choice of diagnostic test(s) for primary and secondary screening influences the effectiveness and efficiency of screening. In accordance with the discussion of diagnostic accuracy in Section 2.2.1, the findings of this thesis highlights the value of a sensitive primary test (i.e., HPV DNA test) and how varying the triage algorithm can help improve the overall accuracy of the screening program. For example, increasing the number of persistent positive test results required before prompting diagnostic work-up, as well as increasing the wait-time between those tests, can help balance the health benefits and resource use trade-offs. For primary screening, decision-makers may consider algorithm levers such as the age to start screening, the primary screening frequency and, if relevant, the age of switching from one algorithm to another (e.g., from cytology-based to HPV-based primary screening). The results of this thesis suggested that while candidate strategies provide nominal differences in health gains (especially when evaluating life years or QALYs gained), differences in resource use were often considerable. This underpins the need to quantify health and economic consequences of candidate strategies to inform the decision-making process and aid decision-makers in adopting strategies that are aligned policy objectives (e.g., maximizing health benefits while minimizing unnecessary screening procedures and potential harms) and local capacity constraints.

7.2 Methodological considerations

By relying on two different decision-analytic models, *Papers I-IV* provide an example of how different model types have different strengths and weaknesses. The input parameters used to inform the models were based on best-available evidence (Norwegian-specific when available), and estimated using calibration techniques for unobservable parameters. Together, the model structure, input parameters and assumptions influence the results of *Papers I-IV*, which will be discussed in the following section.

7.2.1 Model input parameters and assumptions

For *Paper I*, a decision-tree model was utilized to evaluate the short-term outcomes associated with using novel biomarkers to triage younger adult women with minor cervical cytological lesions. A decision-tree structure is appropriate for decision problems characterized by a short time horizon and with a few, defined outcomes [127]; in this context involving a time horizon of a single screening round (i.e., 3 years) and cases of CIN2+ detected as the health outcome. The model relied on empirical positivity rates to govern screening outcomes and diagnostic sensitivity of the different tests to detect CIN2+ given the presence of disease, which represents an alternative approach to modeling when the time horizon is short enough to conceptualize all possible screening pathways (which become infinitely large as the number of screening ‘rounds’ increase). Positivity rates in the model were informed using population-based data from the Cancer Registry of Norway and the University Hospital of North Norway, which represents a major strength of the analysis. Nevertheless, an important limitation of the model is the simplified approach to reflect underlying natural history of disease (i.e., representing CIN2+ as a composite outcome and not reflecting individual HPV types). This simplification may explain why type-specific management (i.e., the strategy identified as cost-effective in *Paper II*) was not identified as cost-efficient in *Paper I*. At the same time, the simplified approach allowed evaluating biomarkers other than HPV DNA tests, which have not been extensively evaluated in policy analyses previously. As more data on novel biomarkers accumulate, decision-

analytic models can be expanded to evaluate the long-term consequences of using these biomarkers in both primary and secondary screening. In particular, with a better understanding of how these biomarkers interact with the underlying natural history of disease, the microsimulation model employed in *Papers II-IV* can be adapted to enable evaluation of candidate biomarkers and the analysis in *Paper I* can be re-evaluated to include long-term outcomes.

The microsimulation model used in *Papers II-IV* was first developed in 1999 [170] and has since been refined and updated as new knowledge of HPV and CC have emerged. Our understanding of HPV epidemiology is rapidly evolving, which can allow further improvements to the model. For example, the current model reflects squamous cell carcinoma only, which is the histologic type of CC that is most readily prevented by screening [2] and for which most data is available. However, a Norwegian study evaluating CC incidence trends in Norway in the period 1956-2010 showed that while squamous cell carcinoma incidence rates have decreased since widespread introduction of cytology-based screening, the incidence of adenocarcinoma continues to increase [2]. As more data on adenocarcinoma accumulates, the model can be extended to reflect both histologic CC types by incorporating health states of adenocarcinoma and its precursor adenocarcinoma-in-situ. This becomes even more important as CC screening transitions to involve primary HPV-testing, which is expected to be more effective in preventing cervical adenocarcinoma than cytology-based screening [2, 41]. Furthermore, although HPV has been established as a causal agent of CC, emerging studies indicate that a proportion of CC cases are HPV-negative, and that this is relatively more common in adenocarcinomas than in squamous cell carcinomas [24]. In addition, some of the uncertainties regarding the causal web of CC (see Section 2.1), such as the impact of smoking and other behavioral factors, is not captured explicitly by the model, but is inherently captured in the calibration targets (e.g., HPV prevalence) and thus reflected in the incidence and progression rates. Although these factors are unlikely to have influenced the results of *Papers II-IV*, the model can be updated and analyses re-evaluated as more evidence concerning the interaction between behavioral factors and CC risk accumulates.

The parameters used to inform the underlying natural history of cervical carcinogenesis are unobservable and were estimated using calibration techniques. As discussed in Section 3.2.2, different calibration techniques are available and the choice of method depends on factors such as the model structure complexity and the uncertainty concerning the natural history of the disease [146]. In *Paper I*, a simple calibration process using ‘acceptable windows’ and a ‘trial-and-error’ search approach was employed to estimate the baseline prevalence of CIN2+ among women with ASC-US/LSIL at their primary cytology. Despite using a simple approach, the estimated parameter value was in accordance with the empirical target and varying this parameter in one-way uncertainty analysis did not change the rank order of the cost-efficient strategies. In contrast, the microsimulation model used in *Papers II-IV* relied on a more complex, multi-parameter calibration approach using log-likelihood functions. The limitations of using this approach have been described previously [145, 146, 148], but two of them will be discussed in more detail here. One limitation is the use of random sampling (which may not have searched the parameter space sufficiently) rather than grid search or a ‘learning’ sampling method such as the Nelder-Mead method. However, comprehensive computing power and storage space allowed a large repository of samples (i.e., ~1 million samples) to be generated. This approach avoids restricting the search space to certain settings, which is beneficial as the model is applied to multiple settings (i.e., both developed and developing countries) with large variations in HPV and CC epidemiology. Another limitation is that the data used to inform the calibration targets of age-specific HPV incidence and HPV type-distribution in CIN3 and SCC were based on samples from larger Norwegian cities and may not be representative of the entire population. Nevertheless, all calibration targets used in the calibration processes for *Papers I-IV* were estimated using primary data from Norway (e.g., the Cancer Registry of Norway) which improves validity of the model to the Norwegian context. Furthermore, the comprehensive validation exercise (described in Section 5.2.2) showed that the model predicted outcomes that corresponded well with Norwegian epidemiologic data not used to inform the model (e.g., screening behavior estimated using population-based data from the Cancer Registry of Norway covering two decades of cytology-based screening).

The costing approach for *Papers I, II* and *IV* were based on micro-costing from Norwegian pathology laboratories as well as aggregate costing using DRGs and other Norwegian fee schedules. In Norway, linkage of registry data (e.g., from the Norwegian Patient Registry and the Norwegian Prescription Database) using personal identification numbers provides opportunities to improve the cost estimates used in *Papers I, II* and *IV*. Furthermore, several guidance bodies for economic evaluation recommend to include spillover effects (e.g., economic impacts on the lives of those close to the patient) [98]; these factors can influence the cost-effectiveness of an intervention yet are often ignored [171]. Spillover effects may be particularly influential in evaluations of interventions to prevent CC, which affects younger adult women in their active productive and social years. Although the 2016 Norwegian priority setting white paper [97] advice not to include productivity losses among patients and their spillover effects, quantifying spillover effects and the impact of considering these costs on optimal prevention approaches is necessary to evaluate the potential bias introduced by disregarding the economic burden of informal care and productivity losses.

