

Economic evaluation of screening

Eline Aas, Emily Burger, Kine Pedersen

Summary

The objective of medical screening is to prevent future disease (secondary prevention) or to improve prognosis by detecting the disease at an earlier stage (early detection). Screening as a prevention approach has gained increased acceptances following advances in screening technologies. Introducing a screening programme is resource demanding, therefore stakeholders emphasize the need for comprehensive evaluation, where costs and health outcomes are reasonably balanced, before any programme is offered to the population.

Economic evaluation of population-based screening programmes, defined as examination of individuals with no symptoms of disease in order to detect a disease or condition, include evaluation of health benefits and costs of all relevant strategies. The alternative screening strategies will be determined by starting- and stopping age, frequency of the screening (e.g. once only, biannual or every five years) and follow-up regimens after a positive test result. Based on the evaluation of health benefits and costs of all alternative strategies, the efficiency frontier would display the efficient strategies. The cost-effectiveness threshold would determine which of the efficient strategies that would be the recommended cost-effective screening strategy.

Similar to preventive interventions in general, estimating the health benefit (life-years and QALYs gained) of screening is challenging since the health benefit accumulate after several years. As the cost of screening are initial costs, the imbalance in cost- and health benefit accumulation when estimating the net present value (NPV) of each strategy, the effect of

discounting on health benefits are substantial. Reporting both discounting and undiscounted results would identify the effect of discounting. In addition, intermediate outcome measures, such as number of positive tests, cases detected and events prevented could be valuable supplement outcomes to report.

Estimating the cost-effectiveness of screening strategies are often based on decision analytic models, synthesising all relevant evidence from clinical trials, literature, guidelines and registers. In decision analytic models, it is possible to include intermediate results from trials and extrapolate to primary endpoints, such as incidence and mortality, by means of sophisticated calibration methods. Further, decision analytic models are also valuable, as a large amount of screening alternatives can be evaluated simultaneously, which are not possible in a randomised controlled trial (RCT) setting. Evaluation of screening based on RCT data are valuable as both costs and health benefits are measured for the same individual, enabling more advanced analysis of the interaction of costs and health benefits.

Screening is a complex system that usually involves multiple stakeholders and multiple decisions to be made along the screening pathway; as such, screening evaluation is a challenging task. In a societal perspective, other outcomes than the cost-effectiveness could be of value for decision makers, such as distributional concerns, severity of the disease and capacity (number of examination following a positive screening test). Analysis of harm-benefit trade-offs could be a useful tool to supplement the cost-effectiveness analysis. Decision analytic models are often based on 100% participation, which is rarely the case in practice. If those participating are different from those not choosing to participate, with regard to for instance risk of the disease or condition, this would result in selection bias in the results based on 100% participation. The development of new methods require re-evaluation of the cost-effectiveness

of screening. If treatment of a disease becomes more efficient, screening becomes less cost-effective, and the introduction of vaccines, such as the HPV-vaccination for cervical cancer, would influence the cost-effectiveness of screening. With the access to individual level data from registers, there is an opportunity to refine the analysis better representing heterogeneity and long-term consequences of screening on health behaviour.

Keyword: Population-based screening programme, cost-effectiveness, decision analytic models, screening strategies, clinical trials, uncertainty

Introduction

Medical screening involves identifying the likely presence of a disease or condition in individuals perceived to be healthy, using diagnostic tests, examinations or other medical procedures (Wilson & Jugner, 1968). The objective of medical screening is thus to prevent future disease (secondary prevention) or to improve prognosis by detecting the disease at an earlier stage (early detection). Screening as a prevention approach has gained increased acceptance following advances in screening technologies and the success of screening programmes (Vaccarella et al. 2014, Elstrom et al., 2015, Arbyn et al., 2010). There are generally two types of screening: Population-based, which involves inviting all individuals in a target group (for instance selected by age) to an examination or screening test; and, high-risk screening, which involves restricting screening to higher-risk groups. While population-based screening includes both prevention and early detection (such as screening for cancer), high-risk screening mainly involves early detection (e.g. screening for tuberculosis among immigrants).

According to the World Health Organization (Wilson & Jugner, 1968, Andermann et al, 2008, Dobrow et al., 2018), several principles and practice of screening should be considered prior to implementing population-based screening for a disease or condition. From an economics perspective, some principles are of particular interest; the target population is unambiguously defined (e.g. all new-born infants, women aged 60 years, or recent immigrants from countries with a high prevalence of tuberculosis); the test is acceptable for use on the population (e.g. not painful, sufficiently accurate); evidence for the effectiveness has been established in the scientific literature (e.g. risk reduction in cancer mortality and/or cancer incidence); the expected benefits of screening outweigh the harms (i.e. balancing the health gains with adverse events, costs and other potential negative consequences); and planned evaluation of the program from the outset. Based on these principles, a broad variety of screening programs across

different disease areas have been implemented. For example, in England, ten different population-based screening programmes are offered (Table 1).

Table 1. List of population-based screening programmes offered in England^a

Screening Programme	Screening test(s)	Target Population/ Age(s)	Target Disease(s)	Number of International CEAs (Year)^b
Abdominal Aortic Aneurysm Screening Programme	Ultrasound test	Men aged 65 and older	Abdominal Aortic Aneurysm	6 (2007-2016)
Cancer Screening				
Bowel Cancer Screening Programme	Faecal occult blood	Every 2 years to all men and women aged 60-74 years	Early detection and prevention of colorectal cancer	20 (1998-2016)
Breast Screening Programme	Mammography	Women aged 50-70 years	Early detection of breast cancer	24 (1991-2016)
Cervical Screening Programme	Liquid-based cytology; human papillomavirus testing (triage)	Every 3 years for women aged 25-49; every 5 years for women aged 50-64 years	Early detection and prevention of cervical cancer	52 (1990-2016)
Prenatal and Infant Programmes				
Fetal Anomaly Screening Programme	Ultrasound to measure fluid at the back of the fetus' neck combined with blood test	Expectant mothers at between 10+0 and 14+1 weeks of their pregnancy	Down's syndrome (T21), Edwards' syndrome (T18), Patau's syndrome (T13), and 9 additional conditions	2 (2007-2015)
Infectious Diseases in Pregnancy	Blood sample	All expectant mothers	HIV, hepatitis B, syphilis	None

Screening Programme				
Newborn and Infant Physical Examination Screening Programme	Physical examination	All infants within 72 hours of birth, and again between 6 to 8 weeks	Congenital heart disease, developmental dysplasia of the hip, congenital cataracts, cryptorchidism (undescended testes)	None
Newborn Blood Spot Screening Programme	Blood from child's heel	All infants 5 days old	Sickle cell disease, cystic fibrosis, congenital hypothyroidism, inherited metabolic diseases (IMDs).	None
Newborn Hearing Screening Programme	Automated otoacoustic emission (AOAE), automated auditory brainstem response (AABR)	All eligible infants ≤5 weeks old	Hearing impairment	None
Sickle Cell and Thalassaemia Screening Programme	Blood sample	All expectant mothers, father-to-be if mother is genetic carrier	Genetic carriers for sickle cell, thalassaemia and other haemoglobin disorders sickle cell disease thalassaemia haemoglobin disorders	None
<p>^ahttps://www.gov.uk/topic/population-screening-programmes;</p> <p>^bhttp://healtheconomics.tuftsmedicalcenter.org/cear4/ (12 Mar 2018) performed world-wide Abbreviations. CEA: Cost-effectiveness Analysis</p>				

Although screening is increasingly common, it has often been implemented without thorough evaluation. Because the introduction of a screening programme is resource demanding, stakeholders emphasize the need for comprehensive evaluation, where the health benefits and

monetary costs are reasonably balanced, prior to offering any programme to the population. (NICE, 2017, CADTH, 2017, Dutch guidelines, 2016 and Holland et al., 2006).

