
Article:
Mandrik, O. orcid.org/0000-0003-3755-3031, Hahn, A.I., Catto, J.W.F. et al. (3 more authors) (2023) Critical appraisal of decision models used for the economic evaluation of bladder cancer screening and diagnosis: a systematic review. PharmacoEconomics, 41 (6). pp. 633-650. ISSN 1170-7690

https://doi.org/10.1007/s40273-023-01256-9

Critical Appraisal of Decision Models Used for the Economic Evaluation of Bladder Cancer Screening and Diagnosis: A Systematic Review

Running heading: Appraisal of bladder cancer screening models

Co-authors:

Olena Mandrik, PhD, Health Economics and Decision Science, School of Health and Related Research, The University of Sheffield.

Anne I. Hahn, MPH, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York.

James W.F. Catto, MB ChB, FRCS(Urol), PhD, Department of Oncology and Metabolism, University of Sheffield, S10 2RX, UK and Department of Urology, Sheffield Teaching Hospitals NHS Foundation Trust, S10 2JF, UK

Ann G. Zauber, PhD, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York.

Marcus Cumberbatch, Department of Oncology and Metabolism, University of Sheffield, S10 2RX, UK; Department of Urology, Sheffield Teaching Hospitals NHS Foundation Trust, S10 2JF, UK.

James Chilcott, MSc, Health Economics and Decision Science, School of Health and Related Research, The University of Sheffield.

Contact Information for Corresponding Author: Lena (Olena) Mandrik, PhD, Research Fellow ScHARR, HEDS, The University of Sheffield; Regent Court, 30 Regent Street, Sheffield, S1 4DA; Tel: +44 (0) 7 487 29 22 40; Email: o.mandrik@sheffield.ac.uk

Keywords: bladder cancer, diagnosis, screening, model, cost-effectiveness

Article Type: systematic literature review

Abbreviations

BC - bladder cancer
NMI - non-muscle-invasive
LYS - life-years saved
QALY - quality adjusted life-years
HAL BLC - hexaminolevulinate blue light cystoscopy
WLC - white light cystoscopy
TNM - tumour, node, metastasis

Financial Disclosure: This study was supported by the Yorkshire Cancer Research UK (Grant number RA/2019/R1/004). JWFC is funded by an NIHR Research Professorship.

Competing interest: J. Catto reported receiving reimbursement for consultancy from AstraZeneca, Ferring, Roche, and Janssen; speaker fees from Bristol Myers Squibb, Merck Sharp & Dohme, Janssen, Astellas, Nucleix, and Roche; honoraria for membership in advisory boards from Ferring, Roche, Gilead, Photocure, Bristol Myers Squibb, QED Therapeutics, and Janssen; and research funding from Roche. O. Mandrik, A. Hahn, A. Zauber, M. Cumberbatch, and J. Chilcott reported no conflict of interest.

Authors' contribution: The work reported in the paper has been performed by the authors, unless clearly specified in the text. Conceptualization: Mandrik, Hahn, Zauber, Chilcott;

Methodology: Mandrik, Chilcott; Analysis and interpretation of data: Mandrik, Hahn; Drafting of the manuscript: Mandrik, Hahn, Catto, Zauber, Cumberbatch, Chilcott; Funding acquisition: Catto, Chilcott.

Abstract

Background

Bladder cancer (BC) is common among current and former smokers. High BC mortality may be decreased through early diagnostic and screening. The aim of this study was to appraise decision models used for the economic evaluation of BC screening and diagnosis, and to summarise the main outcomes of these models.

Methods

MEDLINE via PubMed, Embase, EconLit, and Web of Science databases were systematically searched from January 2006 to May 2022 for modelling studies which assessed the cost-effectiveness of BC screening and diagnostic interventions. Articles were appraised according to Patient, Intervention, Comparator, and Outcomes (PICO) characteristics, modelling methods, model structures, and data sources. Quality of the studies was also appraised using the Philips checklist by two independent reviewers.

Results

Searches identified 3,082 potentially relevant studies, which resulted in 18 articles which met our inclusion criteria. Four of these articles were on BC screening, and the remaining 14 were diagnostic or surveillance interventions. Two of the four screening models were individual-level simulations. All screening models (n=4, with three on high-risk and one on general population), concluded that screening is either cost-saving or cost-effective with cost-effectiveness ratios lower than $53,000/life years saved. Disease prevalence was a strong determinant of cost-effectiveness. Diagnostic models (n=14) assessed multiple interventions; white light cystoscopy was the most common intervention and was considered cost-effective in all studies (n=4).

Screening models relied largely on published evidence generalised from other countries and did not report validation of their predictions to external data. Almost all diagnostic models (n=13 out of 14) had a time horizon of five years or less and most of the models (n=11) did not incorporate health related utilities. In both screening and diagnostic models, epidemiological inputs were based on expert elicitation, assumptions, or international evidence of uncertain generalisability.

In modelling disease, seven models did not use a standard classification system to define cancer states, others used risk-based, numerical or a Tumour, Node, Metastasis classification. Despite including certain components of disease onset or progression, no models included a complete and coherent model of the natural history of BC (i.e. simulating the progression of asymptomatic primary BC from cancer onset, i.e. in the absence of treatment).

Discussion

The variation in natural history model structures and lack of data for model parameterisation suggest that research in BC early detection and screening is at an early stage of development. Appropriate characterisation and analysis of uncertainty in BC models should be considered as a priority.

Key points for decision makers

- Evidence on cost-effectiveness of BC screening is consistent but very limited
- BC models rely on data with high uncertainty such as international data and assumptions.
- In the absence of sufficient data for complex models, more trials are needed to inform the parameters of natural history disease models which in turn can inform the protocols of the trials to test the bladder cancer screening interventions.

1. Introduction

Bladder cancer (BC) is a common malignancy with its highest burden falling on economically developed countries [1-3]. Worldwide, BC ranks sixth in men and seventeenth in women with the lifetime incidence risk of 1.1% and 0.27%, respectively [1]. The risk of BC increases with age and the higher risk for men than women reflects higher exposure to carcinogens [1-3]. Tobacco smoking is the strongest risk factor, accounting for an estimated 50–65% of all BC cases [4, 5]. Other common risk factors include occupational exposure [6, 7], contamination of drinking water with arsenic, and family history of BC [1, 2].

BC is usually first suspected due to visible haematuria or urinary symptoms [8, 9]. At time of diagnosis, around 75% of patients have non-muscle-invasive BC (NMIBC) [10] which generally has favourable prognosis. However, around 15% of patients with NMIBC will progress to invasive disease with much lower expected survival [11]. The diagnostic procedures for symptomatic patients may include: cystoscopy, telescopic endoscopy, ultrasound and/or computed tomography [4]. Screening (i.e. detection of asymptomatic cancers) has been demonstrated to provide survival benefits in prospective studies [8]. However, there remains no conclusive evidence on the effectiveness of the implementation of either national or regional BC screening programmes [1, 8].

In clinical trial settings, several BC screening approaches have been explored [12]: urine dipstick is often considered as a screening intervention in primary care settings, with the potential for urinary biomarkers as well as cystoscopy with ultrasound or computed tomography [8, 13]. Guidelines from professional organisations across different countries -- including the US, Canada, the UK, Japan, and the Netherlands -- are consistent in recommending evaluation for asymptomatic microscopic haematuria [14]. However, the recommendations vary regarding

screening interventions, particularly the role of urine dipstick and how to define the target screening population [14].

From an economic perspective, BC is one of the most expensive malignancies to manage, with the follow up costs being twice as high for medium-risk and five times as high for high-risk compared to low-risk (NMIBC) disease [15]. As multiple BC screening options emerge, modelling studies are often used to assess optimal screening regimes and outcomes prior to large-scale recommendations. The aim of this study was to classify the approaches which have been used in cost-effectiveness models in BC screening and early diagnosis with a specific focus on understanding the modelling methods that have been applied, the structure of the economic models, and modelling inputs and parameterisation. This review also summarised the main outcomes of the identified cost-effectiveness models.