Following recommendations for conducting CEA in Norway [64, 97] and elsewhere [98], *Papers II* and *IV* used QALYs gained as the measure of health outcome used to calculate the ICERs of candidate strategies. Due to the lack of studies to inform HRQoL weights associated with prevention and management of cervical disease (as highlighted in a recent systematic review [172]), the weights used to inform HRQoL for CC health states in *Papers II* and *IV* were based on a modeling study evaluating a Norwegian-specific CC prevention policy [161]. In turn, the HRQoL adjustment factors for the CC health states presented in **Table 5** in Section 5.3 are based on a study from the US [163]. As such, the HRQoL weights for CC health states used in *Papers II* and *IV* are highly uncertain and may not be representative for the Norwegian setting, which represent a limitation of the analyses. Lacking empirical data furthermore involved that the analyses did not account for HRQoL associated with screening and precancer treatment. Due to the uncertainty surrounding the HRQoL weights for CC health states, some analysts have opted to use LYG as the primary health outcome used to calculate ICERs [86], and rather present ICERs based on QALYs in uncertainty

analyses. When calculating the ICERs in *Papers II* and *IV* using LYG instead of QALYs, the ranking of the strategies and the ICERs were only moderately influenced (see Sections 6.2 and 6.4). Future research should focus on evaluating the HRQoL associated with health states related to CC screening and disease to help improve the quality of policy analyses related to CC prevention. However, measuring HRQoL in general remains a challenge; thus, future analyses should include a comprehensive assessment of the impact of alternative HRQoL assumptions on optimal prevention approaches.

7.2.2 Analytic approach

The choice of competing strategies is an important component of CEA and essential for appropriate identification of cost-efficient and cost-effective strategies [109]. In line with recommendations for good modeling practice [128], the analyses in *Papers I-IV* included strategies considered relevant for decision-making in Norway. However, inclusion of other candidate strategies may have identified other cost-efficient and cost-effective strategies. For example, the decision-tree model used in *Paper I* provided limited opportunities to vary the wait time prior to repeat testing, which was identified as an important algorithm lever in *Paper II*, while considering other biomarkers than HPV DNA testing may have altered the conclusions of *Paper II*. None of the 216 strategies considered in *Paper III* included HPV-16/-18 genotyping or repeat cytology triage for HPV-positive/cytology-negative women, which has been suggested as a relevant triage method for HPV-based screening elsewhere [168, 173]. Other biomarkers, such as p16/Ki67 dual stain cytology may also be a relevant triage strategy for HPV-positive women, yet longer-term data is still awaiting [174]. Lastly, the analysis in *Paper IV* focused on evaluating strategies that varied by primary algorithm levers (e.g., screening start age and frequency) as the primary screening algorithm for HPV-vaccinated women had not yet been outlined by decision-makers. Future analyses may use the insight gained from *Papers I-III* to evaluate the optimal triage approach for these vaccinated women. As the selection of strategies for inclusion in each analysis was based on discussions with Norwegian clinicians and policy-makers, these represent the strategies most likely to be adopted within a Norwegian setting.

Papers I-III projected outcomes of candidate CC screening strategies for unvaccinated women, while *Paper IV* evaluated optimal CC screening guidelines for women being fully vaccinated against vaccine-targeted HPV infections in adolescence. Because the analysis in *Paper IV* was conditioned on a woman having received the direct effect of the HPV vaccine, a microsimulation model that was able to reflect complex screening strategies and individual screening history was used. Over the next decades, the screening target population will become increasingly heterogeneous with respect to HPV vaccination status. Future analyses can utilize a dynamic model with the ability to capture herd immunity effects of HPV vaccination, such that screening strategies for both vaccinated and unvaccinated women can be evaluated within the context of Norway. For example in the Netherlands, a recent model-based study suggested that as the proportion of HPV-vaccinated individuals in the population increase (and consequently, a larger herd immunity benefit is achieved), a less intensive universal (i.e., for both vaccinated and unvaccinated women) screening algorithm can be adopted [84].

In *Papers I, II* and *IV*, there was decision uncertainty around which strategy was considered optimal given decision-makers' willingness-to-pay for additional health benefits. In *Paper I*, the average cost per CIN2+ detected as projected by the current Norwegian guidelines was used as a benchmark for what decision-makers are currently willing-to-pay, although this may not represent the maximum threshold willingness-to-pay per additional CIN2+ detected. In *Papers II* and *IV*, a commonly cited Norwegian willingness-to-pay threshold of \$100,000 per QALY was used as a benchmark for what constitutes 'good value for money'; however, this threshold has been referred to as a 'reference value' rather than a strict threshold. In addition, the 2016 Norwegian priority setting white paper [97] indicated that the willingness-to-pay should be weighted by the severity of the disease, yet did not provide an explicit approach to perform such weighting. Under these new guidelines for priority setting in Norway, a higher or lower ICER may be considered 'good value for money' for interventions to prevent CC depending on the operationalization of severity weighting in CEA.

Due to the uncertainty surrounding what constitutes ‘good value for money’ in Norway, the presentation of results in *Papers II* and *IV* intended to focus on all cost-efficient strategies so that decision-makers can evaluate these strategies along with other considerations such as resource use and potential harms. For example, colposcopy-directed biopsy has been previously highlighted as a key outcome for CC policy makers in Norway, both due to limited availability of colposcopy resources and because it is considered a semi-invasive procedure that should be kept at an acceptable level [143]. Although it may be challenging to identify the exact capacity constraint of a particular resource, as well as women’s acceptability for a diagnostic procedure, quantifying the expected change in resource use and referrals associated with candidate interventions can help inform the implementation process. Furthermore, the 2016 Norwegian priority setting white paper [97] emphasized that individual patients should be involved in decisions regarding themselves. Specifically, the white paper outlined a health policy objective of creating “*the patient’s healthcare, where everyone should receive help when they need it, be taken care of and informed, and feel that they can influence and have ownership of choices regarding themselves*” [97, page 3]. Although a screening strategy may be optimal on a population-level, it may not reflect the optimal strategy for an individual woman. Ultimately, the optimal strategy depends on a compendium of factors, such as decision-makers’ willingness-to-pay for additional health benefits and acceptance of additional resource use, the feasibility of implementing these strategies, and the acceptability of the screening strategies to individual women.

7.3 Policy implications

In *Papers I, II* and *IV*, the current Norwegian guidelines were identified as less effective and more costly than candidate strategies. Thus, from a cost-effectiveness standpoint, there are clear opportunities to improve CC prevention in Norway. For the next years, cytology-based screening will remain an important component of the Norwegian CC screening program, particularly for younger women aged 25-33 years unlikely to be offered primary HPV testing under the future HPV-based screening program in Norway, as well as for all women of screening target age until the HPV-based screening

is rolled-out nationwide. When considering changes to the screening algorithm in the short-term, quantifying the impact a strategy has on resource use is essential as immediate substantial changes to the system may not be possible (e.g., availability of colposcopists). Thus, to improve the effectiveness and cost-effectiveness of CC screening in the short-term, the strategies identified as cost-effective in *Papers I and II* can be considered. For example, as reflex HPV DNA testing is currently in use in Norway and several laboratories use an assay that provides genotype-specific information [175], decision-makers could consider referring women positive for HPV types -16 or -18 directly to colposcopy rather than repeat cytology and HPV DNA co-testing in 12 months (i.e., current Norwegian guidelines). *Paper II* projected this strategy to provide additional health benefits for a similar monetary cost, compared to current guidelines. As primary HPV DNA testing is rolled out nationwide, decision-makers could consider revising the triage algorithm to allow an earlier switch to primary HPV testing. The findings of *Paper III* suggest that decision-makers could even consider implementing primary HPV testing starting at age 25 years (an algorithm that was recently approved in the US [56]), and rather use the triage algorithm to limit colposcopy referrals for these younger women. Streamlining the program to involve primary HPV testing would benefit future generations of HPV-vaccinated women, as *Paper IV* identified strategies involving primary HPV testing as the dominant testing method for these women.

By quantifying the short- and long-term health and economic consequences of multiple courses of action, decision-analytic modeling provides a powerful tool to aid decision-makers in choosing strategies that maximize health benefits while providing efficient and feasible use of societal resources and minimize potential harms to women. In the era of ‘the patient’s healthcare’ and informed choice, decision-analytic modeling can be used to quantify and present trade-offs associated with candidate screening algorithms to inform decision-making at both the individual- and system-levels. Given the potential usefulness of decision analyses, policy-makers may consider to include such analyses as a formal component of the decision-making process. Although model-based analyses can never replace real-world data, Weinstein and Stason concluded in their seminal paper on the foundations of cost-effectiveness

analysis for health and medical practices in 1977 that: “Nevertheless, resource allocation decisions do have to be made, and the choice is often between relying upon a responsible analysis, with all its imperfections, and no analysis at all. The former, in these times of increasingly complex decisions, difficult tradeoffs and limited resources, is by far the preferred choice” [176].

