Applied Health Economics and Health Policy (2024) 22:131–144
https://doi.org/10.1007/s40258-023-00855-z

PRACTICAL APPLICATION

Evidence Synthesis and Linkage for Modelling the Cost-Effectiveness of Diagnostic Tests: Preliminary Good Practice Recommendations

Bethany Shinkins[1,2,3] · A. Joy Allen[4] · James Karichu[5] · Louis P. Garrison[6] · Brigitta U. Monz[7]

Accepted: 5 November 2023 / Published online: 5 February 2024

Abstract

Objectives To develop preliminary good practice recommendations for synthesising and linking evidence of treatment effectiveness when modelling the cost-effectiveness of diagnostic tests.

Methods We conducted a targeted review of guidance from key Health Technology Assessment (HTA) bodies to summarise current recommendations on synthesis and linkage of treatment effectiveness evidence within economic evaluations of diagnostic tests. We then focused on a specific case study, the cost-effectiveness of troponin for the diagnosis of myocardial infarction, and reviewed the approach taken to synthesise and link treatment effectiveness evidence in different modelling studies.

Results The Australian and UK HTA bodies provided advice for synthesising and linking treatment effectiveness in diagnostic models, acknowledging that linking test results to treatment options and their outcomes is common. Across all reviewed models for the case study, uniform test-directed treatment decision making was assumed, i.e., all those who tested positive were treated. Treatment outcome data from a variety of sources, including expert opinion, were utilised for linked clinical outcomes. Preliminary good practice recommendations for data identification, integration and description are proposed.

Conclusion Modelling the cost-effectiveness of diagnostic tests poses unique challenges in linking evidence on test accuracy to treatment effectiveness data to understand how a test impacts patient outcomes and costs. Upfront consideration of how a test and its results will likely be incorporated into patient diagnostic pathways is key to exploring the optimal design of such models. We propose some preliminary good practice recommendations to improve the quality of cost-effectiveness evaluations of diagnostics tests going forward.

Key Points for Decision Makers

Cost-effectiveness models to evaluate diagnostic tests mostly rely on linking outcomes related to a treatment. In turn, that has been assumed to be a given, based on a test result. Understanding the data and assumptions used to link interventions and their outcomes to test results is key to assess the validity of the cost-effectiveness results. Good practice recommendations can provide guidance for developing and appraising such cost-effectiveness models.

1 Introduction

Like treatments, evaluating the cost-effectiveness of diagnostic technologies is undertaken to ensure that the benefits to patients warrant any additional associated cost.

✉ Bethany Shinkins
bethany.shinkins@warwick.ac.uk

1 Health Sciences, Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK

2 Leeds Institute for Health Sciences, University of Leeds, Clarendon Way, Leeds, UK

3 NIHR Leeds In Vitro Diagnostics Co-operative (MIC), Leeds, UK

4 Access and Innovation, Roche Diagnostics UK and Ireland, Burgess Hill, UK

5 Global Access & Policy, Roche Diagnostics Solutions, Pleasanton, California, USA

6 The Comparative Health Outcomes, Policy, and Economics (CHOICE) Institute, University of Washington, Seattle, USA

7 Global Access & Policy, Roche Diagnostics International AG, Forrenstrasse 2, 6343 Rotkreuz, Switzerland

The impact that a test has on patient outcomes is typically indirect; the mechanism of benefit is through a change in patient management and the effectiveness of that patient management in improving patient health. Ideally, the data to inform a cost-effectiveness analysis of a diagnostic technology would come from an "end-to-end" study, i.e., a study which follows patients from the point of testing, through any patient management or treatment given, to the measurement of clinically relevant final outcomes [1].