2. Materials and Methods

An initial scoping search was conducted in September 2021 to identify existing reviews. No reviews of BC natural history or cost-effectiveness models were identified, however search strategies from previous reviews of diagnostic and treatment interventions, and a review of economics of BC were used to define the most appropriate search terms [16-19]. Since the scoping search identified few studies, the literature scope was then expanded to include diagnostic and surveillance models to provide a comprehensive understanding of BC modelling. The International Society for Pharmacoeconomics and Outcomes Research Good Practices Task Force Report on Critical Appraisal of Systematic Reviews With Costs and Cost-Effectiveness Outcomes was followed in the development of the protocol and reporting of these studies [20].

The protocol registration number in Prospective Register of Ongoing Systematic Reviews (PROSPERO) is CRD42021281256.

Based on the initial scoping review, a systematic search was conducted in MEDLINE via PubMed, Embase, EconLit, and the Web of Science databases. This search was supplemented by searching the HTA database of the Centre of Reviews and Dissemination of the University of York, the National Institute for Health and Care Excellence appraisal system, the Open Access Theses and Dissertations (https://oatd.org), Google Scholar (the first 300 hits in the search for "bladder cancer", "cost-effectiveness", "model"), and the references of the included studies. The search period in the review was restricted from 01/01/2006 to 08/09/2021 to reflect current practice both with cost-effectiveness modelling methods and early detection pathways. The development of the search strategy was based on the recommendations of the UK InterTASC Information Specialists' Sub-Group [21]. The search strategy was validated on the modelling studies identified through targeted search. An example of the search strategy developed for one of the databases is reported in the Supporting Material 1. An update of the literature search was conducted in May 2022.

Studies in any language were included if they met the following criteria:

- Population: Human adult population;
- Intervention: Bladder cancer screening or diagnostic interventions;
- Design: Model-based research (either cost-effectiveness models or natural history models of bladder cancer);
- Perspective/time horizon: Any;
- Publication type: Original studies; Full-text publications or reports.

Exclusion list:

- Risk models, animal models, lab models, in vitro models, regression statistical models assessing relationships between the parameters, or only cost assessments;
- Reviews of the literature, protocols, commentaries, conference abstracts.

Titles and abstracts were screened by the first author (OM) using the Rayyan tool to synthesize the studies which fit the inclusion criteria [22]. The full texts of the articles were independently evaluated by a second researcher (AIH), who also validated the data extraction and duplicated the quality assessment for each of the included study.

The extraction tables included the categories on several dimensions: (1) General information (authors, publication year, country, setting, funding) and PICO - Population, intervention, comparators, and the outcomes; (2) Modelling methods (model type according to the taxonomy of model structures for economic evaluations of health technologies [23], software, cycle, time horizon, disease states, discounting, inflation, methods used for costs and outcomes, parametrisation approach and sensitivity analysis); (3) Data sources; (4) Choices in modelling BC; and (5) Quality of the studies using the Philips checklist [24] and Bilcke et al. (2011) guide on uncertainty evaluation [25].

The standardised evaluation of the included models was based on two instruments: the Philips checklist [24] and the guide on uncertainty evaluation by Bilcke et al. (2011) [25]. The Philips checklist included the questions on the Structure (S1-S9), Data (D1-D3), and Consistency (C1,C2) [24]. The questions on uncertainty (D4) were excluded from the Philips checklist, while have been guided by the Bilcke et al. (2011) methodology [25] to avoid incompatibility between the instruments (this approach was selected as more detailed and explicit, see the Supporting Material 2 for the details). The ranking options of the Philips checklist included “yes”; “partially”, “can’t tell” and “no” (all treated as “no”); or “NA”.

The approach for data synthesis was consistent with the International Society for Pharmacoeconomics and Outcomes Research Good Practices for systematic reviews with cost and cost-effectiveness outcomes [20]. Narrative synthesis was used to address qualitative aspects of model design, including model scope, methods and choices in modelling BC. For screening studies, graphical synthesis reported standardised (inflated to 2022 and converted to international dollars) incremental cost-effectiveness ratios to visualise the cost-effectiveness outcomes by underlying disease prevalence, using the consumer price index and purchasing power parities to standardise the values [26-28]. Graphical synthesis of the outcomes for the diagnostic studies was not undertaken due to heterogeneity in PICO, methods, and health settings [20].

3. Results

3.1. General description and PICO

Our search identified 3,082 records, of which 18 models - four on BC screening and the remaining on BC diagnostic or surveillance interventions (Figure S1) - met our inclusion criteria. The excluded full text articles are reported in the Supporting material 3.

All included models were developed in high-income countries, with nine of them within the US context (Tables 1 and 2). Payer perspective was mentioned in the majority of the studies (n=12) with two studies stating the societal perspective but reporting the inputs for the direct medical costs only [29, 30].

Three and two out of four screening models simulated high-risk and general-risk populations respectively [29, 31-33]. High-risk groups were defined in the models as heavy smokers and those with occupational exposure, and as any male above the specified age. Two related cost-effectiveness studies assessed biochemical bladder markers [32, 34] as an intervention for BC

screening, and two assessed dipstick haematuria testing [29, 33] (all compared to no-screening, Table 1).

The diagnostic models included patients with haematuria (n=5), NMIBC (n=8), and muscle-invasive BC (n=1). A range of different diagnostic and surveillance interventions were assessed in the models. Hexaminolevulinate blue light cystoscopy (HAL-BLC) and white light cystoscopy (WLC) were the most frequently compared interventions, followed by cystoscopy as a stand-alone or a combination of the interventions (Table 2).

In screening models, two out of four studies reported quality adjusted life-years (QALYs) [29, 33] and one more life-years saved (LYS) [31] (Table 1). In diagnostic models, QALYs were reported only in four out of 14 studies [35-38] and two more studies reported LYS only [39, 40], with cases detected and resource utilization used as the primary modelling outcomes (Table 2).

3.2. Screening models: outcomes

The models which evaluated haematuria tests included the impact on bladder and kidney cancers, as well as other urological diseases. All studies concluded that BC screening is cost-effective in either all (n=1) or only high risk (n=3, as defined using BC demographic features) population groups (Table 1).

All studies concluded that screening is more cost-effective with higher incidence or prevalence of the disease (Figure 1). There was no homogeneity in a value of BC prevalence or incidence that would define when screening becomes a cost-effective intervention. Cost per cancer detected was the lowest in the older age group (71-80 years old) with the highest disease prevalence [24]; though, no studies compared cost per QALY for populations among different ages to examine how cost-effectiveness of screening varies by age.

3.3. Diagnostic models: outcomes

WLC dominated computed tomography (CT) scan [41], the protocol including a microsatellite analysis with control cystoscopy at 3, 12, and 24 months [30], and the protocol using virtual cystoscopy followed up by cystoscopy if the first test is positive [42]. Interventions which supplemented cystoscopy had higher costs and effects, while tumour markers had higher costs and varied values for clinical effects [30, 39, 41, 42]. The strategy of using the cystoscopy only for positive cases with other primary diagnostic tools (such as urine cytology or cystosonography) had lower costs and effects [38, 42]. Compared to HAL-BLC, WLC had higher costs in two out of four studies [36, 43-45]. These studies concluded that HAL-BLC had higher therapeutic effects than WLC, and is therefore likely to be cost-effective [36, 43-45]. Only one of the included studies [28], assessed incremental cost-effectiveness ratio as costs per QALY (the intervention was considered as dominating). Three other studies [32, 35, 36] assessed cost per progression, recurrence, or resource use, leaving a high uncertainty around interpretation of their results. The heterogeneity in the choice of other evaluated diagnostic interventions and their comparators was too large to support systematic comparison (Table 2).

3.4. Screening models: methods

Two screening models used a decision tree and two others used Markov model structures [29, 34] (Table S1). All screening models were cohorts rather than individual patient level models.

The models with decision tree structures predicted the potential health and cost impact of screening interventions by combining the characteristics of the screening tests (such as sensitivity and specificity) with underlying BC prevalence data [32, 33]. Average life expectancy by stage among the modelled population group (75-year-old men) was used in the decision tree model predicting life-years saved (LYS) and QALYs over the lifetime [33].

The models with Markov structures (one with a lifetime and another one with a five-year horizon) used a decision tree to model the screening and diagnostic pathways leading to the detection of BC; patients with the diagnosed BC, entered one of the BC states (Markov model) and could undergo recurrence, surveillance, progression or death [26,27].