7.4 Future research

The methodological considerations discussed in Section 7.2 highlighted several opportunities to improve the simulation models, the inputs, and the analytic approach employed in *Papers I-IV*. This section focuses on future research within the broader area of CC prevention in times of emerging personalized screening interventions and HPV vaccination.

First, there may be opportunities to tailor CC screening algorithms according to individual risk of developing CC (i.e., risk-based screening) beyond stratifying guidelines by an individual’s HPV vaccination status as evaluated in *Paper IV*. In particular, using the risk-based framework initially proposed by Castle et al. [16] (discussed in Section 2.2.1), future studies may use decision-analytic models to evaluate the optimal risk thresholds to prompt surveillance, colposcopy referral and treatment. For example, the microsimulation model used in *Papers II-IV* can be adapted to update an individual’s risk profile (e.g., low-risk, intermediate risk, high-risk) according to screening history over the simulated lifetime. In turn, the model can be informed by synthesizing data from empirical studies and large registry databases (e.g., from the Cancer Registry of Norway or Kaiser Permanente in the US [177]) on the cumulative risk of developing high-grade precancers within a defined period (e.g., 3 or 5 years). In Norway, there is also an increased focus on patient-centered care, shared decision-making and informed choice in healthcare. Thus, in addition to designing risk-based screening algorithms, the era of ‘the patient’s healthcare’ encourages future studies that evaluate screening algorithms based on individuals’ preferences (e.g., acceptance of unnecessary screening procedures versus reduced risk of developing CC). However, while tailored interventions may be an attractive approach, there is a

cost to identifying the personal characteristics of screen-eligible women. Future research could also evaluate the value of identifying these women using the INMB framework proposed in *Paper IV*.

Novel screening technologies continue to provide opportunities for further improvements in CC prevention. For example, mailing an HPV self-sampling device kit (i.e., self-sampling) to non-attenders has shown promising potential to increase participation in Norway [178] and elsewhere [179]. This intervention has also demonstrated ‘good value for money’ in Norway [152]. Self-sampling has the potential to eliminate the burden, anxiety and economic costs associated with traditional office-based pelvic exams, especially considering a recent review concluding that these exams do not provide additional benefit to patients beyond targeted interventions such as CC screening [180]. Accordingly, future studies should evaluate the cost-effectiveness of expanded use of self-sampling as a primary screening approach for all women in the screening target population, which may be of particular relevance as the proportion of HPV-vaccinated women in the population increase and the screening program is scaled down.

Furthermore, given the need to considerably revise the CC screening program as the proportion of vaccinated individuals increase, future studies may expand the analysis performed in *Paper IV* to evaluate optimal screening in HPV-vaccinated women by age of vaccination (e.g., for women vaccinated as part of ‘catch-up’ programs). Recently, several international advisory groups have proposed the expanded use of the HPV vaccines to include older women up to age 50 years (a protocol coined by Bosch et al. [181] as HPV-FASTER). However, such policies are less likely to provide good value for money in countries with well-established screening and vaccination programs. In addition, a recent model-based analysis projected that 75% of women who developed CC acquired their causal HPV infection by age 31 years; as such, vaccinating mid-adult women may provide limited opportunities to prevent additional cancers. Analyses are needed to inform whether and how (e.g., upper age limit to vaccinate) an expanded vaccination policy for older women in high-income settings can contribute to accelerating reductions in the cancer burden among both men and women, while providing ‘good value for money’. This policy should also be evaluated for women who

recently immigrated to Norway, especially for young adult women less than age 26 years who immigrate after the completion of the temporary 2016-2018 Norwegian HPV vaccine 'catch-up' campaign, which may be a high-value prevention policy for these women by avoiding the barriers induced by office-based screening [182].

In sum, these technologies and prevention approaches provide further opportunities to reduce the burden of CC (and other HPV-related cancers) while ensuring that prevention efforts continue to balance benefits and harms, provide efficient use of resources, maximize healthy years of life, and are acceptable to individuals targeted by prevention programs.

8 CONCLUSIONS

This thesis used a decision-analytic approach to address knowledge gaps related to the current and future CC screening policies in Norway, and highlights several opportunities to improve the cost-effectiveness and balance the health benefits and resource use associated with these policies.

For the current cytology-based screening program, *Paper I* indicates that the efficiency and effectiveness (in terms of precancer detection) can be improved using reflex HPV mRNA testing to triage women with minor cervical cytological lesions. However, the long-term consequences of this strategy should be demonstrated prior to implementation. In order to improve the long-term efficiency and effectiveness (in terms of CC incidence), *Paper II* suggests to triage women with minor cervical cytological lesions using reflex HPV DNA testing with direct colposcopy referral for women positive for HPV-16/-18 infections.

For the future CC screening policies in Norway, *Paper III* indicates that in order to maximize the CC preventive benefits of primary HPV-based screening while controlling colposcopy referral rates, HPV-based screening should start at an earlier age and rather utilize a less intensive triage algorithm for HPV-positive/cytology-negative women. Finally, in order for screening to remain cost-effective for women who received the HPV vaccine in adolescence, *Paper IV* suggests that a de-intensified HPV-based screening strategy (e.g., screening once or twice over a lifetime) is required.

Ultimately, decision-makers can use the findings from these analyses together with other factors that influence priority setting to shape CC screening policies in Norway.

9 REFERENCES

1. Cancer Registry of Norway. Cancer in Norway 2015 - Cancer incidence, mortality, survival and prevalence in Norway. Accessed at <https://www.kreftregisteret.no/globalassets/cancer-in-norway/2015/cin-2015-special-issue.pdf> on April 9, 2017. Oslo: Cancer Registry of Norway, 2016.
2. Lonnberg S, Hansen BT, Haldorsen T, Campbell S, Schee K, and Nygard M, Cervical cancer prevented by screening: Long-term incidence trends by morphology in Norway. *Int J Cancer*, 2015. 137(7): p. 1758-64.
3. zur Hausen H, Papillomaviruses in the causation of human cancers - a brief historical account. *Virology*, 2009. 384(2): p. 260-5.
4. Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, et al., Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol*, 1999. 189(1): p. 12-9.
5. Biological agents. Volume 100 B. A review of human carcinogens. IARC Monogr Eval Carcinog Risks Hum, 2012. 100(Pt B): p. 1-441.
6. Schiffman M, Doorbar J, Wentzensen N, de Sanjose S, Fakhry C, Monk BJ, et al., Carcinogenic human papillomavirus infection. *Nat Rev Dis Primers*, 2016. 2: p. 16086.
7. World Health Organization, IARC monographs on the evaluation of carcinogenic risks to humans. Human papillomaviruses, 2007. 90: p. 1-636. Available at: <http://monographs.iarc.fr/ENG/Monographs/vol90/index.php>.
8. Guan P, Howell-Jones R, Li N, Bruni L, de Sanjose S, Franceschi S, et al., Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer. *Int J Cancer*, 2012. 131(10): p. 2349-59.
9. de Martel C, Plummer M, Vignat J, and Franceschi S, Worldwide burden of cancer attributable to HPV by site, country and HPV type. *Int J Cancer*, 2017.
10. Hansen BT, Campbell S, Nygård M, [HPV-related cancer in Norway]. In: Cancer in Norway 2014 - Cancer incidence, mortality, survival, and prevalence in Norway. Accessed at https://www.kreftregisteret.no/globalassets/cancer-in-norway/2014/special_issue.pdf on December 9, 2015. Oslo: Cancer Registry of Norway, 2015.
11. Crosbie EJ, Einstein MH, Franceschi S, and Kitchener HC, Human papillomavirus and cervical cancer. *The Lancet*. 382(9895): p. 889-899.
12. Wentzensen N, Nason M, Schiffman M, Dodd L, Hunt WC, and Wheeler CM, No evidence for synergy between human papillomavirus genotypes for the risk of high-grade squamous intraepithelial lesions in a large population-based study. *J Infect Dis*, 2014. 209(6): p. 855-64.