While such end-to-end studies may be possible in some situations, they may not be feasible, ethical, or advisable in others. For example, consider a new diagnostic test used to identify whether a patient with atrial fibrillation should receive an oral anticoagulant for stroke prevention. Arguably, this diagnostic test should not be assessed in an end-to-end study with stroke as an endpoint, given the very extensive body of evidence demonstrating the effectiveness of anticoagulation in this indication, from both randomised and observational studies, as well as mixed treatment comparisons and meta-analyses [2, 3]. Another possibility is that an end-to-end study is feasible, but could only represent some of the many different possible diagnostic strategies. This is common where a sequence of diagnostic tests is being evaluated, such as in the cost-effectiveness analysis by Faria et al, where there were 32 clinically feasible combinations of tests for prostate cancer [4]. In such complex scenarios, decision analytic models can provide a more useful framework to evaluate the cost-effectiveness of diagnostic tests [5, 6].

Most decision-analytic models of diagnostic tests require linking diagnostic accuracy data to treatment efficacy data to estimate the impact that a test will have on patient outcomes and costs [6]. Currently, there is no specific methodological guidance on how this should be done [7, 8]. Often, modellers may be tempted to assume a 'uniform' action to a test result, such as that all patients receive treatment if they test positive, for example, for a certain infectious disease. However, in clinical practice there will be a probability distribution of how many individuals who test positive actually receive treatment, and there may be differences in treatment strategies based on other clinical factors. Assuming completely test-directed decision making is only likely to be appropriate for very specific situations, such as for companion diagnostics.

A review of Health Technology Assessments (HTAs) in the UK noted that the rigour in which the evidence on treatment efficacy was identified, quality assessed and synthesised within model-based economic evaluations of diagnostic tests was poor and that evidence synthesis efforts were largely focused on diagnostic accuracy [7, 9]. An earlier review of 149 HTAs from eight countries found that intermediate outcomes, such as the impact of test results on patient management, are frequently assessed in medical test HTAs, but interpretation of this evidence is inconsistently reported [10]. It was recommended that evaluators explain the rationale for using intermediate outcomes, identify the assumptions required to link intermediate outcomes to patient health outcomes, and assess the quality of included studies [10].

This paper will build on and expand these recommendations by reviewing the guidance from selected HTA bodies to summarise current recommendations on evidence synthesis and linkage of treatment effectiveness evidence within economic evaluations of diagnostic tests. We will then explore a case study focused on a specific decision problem to better understand current practice. Based on the findings, we derive a set of proposed preliminary good practice recommendations with the aim of advancing the methodological rigour of future cost-effectiveness analyses of diagnostic tests (note: there are likely additional considerations when evaluating screening, monitoring or prognostic tests which are not covered in this paper).

2 Methods

To understand current recommended best practice in terms of evidence synthesis and linkage of treatment-effectiveness data, we reviewed modelling guidance from HTA bodies. We focused on two specific questions: (i) how are test results linked to patient management decisions, and (ii) what evidence is used to translate patient management into clinical outcomes (Fig. 1). We focused on HTA bodies with well-established cost-effectiveness requirements, such as, but not limited to, Australia (Medical Services Advisory Committee, MSAC), Canada (Canadian Agency for Drugs and Technologies in Health, CADTH), and the UK (e.g. The National Institute for Health and Care Excellence, NICE). Ten different guidelines (three for the UK, one in the EU, one for Canada, one for the United States, two for Australia, one for Sweden, and one for the Netherlands) were reviewed in total. The guideline documents were accessed in September 2022. Our aim was to identify any specific guidance on how to identify and synthesise appropriate evidence on treatment effectiveness for the inclusion in a diagnostic cost-effectiveness model, and how to link these data to diagnostic test results.

The second part of this paper focuses on reviewing different cost-effectiveness models for a specific decision problem. We chose the evaluation of the biomarker troponin for the diagnosis of myocardial infarction (MI) as this test is well established, and the recommended actions and treatments are well documented [11], as are suitable study endpoints, such as 30-day mortality in clinical studies [12]. The search was conducted in PubMed, EMBASE, The Cochrane Library, The International HTA Database and EconLit, on July 13, 2022. The search was restricted to English language,