3.5. Diagnostic models: methods

All but one diagnostic model [39] had a time horizon of five years or less. Five out of 14 diagnostic and surveillance models had a decision tree cohort structure [36, 40-42, 45], and one model was a simulated patient-level decision tree model [46] (Table S2). The decision tree structure was applied mainly in the diagnostic and surveillance models with the focus on clinical or healthcare outcomes (e.g., cancers detected, or healthcare resources used, and not LYS or QALYs); similar to screening models, the decision tree structure was used to model the diagnostic and treatment pathways based on sensitivities of the tests. In the simulated patient-level decision tree model of Georgieva et al. (2019), patients were assigned individual characteristics (including sex, age, smoking status, and history of gross haematuria), and the probabilities of different types of urinary tract cancers were based on these characteristics at diagnosis [46]. This model predicted the number of detected and missed cancers, which allowed for the assessment of costs and cost-effectiveness of each intervention based on the sensitivity and specificity of each diagnostic test.

Seven diagnostic models were cohort-level Markov models [30, 35, 37, 39, 43, 44, 47] (one of them was a semi-Markov model [30]), and one model was an individual-level Markov model [40]. Markov states were used in the cohort models to simulate the transitions during surveillance period (i.e. after the diagnosis), such as progression, recurrence of the disease, or death. The simulated patient-level Markov model of Yuan (2012) simulated the natural history of

secondary BC to assess the impact of different diagnostic guidelines, with the Markov states including natural history of BC, treatment, and death [38].

3.6. Screening models: sources of data

Screening models were directly parameterised from published sources and/or registers and were based on assumptions on the disease incidence, prevalence, and screening effect (e.g. downstaging) [32-34] (Table S3). Base-case epidemiological inputs, such as incidence, were based on experts' or researchers' assumptions. The definition of high-risk populations varied by study, from 2% for prevalence to 10% for incidence [29, 31-33]. Data on costs were retrieved from the databases (Medicare, NHS reference costs, and the National Health Insurance) and supplemented with data from local hospitals and expert opinions [24-27]. Three studies [31-33] used other inputs from published sources; the screening accuracy and downstaging data were retrieved from meta-analyses of international studies, individual publications, clinical experts' and authors' opinions. Models had differing assumptions on screening tests sensitivities, which ranged from 60 to100% for different tests (dipstick tests or biomarkers) and population groups (average-risk or high-risk) [32-34]. Two studies (with the UK and Japan context) reported QALYs as the outcome measures and both retrieved utility values from previous cost-effectiveness analyses, including those conducted in other countries (from Canada and the USA respectively). A recent study by Okubo et al. (2022), evaluating the cost-effectiveness of combining haematuria screening with a Specific Health Checkup (where haematuria test is already performed for around 38% of participants), informed the transition probabilities by the Specific Health Checkup report and the National Cancer Registry data [29].

Specificity of the primary tests in the screening models (with values ranging from 60 to 99.9%) impacted the follow up interventions and costs of diagnosis [24-27]. None of the models reported screen-induced overdiagnosis, overtreatment or other potential screening-related harms.

3.7. Diagnostic models: sources of data

Most of the models were directly parameterised from published sources (i.e. used published data as direct model inputs) with one study also using within clinical trial assessment [30] and two others manually calibrating some of the disease parameters by using the data from the European Organization for Research and Treatment Center as calibration targets [38, 47] (Table S4). Expert elicitation, assumptions, and published sources were used for epidemiological data, with all but three studies referenced international data for some of the parameters including sensitivities, disease severity, and progression [35, 37-39, 41-47]. National datasets (such as Medicare for all the US studies, National Health Service reference costs, or the National formularies) were used in all but two studies with in-hospital cost calculations [30, 44] to estimate the direct medical costs. Variable uptake for the diagnostic and surveillance interventions was not considered in the included models, as it was not measured empirically for the evaluated interventions. Diagnostic studies included harms (n=7) related to unnecessary tests for those with false positive diagnoses, complications from invasive diagnostic and treatment procedures, including mortality from radiation-induced tumours and anaesthesia based on published data [30, 37-39, 41, 42, 46].

Three out of four studies reporting QALYs retrieved health related utility values from previous cost-effectiveness studies [35, 37, 38]; all three studies (two from the US and one from the UK) referenced a cost-effectiveness analysis of radical cystectomy in Canada that evaluated related utilities based on a standard gamble approach involving 25 urologists [48]. Mowatt et all (2010)

used utility values from the other urological cancers [39] stating that the modellers selected the best available source of the evidence to inform health related utility values. While the study of Mowatt et al. (2010)[39] is not recent, the reliance of the later studies on qualitative data from the previous model suggests that scarcity in utility values may still be an issue.

3.8. Modelling bladder cancer

The identified models defined BC states in the following ways (Table 3):

(1) Without a standard classification system defining the cancer as detected, progressed and/or recurrent [30, 35, 40, 42, 45-47];

(2) Using Tumour, Node, Metastasis (TNM) system or its elements [34, 36] or numerical staging [29];

(3) Using risk-based classification states, such as NMIBC of low-, intermediate-, and high-risk, and non-metastatic and metastatic muscle invasive BC [33, 37, 39, 43, 44].

Some of the diagnostic and screening models simulated population groups including patients with asymptomatic microscopic haematuria [41], microscopic haematuria [39, 42, 46], or suspected haematuria [42], while predicting outcomes from the time of the diagnosis. However, none of the models simulated a complete natural history (i.e. progression of asymptomatic disease from primary cancer onset). Screening or diagnostic models can be divided into several types according to the inclusion of the natural history components.

1. Models without progression of undiagnosed cancer

Models of this type simulate effects and costs based on stage at diagnosis for screen-detected and symptomatic disease and did not consider cancer progression [33-36, 40-42, 45, 46] or considered only progression for diagnosed disease [29, 30, 47]. These models were informed by the assumed or evidenced incidence rates and test sensitivities. When modelling the

consequences of a false negative test instead of disease progression, these models assessed incremental costs. For example, the diagnostic study of Rodgers et al. (2006) [42] considered costs of repeat testing for microscopic haematuria with false negative diagnosis. Teoh et al. (2018) [33] applied higher lifetime treatment costs to false negative screened patients, similar to those detected symptomatically.

2. Models with progression of undiagnosed cancer as a result of false negative test

These models simulate progression to more advanced BC states for patients with a false negative test result by combining prevalence data and characteristics of screening tests [37, 39, 43, 44]. For example, the diagnostic model of Sutton et al. (2018) included an undiagnosed state for patients with false negative results and assumed that these patients will be diagnosed within the next two years; patients in an undiagnosed state could progress to low-risk, high-risk or metastatic states and could then be diagnosed [37].

3. Models with progression of asymptomatic cancer

The only model that included undiagnosed states for BC that were not related to testing false negative (i.e. asymptomatic cancer) was a clinical surveillance model of Yuan et al. (2012) [38]. This model simulated the natural history of secondary BC for patients defined as low-risk at time of diagnosis and were disease-free following the treatment. This model assumed a progression of patients from treated low-risk BC, to asymptomatic intermediate risk, and then finally high-risk. At each of these states, patients could transit to the detected state following the surveillance intervention. Diagnosed patients could not progress to more advanced disease but could progress to the death state as a result of BC death or age-specific death from other causes. The progression of asymptomatic disease was estimated by comparing the predicted disease rates to the one

observed in the European Organisation for Research and Treatment of Cancer trials [49]. The process of calibration is not described in the manuscript.

3.9. Quality ranking using the Philips checklist

In general, studies addressed most of the evaluated quality criteria of the Philips checklist (Table S5), with 14 studies were scored "no" only on 30% or less questions. Meanwhile, assessment of internal and external consistency was not reported in 17 and 14 studies respectively (possibly being reported in separate publications or reports). Short time horizon was also a frequent concern (n=8 out of 18 studies) in the models (Figure 2). Quality of two older screening studies was lower than the quality of later screening models and the diagnostic studies, however, because only a few studies were identified, no meaningful comparison can be provided. Agreement between the two reviewers for each category of the Philips checklist [24] was very high at 92%.