13. Rositch AF, Koshiol J, Hudgens MG, Razzaghi H, Backes DM, Pimenta JM, et al., Patterns of persistent genital human papillomavirus infection among women worldwide: a literature review and meta-analysis. *Int J Cancer*, 2013. 133(6): p. 1271-85.
14. Plummer M, Schiffman M, Castle PE, Maucort-Boulch D, and Wheeler CM, A 2-year prospective study of human papillomavirus persistence among women with a cytological diagnosis of atypical squamous cells of undetermined significance or low-grade squamous intraepithelial lesion. *J Infect Dis*, 2007. 195(11): p. 1582-9.
15. Beachler DC, Jenkins G, Safaeian M, Kreimer AR, and Wentzensen N, Natural Acquired Immunity Against Subsequent Genital Human Papillomavirus Infection: A Systematic Review and Meta-analysis. *J Infect Dis*, 2016. 213(9): p. 1444-54.
16. Castle PE, Sideri M, Jeronimo J, Solomon D, and Schiffman M, Risk assessment to guide the prevention of cervical cancer. *Am J Obstet Gynecol*, 2007. 197(4): p. 356.e1-6.
17. Denslow SA, Rositch AF, Firnhaber C, Ting J, and Smith JS, Incidence and progression of cervical lesions in women with HIV: a systematic global review. *Int J STD AIDS*, 2014. 25(3): p. 163-77.
18. Abraham AG, D'Souza G, Jing Y, Gange SJ, Sterling TR, Silverberg MJ, et al., Invasive cervical cancer risk among HIV-infected women: a North American multicohort collaboration prospective study. *J Acquir Immune Defic Syndr*, 2013. 62(4): p. 405-13.
19. Appleby P, Beral V, Berrington de Gonzalez A, Colin D, Franceschi S, Goodill 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(6): p. 1481-95.
20. International Collaboration of Epidemiological Studies of Cervical Cancer, Cervical carcinoma and reproductive factors: collaborative reanalysis of individual data on 16,563 women with cervical carcinoma and 33,542 women without cervical carcinoma from 25 epidemiological studies. *Int J Cancer*, 2006. 119(5): p. 1108-24.
21. Appleby P, Beral V, Berrington de Gonzalez A, Colin D, Franceschi S, Goodhill A, et al., Cervical cancer and hormonal contraceptives: collaborative reanalysis of individual data for 16,573 women with cervical cancer and 35,509 women without cervical cancer from 24 epidemiological studies. *Lancet*, 2007. 370(9599): p. 1609-21.
22. McCredie MR, Sharples KJ, Paul C, Baranyai J, Medley G, Jones RW, 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(5): p. 425-34.
23. Schiffman M and Wentzensen N, Human papillomavirus infection and the multistage carcinogenesis of cervical cancer. *Cancer Epidemiol Biomarkers Prev*, 2013. 22(4): p. 553-60.
24. Rodriguez-Carunchio L, Soveral I, Steenbergen RD, Torne A, Martinez S, Fuste P, et al., HPV-negative carcinoma of the uterine cervix: a distinct type of cervical cancer with poor prognosis. *Bjog*, 2015. 122(1): p. 119-27.

25. Castle PE and Maza M, Prophylactic HPV vaccination: past, present, and future. *Epidemiol Infect*, 2016. 144(3): p. 449-68.
26. Hill AB, The environment and disease: association or causation? *Proc R Soc Med*, 1965. 58: p. 295-300.
27. Wilson JMG and Jungner G, Principles and practice of screening for disease. World Health Organization: Geneva, 1968. Available at: http://apps.who.int/iris/bitstream/10665/37650/17/WHO_PHP_34.pdf.
28. World Health Organization, WHO Guidelines for Screening and Treatment of Precancerous Lesions for Cervical Cancer Prevention. World Health Organization: Geneva, 2013. Available at: http://apps.who.int/iris/bitstream/10665/94830/1/9789241548694_eng.pdf.
29. Sharp L, Cotton S, Cochran C, Gray N, Little J, Neal K, et al., After-effects reported by women following colposcopy, cervical biopsies and LLETZ: results from the TOMBOLA trial. *Bjog*, 2009. 116(11): p. 1506-14.
30. Kyrgiou M, Athanasiou A, Paraskevaïdi M, Mitra A, Kalliala I, Martin-Hirsch P, 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: p. i3633.
31. Jin G, LanLan Z, Li C, and Dan Z, Pregnancy outcome following loop electrosurgical excision procedure (LEEP) a systematic review and meta-analysis. *Arch Gynecol Obstet*, 2014. 289(1): p. 85-99.
32. Mayrand MH, Duarte-Franco E, Rodrigues I, Walter SD, Hanley J, Ferenczy A, et al., Human papillomavirus DNA versus Papanicolaou screening tests for cervical cancer. *N Engl J Med*, 2007. 357(16): p. 1579-88.
33. Solomon D, Chapter 14: Role of Triage Testing in Cervical Cancer Screening. *JNCI Monographs*, 2003. 2003(31): p. 97-101.
34. Nanda K, McCrory DC, Myers ER, Bastian LA, Hasselblad V, Hickey JD, 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(10): p. 810-9.
35. Arbyn M, Bergeron C, Klinkhamer P, Martin-Hirsch P, Siebers AG, and Bulten J, Liquid compared with conventional cervical cytology: a systematic review and meta-analysis. *Obstet Gynecol*, 2008. 111(1): p. 167-77.
36. Arbyn M, Roelens J, Simoens C, Buntinx F, Paraskevaïdis E, Martin-Hirsch PP, et al., Human papillomavirus testing versus repeat cytology for triage of minor cytological cervical lesions. *Cochrane Database Syst Rev*, 2013. 3: p. Cd008054.
37. Krahn M, McLachlin M, Pham B, Rosen B, Sander B, Grootendorst P, et al., Liquid-based techniques for cervical cancer screening: systematic review and cost-effectiveness analysis. 2008.

38. Rozemeijer K, Naber SK, Penning C, Overbeek LI, Looman CW, de Kok IM, et al., Cervical cancer incidence after normal cytological sample in routine screening using SurePath, ThinPrep, and conventional cytology: population based study. *Bmj*, 2017. 356: p. j504.
39. 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*, 2016. 27(1): p. 15-25.
40. National Cancer Institute, NCI Dictionary of Cancer Terms: biomarker. Accessed at <https://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=45618> on June 20, 2017.
41. Ronco G, Dillner J, Elfstrom KM, Tunesi S, Snijders PJ, Arbyn M, et al., Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. *Lancet*, 2014. 383(9916): p. 524-32.
42. Arbyn M, Ronco G, Anttila A, Meijer CJ, Poljak M, Ogilvie G, et al., Evidence regarding human papillomavirus testing in secondary prevention of cervical cancer. *Vaccine*, 2012. 30 Suppl 5: p. F88-99.
43. Wentzensen N, Schiffman M, Palmer T, and Arbyn M, Triage of HPV positive women in cervical cancer screening. *J Clin Virol*, 2016. 76 Suppl 1: p. S49-s55.
44. Sahasrabudde VV, Luhn P, and Wentzensen N, Human papillomavirus and cervical cancer: biomarkers for improved prevention efforts. *Future Microbiol*, 2011. 6(9): p. 1083-98.
45. Peirson L, Fitzpatrick-Lewis D, Ciliska D, and Warren R, Screening for cervical cancer: a systematic review and meta-analysis. *Syst Rev*, 2013. 2: p. 35.
46. Bray F, Loos AH, McCarron P, Weiderpass E, Arbyn M, Moller H, et al., Trends in cervical squamous cell carcinoma incidence in 13 European countries: changing risk and the effects of screening. *Cancer Epidemiol Biomarkers Prev*, 2005. 14(3): p. 677-86.
47. Vaccarella S, Franceschi S, Engholm G, Lonnberg S, Khan S, and Bray F, 50 years of screening in the Nordic countries: quantifying the effects on cervical cancer incidence. *Br J Cancer*, 2014. 111(5): p. 965-9.
48. Elfstrom KM, Arnheim-Dahlstrom L, von Karsa L, and Dillner J, Cervical cancer screening in Europe: Quality assurance and organisation of programmes. *Eur J Cancer*, 2015. 51(8): p. 950-68.
49. Arbyn M, Anttila A, Jordan J, Ronco G, Schenck U, Segnan N, et al., European Guidelines for Quality Assurance in Cervical Cancer Screening. Second edition--summary document. *Ann Oncol*, 2010. 21(3): p. 448-58.
50. Andrae B, Kemetli L, Sparen P, Silfverdal L, Strander B, Ryd W, et al., Screening-preventable cervical cancer risks: evidence from a nationwide audit in Sweden. *J Natl Cancer Inst*, 2008. 100(9): p. 622-9.