Evaluation of population based screening programmes should include both the long-term health and economic implications of a programme as well as evaluate all variations of relevant strategies. Compared to pharmaceuticals or new technologies, there are specific properties of screening that require special consideration during the evaluation process, and complicate the sole use of randomized clinical trials (RCTs) to complete the evaluation process. For example, there are often multiple screening approaches that needs to be considered, such as type of screening test, the screening target groups (e.g., ages for when to start and stop screening); the screening frequency (e.g., once-only, annual, biannual) and how to follow-up screen-positive individuals (the triage algorithm). In addition, the benefits and costs of the programme should be evaluated over a sufficiently long time horizon (often involving decades or a lifetime) to capture all relevant outcomes of the screening programme. Together, these challenges complicate the evaluation of screening programs and warrant the use of complex methods (such a decision-analytic/disease simulation models) to comprehensively estimate the value of a screening programme.

This paper provides an overview of evaluation of population based screening programmes, for any type of disease or condition, within the framework of economic evaluation. Due to the characteristics and complexities of screening, the scope of the paper is on economic evaluation of screening programmes using decision-analytic modelling, which allows evaluation an array of mutual exclusive screening algorithms. In addition, a discussion of how screening can be evaluated using data from RCTs and observational studies (primary sources of data used to inform decision-analytic models) is provided. Evaluation of screening for cancer is used as an

examples, as there exists several good examples of economic evaluation based on decision-analytic models in this disease area.

The properties of screening and how screening can influence the natural history of disease will be defined in the next section, followed by a presentation of the framework of cost-effectiveness analysis. Then the analytic frameworks will be described, including both the use of individual level data and decision-analytic models, followed by a discussion of the challenges in evaluation of population-based screening programmes. Finally, some concluding remarks will be provided.

Screening

The natural history of a disease and how medical screening could affect the disease pathway, is illustrated in **Figure 1**. The disease in question has a biological onset; without screening, the disease would be detected during the clinical phase, based on symptoms. The characteristics of a screening test (i.e., sensitivity and specificity) determine the ability of the screening test to detect the disease being screened for. The sensitivity of a test is defined as the probability of having a positive test result given that you have the disease, while the specificity of a test is defined as probability of having a negative test result given that you do not have the disease. The earlier a disease is detected by screening, the shorter is the delay time. Earlier detection of the disease also increases the lead time, i.e. the amount of time the diagnosis of the disease is advanced due to screening.

Figure 1: *The natural history of disease (panel A) and illustration of where screening would affect the disease pathway, with screening for cervical cancer as an example (panel B)*

Two types of biases can occur in evaluation of the effectiveness of a screening programme: Length bias and lead-time bias. Length bias occurs because screen-detected diseases may inherently be less aggressive forms of the disease. In contrast, an aggressive disease may have a shorter screen-detectable period, defined as the period in which the disease can be diagnosed with a screening test (**Figure 1**). Therefore, the most aggressive forms of diseases are diagnosed due to the development of symptoms. The less aggressive diseases may be over-represented in the screening cohort because they have a longer asymptomatic period of time (time of the screening examination(s)). If the probability of cure and survival is higher for a patient with a less aggressive form of disease than a patient with more aggressive disease, a screen-detected aggressive disease may not face the same survival improvements as a screen-detected less aggressive disease. Lead-time bias can occur when screening falsely prolongs survival. Due to screening, the disease is diagnosed earlier than when diagnosed due to symptoms, but the outcome in terms of time of death, remains unchanged by the screening intervention. Consequently, the patient will not benefit from screening (no additional life years), but because of an earlier diagnosis, the individual will be aware of and anxious about the disease for a longer period of time. Furthermore, screening may also induce overdiagnosis, i.e. detect diseases that would never have been diagnosed due to symptoms. The challenge of overdiagnosis has been addressed in several screening programmes, such as breast- and prostate cancer screening (Vickers et al, 2014, Etzioni et al, 2002, Jørgensen et al, 2017), which has been shown to result in unnecessary treatments and follow-ups and reduce overall quality of life and.

The majority of screening tests are imperfect (i.e., the sensitivity and specificity is less than 100%); consequently, some individuals who have the disease are falsely negative and sent back to routine screening, while some healthy individuals will have a false positive test result and will be advised additional follow-up. A positive test can trigger a cascade of additional testing

including potentially anxiety-inducing and painful diagnostic testing (e.g., biopsies, performance of the colon during colonoscopy) to ascertain the diagnosis (Lin et al, 2016, Henderson et al 2018, Habbema et al 2017). These potential negative consequences of screening are often overlooked and under-researched, but should play an important role in comprehensive screening evaluations (Brewer et al, 2007, Gareen et al 2014).

As screening involves inviting a healthy population without any symptoms to prevent or detect a few number of cases, it is essential that the test is deemed acceptable by individuals in the target population. Acceptance is important to ensure a high participation rate – the benefits of screening can only be achieved if individuals actually participate in screening. In sum, the choice of screening test may depend on availability of required resources, existing qualifications of health personnel, and individual preferences.

In a population-based screening programme, cohorts are invited to a standardized screening programme with a specific follow-up algorithm, which is carefully evaluated for key-performance indicators such as coverage, timely assessment, quality and standardization of laboratories (Public Health England, 2017). Contrary to organized programmes, opportunistic screening is unorganized screening of people and relies on an individual's own initiative to participation and follow-up recommendations. Opportunistic screening could therefore contain several non-standardised screening methods (combination of screening test and age) and result in heterogeneity in the follow-up algorithm. Evaluation of the costs and health outcomes of a screening programme may be impacted by opportunistic screening, particularly if it is frequently used and of relatively high quality. A large proportion of opportunistic screening may reduce the incremental health effect of a screening programme.

Cost-effectiveness framework

In economic evaluations the net present value (NPV) of all relevant (often long-term) health benefits and monetary costs of alternative screening strategies are calculated. The two most common types of evaluations are cost-effectiveness analysis (CEA) and cost-utility analysis (CUA), both measuring the health benefits in natural units, but the latter in quality-adjusted life-years (QALYs). CEAs, particularly those evaluating public health programmes, such as population based screening programmes, are used to inform population-level decisions, but are not aimed at directly informing individual-level decisions.

The valuation of each screening strategy is represented by the incremental cost-effectiveness ratio (ICER), calculated by taking the differences in total costs between one screening strategy compared to the next least costly strategy divided by the difference in health benefits between the two strategies (**Figure 2**). Screening strategies that are more costly but provide less health benefits than other strategies are removed from further consideration (i.e., strongly dominated). Similarly, strategies that are more costly, but less cost-efficient (i.e., higher ICER), than the next strategy are considered weakly dominated are also removed from further consideration. The remaining non-dominated strategies are considered “cost-efficient” and can be plotted on the efficiency frontier (**Figure 2**). While all strategies on the efficiency frontier are considered “cost-efficient”, only one strategy would be considered optimal, or “cost-effective”.

Figure 2: The efficiency frontier and the ICER when comparison several screening strategies.