3.10. Structural and parameter uncertainty in BC models

Structural uncertainty in screening models was related to different structural assumptions, such as using a decision tree structure to ascertain long-term outcomes, choice of static probabilities, using the BC cases detected as the modelling outcome (instead of the LYS or QALYs), methods and assumptions on BC mortality/survival use in modelling, and mismatch between the selected perspective and costs [29, 31-33]. Structural uncertainty was not fully addressed in screening models (Table S6), with the study of Teoh et al. (2018) partially exploring structural uncertainty by specifying the availability of sources of evidence, their appropriateness, and the limitations. Parameter uncertainty (related to the assumed epidemiologic values, unrepresentative populations [using international data or small sample samples], or unspecified sources) was present in all screening studies (Tables S3, S6) [29, 31-33]. None of the published articles

mentioned the model validation. Only the most recent study by Okubo et al. (2022) fully addressed parameter uncertainty by the explicit probabilistic sensitivity analysis [29].

In diagnostic models, short-term time horizon, static transition rates, choice of the outcomes (only BC cases detected or resource use and not LYS or QALYs), the approaches to test sensitivity, incidence, disease progression, recurrence, and BC mortality evaluation, were recorded among the other sources of structural uncertainty (Table S6) [7, 30, 35-47]. While most models did not report on structural uncertainty, Klaassen et al. (2017)[44] and Mowatt et al. (2010)[39] addressed the structural uncertainty by conducting scenario analyses, while three other studies [37, 38, 42] partially addressed structural uncertainty by explicitly specifying the accepted and the alternative assumptions. Similar to the screening models, parameter uncertainty was identified in all included studies while was addressed through the probabilistic sensitivity analysis in five diagnostic studies [37, 39, 41, 42, 46]. Two publications described a validation conducted for the diagnostic models (one of them with the calibrated parameters) using survival data or the risk-distribution [38, 46].

3.11. Summary from studies with low uncertainty

Studies that addressed at least partially both structural and parameter uncertainty [25] also were ranked high on Philips checklist criteria [24]. All three studies with explicitly addressed structural and parameter uncertainty (two of which were reports) were the diagnostic studies [37-39]. Mowatt et al. (2010)[39] analysed multiple diagnostic interventions, concluding that cytology followed by WLC in initial diagnosis and follow up while the least effective strategy is the most cost-effective approach in the UK setting. Sutton et al. (2018) concluded that a diagnostic classifier for risk stratification of haematuria patients is cost-effective in the UK with the probability of 68% [37]. Yuan et al. (2012) compared long-term clinical effect of different guidelines in the US setting and concluded that none of the comparators dominate each other [38].

4. Discussion

This review explored methods used in modelling cost-effectiveness of diagnostic, surveillance and screening interventions in BC. The screening models evaluated cost-effectiveness of biomarkers and urine dipstick tests in general-risk and high-risk population; all screening studies concluded that screening is cost-effective with the underlying disease prevalence being its important determinant. The earlier models evaluating cost-effectiveness of biomarkers [31, 32] though had low quality and high structural and parameter uncertainties.

Diagnostic models assessed a wide range of interventions. In studies of variable quality, HAL-BLC was consistently considered as a cost-effective intervention compared to WLC. The studies with low structural and parameter uncertainty concluded on cost-effectiveness of cytology followed by WLC in initial diagnosis (compared to multiple alternatives)[39] and a risk stratification approach for haematuria patients in the UK [37]. Diagnostic models had variable predictions on cost-effectiveness of urine biomarkers in BC diagnosis (reporting higher costs and variable effects compared to their alternatives), with a high-quality model with low uncertainty reporting that tumour markers are not cost-effective in the UK setting [39].

The conclusions of the cost-effectiveness analyses are subject to provisos regarding limitations of the methods used and available data constraints, with the following discussion points identified:

(1) *Correspondence of the PICO to the decision problem*

The description of the population (asymptomatic, symptomatic, or diagnosed with NMIBC or muscle invasive BC) defined the initial and the following states of the models. The choice of the intervention will affect the model design since some screening and diagnostic tests, such as urine dipstick test, may also lead to diagnosis of other diseases (e.g. kidney cancer or other urological conditions). As such, BC models should assess a need to include simulation of other relevant

health conditions to avoid underestimating the potential benefits of screening and diagnostic interventions.

While patients, interventions, and comparators were well-defined in BC models, the economic outcomes investigated were more inconsistent. BC models frequently reported cost per detection, recurrence, progression, or resources used as the main outcome. While these outcomes may be interesting in their own right, they are inadequate in two regards: first, they do not capture the long-term mortality or health related quality of life impacts of early or delayed detection; second, they do not allow comparative economic analyses across different health conditions and so can not inform policy decisions [50].

(2) *Choice of the model structure*

Selection of the model should be based on the simplest structure which addresses the objectives of the study, the structure of the disease, and the clinical guidelines or treatment pathways [23]. The healthcare decisions, particularly large investments such as national screening programmes, should consider uncertainty that cannot be reflected in deterministic models. In cancer modelling, timing is important for costs and health outcomes, as costs are commonly higher the first year of diagnosis than the following years [51] and cancer-related decrements in health related utilities vary over time [52]. While stochastic timed models without interaction would be the expected choice for most BC screening and diagnostic models, in our review, most of the included models were deterministic, and the decision tree structure was used in more than third of all the analysed models.

(3) *Modelling natural history of bladder cancer in screening models*

In comparison to breast, cervical and colorectal cancers [53-55], the evidence pertaining to cost-effectiveness of BC screening is currently limited. As such, BC models are less sophisticated and

453 have a much greater reliance on expert judgement than models for cancers with well-established
454 screening programmes. Only one natural history model -- without a cost-effectiveness
455 component -- was identified. However, since this model simulated only secondary BC cancer, it
456 is not directly applicable to a screening population [38]. None of the cost-effectiveness models
457 simulated a complete natural history (i.e. a progression of asymptomatic primary BC from cancer
458 onset) which hinders cross-comparisons between modelling predictions. While there is some
459 understanding of the BC risk factors, onset, progression and recurrence [11, 56], modelling
460 natural history of BC is constrained by a lack of direct or indirect data which is able to: (a)
461 inform the progression of asymptomatic disease (e.g. dwell time), and (b) inform long-term
462 clinical outcomes (eg. survival) in complex individual-level models or when the model states are
463 consistent with the detailed histology of the disease. The absence of the natural history modelling
464 leads to a general limitation of published BC screening models. Such models are not nimble
465 enough to compare different designs of screening programmes, accurately predict a long-term
466 effect of repeated screenings, or the impact of screening on screening-related harms, such as
467 overdiagnosis.
468 Modelling a complete natural history in screening models requires complex structure and life-
469 time horizon to capture the long-term effect and harms. There is a high requirement in data for
470 indirect parametrisation of such models (i.e. calibration of the parameters to inform the
471 transitions in unobserved health states), including prevalence of undiagnosed cancer, speed of
472 cancer growth or sojourn time, and the probability for cancer spontaneous regression or
473 recurrence [47], which in turn implies that more modelling inputs need to be evaluated for their
474 quality in screening models comparing to diagnostic models.
475 (4) *Uncertainty in bladder cancer modelling*

Structural and parameter uncertainty is common in screening and diagnostic BC models. This uncertainty relates to both the epistemic uncertainty in the applicability of data (e.g. using the international data or assumptions), an aleatory uncertainty with a frequent reliance on deterministic analysis, and a lack of validation or scenario analyses to explore uncertainty in model structures.

The parameter uncertainty in the identified models suggests a possible scarcity of sources to inform country-specific parameters and a need to assess the transferability of sources available for modelling. In particular, the data needs to be improved to inform health related utility values in BC models.

While a clinical effect of medical interventions is generally considered to be generalisable, there may be specific considerations that make this less so for diagnostic tests, especially for screening interventions [57, 58]. It is common for cancer screening models to assume that disease onset is a setting-specific transition relying on a set of risk factors, while cancer progression consists of generalisable parameters [59]. This assumption, mainly based on a lack of data to state otherwise suggest that careful consideration should be taken to generalise the baseline disease risk from other settings [57]. Considering that all models were developed within the context of high-income countries, neither their outcomes nor their inputs are generalizable to the middle- or lower-income settings.