51. Andrae B, Andersson TM, Lambert PC, Kemetli L, Silfverdal L, Strander B, et al., Screening and cervical cancer cure: population based cohort study. *Bmj*, 2012. 344: p. e900.
52. Nygard M, Andreassen T, Berland J, Hagen B, Hagmar B, Iversen O-E, et al., [HPV testing in primary cervical cancer screening. Controlled implementation and evaluation of improved health service]. The Norwegian Directorate of Health: Oslo, 2013.
53. Health Council of the Netherlands, Population screening for cervical cancer. Accessed at https://www.gezondheidsraad.nl/sites/default/files/201107E_PopulationSCC_0.pdf on August 7, 2017. Health Council of the Netherlands: The Hague, 2011.
54. Medical Services Advisory Committee, MSAC Outcomes Public Summary Document: Application No. 1276 - Renewal of the National Cervical Screening Program 2014. Accessed at [http://www.msac.gov.au/internet/msac/publishing.nsf/Content/D924E2F768B13C4BCA25801000123B9E/\\$File/1276%20-%20Final%20MSAC%20PSD%20-%20NCSP%20Renewal.pdf](http://www.msac.gov.au/internet/msac/publishing.nsf/Content/D924E2F768B13C4BCA25801000123B9E/$File/1276%20-%20Final%20MSAC%20PSD%20-%20NCSP%20Renewal.pdf) on August 7, 2017. Australian Government, 2014.
55. Ronco G, Biggeri A, Confortini M, Naldoni C, Segnan N, Sideri M, et al., [Health technology assessment report: HPV DNA based primary screening for cervical cancer precursors]. *Epidemiol Prev*, 2012. 36(3-4 Suppl 1): p. e1-72.
56. Huh WK, Ault KA, Chelmow D, Davey DD, Goulart RA, Garcia FA, et al., Use of primary high-risk human papillomavirus testing for cervical cancer screening: interim clinical guidance. *Gynecol Oncol*, 2015. 136(2): p. 178-82.
57. Cancer Registry of Norway, [Annual Report of the Norwegian Cervical Cancer Screening Program, 2013-14]. Accessed at https://www.kreftregisteret.no/globalassets/publikasjoner-og-rapporter/livmorhalskreft/arsrapport/livmorhals_2015.pdf on October 16, 2015. Cancer Registry of Norway: Oslo, 2015.
58. Pedersen K, Burger EA, Campbell S, Nygård M, Aas E, and Lönnberg S, Advancing the evaluation of cervical cancer screening: development and application of a longitudinal adherence metric. *European Journal of Public Health*, 2017. doi:10.1093/eurpub/ckx073.
59. Cancer Registry of Norway, [Secondary screening with HPV testing in the Norwegian Cervical Cancer Screening Program]. Accessed at https://www.kreftregisteret.no/globalassets/publikasjoner-og-rapporter/rapport_sekundaerscreening_med_hpv_tester_i_masseundersokelsen_081211.pdf on January 20, 2013. Cancer Registry of Norway: Oslo, 2011.
60. Cancer Registry of Norway, [New national guidelines for HPV testing in secondary screening]. Accessed at <https://www.kreftregisteret.no/Generelt/Nyheter/Nye-nasjonale-retningslinjer-for-HPV-test-i-sekundarscreening/> on August 7, 2017.
61. Bosch FX and de Sanjose S, The epidemiology of human papillomavirus infection and cervical cancer. *Dis Markers*, 2007. 23(4): p. 213-27.

62. Pedersen K, Lonnberg S, Skare GB, Sorbye SW, Burger EA, and Kristiansen IS, [Costs of the Norwegian Cervical Cancer Screening Program]. *Sykepleien Forskning*, 2015. nb. 1, 2015(10): p. 62-71.
63. Burger EA, Ortendahl JD, Sy S, Kristiansen IS, and Kim JJ, Cost-effectiveness of cervical cancer screening with primary human papillomavirus testing in Norway. *Br J Cancer*, 2012. 106(9): p. 1571-8.
64. Norwegian Directorate of Health, [Economic evaluation of health interventions – A guide]. Accessed at <http://helsedirektoratet.no/publikasjoner/okonomisk-evaluering-av-helsetiltak--en-veileder/Publikasjoner/IS-1985.pdf> on September 9, 2014. 15.09.2014). Norwegian Directorate of Health: Oslo, 2012.
65. Nygard M, Roysland K, Campbell S, and Dillner J, Comparative effectiveness study on human papillomavirus detection methods used in the cervical cancer screening programme. *BMJ Open*, 2014. 4(1): p. e003460.
66. World Health Organization, Vaccine safety basics e-learning course. Module 1: Introduction to vaccine safety. Balancing efficacy and safety. Accessed at <http://vaccine-safety-training.org/balancing-efficacy-and-safety.html> on August 7, 2018.
67. Moreira ED, Jr., Block SL, Ferris D, Giuliano AR, Iversen OE, Joura EA, et al., Safety Profile of the 9-Valent HPV Vaccine: A Combined Analysis of 7 Phase III Clinical Trials. *Pediatrics*, 2016. 138(2).
68. Munoz N, Kjaer SK, Sigurdsson K, Iversen OE, Hernandez-Avila M, Wheeler CM, 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(5): p. 325-39.
69. Lehtinen M, Paavonen J, Wheeler CM, Jaisamrarn U, Garland SM, Castellsague X, 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(1): p. 89-99.
70. Joura EA, Giuliano AR, Iversen OE, Bouchard C, Mao C, Mehlsen J, et al., A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women. *N Engl J Med*, 2015. 372(8): p. 711-23.
71. Naud PS, Roteli-Martins CM, De Carvalho NS, Teixeira JC, de Borba PC, Sanchez N, et al., Sustained efficacy, immunogenicity, and safety of the HPV-16/18 AS04-adjuvanted vaccine: final analysis of a long-term follow-up study up to 9.4 years post-vaccination. *Hum Vaccin Immunother*, 2014. 10(8): p. 2147-62.
72. Paavonen J, Naud P, Salmeron J, Wheeler CM, Chow SN, Apter D, 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(9686): p. 301-14.
73. Garland SM, Kjaer SK, Munoz N, Block SL, Brown DR, DiNubile MJ, et al., Impact and Effectiveness of the Quadrivalent Human Papillomavirus Vaccine: A Systematic Review of 10 Years of Real-world Experience. *Clin Infect Dis*, 2016. 63(4): p. 519-27.