In a recent review study of the Tufts Medical Center CEA Registry (<http://healtheconomics.tuftsmedicalcenter.org/cear4/>) and the Global Cost-Effectiveness Analysis (GHCEA) Registry, 12% of all CEAs evaluated screening programmes (Neumann 2018). In contrast, the majority of CEAs (44%) evaluate pharmaceuticals (Neumann 2018),

usually with the aim to inform market access and reimbursement decisions. For evaluation of screening, the aim of the CEA may be broader; that is, a screening programme may have already been implemented, and the aim of the CEA is to inform optimal refinement of the programme following advances in screening technologies (e.g. Mendes et al 2015). In refining an existing screening programme, decision makers are faced with the challenging task of considering a large number of candidate strategies. In this context, a CEA can help inform which strategies are more efficient than others, and thus help narrow down the number of strategies decision-makers need to consider (Pedersen et al 2015).

To identify the optimal or most “cost-effective” intervention, an external decision rule is applied in order to determine where on the efficiency frontier to operate. The decision rule indicates the amount society is willing to pay for an additional health benefit, commonly referred to as the “willingness to pay threshold” or “cost-effectiveness threshold” (CE threshold), often expressed as the additional costs per additional QALY gained (Weinstein et al. 2009). The preferred intervention is the intervention with an ICER just below the CE threshold. While there is no universal criterion for what defines a CE threshold, benchmarks exist for some countries (e.g. Sweden and Denmark). For example, the National Institute for Health and Care Excellence (NICE), which issues guidance on high-quality and value-driven healthcare in England and Wales, has defined interventions with an ICER between £20,000 to £30,000 per QALY gained to be considered cost-effective (NICE 2017). Importantly, the optimal strategy identified by the CE threshold is rarely the sole contributor to decision-making; other factors such as distributional and ethical concerns, feasibility, affordability, political discourse may contribute to an alternative choice of optimal strategy. Even so, quantifying the trade-offs in health benefits and resource use improve the accountability and transparency of decision-making.

In order to increase transparency of analyses, The Second Panel on Cost-effectiveness analysis in the US (Neumann et al., 2016) recommended to include an impact inventory table as part of the analysis. Within the context of screening evaluation, different analytic viewpoints can inform different aspects of the decision-making process. For example, while a societal perspective can inform which strategy is optimal for the society, using a healthcare perspective can inform whether a screening strategy is feasible. Even though a screening strategy may be cost-effective, it may not be feasible within current capacity constraints of the healthcare system. For example, a more intensive screening strategy may cause more screen-positive results, requiring more follow-up tests and thus more pressure on pathology laboratories. However, pathology resources are often limited and changes in a screening programme may cause bottlenecks in the system.

Health benefits and costs

Health benefits in screening are most commonly measured in life-years or QALYs, but also with intermediate outcomes, such as cases detected or events avoided (e.g. reduction in cancer mortality or avoided stroke). In order to calculate of the total health benefit of screening, requires monitoring of survival until everybody exposed and not exposed to screening have died. For a trial among 50 years old, this would result in monitoring survival for more than 40 years before the total benefit is revealed. For this reason, the use of intermediate outcomes are important, for calibrating long-term health benefits of screening. Further, similar as for other preventive initiatives, where the main objective is to prevent future disease, the health benefit will not occur immediately, but often after several years, which implies that time horizon is required to be a lifetime perspective.

Evaluating screening in a health care perspective would typically include cost of the screening test, adverse events, follow-up procedures, and treatment costs. Extending to a societal perspective, would potentially have great impact on the estimation of costs. Including the cost of travelling and the cost of time related to participating in screening and additional examinations as consequence of a positive test result, would increase the cost of screening considerable. As guidelines in several countries are recommending a health care perspective (e.g. UK and Norway), these costs will not be included. As population-based screening programmes invite individuals without any symptoms of disease, the majority are likely to be working. Hence, in order to participate, the time allocated to travelling and the screening test, would either be working hours or leisure, both with an opportunity cost of time associated with it.

Calculating the NPV of screening strategies, it is recommended that both health benefits and costs are discounted at an equal rate (Neumann et al 2016); however, this has been subject to debate (Paulden et al 2017). In the calculation of NPV, another challenge emerge, the imbalance in timing of the accumulation of costs and health benefits. While the cost of screening occur initially, the health benefits accumulate after several years. Hence, the effect of discounting is therefore substantial, and reporting both discounted and undiscounted health benefits could be relevant to identify the health benefits in absolute terms. Supplementing the result section with intermediate health outcomes, that are highly correlated with the main health outcome (avoided cases), may overcome some of the challenges with measuring long-term health benefits.

Analytic framework

Clinical trial based evaluations

The preferred methodology for evaluating the effect of a new technology, screening methods included, is RCT. Because the intervention is assigned to individuals at random, and blinded when possible, the RCT framework ensures that the group invited to screening is statistically equivalent to the control group, and any resulting differences in costs and health benefits are directly caused by screening alone (Wooldridge, 2010, Angrist and Pischke, 2015). In an RCT, costs and health outcomes should be collected simultaneously for all individuals, which provide an opportunity for advanced analysis on the interaction of costs and benefits. The analytic approach is intention-to-treat (ITT), where the average costs and health outcomes for all invited to screening is compared with the average outcomes in the control group. The discounted health benefit is measured by area under the curve (AUC) for both life-years gained and QALYs, discounted over the time horizon, while costs will be estimated as NPV of costs accumulated over the observation period. To account for uncertainty, bootstrap is typically applied to identify the variation in costs and health benefits between the control and the screening group. Importantly, evaluators should consider the potential impact of length- and lead time biases, which can conflate survival and QALY gains, particularly when interventions are evaluated before all individuals have died. See Glick et al., (2007) for additional information on conducting a CEA in a trial setting.

An advantage with individual level data from an RCT is the opportunity to more detailed analysis on heterogeneity, which could potentially give recommendations for sub-groups and not only for the average population. For instance, with individual level data the effect of gender and comorbidity on costs and health benefits could be calculated, further differences in costs between cases detected at screening and cases detected later could also be estimated. Adverse

events of screening are easier to identify in trials and could also be linked to short- and long-term costs and health benefits.

Further, individual-level data from an RCT enable more in-depth analysis on heterogeneity, which could potentially provide recommendations for sub-groups, rather than only for the average population. For example, individual-level data can explore the effects of gender and comorbidity on costs and health benefits. Adverse events of screening may also be easier to identify in trials, which provides opportunities to link the short- and long-term costs and health benefits.

When an RCT has not been conducted to establish the effect of screening, specific statistical methods can be applied to observational data to identify the causal effect of screening on costs and health outcomes (Wooldridge, 2010, Angrist and Pischke, 2015). This is, however, challenging due to potential selection bias. Analysts can apply instrumental variable analysis to adjust for selection bias. The instrument must fulfil certain properties: 1) needs to have a causal effect on being allocated to screening, 2) the instrument must be randomly assigned and independent of other variables (independence assumption), and 3) the instrument cannot affect the main outcome variables (life-years gained, QALYs or costs) other than through screening (exclusion restriction).

Another approach to evaluate empirical, non-randomized screening data is to use regression of discontinuity design (Angrist and Pischke, 2015). This method estimates the effect of the introduction of screening on the health benefit, such as mortality, by estimating whether there is a discontinuity in trend after introduction of screening. Introduction of screening will be identified by a dummy variable, defined as 1 after screening and 0 before the screening. One

important assumption for using this method is that it requires a sharp introduction of screening, requiring that there was very little or no screening (including opportunistic screening) prior to the introduction of the programme.