from 2012 onwards, and the following countries: UK, Canada, US, Australia, Sweden and The Netherlands. Search terms included: Modelling studies, troponin, myocardial infarction, and diagnostics. The search resulted in four unique cost-effectiveness models, with additional publications using variations of these unique models. Our aim was to provide a snapshot of current methodological practice in terms of evidence synthesis of treatment effectiveness and linkage of this evidence within the identified models. The focus of data extraction was, therefore (apart from some standard information [e.g., type of model, setting and perspective]), the information that was used to link test results with clinical actions, and to link clinical actions with short- and long-term outcomes. We were not specifically interested in the treatment that was undertaken (e.g., in case of a positive test result), but rather in the clinical outcomes that were modelled, and how they were linked to the patient management decision.

Last, based on information from the HTA guidelines and the observed patterns in our example, we created initial proposed recommendations for linking test results to patient management and for translating patient management to clinical outcomes in model-based economic evaluations of diagnostic tests. For each recommendation, we provide a brief description of the problem to be considered, and specific actions as to how these should be dealt with when developing your own model.

3 Results

3.1 HTA Body Recommendations on Linked Evidence for Diagnostic Models

The Australian HTA Guidelines from the MSAC had the most extensive guidance for evidence linkage containing several specific sections on linked evidence in diagnostic technology assessments (Technical Guidance [TG] 12 and 13) [13]. These sections feature in the clinical evaluation section as, in the scenario that end-to-end studies are not feasible, a linked-evidence approach to clinical evaluation may be adopted instead. They provide a very thorough overview of the evidence requirements in this context, and thus are very informative when thinking through how to link this evidence within a decision-analytic model for a diagnostic test.

Particularly useful for our first question is that the guidelines separate the actions that may follow a test result into three components, each of which will impact the resulting adoption rate of an action (TG 12): "*change in diagnostic thinking, change in recommended management and actual management*" [13]. Furthermore, it is stated that consideration should be given to whether the tests under comparison should have different actions following the same results (i.e., a positive result leads to a different action in one test vs the other) and to provide justification either way.

With regard to our second question, TG 13 details the thought process for identifying the most suitable linked evidence, by working through four considerations: (a) availability of management strategy/treatment, (b) effectiveness of management strategy/treatment, (c) what happens to wrongly classified patients (false positive and false negatives), and (d) is the available evidence (i.e., created under current tests) likely applicable to the population selected with the new test [13].

The NICE Health Technology Evaluations Manual in the UK provides guidance throughout the document on the importance of evidence linkage for cost-effectiveness models of diagnostic interventions, as direct evidence leading from test result to relevant clinical outcomes are mostly not available [1]. As potential data sources, NICE recommends the use of study data, clinical guidelines, or rely on expert clinical input, if needed. We could not identify specific recommendations relating to our two questions in the Canadian guidelines [14].