74. Malagon T, Drolet M, Boily MC, Franco EL, Jit M, Brisson J, et al., Cross-protective efficacy of two human papillomavirus vaccines: a systematic review and meta-analysis. *Lancet Infect Dis*, 2012. 12(10): p. 781-9.
75. Drolet M, Benard E, Boily MC, Ali H, Baandrup L, Bauer H, et al., Population-level impact and herd effects following human papillomavirus vaccination programmes: a systematic review and meta-analysis. *Lancet Infect Dis*, 2015. 15(5): p. 565-80.
76. Dobson SR, McNeil S, Dionne M, Dawar M, Ogilvie G, Krajden M, et al., Immunogenicity of 2 doses of HPV vaccine in younger adolescents vs 3 doses in young women: a randomized clinical trial. *Jama*, 2013. 309(17): p. 1793-802.
77. Romanowski B, Schwarz TF, Ferguson LM, Peters K, Dionne M, Schulze K, et al., Immunogenicity and safety of the HPV-16/18 AS04-adjuvanted vaccine administered as a 2-dose schedule compared with the licensed 3-dose schedule: results from a randomized study. *Hum Vaccin*, 2011. 7(12): p. 1374-86.
78. World Health Organization, Human papillomavirus vaccines: WHO position paper, May 2017 - Recommendations. *Vaccine*, 2017. doi: 10.1016/j.vaccine.2017.05.069.
79. Bruni L, Diaz M, Barrionuevo-Rosas L, Herrero R, Bray F, Bosch FX, et al., Global estimates of human papillomavirus vaccination coverage by region and income level: a pooled analysis. *Lancet Glob Health*, 2016. 4(7): p. e453-63.
80. Norwegian Institute of Public Health, [Change of HPV vaccine in the childhood immunization program]. Press release, March 13, 2017. Accessed at <https://www.fhi.no/nyheter/2017/ending-av-hpv-vaksine/> on May 23, 2017.
81. Norwegian Institute of Public Health, [Change from 3 to 2 doses of the HPV vaccine in the childhood immunization program]. Press release, March 15, 2017. Accessed at <https://www.fhi.no/meldinger/hpv-vaksine-2-dose/> on May 23, 2017.
82. Norwegian Ministry of Health and Care Services, [Free HPV vaccine to boys]. Press release (No: 29/2017), May 11, 2017. Accessed at <https://www.regjeringen.no/en/aktuelt/free-hpv-vaccine-to-boys/id2552726/> on May 23, 2017.
83. Kim JJ, Burger EA, Sy S, and Campos NG, Optimal Cervical Cancer Screening in Women Vaccinated Against Human Papillomavirus. *J Natl Cancer Inst*, 2017. 109(2).
84. Naber SK, Matthijsse SM, Rozemeijer K, Penning C, de Kok IM, and van Ballegooijen M, Cervical Cancer Screening in Partly HPV Vaccinated Cohorts - A Cost-Effectiveness Analysis. *PLoS One*, 2016. 11(1): p. e0145548.
85. Goldhaber-Fiebert JD, Stout NK, Salomon JA, Kuntz KM, and Goldie SJ, Cost-effectiveness of cervical cancer screening with human papillomavirus DNA testing and HPV-16,18 vaccination. *J Natl Cancer Inst*, 2008. 100(5): p. 308-20.
86. Simms KT, Smith MA, Lew JB, Kitchener HC, Castle PE, and Canfell K, Will cervical screening remain cost-effective in women offered the next generation nonavalent HPV vaccine? Results for four developed countries. *Int J Cancer*, 2016. 139(12): p. 2771-2780.

87. Norwegian Directorate of Health, [National goals and prioritization within health and care in 2015]. Accessed at <https://helsedirektoratet.no/Lists/Publikasjoner/Attachments/884/IS-1%202015.pdf> on June 25, 2017. Norwegian Directorate of Health: Oslo, 2015.
88. Norwegian Ministry of Health and Care Services. White paper: 19 (2014-2015). White paper on public health. Accessed at <https://www.regjeringen.no/contentassets/7fe0d990020b4e0fb61f35e1e05c84fe/no/pdfs/stm201420150019000dddpdfs.pdf> on June 25, 2017. Norwegian Ministry of Health and Care Services: Oslo, 2015.
89. Norwegian Ministry of Social Affairs, NOU 1991:10 [More healthy years of life for the population as a whole]. Accessed at <https://www.regjeringen.no/globalassets/upload/kilde/odn/tmp/2002/0034/ddd/pdfv/154756-nou1991-10.pdf> on June 25, 2017. Norwegian Ministry Social Affairs: Oslo, 1991.
90. Norwegian Ministry of Health and Care Services. Patients' Right Act §2.1. Accessed at https://lovdata.no/dokument/NL/lov/1999-07-02-63#KAPITTEL_2 on June 26, 2017.
91. Norwegian Ministry of Health and Care Services, Regulations for prioritization of healthcare services, right to necessary healthcare from specialist care, right to treatment abroad and about appeal board (The Regulation of Priority Setting). Last updated November 11, 2015. Accessed at <https://lovdata.no/dokument/SF/forskrift/2000-12-01-1208> on June 26, 2017.
92. Norwegian Ministry of Social Affairs, NOU 1987: 23 (1987). [Guidelines for priority setting within Norwegian healthcare]. Accessed at <https://www.regjeringen.no/globalassets/upload/kilde/odn/tmp/2002/0034/ddd/pdfv/154646-nou1987-23.pdf> on June 26, 2017. Norwegian Ministry of Social Affairs: Oslo, 1987.
93. Norwegian Ministry of Social Affairs. NOU 1997: 18 (1997). [Priority setting again]. Accessed at <https://www.regjeringen.no/contentassets/e2c9e765bf484b2ba4e05dd545a6302d/no/pdfa/nou199719970018000dddpdfa.pdf> on June 26, 2017. Norwegian Ministry of Social Affairs: Oslo, 1997.
94. Ottersen T, Førde R, Kakad M, Kjellevoid A, Melberg HO, Moen A, et al., A new proposal for priority setting in Norway: Open and fair. Health Policy, 2016. 120(3): p. 246-251.
95. Norwegian Ministry of Health and Care Services. NOU 2014:12 (2014). [Open and fair - priority setting in healthcare]. Accessed at <https://www.regjeringen.no/contentassets/16a0834c9c3e43fab452aelb6d8cd3f6/no/pdfs/nou201420140012000dddpdfs.pdf> on June 26, 2017. Norwegian Ministry of Health and Care Services: Oslo, 2014.
96. Norwegian Ministry of Health and Care Services. [Severity and priority setting]. Accessed at https://www.regjeringen.no/contentassets/d5da48ca5d1a4b128c72fc5daa3b4fd8/paa_ramme_alvor.pdf on June 26, 2017. Norwegian Ministry of Health and Care Services: Oslo, 2015.

97. Norwegian Ministry of Health and Care Services, White paper 34: 2015-2016. [Values in the patient's healthcare. Report on priority setting]. Accessed at <https://www.regjeringen.no/contentassets/439a420e01914a18b21f351143ccc6af/no/pdfs/stm201520160034000dddpdfs.pdf> on June 6, 2017. Norwegian Ministry of Health and Care Services: Oslo, 2016.
98. Neumann PJ, Sanders GD, Russell LB, Siegel JE, and Ganiats TG, Cost-effectiveness in health and medicine. 2016: Oxford University Press.
99. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, and Greg LS, Methods for the Economic Evaluation of Health Care Programmes. 3rd edition. Oxford, New York: Oxford University Press, 2005.
100. Drummond, Michael F., Claxton K., Stoddart G.L., Torrance, G.W. Methods for the economic evaluation of health care programmes. 4th edition. Oxford university press, 2015.
101. Weinstein MC, Torrance G, and McGuire A, QALYs: the basics. Value Health, 2009. 12 Suppl 1: p. S5-9.
102. Gold MR, Cost-effectiveness in health and medicine. 1996: Oxford university press.
103. Brouwer WB, Culyer AJ, van Exel NJ, and Rutten FF, Welfarism vs. extra-welfarism. J Health Econ, 2008. 27(2): p. 325-38.
104. Kahneman D and Tversky A, Prospect theory: An analysis of decision under risk. Econometrica: Journal of the econometric society, 1979: p. 263-291.
105. Loomes G and Sugden R, Regret theory: An alternative theory of rational choice under uncertainty. The economic journal, 1982. 92(368): p. 805-824.
106. Sanders GD, Neumann PJ, Basu A, and et al., Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second panel on cost-effectiveness in health and medicine. JAMA, 2016. 316(10): p. 1093-1103.
107. National Institute for Health and Care Excellence, Guide to the methods of technology appraisal 2013. Accessed at <https://www.nice.org.uk/guidance/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781> on August 8, 2017. National Institute for Health and Care Excellence: London, 2013.
108. Norwegian Medicines Agency. [Guidelines for pharmacoeconomic analyses]. Accessed at <https://legemiddelverket.no/Documents/Refusjon%20og%20pris/Refusjonss%C3%B8knad/Retningslinjer%20per%201%20mars%202012.pdf> on April 18, 2014. Norwegian Medicines Agency: Oslo, 2012.
109. O'Mahony JF, Naber SK, Normand C, Sharp L, O'Leary JJ, and de Kok IM, 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(8): p. 1138-51.