Lastly, the difference in difference (DID) approach can also be used to evaluate the benefit of screening (Angrist and Pischke, 2015). DID compare two groups before and after introduction of screening, where one group is allocated to screening and the other is not offered screening (e.g. DID can be applied in two different geographical areas). DID compared the trends in an outcome, such as mortality, before and after introduction of screening. The health benefit of screening on mortality is measured as the interaction between being allocated to screening and the post screening period (referred to as the DID estimator). A primary assumption of DID requires that the trend in the two groups would have been similar if screening was not introduced in one group.

Decision-analytic models

In the context of evaluating screening programmes, in which multiple competing strategies are evaluated and the health gains may not be observed for several decades, the use of decision-analytic modelling (simulation or mathematical models) may be the preferred approach. In contrast to statistical models, mathematical models are often designed to capture some underlying natural history process (e.g., progression from healthy to precancer to cancer). (**Figure 1B**). No single clinical trial can capture all the short and long-term health and economic consequences of all possible strategies needed to inform complex policy decisions, for example, surrounding the alternative preventive strategies for colorectal or cervical cancer screening. An advantage of modelling is the ability to synthesize available evidence (e.g., RCTs, observational studies, registry data) from multiple sources, extrapolate data beyond the time horizon of

studies, reflect parameter and process uncertainty, and identify the most influential factors on the decision. As such, decision-analytic modelling allows evaluation of scenarios not evaluated in an RCT, such as other target groups, screening intervals or follow-up of screen-positive results.

As decisions related to health outcomes involve choices between risky or uncertain prospects, expected utility theory provides the foundation behind decision-analytic modeling (Neumann 2016). Within the framework of expected utility theory, alternative pathways are characterized by numerical representations of the outcomes and the respective probability of achieving each outcome. The use of decision-analytic modeling in health policy is not without controversy, stemming from accusations of less transparent methods (Buxton 1997) and strong assumptions. However, following recommendations for best practices (e.g., ISPOR-SMDM Modeling Good Research Practices) and using standardized reporting guidelines (Drummond 2005, Husereau 2013) instils greater confidence in model-based analyses. Subsequently, there is a growing dependence on decision-analytic methods to help guide drug and technology reimbursement decisions in, for example, England, Wales, Australia and Scotland (Erntoft 2011). The U.S. Preventive Services Task Force (USPSTF) has warned that “failure to use models to extrapolate from primary data can lead to greater errors than the models themselves would introduce” (Weinstein 1996).

There are several types of decision-analytic models each defined by a set of specific characteristics. In a simplistic classification, simulation models can be static (i.e., do not allow for interactions between individuals) or dynamic (e.g., one individual can infect another individual), and they simulate a population at the cohort- or individual-level. Kuntz and colleagues provide a full overview of model taxonomy (Kuntz et al. 2013). The suitability of a

particular model is governed by matching the attributes and characteristics of a particular type of model (**Table 2**) to the needs of the decision problem at hand (Roberts et al. 2012).

Table 2. Overview of types of models, attributes and examples

Type of model (ISPOR-SMDM Modeling Good Research Practices Reference)	Attributes and Characteristics of Model Types	Strengths and Weaknesses	Example Application to Screening Programme
Decision Trees (Siebert et al., 2012)	Diagrams the probability of events over a fixed time horizon	Appropriate for a short time horizon, does	Abdominal aortic aneurysm (Ehlers, 2009)
State-transition Models (Siebert et al., 2012)	Simulates individuals or a cohort of individuals through a series of mutually exclusive and collectively exhaustive health states	Appropriate for a longer time horizon, time-dependent transition probabilities, does not allow for interactions between individuals	Cervical cancer screening (Goldhaber-Fiebert, 2008)
Dynamic Transmission Models (Pitman et al., 2012)	Compartmentalizes individuals or a cohort of individuals by infection status	Appropriate for detailed analysis of disease transmission patterns	Chlamydia trachomatis (CT) screening programme (De Vries, 2006)
Discrete Event Simulation Models (Karnon et al., 2012)	Analysis of individuals' interactions with each other and/or within systems of constrained resources	Computationally efficient	Breast cancer screening (Wisconsin Breast Cancer Epidemiology Simulation Model; Fryback, 2006)

The selection of model structure and input parameters requires a multidisciplinary approach (e.g., decision analysts, clinicians, and epidemiologists). Importantly, those input parameters that inform transitions between natural history health states (e.g., progression to or from cervical

precancer) should be informed using comprehensive literature reviews to guard against biased results. For transitions that are often unobservable (e.g., progression from cervical precancer to invasive cancer) or may vary from setting to setting, calibration (model fitting or dependent validation). Calibration involves a multi-step process to specify the value of an input parameter that corresponds, or generates good-fit, to available empirical data (i.e., a calibration target) (Vanni et al., 2011). For example, due to variations in sexual behaviour across settings, the incidence and clearance of a sexually transmitting infection such as human papillomavirus (HPV), can be fit to observed data on HPV prevalence (Campos et al., 2014). A goodness-of-fit measure, of which several exist (Vanni et al., 2011), is used to evaluate the fit of model outcomes to the observed data. Bayesian or approximate Bayesian calibration methods are also gaining traction (e.g., Menzies 2017).

Confidence in the model's projections can be strengthened through an independent validation process that involves comparing model outputs to 1) to other simulation models, 2) retrospective data not used to inform model inputs or calibration, and 3) prospective data (Eddy 2012). Convergent validity compares two or more models aimed at answering similar questions and reconciles reasons why different simulation models may come to different conclusions. A primary example of comparative modelling is the Cancer Intervention and Surveillance Modeling Network (CISNET, <http://cisnet.cancer.gov/>). CISNET is a simulation modelling consortium aimed at improving the understanding of important factors related to cancer control strategies by comparing multiple simulation models for six cancer sites: Breast, cervical, colorectal, oesophagus, lung, and prostate. Ongoing work of CISNET teams involves independent model validation to forthcoming RCTs (Rutter 2016).

Finally, the ISPOR-SMDM Modeling Task Force (Eddy 2012) emphasizes the need for transparent reporting of model structure, inputs, model-specific methods, and model validation, often in the form of technical appendices, which can include supplementary results. Appendices also provide the required space to include standardized checklists for economic evaluation, such as the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist and the impact inventory (see above). Calibrating decision-analytic models to empirical data, independently validating the models to external data, and reporting detailed methods and results transparently comprise a few of the necessary steps required to inform health policy decisions.

Uncertainty

When evaluating screening interventions, the amount of evidence may be either large, small, or not existent, and there may be uncertainty about effectiveness and long-term outcomes of screening technologies. A challenge of screening evaluation is to synthesize and assess all available evidence and evaluate the impact of uncertainty on outcomes, even when extensive model-based calibration approaches have been applied. When using economic evaluation and decision analytic modeling to evaluate screening, there are multiple types of uncertainty analyses that can be conducted to help inform decision making. Specifically, the ISPOR-SMDM Task Force distinguishes between four main types of uncertainty for decision analytic modelling: parameter uncertainty, stochastic uncertainty, heterogeneity and structural uncertainty (Briggs et al, 2012).