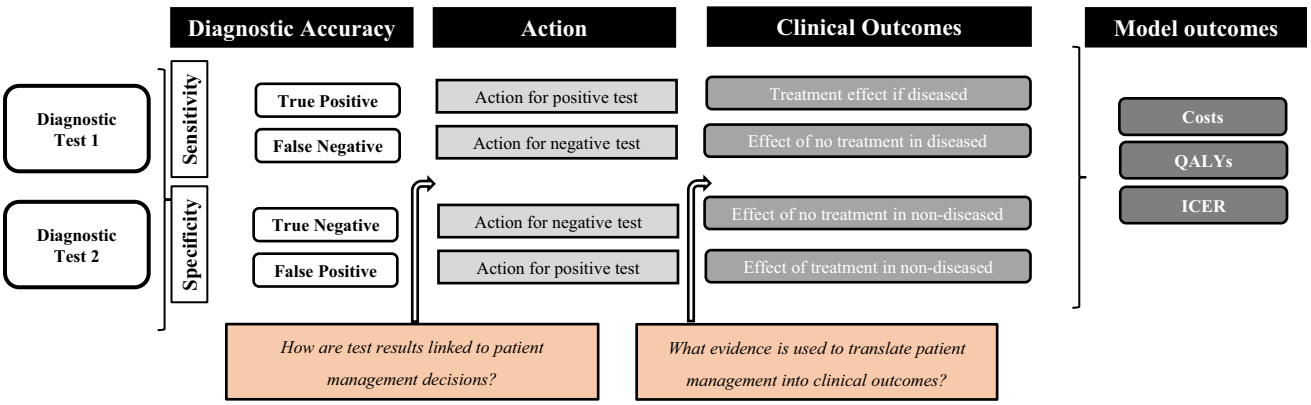

Fig. 1 Schematic of a cost-effectiveness model for comparing two diagnostic tests, showing two modelling questions on linked evidence. *ICER* incremental cost-effectiveness ratio, *QALYs* quality-adjusted life years

3.2 Example of Troponin for Diagnosis of Myocardial Infarction

Tables 1 and 2 summarise the publications reviewed for data extraction (base-case models for each) and provide details on the data linkages performed for each model [15–18]. Table 1 provides a summary of key model characteristics and patient utility data, while Table 2 shows details with regard to linked clinical data.

The model structure from Thokala et al. [15] was applied to other decision contexts [19, 20]. Westwood et al. [16] expanded the model structure by Thokala et al. [15] including, among others, an additional Markov model with more health states and using outcomes from larger studies [21, 22]. In a recent publication, Westwood et al. [23] utilised the same model structure as in their 2015 publication [16].

None of the described models included actual details of the treatment, but Vaidya et al. assumed that in the Dutch setting percutaneous coronary interventions (PCI) would be used for all patients who tested positive [17]. In the CADTH model, treatment was implicitly included by using a weighted cost, based on the observed mixture of codes for MIs treated with either bypass surgery, PCI or non-invasively [18]. Overall, the models worked by directly linking outcomes to test results.

The Dutch model [17] linked outcomes from patients having undergone PCI [24, 25], given their model assumed that all MI patients will undergo PCI. Those came from studies in US hospitals. Others aimed to link outcomes data on mortality and re-infarction mostly from the country where the model was based [15, 16]. One study from the US [21] on the outcomes of false negative tested patients discharged from hospital untreated, was used in two models [16, 17]. In all models where the outcomes (mainly mortality) have been tested in one-way sensitivity analyses, they had significant impact on the findings of the model [16–18].

Patient utility data for the four investigated models predominantly came from the UK (Table 1). In summary, model input on how treatment was implemented based on test results was assumption based, and in a way that ALL patients uniformly received the action (e.g., all patients with positive test result received the treatment; all patients with negative test result were discharged).

For the linked clinical outcomes, a variety of sources have been utilised, including clinical and observation studies (e.g., disease and procedure registries), meta-analysis, inputs and outputs from other published cost-effectiveness models, national statistics (for overall life expectancy) and expert opinion. The data may have come from the country where the model was based, or from another country. The models used published data, or re-analysed patient-level data to fit the model population [15]. For some of the linked outcomes, there was a significant time period between the outcomes data collection and the model publication. For example, clinical data collected in 1993 [21] was used in a model published in 2015 [16], representing more than 20 years of time difference.

For false positive patients it was mostly assumed that no harm was done and normal life expectancy was modelled, with the exception of the Dutch model that accounted for the risk of an invasive procedure. There was wider variation in the modelling approach for false negative patients. Data from clinical studies were used to model short-term outcomes, while assumptions and previous model inputs and outputs were used to extrapolate beyond the initial period.

4 Proposed Recommendations

Given the highly relevant details provided in the Australian guidelines [13], we aimed to translate this information into practical considerations for use in economic modelling.

Our proposed recommendations regarding research question 1 (how is the action based on the test result modelled, see Fig. 1) build on the Australian framework (TG 12) [13], which describes three consequences that may follow a diagnostic test result: (a) change in diagnostic thinking, i.e., how a test is interpreted, (b) change in recommended management, i.e., what recommendations are made in response to the test results, and (c) change in actual management, i.e., what patient management, if any, is adopted.