110. National Health Care Institute. Guideline for economic evaluations in healthcare. Accessed at <https://english.zorginstituutnederland.nl/publications/reports/2016/06/16/guideline-for-economic-evaluations-in-healthcare> on August 4, 2017. The Netherlands: National Health Care Institute, 2016.
111. Neumann PJ, Cohen JT, and Weinstein MC, Updating cost-effectiveness--the curious resilience of the \$50,000-per-QALY threshold. *N Engl J Med*, 2014. 371(9): p. 796-7.
112. Norwegian Directorate of Health. [Health effects of socio-economic analyses]. Norwegian Directorate of Health: Oslo, 2007. Available at: <https://helsedirektoratet.no/Lists/Publikasjoner/Attachments/643/Helseeffekter-i-samfunnsokonomiske-analyser-IS-1435.pdf>
113. Federal Reserve. Average rates of exchange 2014. Accessed at <http://www.federalreserve.gov/releases/G5A/current/> on November 22, 2015.
114. Statistics Norway, 2015. Accessed at <https://www.ssb.no/nasjonalregnskap-og-konjunkturer/artikler-og-publikasjoner/norsk-okonomi-i-moderat-fart--180967?tabell=180881> on December 11, 2015.
115. Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, et al., Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technol Assess*, 2015. 19(14): p. 1-503, v-vi.
116. O'Mahony JF and Coughlan D, The Irish Cost-Effectiveness Threshold: Does It Support Rational Rationing or Might It Lead to Unintended Harm of Ireland's Health System? *Value Health*, 2015. 18(7): p. A570.
117. Torrance GW and Feeny D, Utilities and quality-adjusted life years. *Int J Technol Assess Health Care*, 1989. 5(4): p. 559-75.
118. Keeney RL and Raiffa H, *Decisions with multiple objectives: preferences and value trade-offs*. 1993: Cambridge university press.
119. Krahn M, Bremner KE, Tomlinson G, Ritvo P, Irvine J, and Naglie G, Responsiveness of disease-specific and generic utility instruments in prostate cancer patients. *Qual Life Res*, 2007. 16(3): p. 509-22.
120. Menzel P, Dolan P, Richardson J, and Olsen JA, The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. *Soc Sci Med*, 2002. 55(12): p. 2149-58.
121. Gold MR, Stevenson D, and Fryback DG, HALYS and QALYS and DALYS, Oh My: similarities and differences in summary measures of population Health. *Annu Rev Public Health*, 2002. 23: p. 115-34.
122. Whitehurst DG and Bryan S, Another study showing that two preference-based measures of health-related quality of life (EQ-5D and SF-6D) are not interchangeable. But why should we expect them to be? *Value Health*, 2011. 14(4): p. 531-8.

123. Augestad LA, Stavem K, Kristiansen IS, Samuelsen CH, and Rand-Hendriksen K, Influenced from the start: anchoring bias in time trade-off valuations. *Qual Life Res*, 2016. 25(9): p. 2179-91.
124. Owens DK, Whitlock EP, Henderson J, Pignone MP, Krist AH, Bibbins-Domingo K, et al., Use of Decision Models in the Development of Evidence-Based Clinical Preventive Services Recommendations: Methods of the U.S. Preventive Services Task Force. *Ann Intern Med*, 2016. 165(7): p. 501-508.
125. Briggs A, Claxton K, and Sculpher M, Decision Modelling for Health Economic Evaluation. *Handbooks in Health Economic Evaluation Series*, ed. A.G.a.A. Briggs. 2006, New York, Us: Oxford University Press.
126. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, and Wong JB, Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. *Value Health*, 2012. 15(6): p. 843-50.
127. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, and Krahn M, Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--2. *Value Health*, 2012. 15(6): p. 804-11.
128. Caro JJ, Briggs AH, Siebert U, and Kuntz KM, Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--1. *Value Health*, 2012. 15(6): p. 796-803.
129. Pitman R, Fisman D, Zaric GS, Postma M, Kretzschmar M, Edmunds J, et al., Dynamic transmission modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--5. *Value Health*, 2012. 15(6): p. 828-34.
130. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, and Moller J, Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--4. *Value Health*, 2012. 15(6): p. 821-7.
131. Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ, et al., State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--3. *Med Decis Making*, 2012. 32(5): p. 690-700.
132. Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, and Paltiel AD, Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--6. *Value Health*, 2012. 15(6): p. 835-42.
133. Brennan A, Chick SE, and Davies R, A taxonomy of model structures for economic evaluation of health technologies. *Health Econ*, 2006. 15(12): p. 1295-310.
134. Scholz S and Mittendorf T, Modeling rheumatoid arthritis using different techniques - a review of model construction and results. *Health Econ Rev*, 2014. 4(1): p. 18.
135. Auguste P, Tsertsvadze A, Court R, and Pink J, A systematic review of economic models used to assess the cost-effectiveness of strategies for identifying latent tuberculosis in high-risk groups. *Tuberculosis (Edinb)*, 2016. 99: p. 81-91.

136. Hoang VP, Shanahan M, Shukla N, Perez P, Farrell M, and Ritter A, A systematic review of modelling approaches in economic evaluations of health interventions for drug and alcohol problems. *BMC Health Serv Res*, 2016. 16: p. 127.
137. O'Mahony JF, van Rosmalen J, Zauber AG, and van Ballegooijen M, Multicohort models in cost-effectiveness analysis: why aggregating estimates over multiple cohorts can hide useful information. *Med Decis Making*, 2013. 33(3): p. 407-14.
138. Vanni T, Karnon J, Madan J, White RG, Edmunds WJ, Foss AM, et al., Calibrating models in economic evaluation: a seven-step approach. *Pharmacoeconomics*, 2011. 29(1): p. 35-49.
139. Stout NK, Knudsen AB, Kong CY, McMahon PM, and Gazelle GS, Calibration methods used in cancer simulation models and suggested reporting guidelines. *Pharmacoeconomics*, 2009. 27(7): p. 533-45.
140. Le QA, Structural Uncertainty of Markov Models for Advanced Breast Cancer: A Simulation Study of Lapatinib. *Med Decis Making*, 2016. 36(5): p. 629-40.
141. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al., Consolidated Health Economic Evaluation Reporting Standards (CHEERS)-- explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value Health*, 2013. 16(2): p. 231-50.
142. The Central Bank of Norway: Average exchange rates 2014. Accessed at <http://www.norgesbank.no/Statistikk/Valutakurser/valuta/EUR/> on January 19, 2016.
143. Pedersen K, Sørbye SW, Burger EA, Lönnberg S, and Kristiansen IS, Using Decision-Analytic Modeling to Isolate Interventions That Are Feasible, Efficient and Optimal: An Application from the Norwegian Cervical Cancer Screening Program. *Value in Health*, 2015. 18(8): p. 1088-1097.
144. Castle PE, Schiffman M, Wheeler CM, and Solomon D, Evidence for frequent regression of cervical intraepithelial neoplasia-grade 2. *Obstet Gynecol*, 2009. 113(1): p. 18-25.
145. Campos NG, Burger EA, Sy S, Sharma M, Schiffman M, Rodriguez AC, et al., An updated natural history model of cervical cancer: derivation of model parameters. *Am J Epidemiol*, 2014. 180(5): p. 545-55.
146. Kim JJ, Kuntz KM, Stout NK, Mahmud S, Villa LL, Franco EL, et al., Multiparameter calibration of a natural history model of cervical cancer. *Am J Epidemiol*, 2007. 166(2): p. 137-50.
147. Goldie SJ, Grima D, Kohli M, Wright TC, Weinstein M, and Franco E, A comprehensive natural history model of HPV infection and cervical cancer to estimate the clinical impact of a prophylactic HPV-16/18 vaccine. *Int J Cancer*, 2003. 106(6): p. 896-904.
148. Goldhaber-Fiebert JD, Stout NK, Ortendahl J, Kuntz KM, Goldie SJ, and Salomon JA, Modeling human papillomavirus and cervical cancer in the United States for analyses of screening and vaccination. *Popul Health Metr*, 2007. 5: p. 11.