First, parameter uncertainty refers to the uncertainty stemming from estimation of model parameters that are inherently uncertain, such as the probability of experiencing a specific event or the accuracy of a diagnostic test. When incorporating multiple sources of data, these parameters may take a range of different values which should be accounted for when evaluating

screening strategies. For diagnostic accuracy estimates, two common biases should be considered: (i) spectrum bias, occurring due to different disease severity of different populations, and (ii) verification bias, occurring if the gold standard test has not been performed in test-negative individuals. Estimates of diagnostic accuracy that are adjusted for these biases may differ from unadjusted estimates (e.g. Mayrand et al., 2007).

Second, variability between individuals or subgroups of individuals can either be attributed to overall random variability (referred to as stochastic uncertainty) or to individual's characteristics such as age, gender, socioeconomic status and other indicators of disease risk (referred to as heterogeneity). Considering heterogeneity is essential for screening evaluations in order to determine the optimal target population for the screening programme.

Lastly, for screening evaluations that rely on decision-analytic modelling, structural uncertainty relates to the structural assumptions of the model. One example is the conceptualization of the natural history of the disease that is being screened for (i.e., which health states are included in the model and how individuals can transition between the health states). Although cost-effectiveness results may differ considerably between alternative structural assumptions (Le et al., 2016), structural uncertainty is not usually formally quantified as part of the analysis because of the programming and computation time required to perform such analyses.

To assess the impact of these different types of uncertainties as part of the screening evaluation, a range of sensitivity analyses can be performed. A simple approach is to perform deterministic sensitivity analysis and manually vary one, two or more parameters at a time (referred to as one-way, two-way and multi-way sensitivity analysis, respectively), while holding all other parameters constant. An alternative approach is to vary all parameters simultaneously by

assigning predefined probability distributions (e.g., beta, gamma) to each parameter and sample multiple sets of parameter values (referred to as probabilistic sensitivity analysis, PSA). Model output can then be reported as the mean values across all parameter sets with uncertainty bounds. 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'* (Briggs et al., 2013). Using the results of PSA, analysts can calculate the probability each screening strategy is cost-effective for a given cost-effectiveness threshold (i.e., cost-effectiveness acceptability curves), as well as quantifying the value of acquiring additional information (i.e., expected value of perfect information).

Challenges of screening evaluation

Screening is a complex system that usually involves multiple stakeholders and multiple decisions to be made along the screening pathway; as such, comprehensive screening evaluation is a challenging task.

Further, the prior presentation of evaluation of screening has only considered the health benefit and costs of screening. The optimal use of resources depends on the insurer's objective function (i.e. a composite of factors the insurer seeks to achieve). The objective function for an insurer in the health care sector may include measures such as population health (e.g. mortality and morbidity), profit and equity. Consequently, optimal resource allocation in the sector depend on how new interventions affect the insurer's objective function relative to the cost of the new intervention.

Multiple aspects need to be addressed when evaluating screening interventions, which could be part of the insurer's objective function. Certain challenges unique to screening evaluation are important to consider, including harm-benefit tradeoffs, adverse lifestyle effects of screening, the participation rate, the heterogeneity of the target population, and 'prevention versus cure'.

Harm-benefits

As mentioned previously, all screening programmes should ensure that the benefits of screening outweigh the harms, thus the harm-benefit tradeoff should be considered alongside the feasibility and cost-effectiveness of a programme prior to implementation. Relevant harm-benefit outcomes associated with a particular screening strategy should be identified by stakeholders and quantified using decision-analytic modelling. For example, a study evaluating tradeoffs in health benefits and resource use associated with candidate cervical cancer screening strategies quantified resource use or 'harms' as the number of referrals for diagnostic colposcopy with biopsy (a semi-invasive procedure required to confirm presence of high-grade precancer), while the reduction in cancer incidence associated with each strategy represented health benefits (Burger et al., 2017). In addition to quantifying relevant benefit and harm outcomes in absolute terms, some studies have used a framework similar to the one used in CEA to evaluate incremental harm-benefit ratios (a metric analogous to the ICER). For example, a study evaluating candidate cervical cancer screening strategies calculated incremental harm-benefit ratios defined as the additional number of colposcopy referrals per additional precancers detected (Pedersen et al., 2015). Similarly, a study evaluating alternative breast cancer screening strategies evaluated incremental harm-benefit ratios defined as the additional number of false-positive findings per additional life-year gained (Van Ravesteyn et al., 2012). Quantifying such benefit-harm outcomes and metrics may help inform the non-

economic efficiency of a strategy, and may help decision-makers and individuals alike decide which strategies are optimal given their preferences for harms versus benefits of screening.

Adverse lifestyle effects of screening

An evaluation of screening from a societal perspective, aims to quantify all consequences of screening. One feature of screening programmes addressed in the literature is the “Health Certificate Effect” (Larsen et al., 2007, and van der Alst et al., 2010). The “Health Certificate Effect” is a result of participants’ misinterpretation of the screening result, with a consequence on lifestyle decisions. For instance, if a negative test result is misinterpreted as a verification of being in perfect health, fewer measures may be taken to prevent future illness (allocative inefficiency of health production). The “Health Certificate Effect” can be interpreted within the framework of Bayesian updating of disease probabilities from test results. Several studies have identified the problems with understanding information both among health professionals and the general population (Bramwell et al., 2006, Whiting et al., 2015). In a RCT for screening for colorectal cancer, the percentage with a lifestyle related disease increased more among the group invited to screening than in the control group, which provides support for the “Health Certificate effect” (Aas, Iversen and Hoff, 2017).

In order to account for the “Health Certificate Effect” when evaluating the cost-effectiveness of screening, unrelated health care costs need to be included in addition to related health care costs (Neumann et al., 2016). In addition, lifestyle changes could potentially influence survival.

Participation

Evaluating the cost-effectiveness of screening within a model-based framework often relies on specific assumptions; for example assumption about perfect participation in the program. Data from trials, implementation studies and quality indicators of screening programmes show that 100% participation rarely is achieved (Elfstrøm et al, 2015, Aas, 2009). Whether or not imperfect screening participation affects the cost-effectiveness of screening depends on the characteristics of those not participating compared to those participating. In particular, a specific challenge is related to the healthy screenee bias, a concept suggesting that those who choose participating in screening are those at the lowest risk of having the disease or condition that is being screened for. For example, in a Danish study, cervical cancer screening participants had lower all-cause mortality rates than non-participants (Kranse et al., 2013). In a Norwegian study, the frequency of lifestyle-related diseases was substantially higher among non-participants (Aas et al, 2017). Accounting for higher risk among non-participants could be captured in simulation models by applying information based in individual level data or by exploring different risk profiles in sensitivity analysis. Individual level data allow for estimation of the correlation between risk profiles, survival and costs. The consequence of identifying the effect of different risk profiles on the cost-effectiveness of screening is ambiguous. For example, if the risk of the disease screened for is higher among the non-participants, increasing participation would increase life-years gained; however, if all-cause mortality is higher among the non-attendees, the expected health gain of screening is lower than among the attendees. Finally, if increased risk of the disease being screened for and/or disease-specific mortality is associated with costs of screening (adverse events), treatment, and follow-up, the incremental cost of screening will likely be affected.