The recommendations are therefore focused on linking diagnostic test results to actions and should be applied for each testing strategy. If the same linkage assumptions are made across different testing strategies, then this should be explicitly stated. Details of the six proposed recommendations can be found in Box 1. Fundamental to many of these recommendations is the identification and synthesis of 'change in management' studies. These types of studies are well described in the Australian Framework (see TG 12.2). Given the importance of these studies to justify the linking of evidence in decision-analytic models for diagnostic tests, the search and screening of studies, risk of bias assessment, presentation of results and meta-analysis (if appropriate) should follow the same methodological rigour as when synthesising evidence on diagnostic accuracy or treatment effectiveness.

With regard to our second question (what evidence is used to translate the patient management into clinical outcomes, see Fig. 1), TG 13 in the Australian guidelines [13] is centred around four relevant aspects (a) availability of treatment, (b) treatment effectiveness, (c) outcomes of wrongly classified patients, and d) the applicability of the evidence for treatment effectiveness. Again, given the high relevance of this information, we have

Table 1 Overview of cost-effectiveness models for early diagnosis of myocardial infarction

	Thokala et al. (2012) [15]	Westwood et al. (2015) [16]	Vaidya et al. (2014) [17]	CADTH (2013) [18]
Model structure	Decision tree	Decision tree (30 days), followed by Markov model	Decision tree	Decision tree (1-year), followed by long-term part (type not specified)
Clinical outcome(s)	Death and re-infarction	Death and re-infarction	Death	Death
Model output	Cost/QALY	Cost/QALY	Cost/QALY	Cost/QALY
Health states	3 states: Post-MI, post re-infarction, general population	3 states: Post-MI, unstable angina, general population	1 state: Post-MI	2 states: Post-MI, general population
Source of data for utilities	Post-MI: Single-centre study in ER in UK, $n = 972$, year 2001/2002, EQ-5D; post re-infraction: expert opinion (based on post-MI utility multiplied with correction factor) [29]; general population: 5-year age-band utilities, EQ-5D, UK general population [30]	Post-MI: Utility decrements applied to general population values, decrements based on registry-nested survey in German survivors of MI (survival time between 2 and 20 years), $n = 2950$ patients, EQ-5D [33] General population and unstable angina [30]	Post-MI: One single UK-based utility value from patients surviving MI [31]	Post-MI: Utility decrements applied to general population values, decrements based on registry-nested survey in German survivors of MI (survival time between 2 and 20 years), $n = 2950$ patients, EQ-5D [33] General population: 10-year age-band utilities by gender, EQ-5D, UK general population [32]
Time horizon	Lifetime	Lifetime	Lifetime	Lifetime
Perspective	NHS, UK	NHS, UK	Dutch health care system	Canadian public health care system
No. of testing strategies compared	5	7	3	3
Same test result = same action[a]	Yes	Yes	Yes	Yes

ER emergency room, *MI* myocardial infarction, *NHS* National Health Service, *QALY* quality-adjusted life-year, *UK* United Kingdom

[a]Meaning regardless of test strategy investigated; e.g., a true positive result leads to the same action across all investigated test strategies

Table 2 Cost-effectiveness models for early diagnosis of myocardial infarction: extracted data on patient management and clinical outcome(s)