Implications for research

While empirical evidence is necessary to inform the modelling parameters and to improve predictions, mathematical disease models are also used to inform the trials design [60, 61]. As such, development and implementation of trials needed to inform the models and models to inform the trials should be an iterative process. This also suggests that BC models informed by

499 the limited trial data should be flexible enough to incorporate this iterative process when the new
500 data appears, especially where this has the potential to inform developments to model structure
501 in addition to simple parameter updates.
502 The utility values for BC health states as well as population preferences for different diagnostic
503 and screening interventions, currently not considered in the mathematical disease models, should
504 be explored in the future studies.
505 *Limitations of the review*
506 While this review sought to search comprehensively the literature, there are limitations to note as
507 well. Only one reviewer screened the initial abstracts, which may have resulted in missed studies
508 or an unintentional bias in the initial search. To assuage any further bias, two independent
509 reviewers assessed the full texts of the included publications and the quality of studies.
510 Moreover, two of the included publications were grey-literature reports (i.e. publications that did
511 not go through the formal peer-review process) which may not appear in a systematic search if
512 the reproduction of the search strategy is attempted. To standardise the quality assessment, the
513 Philips checklist [23] was used, with very high average agreement rate among the raters (92%).
514 However, some of its components - such as short time horizon - are better suited for screening
515 studies rather than diagnostic studies. Moreover, some limitations of the appraised health
516 economic studies may be reasoned by compliance to local guidelines. Finally, all the models
517 which were included in this review were from high-income countries, and therefore may not be
518 generalisable to other populations across the globe.
519 *Conclusions*
520 Although the evidence pertaining to cost-effectiveness of BC screening is consistent, it is still in its
521 nascent stages. More data is needed to systemically address uncertainties in models, as well as the natural
522 history of BC. This suggests that BC models are not nimble enough to compare different designs of

screening programmes, or to predict screening-related harms such as overdiagnosis. Future clinical trials may help to decrease uncertainty in structures and parameters of BC models, as all models rely on data. Once natural history of BC models are established, these models can then inform optimal population screening and surveillance strategies which may not be possible to evaluate in the scope of clinical trials.

5. References

1. Richters A, Aben KKH, Kiemeney LALM. The global burden of urinary bladder cancer: an update. World Journal of Urology. 2020;38(8):1895-904.
2. Wong MCS, Fung FDH, Leung C, Cheung WWL, Goggins WB, Ng CF. The global epidemiology of bladder cancer: a joinpoint regression analysis of its incidence and mortality trends and projection. Scientific Reports. 2018;8(1):1129.
3. Cumberbatch MGK, Jubber I, Black PC, Esperto F, Figueroa JD, Kamat AM, et al. Epidemiology of Bladder Cancer: A Systematic Review and Contemporary Update of Risk Factors in 2018. Eur Urol. 2018;74(6):784-95.
4. Saginala K, Barsouk A, Aluru JS, Rawla P, Padala SA, Barsouk A. Epidemiology of Bladder Cancer. Medical Sciences. 2020;8(1):15.
5. Cumberbatch MG, Rota M, Catto JW, La Vecchia C. The Role of Tobacco Smoke in Bladder and Kidney Carcinogenesis: A Comparison of Exposures and Meta-analysis of Incidence and Mortality Risks. Eur Urol. 2016;70(3):458-66.
6. Reed O, Jubber I, Griffin J, Noon AP, Goodwin L, Hussain S, et al. Occupational bladder cancer: A cross section survey of previous employments, tasks and exposures matched to cancer phenotypes. PLoS One. 2020;15(10):e0239338.
7. Cumberbatch MG, Cox A, Teare D, Catto JW. Contemporary Occupational Carcinogen Exposure and Bladder Cancer: A Systematic Review and Meta-analysis. JAMA Oncol. 2015;1(9):1282-90.
8. Cumberbatch MGK, Noon AP. Epidemiology, aetiology and screening of bladder cancer. Transl Androl Urol. 2019;8(1):5-11.
9. Jubber I, Shariat SF, Conroy S, Tan WS, Gordon PC, Lotan Y, et al. Non-visible haematuria for the Detection of Bladder, Upper Tract, and Kidney Cancer: An Updated Systematic Review and Meta-analysis. Eur Urol. 2020;77(5):583-98.
10. Kamat AM, Hahn NM, Efstathiou JA, Lerner SP, Malmström P-U, Choi W, et al. Bladder cancer. The Lancet. 2016;388(10061):2796-810.
11. Fujii Y. Prediction models for progression of non-muscle-invasive bladder cancer: A review. Int J Urol. 2018;25(3):212-8.
12. Larré S, Catto JW, Cookson MS, Messing EM, Shariat SF, Soloway MS, et al. Screening for bladder cancer: rationale, limitations, whom to target, and perspectives. Eur Urol. 2013;63(6):1049-58.
13. Soria F, Krabbe L-M, Todenhöfer T, Dobruch J, Mitra AP, Inman BA, et al. Molecular markers in bladder cancer. World Journal of Urology. 2019;37(1):31-40.
14. Linder BJ, Bass EJ, Mostafid H, Boorjian SA. Guideline of guidelines: asymptomatic microscopic haematuria. BJU Int. 2018;121(2):176-83.

15. Michaeli JC, Boch T, Albers S, Michaeli T, Michaeli DT. Socio-economic burden of disease: Survivorship costs for bladder cancer. Journal of Cancer Policy. 2022;32:100326.
16. Yeung C, Dinh T, Lee J. The health economics of bladder cancer: an updated review of the published literature. Pharmacoeconomics. 2014;32(11):1093-104.
17. van Rhijn BWG, van der Poel HG, van der Kwast TH. Urine Markers for Bladder Cancer Surveillance: A Systematic Review. European Urology. 2005;47(6):736-48.
18. Wang Z, Que H, Suo C, Han Z, Tao J, Huang Z, et al. Evaluation of the NMP22 BladderChek test for detecting bladder cancer: a systematic review and meta-analysis. Oncotarget. 2017;8(59):100648-56.
19. Gandhi N, Krishna S, Booth CM, Breau RH, Flood TA, Morgan SC, et al. Diagnostic accuracy of magnetic resonance imaging for tumour staging of bladder cancer: systematic review and meta-analysis. BJU Int. 2018;122(5):744-53.
20. Mandrik O, Severens JL, Bardach A, Ghabri S, Hamel C, Mathes T, et al. Critical Appraisal of Systematic Reviews With Costs and Cost-Effectiveness Outcomes: An ISPOR Good Practices Task Force Report. Value in Health. 2021;24(4):463-72.
21. Glanville J, Bayliss S, Booth A, Dundar Y, Fernandes H, Fleeman ND, et al. So many filters, so little time: the development of a search filter appraisal checklist. J Med Libr Assoc. 2008;96(4):356-61.
22. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Systematic Reviews. 2016;5(1):210.
23. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health Econ. 2006;15(12):1295-310.
24. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. Health Technol Assess. 2004;8(36):iii-iv, ix-xi, 1-158.
25. Bilcke J, Beutels P, Brisson M, Jit M. Accounting for Methodological, Structural, and Parameter Uncertainty in Decision-Analytic Models:A Practical Guide. Medical Decision Making. 2011;31(4):675-92.
26. Consumer price inflation tables. Office for National Statistics; 2022. https://www.ons.gov.uk/economy/inflationandpriceindices/datasets/consumerpriceinflation. Accessed 21 Sep 2022.
27. Consumer Price Index. US Bureau of labor statistics; 2022. https://www.bls.gov/cpi. Accessed 21 Sep 2022.
28. Purchasing power parities (PPP). Organisation for Economic Co-operation and Development; 2021. https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm. Accessed 21 Sep 2022.
29. Okubo R, Hoshi SL, Kimura T, Kondo M, Asahi K, Iseki C, et al. Cost-effectiveness of mass screening for dipstick hematuria in Japan. Clin Exp Nephrol. 2022;26(5):398-412.
30. de Bekker-Grob EW, van der Aa MN, Zwarthoff EC, Eijkemans MJ, van Rhijn BW, van der Kwast TH, et al. Non-muscle-invasive bladder cancer surveillance for which cystoscopy is partly replaced by microsatellite analysis of urine: a cost-effective alternative? BJU Int. 2009;104(1):41-7.
31. Lotan Y, Svatek RS. Cost-effectiveness of bladder cancer screening. Expert Rev Pharmacoecon Outcomes Res. 2007;7(6):627-32.
32. Svatek RS, Sagalowsky AI, Lotan Y. Economic impact of screening for bladder cancer using bladder tumor markers: a decision analysis. Urol Oncol. 2006;24(4):338-43.