Socioeconomic status is also positively related to attendance in some screening programmes (Marlow et al., 2017, Vernon, 1997; Petersen, 2002; McCaffery *et al.*, 2002; Aas, 2009). Individuals with low socioeconomic status are often less likely to participate in screening, and if this is due to a lack of knowledge of the programme, it may contribute to increasing inequities in health care, a founding principle of in many countries. A consequence of bias in participation is a negative distributional effect on health in the population. To identify the distributional effect, methods have been developed to balance distributional- and efficiency effects of screening (Asaria et al., 2015). In an evaluation of increased participation in the bowel cancer screening programme in the UK, three interventions were compared to no screening; standard screening, targeted reminder in deprived areas and universal reminder. From a cost-effectiveness standpoint a universal reminder was the preferred alternative, while a targeted reminder was the optimal choice when considering minimising unfair distribution. Several measures for inequality were applied in the paper.

Prevention versus cure

The demand for a preventive service, such as screening, is closely, but complexly related to the cure, see **Figure 1**. 1) The impact of the price of a screening method on the demand for screening, 2) the influence of the development in screening technology on the demand of screening, and 3) whether technological developments in treatment could potentially affect the demand for the preventive services are important considerations for insurers (discussed in Hey and Patel's (1983) study of the interaction between prevention and cure). Hey and Patel focus on the "*appropriate allocation of expenditure in health care*" (p. 119) and present a two-state model (health and sickness), where the movements between the states not depend on time, only on the consumption in each health state, prevention or cure. Optimal resource allocation of prevention is defined where the marginal cost of prevention is equal to the marginal utility of

prevention. Within the framework suggested by Hey and Patel, it is possible to evaluate how changes in prices and technology affect the demand for the preventive service (such as screening) and cure (treatment). Hence, with continuous developments in both treatment and screening technologies, re-evaluations of the benefits, harms and the economic costs of a screening programme are required.

When new and better screening methods becomes available, the increased health gain has to be evaluated against the costs. When the treatment of a disease is improving, there will be a need for re-evaluating the existing screening programme as screening will provide a lower incremental health gain. Furthermore, introduction of new preventive technology, such as HPV vaccines for cervical cancer (**Figure 1B**), will change the existing screening programme and new evaluation of the cost-effectiveness are warranted (Kim et al., 2016; Pedersen et al., 2017; Burger et al., 2012).

Heterogeneity

Evaluation of population-based screening programme includes estimates of expected health benefits and costs for the whole population. Offering screening to sub-groups, in which screening is not cost-effective, would be an inefficient use of resources. For screening methods aimed at preventing future disease, life expectancy without screening should be longer than the expected time-period the individual would have to live before benefiting from screening. Braithwaite (2012) addressed this issue in screening for aorta aneurism and screening for colorectal cancer by comparing risks of complication during screening (short-term harms) with the long-term benefit of screening, and compare this scenario with the life-expectancy without screening. The risk of dying due to screen-related complications and the life expectancy without screening were important components in the recommendation of screening for specific sub-

groups. These individual attributes may not be apparent when evaluating a screening programme on a population-level.

Screening for colorectal cancer is a population-based intervention where patient heterogeneity, except for age, is generally not considered in decision analyses but may be important (Lansdorp-Vogelaar et al., 2011). For this cancer type, it is known that several subgroups are at increased risk of developing and dying from colorectal cancer due to the presence of certain comorbidities (e.g., diabetes, obesity and inflammatory bowel disease), lifestyle-related factors (e.g., smoking and heavy alcohol use), and familial history and genetic predisposition (American Cancer Society, 2016). In addition, the risk of screening-related adverse events may be different among patient subgroups (Warren, 2009). In some studies, they have considered personal characteristics in determining the optimal stopping age for screening (Dinh et al., 2012; Lansdorp-Vogelaar et al., 2014; Lansdorp-Vogelaar et al., 2009; Van Hees et al., 2015). In a recent study, different effect of gender on the health outcome of screening was identified (Holme et al., 2018). Compared to an evaluation for the entire population, differences in relative risk between genders would influence the cost-effectiveness of screening women and men differentially.

Personalised screening

In contrast to the disease simulation models utilized for cost-effectiveness analyses to address population-level policies, clinical prediction models are a tool used to assess personalized risk at the individual level, and have been successfully applied to a range of medical conditions such as coronary heart disease (Genders et al., 2012), osteoporosis (Collins et al., 2011), type 2 diabetes (Abbasi et al., 2012) and various types of cancer (e.g., Beane et al., 2008). Within cervical cancer screening, it has been proposed (Castle et al., 2007) that women should not return

to routine screening if the risk of high-grade precancer exceeds 2% before the next screening round; as such, a formal tool to help clinicians assess the risk of precancer and cancer is needed in clinical practice. Similarly, for other screening programmes more personalised screening algorithms are currently being discussed (www.assure-project.eu, Seibert et al., 2018). However, as algorithms become more personalised, the monetary costs of such strategies may also increase, necessitating full evaluations of these personalised approaches.

Concluding remarks

Evaluation of medical screening programmes requires robust and complex methods that should capture the unique properties of disease and screening approach, including multiple screening strategies: The screening target ages, the screening frequency, and how to follow-up screening positive individuals (the triage algorithm). In addition, capturing all potential health benefits, harms and monetary costs of a screening programme require a lifetime horizon. While critical to inform clinical effectiveness of screening programs, randomised trials often do not capture all necessary components of screening evaluation. Alternatively, constructing decision-analytic models, which incorporate empirical data, is the preferred method for a more comprehensive evaluation of a screening programme and can be used to inform the cost-effectiveness of screening. However, cost-effectiveness analyses may not inform all aspects of the decision-making process.

References

Aas, Eline; Iversen, Tor & Hoff, Geir (2017). The effect of education on health behavior after screening for colorectal cancer., In Kristian Bolin (ed.), *Human Capital and Health Behavior. Advances in Health Economics and Health Services Research*. Emerald Group Publishing Limited. ISBN 978-1-78635-466-2. Part IV Information and health behavior. s 207 - 242

Aas, E. 2009. Pecuniary compensation increases participation in screening for colorectal cancer. *Health Economics*. ISSN 1057-9230. 18(3), s 337- 354 . doi: 10.1002/hec.1371

American Cancer Society. (2016). *Colorectal cancer* [Online]. Available: <http://www.cancer.org/cancer/colonandrectumcancer/detailedguide/colorectal-cancer-risk-factors> [Accessed 31-10-2016].

Andermann, Anne, Blancquaert, I, Beauchamp, S, Déry, V. (2008). Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years: *Bulletin of the World Health Organization*; 86(4), 241-320
<http://www.who.int/bulletin/volumes/86/4/07-050112/en/>

Angrist JD and Pischke JS. 2015. *Mastering Metrics: The path from cause to effect*. Princeton University Press. Princeton and Oxford.

Arbyn M, Anttiila A, Jordan J, Ronco G, Schenck U, Segnan N, et al. 2010. European Guidelines for Quality assurance in Cervical Cancer Screening. Second edition – summary document. *Annals of Oncology* 21(3):p 448-58.

Asaria M, Griffin S, Cookson R, Whyte S, Tappenden P. 2015. Distributional cost-effectiveness analysis of health care programmes--a methodological case study of the UK Bowel Cancer Screening Programme. *Health Econ*.24(6):742-54. doi: 10.1002/hec.3058.

Braithwaite, R. S. (2011). Can life expectancy and QALYs be improved by a framework for deciding whether to apply clinical guidelines to patients with severe comorbid disease? *Med Decis Making*, 31, 582-95.