149. Kim JJ, Campos NG, Sy S, Burger EA, Cuzick J, Castle PE, et al., Inefficiencies and High-Value Improvements in U.S. Cervical Cancer Screening Practice: A Cost-Effectiveness Analysis. *Ann Intern Med*, 2015.
150. Kim JJ, Wright TC, and Goldie SJ, Cost-effectiveness of alternative triage strategies for atypical squamous cells of undetermined significance. *Jama*, 2002. 287(18): p. 2382-90.
151. Stout NK, Goldhaber-Fiebert JD, Ortendahl JD, and Goldie SJ, Trade-offs in cervical cancer prevention: balancing benefits and risks. *Arch Intern Med*, 2008. 168(17): p. 1881-9.
152. Burger EA, Sy S, Nygard M, and Kim JJ, The cost-effectiveness of cervical self-sampling to improve routine cervical cancer screening: The importance of respondent screening history and compliance. *Cancer Epidemiol Biomarkers Prev*, 2016.
153. Burger EA and Kim JJ, The value of improving failures within a cervical cancer screening program: an example from Norway. *Int J Cancer*, 2014. 135(8): p. 1931-9.
154. Kim JJ, Wright TC, and Goldie SJ, Cost-effectiveness of human papillomavirus DNA testing in the United Kingdom, The Netherlands, France, and Italy. *J Natl Cancer Inst*, 2005. 97(12): p. 888-95.
155. Campos NG, Tsu V, Jeronimo J, Njama-Meya D, Mvundura M, and Kim JJ, Cost-effectiveness of an HPV self-collection campaign in Uganda: comparing models for delivery of cervical cancer screening in a low-income setting. *Health Policy Plan*, 2017.
156. Campos NG, Mvundura M, Jeronimo J, Holme F, Vodicka E, and Kim JJ, Cost-effectiveness of HPV-based cervical cancer screening in the public health system in Nicaragua. *BMJ Open*, 2017. 7(6): p. e015048.
157. Campos NG, Maza M, Alfaro K, Gage JC, Castle PE, Felix JC, et al., The comparative and cost-effectiveness of HPV-based cervical cancer screening algorithms in El Salvador. *Int J Cancer*, 2015. 137(4): p. 893-902.
158. Statistics Norway, 2015. URL: <http://ssb.no/befolkning/statistikker/dode/aar/2015-03-26?fane=tabell&sort=nummer&tabell=220943> (Accessed: October 16, 2015).
159. Tjalma WA, Fiander A, Reich O, Powell N, Nowakowski AM, Kirschner B, et al., Differences in human papillomavirus type distribution in high-grade cervical intraepithelial neoplasia and invasive cervical cancer in Europe. *Int J Cancer*, 2013. 132(4): p. 854-67.
160. Cancer Registry of Norway. The Norwegian Cervical Cancer Screening Program. 2016: Oslo, Norway.
161. Burger EA, Sy S, Nygard M, Kristiansen IS, and Kim JJ, Prevention of HPV-related cancers in Norway: cost-effectiveness of expanding the HPV vaccination program to include pre-adolescent boys. *PLoS One*, 2014. 9(3): p. e89974.
162. Olsen J and Jepsen MR, Human papillomavirus transmission and cost-effectiveness of introducing quadrivalent HPV vaccination in Denmark. *Int J Technol Assess Health Care*, 2010. 26(2): p. 183-91.

163. Myers E. GS and I. L, Patient preferences for health states related to HPV infection: visual analogue scales versus time trade-off elicitation [abstract]. Poceeding of the 21st International Papillomavirus Conference. Mexico City, Mexico. 2004.
164. Rannestad T and Skjeldestad FE, Pain and quality of life among long-term gynecological cancer survivors: a population-based case-control study. *Acta Obstet Gynecol Scand*, 2007. 86(12): p. 1510-6.
165. Klee M, Thranov I, and Machin Prof D, The patients' perspective on physical symptoms after radiotherapy for cervical cancer. *Gynecol Oncol*, 2000. 76(1): p. 14-23.
166. Massad LS, Einstein MH, Huh WK, Katki HA, Kinney WK, Schiffman M, et al., 2012 updated consensus guidelines for the management of abnormal cervical cancer screening tests and cancer precursors. *Obstet Gynecol*, 2013. 121(4): p. 829-46.
167. Arbyn M, Xu L, Verdoodt F, Cuzick J, Szarewski A, Belinson JL, et al., Genotyping for Human Papillomavirus Types 16 and 18 in Women With Minor Cervical Lesions: A Systematic Review and Meta-analysis. *Ann Intern Med*, 2017. 166(2): p. 118-127.
168. Isidean SD, Mayrand MH, Ramanakumar AV, Rodrigues I, Ferenczy A, Ratnam S, et al., Comparison of Triage Strategies for HPV-Positive Women: Canadian Cervical Cancer Screening Trial Results. *Cancer Epidemiol Biomarkers Prev*, 2017. 26(6): p. 923-929.
169. Nygard JF, Nygard M, Skare GB, and Thoresen SO, Pap smear screening in women under 30 in the Norwegian Coordinated Cervical Cancer Screening Program, with a comparison of immediate biopsy vs Pap smear triage of moderate dysplasia. *Acta Cytol*, 2006. 50(3): p. 295-302.
170. Goldie SJ, Weinstein MC, Kuntz KM, and Freedberg KA, The costs, clinical benefits, and cost-effectiveness of screening for cervical cancer in HIV-infected women. *Ann Intern Med*, 1999. 130(2): p. 97-107.
171. Krol M, Papenburg J, and van Exel J, Does including informal care in economic evaluations matter? A systematic review of inclusion and impact of informal care in cost-effectiveness studies. *Pharmacoeconomics*, 2015. 33(2): p. 123-35.
172. A OC, O'Mahony JF, O'Connor M, O'Leary J, Normand C, Martin C, et al., Health-related quality of life as measured by the EQ-5D in the prevention, screening and management of cervical disease: A systematic review. *Qual Life Res*, 2017.
173. van Rosmalen J, de Kok IM, and van Ballegooijen M, Cost-effectiveness of cervical cancer screening: cytology versus human papillomavirus DNA testing. *Bjog*, 2012. 119(6): p. 699-709.
174. Wentzensen N, Fetterman B, Castle PE, Schiffman M, Wood SN, Stiemerling E, et al., p16/Ki-67 Dual Stain Cytology for Detection of Cervical Precancer in HPV-Positive Women. *J Natl Cancer Inst*, 2015. 107(12): p. djv257.
175. Engesaeter B, van Diermen Hidle B, Hansen M, Moltu P, Staby KM, Borchgrevink-Persen S, et al., Quality assurance of human papillomavirus (HPV) testing in the implementation of HPV primary screening in Norway: an inter-laboratory reproducibility study. *BMC Infect Dis*, 2016. 16(1): p. 698.

176. Weinstein MC and Stason WB, Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med*, 1977. 296(13): p. 716-21.
177. Katki HA, Kinney WK, Fetterman B, Lorey T, Poitras NE, Cheung L, et al., Cervical cancer risk for women undergoing concurrent testing for human papillomavirus and cervical cytology: a population-based study in routine clinical practice. *Lancet Oncol*, 2011. 12(7): p. 663-72.
178. Enerly E, Bonde J, Schee K, Pedersen H, Lonnberg S, and Nygard M, Self-Sampling for Human Papillomavirus Testing among Non-Attenders Increases Attendance to the Norwegian Cervical Cancer Screening Programme. *PLoS One*, 2016. 11(4): p. e0151978.
179. Verdoodt F, Jentschke M, Hillemanns P, Racey CS, Snijders PJ, and Arbyn M, 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(16): p. 2375-85.
180. Guirguis-Blake JM, Henderson JT, and Perdue LA, Periodic Screening Pelvic Examination: Evidence Report and Systematic Review for the US Preventive Services Task Force. *Jama*, 2017. 317(9): p. 954-966.
181. Bosch FX, Robles C, Diaz M, Arbyn M, Baussano I, Clavel C, et al., HPV-FASTER: broadening the scope for prevention of HPV-related cancer. *Nat Rev Clin Oncol*, 2016. 13(2): p. 119-32.
182. Moen KA, Kumar B, Qureshi S, and Diaz E, Differences in cervical cancer screening between immigrants and nonimmigrants in Norway: a primary healthcare register-based study. *Eur J Cancer Prev*, 2016.

10 PAPERS I-IV

Paper I

Paper II

Paper III

Paper IV