33. Sheh KT. An Early Health Economic Model of Targeted Screening for Bladder and Kidney Cancer. In: Sheffield TUo, editor. HAR673 Dissertation. Sheffield, UK2018. p. 79.
34. Lotan Y, Svatek RS, Sagalowsky AI. Should we screen for bladder cancer in a high-risk population?: A cost per life-year saved analysis. Cancer. 2006;107(5):982-90.
35. Al Hussein Al Awamlh B, Lee R, Chughtai B, Donat SM, Sandhu JS, Herr HW. A cost-effectiveness analysis of management of low-risk non-muscle-invasive bladder cancer using office-based fulguration. Urology. 2015;85(2):381-6.
36. Garfield SS, Gavaghan MB, Armstrong SO, Jones JS. The cost-effectiveness of blue light cystoscopy in bladder cancer detection: United States projections based on clinical data showing 4.5 years of follow up after a single hexaminolevulinate hydrochloride instillation. Can J Urol. 2013;20(2):6682-9.
37. Sutton AJ, Lamont JV, Evans RM, Williamson K, O'Rourke D, Duggan B, et al. An early analysis of the cost-effectiveness of a diagnostic classifier for risk stratification of haematuria patients (DCRSHP) compared to flexible cystoscopy in the diagnosis of bladder cancer. PLoS One. 2018;13(8):e0202796.
38. Yuan Z. A Partially Observable Markov Decision Process for Optimal Design of Surveillance Policies for Bladder Cancer. In: University NCS, editor. North Carolina2012. p. 96.
39. Mowatt G, Zhu S, Kilonzo M, Boachie C, Fraser C, Griffiths TR, et al. Systematic review of the clinical effectiveness and cost-effectiveness of photodynamic diagnosis and urine biomarkers (FISH, ImmunoCyt, NMP22) and cytology for the detection and follow-up of bladder cancer. Health Technol Assess. 2010;14(4):1-331, iii-iv.
40. Lotan Y, Woldu SL, Sanli O, Black P, Milowsky MI. Modelling cost-effectiveness of a biomarker-based approach to neoadjuvant chemotherapy for muscle-invasive bladder cancer. BJU Int. 2018;122(3):434-40.
41. Halpern JA, Chughtai B, Ghomrawi H. Cost-effectiveness of Common Diagnostic Approaches for Evaluation of Asymptomatic Microscopic Hematuria. JAMA Intern Med. 2017;177(6):800-7.
42. Rodgers M, Nixon J, Hempel S, Aho T, Kelly J, Neal D, et al. Diagnostic tests and algorithms used in the investigation of haematuria: systematic reviews and economic evaluation. Health Technol Assess. 2006;10(18):iii-iv, xi-259.
43. Dansk V, Malmström P-U, Bläckberg M, Malmenäs M. Hexaminolevulinate hydrochloride blue-light flexible cystoscopy in the detection and follow-up of nonmuscle-invasive bladder cancer: cost consequences during outpatient surveillance in Sweden. Future Oncology. 2016;12(8):1025-38.
44. Klaassen Z, Li K, Kassouf W, Black PC, Dragomir A, Kulkarni GS. Contemporary cost-consequence analysis of blue light cystoscopy with hexaminolevulinate in non-muscle-invasive bladder cancer. Can Urol Assoc J. 2017;11(6):173-81.
45. Rose JB, Armstrong S, Hermann GG, Kjellberg J, Malmström PU. Budget impact of incorporating one instillation of hexaminolevulinate hydrochloride blue-light cytoscopy in transurethral bladder tumour resection for patients with non-muscle-invasive bladder cancer in Sweden. BJU Int. 2016;117(6b):E102-13.
46. Georgieva MV, Wheeler SB, Erim D, Smith-Bindman R, Loo R, Ng C, et al. Comparison of the Harms, Advantages, and Costs Associated With Alternative Guidelines for the Evaluation of Hematuria. JAMA internal medicine. 2019;179(10):1352-62.

47. Mossanen M, Wang Y, Szymaniak J, Tan WS, Huynh MJ, Preston MA, et al. Evaluating the cost of surveillance for non-muscle-invasive bladder cancer: an analysis based on risk categories. World J Urol. 2019;37(10):2059-65.
48. Kulkarni GS, Alibhai SM, Finelli A, Fleshner NE, Jewett MA, Lopushinsky SR, et al. Cost-effectiveness analysis of immediate radical cystectomy versus intravesical Bacillus Calmette-Guerin therapy for high-risk, high-grade (T1G3) bladder cancer. Cancer. 2009;115(23):5450-9.
49. Sylvester RJ, van der Meijden AP, Oosterlinck W, Witjes JA, Bouffioux C, Denis L, et al. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. Eur Urol. 2006;49(3):466-5; discussion 75-7.
50. Kim DD, Basu A. How Does Cost-Effectiveness Analysis Inform Health Care Decisions? AMA J Ethics. 2021;23(8):E639-47.
51. Mariotto AB, Yabroff KR, Shao Y, Feuer EJ, Brown ML. Projections of the cost of cancer care in the United States: 2010-2020. J Natl Cancer Inst. 2011;103(2):117-28.
52. Wang SY, Hsu SH, Gross CP, Sanft T, Davidoff AJ, Ma X, et al. Association between Time since Cancer Diagnosis and Health-Related Quality of Life: A Population-Level Analysis. Value Health. 2016;19(5):631-8.
53. Mandrik O, Ekwunife OI, Meheus F, Severens JL, Lhachimi S, Uyl-de Groot CA, et al. Systematic reviews as a “lens of evidence”: Determinants of cost-effectiveness of breast cancer screening. Cancer Medicine. 2019;8(18):7846-58.
54. Goldie SJ, Kim JJ, Myers E. Chapter 19: Cost-effectiveness of cervical cancer screening. Vaccine. 2006;24:S164-S70.
55. Silva-Illanes N, Espinoza M. Critical Analysis of Markov Models Used for the Economic Evaluation of Colorectal Cancer Screening: A Systematic Review. Value in Health. 2018;21(7):858-73.
56. Martini A, Sfakianos JP, Renström-Koskela L, Mortezavi A, Falagario UG, Egevad L, et al. The natural history of untreated muscle-invasive bladder cancer. BJU Int. 2020;125(2):270-5.
57. Drummond M, Barbieri M, Cook J, Glick HA, Lis J, Malik F, et al. Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force report. Value Health. 2009;12(4):409-18.
58. Daubner-Bendes R, Kovács S, Niewada M, Huic M, Drummond M, Ciani O, et al. Quo Vadis HTA for Medical Devices in Central and Eastern Europe? Recommendations to Address Methodological Challenges. Front Public Health. 2020;8:612410.
59. Gopalappa C, Guo J, Meckoni P, Munkhbat B, Pretorius C, Lauer J, et al. A Two-Step Markov Processes Approach for Parameterization of Cancer State-Transition Models for Low- and Middle-Income Countries. Med Decis Making. 2018;38(4):520-30.
60. Schiffer JT, Schiffer CA. To what extent can mathematical modeling inform the design of clinical trials? The example of safe dose reduction of tyrosine kinase inhibitors in responding patients with chronic myeloid leukemia. Haematologica. 2018;103(11):1756-7.
61. Herzog SA, Blaizot S, Hens N. Mathematical models used to inform study design or surveillance systems in infectious diseases: a systematic review. BMC Infectious Diseases. 2017;17(1):775.