Bramwell, R., West, H., Salmon, P., 2006. Health professionals' and service users' interpretation of screening test results: experimental study. *British Medical Journal* 2006; 333:284.

Brewer NT, Salz T, Lillie SE. Systematic Review: The Long-Term Effects of False-Positive Mammograms. *Ann Intern Med*. 2007;146:502–510. doi: 10.7326/0003-4819-146-7-200704030-00006

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.

Burger EA, Pedersen K, Sy S, Kristiansen IS, Kim JJ (2017). 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*. ISSN 0007-0920. 117(6), s 783- 790 . doi:10.1038/bjc.2017.248

Burger EA, Ortendahl JD, Sy S, Kristiansen IS, Kim JJ. Cost-effectiveness of cervical cancer screening with primary human papillomavirus testing in Norway. *British journal of cancer*. 2012 Apr;106(9):1571.

Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, Vray M (1997) Modelling in Economic Evaluation: An Unavoidable Fact of Life. *Health Economics* 6: 217-227.

Canadian Agency for Drugs and Technologies in Health (CADTH). 2017. Guidelines for economic evaluation of health technologies. Ottawa, Canada. <https://www.cadth.ca/about->

cadth/how-we-do-it/methods-and-guidelines/guidelines-for-the-economic-evaluation-of-health-technologies-canada

De Vries R, Van Bergen JE, Jong-van den Berg D, Lolkje TW, Postma MJ. 2006. Systematic screening for Chlamydia trachomatis: estimating cost-effectiveness using dynamic modeling and Dutch data. *Value in Health*. 1;9(1):1-1.

Campos NG, Burger EA, Sy S, Sharma M, Schiffman M, Rodriguez AC, Hildesheim A, Herrero R, Kim JJ. An updated natural history model of cervical cancer: derivation of model parameters. *American journal of epidemiology*. 2014 Jul 31;180(5):545-55.

Dinh, T. A., Alperin, P., Walter, L. C. and Smith, R. (2012). Impact of comorbidity on colorectal cancer screening cost-effectiveness study in diabetic populations. *J Gen Intern Med*, 27, 730-8.

Dobrow MJ, Hagens V, Chafe R, Sullivan T, Rabeneck L. 2018. Consolidated principles for screening based on a systematic review and consensus process CMAJ 190 (14) E422-E429; **DOI:** 10.1503/cmaj.171154

Drummond, M, Sculpher, MJ, Torrance, GW, O'Brian, BJ, and Stoddart, GL (2005) *Methods for Economic Evaluation in Health Care Programmes*. Oxford University Press, 3rd Edition.

Dutch National Health Care Institute. 2016. Guidelines for the conduct of economic evaluation in Health Care. Diemen, Netherland.
<https://english.zorginstituutnederland.nl/publications/reports/2016/06/16/guideline-for-economic-evaluations-in-healthcare>

Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Value in health*. 2012 Sep 1;15(6):843-50.

Ehlers L, Overvad K, Sørensen J, Christensen S, Bech M, Kjølby M. 2009. Analysis of cost effectiveness of screening Danish men aged 65 for abdominal aortic aneurysm. *BMJ*. 1;338:b2243.

Elfström KM, Arnheim-Dahlström L, von Karsa L, Dillner J. 2015. Cervical cancer screening in Europe: Quality assurance and organisation of programmes. *Eur J Cancer*. 2015 May;51(8):950-68. doi: 10.1016/j.ejca.2015.03.008.

Erntoft S (2011) Pharmaceutical Priority Setting and the Use of Health Economic Evaluations: A Systematic Literature Review. *Value in Health* 14: 587-599.

Etzioni R, Penson DF, Legler JM, di Tommaso D, Boer R, Gann PH, Feuer EJ. 2002. Overdiagnosis Due to Prostate-Specific Antigen Screening: Lessons From U.S. Prostate Cancer Incidence Trends. *Journal of the National Cancer Institute*, Vol. 94, No. 13, July 3, 2002 981-990

Fryback DG, Stout NK, Rosenberg MA, Trentham-Dietz A, Kuruchittham V, Remington PL. Chapter 7: The Wisconsin breast cancer epidemiology simulation model. *JNCI Monographs*. 2006 Oct 1;2006(36):37-47.

Gareen, IF, Duan F, Greco EM, Snyder BS, Boiselle PM, Park ER, Fryback D and Gatsonis C. 2014. "Impact of lung cancer screening results on participant health-related quality of life and state anxiety in the National Lung Screening Trial" *Cancer* 120 (21): 3401-3409. doi:10.1002/cncr.28833.

Glick HA, Doshi JA, Sonnad SS, Polsky D. 2007. *Economic Evaluation in Clinical Trials*. Oxford University Press. Oxford and New York

Goldhaber-Fiebert JD, Stout NK, Salomon JA, Kuntz KM, Goldie SJ. Cost-effectiveness of cervical cancer screening with human papillomavirus DNA testing and HPV-16, 18 vaccination. *Journal of the National Cancer Institute*. 2008 Mar 5;100(5):308-20.

Habbema, D., Weinmann, S., Arbyn, M., Kamineni, A., Williams, A. E., de Kok, I. M., ... Brown, M. (2017). HARMS OF CERVICAL CANCER SCREENING IN THE UNITED STATES AND THE NETHERLANDS. *International Journal of Cancer*, 140(5), 1215–1222. <http://doi.org/10.1002/ijc.30524>

Henderson JT, Webber EM, Sawaya GF. Screening for Ovarian Cancer Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*. 2018;319(6):595–606. doi:10.1001/jama.2017.21421

Hey, J.D., Patel, M., 1993. Prevention and Cure? Or: Is announce of prevention better than a pound of cure? *Journal of Health Economics* (2), 119-138.

Holland, W.W., Stewart, S., Masseria, C., 2006. Policy brief – screening in Europe. World Health Organization on behalf of the European Observatory on Health Systems and Policies.

Holme Ø, Løberg M, Kalager M, Bretthauer M, Hernan MA, Aas E, Eide TJ, Skovlund E, Lekven J, Schneede J, Tveit MK, Vatn MH, Ursin G & Hoff G .2018. Long-Term Effectiveness of Sigmoidoscopy Screening on Colorectal Cancer Incidence and Mortality in Women and Men: A Randomized Trial. *Annals of Internal Medicine*. doi: 10.7326/M17-1441

Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, Augustovski F, Briggs A, Mauskopf J, Loder E, on behalf of the CHEERS Task Force (2013) Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *BMC Medicine* 11: 80.

Jørgensen, KJ, Gøtzsche PC; Kalager M, Zahl PH .2017. Breast cancer screening in Denmark: A cohort study of tumor size and overdiagnosis. *Annals of Internal Medicine*. ISSN 0003-4819. 166(5), s 313- 323 . doi: 10.7326/M16-0270

Kalager, Mette; Adami, Hans-Olov; Bretthauer, Michael & Tamimi, Rulla (2012). Overdiagnosis of Invasive Breast Cancer Due to Mammography Screening:

Results From the Norwegian Screening Program. *Annals of Internal Medicine*. ISSN 0003-4819. 156(7), s 491- U61 . doi: 10.1059/0003-4819-156-7-201204030-00005

Kim JJ, Burger EA, Sy S, Campos NG. Optimal cervical cancer screening in women vaccinated against human papillomavirus. *Journal of the National Cancer Institute*. 2016 Oct 17;109(2):djw216.