Legends to illustrations

Figure 1. Incremental cost-effectiveness ratio for bladder cancer screening with different prevalence rate for the disease

The Legend to Figure 1. Squares reflect the outcomes "per cancer detected", circles reflect the outcomes the life years saved or quality adjusted life years. The ICERs under the axe represent cost-saving outcomes. The grey circle reflects the ICER in the UK study (assumed cost-effectiveness threshold £20,000).

Figure 2. Critical appraisal of the economic models using the Philips et al. checklist

The Legend to Figure 2. Dimensions of Quality in the Philips et al checklist:

S1 Clear statement of decision problem, defined objectives and decision makers; S2 Clear statement, justification, and consistency of scope and perspective; S3 Rationale for Structure explained and based on evidence; S4 Structural Assumptions justified and reasonable; S5 Strategies/comparators defined with all the options considered; S6 Model Type based on decision problem; S7 Sufficient and justified time horizon; S8 Disease states/pathways reflect biological process; S9. Cycle length justified by the nature of the disease; D1 Data identification is transparent, appropriate, justified, and high quality; D2a Baseline data described and justified; D2b Treatment effect based on recognised meta-synthesis, justified extrapolation and survival, with all assumptions documented and justified; D2c Costs and discounting accord with standard guidelines; D2d Quality of life weights (utilities) appropriate, justified, and referenced; D3 Data incorporation justified and transparent; C. Internal and external consistency is evaluated.

The categories used: "Yes", "No" (No, Partially, or Can't tell), "NA" (not applicable).

Table 1. Bladder Cancer Screening Cost-Effectiveness Studies: PICO and the outcomes

Author /year	County/ Perspective /Funding	Population	Intervention(s)/ Comparator	Disease/ Frequency of screening	Health Outcomes	Incremental costs, base-case	Incremental effects, base-case	ICER	Results summary
Lotan, 2006 **[34]**	US / NS / NS	High-risk (heavy smoking or occupational exposure)	Bladder markers NMP22 / No screening	Bladder cancer /1-time screen; annual; biennial	LYS	$101 per 1 screened (cost savings)	0.003 LYS per 1 screened	Intervention resulted in cost-saving even at a 2% annual cancer incidence. An annual incidence of 1% resulted in ICER $35,358 /LYS	Cost-effective in a high-risk population
Svatek, 2006 **[32]**	US / NS / NS	General population and men 50–59, 60–69, and 70–79 years (defined as high-risk)	Bladder markers NMP22 / No screening	Bladder cancer /1-time screen	Cancer detected	NS	NS	Cost per cancer detected: general risk 50–59 years: $783,913, 60–69 years: $269,028; 70–79 years: $139,305; High risk: <40 years: $6,690; 41–50 years: $9,245; 51–60 years: $4,530; 61-70 years: $2,341; 71-80 years: $2,090.	Cost-effective in a high-risk but not the whole population
Teoh, 2018 **[33]**	UK / NHS perspective /no funding	Men aged 65 years and above (defined as high-risk)	Home dipstick haematuria testing / No screening	Bladder and kidney cancers /1-time screen	LYS and QALYs	£37.14 per 1 screened	0.009 LYS per 1 screened; 0.008 QALY per 1 screened	£4198.62/QALY and £4016.80 /LYS (combined bladder and kidney cancers)	Cost-effective in a high-risk population
Okubo, 2022 **[29]**	Japan /Societal/ State funding*	Patients in the SHC (aged 40-79 years)	Dipstick haematuria for all attendants of the SHC / current SHC	Bladder, other kidney or urological diseases / Annual screening	QALYs	-¥97	0.0000098 QALY	Cost-saving	Screening for dipstick haematuria in addition to the standard SHC is cost-saving

The Legend: *ICER – incremental cost-effectiveness ratio; LYS – life-years saved; NHS - National Health Service; NS – not stated; PICO – Population, intervention, comparators, outcomes; SHC - Specific Health Checkup; QALY – quality-adjusted life-years; * Japan Agency for Medical Research and Development and Japanese Grant in Aid for Scientific Research*

Table 2. Bladder Cancer Diagnosis and Surveillance Cost-Effectiveness Models: PICO and the outcomes

Author/year	County/ Perspective /Funding	Population	Intervention(s)/ Comparator	Outcomes	Incremental costs per 1 person*	Incremental outcome per 1 person *	ICER per Outcome**	Conclusions
Awamlh, 2015 [35]	US / NS/ NS	Patients with LR papillary NMIBC	Office-based / Operating Room-based management of LR papillary NMIBC	QALY	-$511	0.03	Dominating	Surveillance cystoscopy coupled with outpatient fulguration was more cost-effective than traditional OR-based TURB for managing patients with LR recurrent NMIBC.
Bekker-Grob, 2009 [30]	Netherlands/ Societal[1]/ Efficiency Research program of the Netherlands OHRD	Patients with NMI urothelial carcinoma	Less invasive conventional surveillance[3] / Conventional surveillance[4]	Probability of losing the bladder, Probability of having a progressive tumour	€ 671	Losing the bladder: 1.8% Having a progressive tumour: 0.03%	Dominated	Currently available markers are not accurate enough to use as a replacement for cystoscopy in the surveillance of BC.
Dansk, 2015 [43]	Sweden / Purchaser (hospitals and other healthcare providers) / NS	Patients with NMIBC receiving a TURBT	HAL BLC+WLC/WLC	Number of rigid cystoscopies, flexible cystocsopies, cystectomies, fulgurations, FP, bed days	SK 218,709 (1.6%)	Number of rigid cystoscopies: -3.7%; Cystectomies: -4.3%; Number of fulgurations: 9.5% False-positive: 8.1% Number of bed days: -4.1%	NA	BLC is likely to be both cost saving and adding clinical benefits
Garfield, 2013 [36]	US / Health care system/ NS	Patients with suspected new or recurrent NMI BC	HAL BLC+WLC/WLC	QALY	- $4,660	0.5	Dominating	Lower clinical and cost burden for BLC
Georgieva, 2019 [46]	US/ Payer / NS	Adults with gross and microscopic haematuria	Guidelines: AUA, KP, CUA, Dutch Guidelines on Haematuria 2010	Number of cancers detected	$ 10-341	0.00001-0.0008	$23,900-1 million per cancer detected	Harms may outweigh the advantages of early diagnosis.
Halpern, 2017 [41]	US/ Payer / NS	(1) Adult patients with asymptomatic haematuria (2)	(1) CT and cystoscopy/	Number of cancers detected	$120-930	0.0001-0.0222	CT: dominated; cystoscopy: $10,000;	Ultrasound and cystoscopy is the most cost-effective

Author/year	County/ Perspective /Funding	Population	Intervention(s)/ Comparator	Outcomes	Incremental costs per 1 person*	Incremental outcome per 1 person *	ICER per Outcome**	Conclusions
		High-risk patients with asymptomatic haematuria (males, smokers, age ≥50 years)	(2) renal ultrasound and cystoscopy/ (3) CT/ (4) Cystoscopy				US+ cystoscopy: $54,000; CT+cystoscopy: $6,500 per cancer detected	
Klaassen, 2017 [44]	Canada/ Canadian universal single-payer/ NS	Patients with NMIBC	HAL BLC+WLC/WLC	Recurrences, bed days, overall costs	$1,236–1,372	Recurrences: -0.1 Bed days: -0.3	$19 354 - $28 463 per recurrence saved	If BLC improves progression rates, there would be improved cost-effectiveness
Lotan, 2018 [40]	US / Medicare/ NIH	Patients with MIBC	Biomarkers to guide the use of NAC /radical cystectomy or NAC followed by radical cystectomy	LYS	$1,496-$10,040 between the biomarkers and traditional approaches	Mean overall survival: -0.21 to 0.43 life years between the biomarkers and traditional approaches	RC alone: $35,259; RC + NAC (100% patients): $32,129; DNA-repair genes: $31,482; ERCC2: $35,072; RNA subtyping: $35,794.	The most cost-effective strategy was based on DNA-repair genes.
Mossanen, 2019 [47]	US/ NS/ NS	65-year-old male compliant patient with NMIBC	No interventions (aims to compare costs across risk categories)	NA	The cumulative 5-year costs for LR, IR, HR NMIBC were $146,250, and $37,427 and $366,143, respectively.	NA	NA	Detection of intermediate or HR disease before it progresses may improve clinical and economic outcomes.

Author/year	County/ Perspective /Funding	Population	Intervention(s)/ Comparator	Outcomes	Incremental costs per 1 person*	Incremental outcome per 1 person *	ICER per Outcome**	Conclusions
Mowatt, 2010 [39]	UK/ Payer/ NICE, UK	Adults with symptoms of bladder cancer (microscopic or gross haematuria or lower urinary tract symptoms)	26 diagnostic strategies (combinations of FC, cytology, three types of biomarkers (NMP22, FISH, ImmunoCyt), WLC and PDD/compared to cytology + WLC	Number of cases detected, LYS	£6-547	Number of cases: -14 to 8 LYG: -0.04-0.02	Not dominated strategies per LYS are: cytology +PDD (£3,423), FISH+WLC (£5,575), FISH+PDD (£2,762), ImmunoCyt +PDD (£28,864), FS+FISH+PDD (£60,284), FS+ ImmunoCyt +PDD (£309,256), FS+ ImmunoCyt +PDD (£237,863)	Cytology followed by WLC in initial diagnosis and follow-up is the most cost-effective with the current threshold
Rodgers, 2006 [42]	UK/ NHS/ NICE, UK	Patients with suspected haematuria	1. Dipstick alone / 2. Dipstick + routine microscopy if positive/ 3. Dipstick + immediate microscopy if positive/ 4. Immediate microscopy/	True cases detected	Increment from cheapest intervention, depending on prevalence: 1. Dipstick: cost-saving -£3.1; 2. Dipstick routine	Increment from cheapest intervention, depending on prevalence: 1. Dipstick: 0 2. Dipstick routine microscopy (-0.016 to -0.087) 3. Dipstick immediate microscopy: 0 4. Immediate microscopy: 0.007-0.036;	ICER compared to cheapest intervention, all dominated except Immediate microscopy: £816-4.9k (depending on prevalence)	At a willingness to pay of £1900 per additional case (the mean ICER), the probability that 'immediate microscopy' will be cost-effective is approximately 0.5.

Author/year	County/ Perspective /Funding	Population	Intervention(s)/ Comparator	Outcomes	Incremental costs per 1 person*	Incremental outcome per 1 person *	ICER per Outcome**	Conclusions
			5. Routine microscopy		microscopy £5 – £12.4; 3. Dipstick immediate microscopy: cost-saving -£2.7; 4. Immediate microscopy: £29.4-£34.9; 5. Routine microscopy: £27.3-£29	5. Routine microscopy: -20 to -105.		
Rodgers, 2006 [42]	UK/ NHS/ NICE, UK	Patients with microscopic haematuria defined as LR	1. Cystoscopy alone/ 2. Cytology + cystoscopy/[5] 3. NMP22 + cystoscopy/[5] 4. BTA + cystoscopy/[5] 5. MCM5 +cystoscopy/[5] 6. FISH +cystoscopy/[5] 7. Cystosonography + cystoscopy/[5] 8. Virtual cystoscopy + cystoscopy[5]	True cases detected	Increment from cheapest intervention (Cytology + cystoscopy): 1. Cystoscopy: £491 2. – 3. NMP22 + cystoscopy: £67 4. BTA + cystoscopy: £77 5. MCM5 +cystoscopy: £45 6. FISH +cystoscopy:£35 7. Cystosonogr	Increment from cheapest intervention (Cytology + cystoscopy): 1. Cystoscopy: 10 2. – 3. NMP22 + cystoscopy: -8 4. BTA + cystoscopy: -10 5. MCM5 +cystoscopy: 30 6. FISH +cystoscopy: -16 7. Cystosonography + cystoscopy: -3 8. Virtual cystoscopy + cystoscopy: -5.	All dominated comparing to Cytology cystoscopy, except MCM5 cystoscopy - £1,500, Cystoscopy alone £49,000	All strategies detect original cancer cases. Cytology cystoscopy and MCM5 cystoscopy are less costly than other strategies

Author/year	County/ Perspective /Funding	Population	Intervention(s)/ Comparator	Outcomes	Incremental costs per 1 person*	Incremental outcome per 1 person *	ICER per Outcome**	Conclusions
					aphy + cystoscopy: £137 8. Virtual cystoscopy + cystoscopy: £21			
Rose, 2016 [45]	Sweden / Swedish healthcare (Payer) /Funding: Photocure ASA	patients with suspected new or recurrent NMIBC	HAL BLC+WLC/WLC	Progression to MIBC, number of cystoscopies	- 9.35 million SEK	Progression to MIBC: -17.2% Number of cystoscopies: -1%; Number of resections: -3% Number of cystectomies: -10%	NR	Instillation of HAL for TURBT is likely to be cost-neutral or cost-negative; however, clinical benefits are likely to be observed.
Sutton, 2018 [37]	UK/ Health-care provider/ SBRI Stratified Medicine Connecting the UK Infrastructure Phase I grant.	Adults with haematuria	DCRSHP / flexible cystoscopy	LYG, QALY	£ 0.76	LYG: 0.013 QALY : 0.0001	£20,088 per QALY	At a threshold of £20,000 per QALY, DCRSHP has a probability of 0.68 of being cost-effective
Yuan, 2012 [38]	US/ NA (clinical model)/ The National Science Foundation	Patients with low grade NMIBC	Surveillance strategies: AUA guidelines / EAU 2009 / optimal policy for cystoscopy	Progression rate, Cystoscopies, LYS, QALYs	NA	AUA/ EAU: Progression rate: -0.4% ; Cystoscopies: -6.48; LYS: 0.02 for males and 0.03 for females; QALYs: 0 Optimal policy dominated in all outcomes	NA	EAU and AUA guidelines didn't dominate each other

The Legend: *BC – bladder cancer; BLC - blue light cystoscopy; DCRSHP - Diagnostic classifier for risk stratification of haematuria patients; FC - flexible cystoscopy; FISH - Fluorescence in situ hybridization diagnostic test; FP - false-positive; HAL - Hexaminolevulinate; HR – high-risk; ImmunoCyt - Immunocyt™; IR – intermediate risk; LR - low-risk; NAC - neoadjuvant chemotherapy; NIH - the National Institute of Health; NMP22 - nuclear matrix protein diagnostic test; NMI - nonmuscle-invasive; NS – not stated; OHRD - Organization for Health Research and Development; PDD - photodynamic diagnosis; WLC - white light cystoscopy.* ***Guidelines:*** *AUA – American Urological Association guideline, 2012 [do not explicitly risk-stratify surveillance recommendations, and outline a schedule of cystoscopy every 3 months for 2 years, every 6 months for the*

next 3 years, and annually thereafter for patients without recurrence]; EAU – guidelines of the European Association of Urology (explicit risk strati cation and, among low risk cases, recommends surveillance cystoscopy at 3 months, 9 months, and annually thereafter for patients without recurrence); KP – Kaiser Permanente Southern California Standardized Haematuria Evaluation, 2012; CUA – Canadian Urological Association guideline, 2009; Dutch – Dutch Guidelines on Haematuria 2010.

**Per patient over model timeframe; ** ICER per QALY or LYG is presented if reported; [1]Stated by the authors, payer perspective since only direct medical costs are reported; [2] Only direct medical costs are included; [3]Every 3 months microsatellite analysis with control cystoscopy at 3, 12, and 24 months; [4]Every 3 months cystoscopy; [5]If positive, otherwise screen patients every 3 months for 1 year.*

Table 3. Modelling bladder cancer in diagnostic and screening models

Characteristics	Types of the models identified	Studies
Bladder cancer states	No classification system for disease progression is used	[30, 35, 40, 42, 45-47]
	Tumour, Node, Metastasis system	[34, 36]
	Risk-based classification states	[33, 37, 39, 43, 44]
	Numerical staging	[29]
Bladder cancer natural history	No progression of cancer	[33-36, 40-42, 45, 46]
	Progression only for diagnosed cancer	[29, 30, 47]
	Progression of undiagnosed cancer if false-negative test	[37, 39, 43, 44]
	Progression of secondary asymptomatic cancer	[38]
	Progression of any asymptomatic cancer	No models