Kranse R, van Leeuwen PJ, Hakulinen T, et al. 2013. Excess all-cause mortality in the evaluation of a screening trial to account for selective participation. *J Med Screen*;20:39–45.

Kuntz K, Sainfort F, Butler M, Taylor B, Kulasingam S, Gregory S, Mann E, Anderson JM, Kane RL. (2013) Decision and Simulation Modeling Alongside Systematic Reviews.

Lansdorp-Vogelaar, I., Gulati, R., Mariotto, A. B., et al. (2014). Personalizing age of cancer screening cessation based on comorbid conditions: model estimates of harms and benefits. *Ann Intern Med*, 161, 104-12.

Lansdorp-Vogelaar, I., Knudsen, A. B. and Brenner, H. (2011). Cost-effectiveness of colorectal cancer screening. *Epidemiol Rev*, 33, 88-100.

Lansdorp-Vogelaar, I., Van Ballegooijen, M., Zauber, A. G., et al. (2009). Individualizing colonoscopy screening by sex and race. *Gastrointest Endosc*, 70, 96-108

Larsen IK, Grotmol T, Almendingen K, Hoff G. 2007. Impact of Colorectal Cancer Screening on Future Lifestyle Choices: A Three-Year Randomized Controlled Trial, *Clinical Gastroenterology and Hepatology*, 5(4), p477-483, <https://doi.org/10.1016/j.cgh.2006.12.011>.

Le QA. 2016. Structural Uncertainty of Markov Models for Advanced Breast Cancer: A Simulation Study of Lapatinib. *Med Decis Making*, 36(5): p. 629-40.

Lin JS, Piper MA, Perdue LA, et al. Screening for Colorectal Cancer Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*. 2016;315(23):2576–2594.

Marlow LAV, Chorley AJ, Haddrell J, Ferrer R, Waller J. 2017. Understanding the heterogeneity of cervical cancer screening non-participants: Data from a national sample of British women, *European Journal of Cancer*, 80, p30-38, doi.org/10.1016/j.ejca.2017.04.017.

McCaffery K, J Wardle, M Nadel, W Atkin. 2002. Socioeconomic variation in participation in colorectal cancer screening. *Journal of Medical Screening*, 9:104, DOI: 10.1136/jms.9.3.104

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.

Mendes D, Bains I, Vanni T, Jit M (2015) Systematic review of model-based cervical screening evaluations. *BMC Cancer* 15: 334

Menzies NA, Soeteman DI, Pandya A, Kim JJ. Bayesian Methods for Calibrating Health Policy Models: A Tutorial. *PharmacoEconomics*. 2017 Jun 1;35(6):613-24.

National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2017. Available at: <https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview>

Neumann PJ, Anderson JE, Panzer AD, Pope EF, D'Cruz BN, Kim DD, Cohen JT. Comparing the cost-per-QALYs gained and cost-per-DALYs averted literatures. *Gates Open Research*. 2018 Jan 18;2.

Neumann PJ, Sanders GD, Russell LB, Siegel JE, and Ganiats TG, *Cost-effectiveness in health and medicine*. 2016: Oxford University Press.

Paulden M; O'Mahony JF; McCabe C. 2017. Discounting the Recommendations of the Second Panel on Cost-Effectiveness in Health and Medicine. *Pharmacoeconomics* 35(1):5-13. doi: 10.1007/s40273-016-0482-0

Pedersen, Kine; Sørbye, Sveinung Wergeland; Burger, Emily; Lönnberg, Stefan & Kristiansen, Ivar Sønbo (2015). 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*. 18(8), s 1088- 1097 . doi: 10.1016/j.jval.2015.08.003.

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. *European Journal of Cancer*. 2018 Mar 1;91:68-75.

Petersen, GM. 2002. Barriers to preventive intervention. *Gastroenterology of Clinics of North America* 31, 1061-1068.

Public Health England. 2017. Key performance indicators for the NHS screening programmes: Definitions and data submission guidance.

Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M (2012) Conceptualizing a Model: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force. *Medical Decision Making* 32: 678-689.

Rutter CM, Knudsen AB, Marsh TL, Doria-Rose VP, Johnson E, Pabiniak C, Kuntz KM, Van Ballegooijen M, Zauber AG, Lansdorp-Vogelaar I. Validation of models used to inform colorectal cancer screening guidelines: accuracy and implications. *Medical Decision Making*. 2016 Jul;36(5):604-14.

Sabik LM, Lie RK. Priority setting in health care: Lessons from the experiences of eight countries. *International Journal for equity in health*. 2008 Dec;7(1):4.

Seibert TM, Chieh FC, Yunpeng W, Verena Z, Roshan K, Kellogg PJ et al. 2018. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts *BMJ* 360 :j5757

Vaccarella S, Franceschi S, Engholm G, Lonnberg S, Khan S and Bray F. 2015. 50 years of screening in the Nordic countries quantifying the effect on cervical cancer incidence. *British Journal of Cancer*. III(5): p965-9

Van der Aalst CM, van Klaveren RJ, de Koning HJ. 2010. Does participation to screening unintentionally influence lifestyle behaviour and thus lifestyle-related morbidity? *Best Practice & Research Clinical Gastroenterology*, 24 (4), p 465-478, <https://doi.org/10.1016/j.bpg.2010.06.001>.

Van Hees, F., Saini, S. D., Lansdorp-Vogelaar, I., et al. (2015). Personalizing colonoscopy screening for elderly individuals based on screening history, cancer risk, and comorbidity status could increase cost effectiveness. *Gastroenterology*, 149, 1425-37.

Van Ravesteyn NT, Miglioretti DL, Stout NK, et al. 2012. What level of risk tips the balance of benefits and harms to favor screening mammography starting at age 40? *Annals of internal medicine*;156(9):609-617. doi:10.1059/0003-4819-156-9-201205010-00002.

Vanni T, Karnon J, Madan J, White R, Edmunds WJ, Foss A, Legood R. 2011. Calibrating Models in Economic Evaluation. *Pharmacoeconomics* 29: 35-49.

Vernon S. 1997. Participation in colorectal cancer screening: a review. *Journal of the National Cancer Institute* (89):1406–22

Vickers, AJ, Sjoberg DD, Ulmert D, Vertosick E, Roobol EJ, Thompson I, Heijnsdijk EAM, De Koning H, Atoria-Swartz C, Scardino PT and Lilja H: (2014). Empirical estimates of prostate cancer overdiagnosis by age and prostate-specific antigen. *BMC Medicine*, 12:26 doi:10.1186/1741-7015-12-26

Warren, J. L. (2009). Adverse Events After Outpatient Colonoscopy in the Medicare Population. *Annals of Internal Medicine*, 150, 849.

Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB (1996) Recommendations of the panel on cost-effectiveness in health and medicine. *Journal of the American Medical Association* 276: 1253-1258.

Weinstein MC, Torrance G, McGuire A. 2009. QALYs: the basics. *Value in health*. Mar 1;12(s1).

Whiting PF, Davenport C, Jameson C, et al., 2015. How well do health professionals interpret diagnostic information? A systematic review. *British Medical Journal Open*; 5: e008155. doi:10.1136/bmjopen-2015-008155

Wilson, JMG; Jungner, G. 1968. "Principles and practice of screening for disease". WHO Chronicle. Geneva: *World Health Organization*. 22 (11): 473Public Health Papers, #34.

Wooldridge JM. 2010. *Econometric analysis of cross sectional and panel data*. The MIT Press. Second edition. Cambridge and London.