	Question 1: Linkage test result to patient care		Question 2: Linkage patient care to clinical outcomes	
	Model input used	Data source	Model input used	Data source
Thokala et al. (2012) [15]				
True positives	All treated	None specified (assumption)	1st year: Re-analysis of patient-level data for re-infarction or death ($n = 80$)	Single-centre study ($n = 2092$, 2008–09, UK) assessing troponin cut-offs for MI diagnosis, including 1-year treatment and outcomes (MI and death) [34]. Used subset with patient characteristics matching model population
			> 1st year: Life expectancy by age and MI re-infarction status	Based on a previously published (1999) cost-effectiveness model [35]. The estimates for that model were taken from a 'health disease policy model' in the US (referenced as 'personal communication' from one of the co-authors)
False negatives	All discharged	None specified (assumption)	1st year: Re-analysis of patient-level data for re-infarction or death ($n = 90$)	Single-centre study ($n = 2092$, 2008–09, UK) assessing troponin cut-offs for MI diagnosis, including 1-year treatment and outcomes (MI and death) [34]. Used subset with patient characteristics matching model population
			> 1st year: Life-expectancy by age and MI re-infarction status	Based on a previously published (1999) cost-effectiveness model [35]. The estimates for that model were taken from a 'health disease policy model' in the US (referenced as 'personal communication' from one of the co-authors)
True negatives	All discharged	None specified (assumption)	Life-expectancy of general population	UK Office of National Statistics
False positives	Confirmatory test 10 hours post admission (assumed to have perfect accuracy) and all discharged	None specified (assumption)	No treatment until confirmatory testing Life-expectancy of general population	Assumption UK Office of National Statistics
Westwood et al. [16]				
True positives	All treated	None specified (assumption)	First 30 days: 30-day mortality	US multicentre study ($n = 10{,}689$, 10 centres, 1993), patients admitted to the ER with symptoms suggestive of MI, followed-up for 30 days [21]
			> Day 30: Study data from hospital and death data record linkage	Record linkage study ($n = 387{,}452$, 2004–10, England) assessing 7-year mortality and re-infarction rates in 30-day survivors of MI [22]

Table 2 (continued)

	Question 1: Linkage test result to patient care		Question 2: Linkage patient care to clinical outcomes	
	Model input used	Data source	Model input used	Data source
False negatives	All discharged	None specified (assumption)	First 30 days: 30-day mortality	US multicentre study (n = 10,689, 10 centres, 1993), patients admitted to the ER with symptoms suggestive of MI, followed-up for 30 days [21]
			Days 30–365: Increased RR of re-infarction and death (compared to treated)	Single-centre study (n = 2092, 2008–09, UK) assessing troponin cut-offs for MI diagnosis, including 1-year treatment and outcomes (MI and death) [34]. Used subset with patient characteristics matching model population
			> 1 year: RR = 1 (compared to treated)	Assumption
True negatives	All discharged	None specified (assumption)	Age-dependent mortality from general population	UK Office of National Statistics
False positives	All treated	None specified (assumption)	Age-dependent mortality from general population	Assumption UK Office of National Statistics
Vaidya et al. [17]				
True positives	All treated	None specified (assumption)	Peri-procedure mortality (PPCI)	Procedure: Based on 2 studies: (1) patient registry study [24] (n = 43,801, 2005–06, US) patients with ST elevation MI undergoing PPCI, (2) PCI patient registry study [25] (n = 559,273 procedures, 2001–04, US)
			Average post MI life-expectancy	Cost-effectiveness model (Norway, 2005) [36]
False negatives	All discharged after repeat testing	None specified (assumption)	Patient-level data with 30-day outcomes	US multicentre study (n = 10,689, 10 centres, 1993), patients admitted to the ER with symptoms suggestive of MI, followed-up for 30 days [21]
			Average post MI life-expectancy	Cost-effectiveness model (Norway, 2005) [36]
True negatives	All discharged after repeat testing	None specified (assumption)	No details	
False positives	All treated	None specified (assumption)	Procedure mortality (PPCI)	PCI Registry in the US (2001–04), n = 559,273 procedures [25]
			Further life expectancy	No details
CADTH [18]				
True positives	All treated	Assumption	1-year mortality: Meta-analysis, stratified by time of treatment initiation	Meta-analysis of three studies: 1) PCI Registry [37] (n = 1,486, 1999–2004, US), 2) Single centre study (n = 760, 1990–91, US), assessing mortality by type of MI [38], 3) single-centre study (n = 1188, 2007, Finland), assessing mortality [39]; stratification by time to treatment: expert opinion
			> 1-year overall mortality: RR compared to general population based on study	Single-centre study (n = 760, 1990–91, US), assessing mortality by type of MI [38] and Canadian life tables

Table 2 (continued)

	Question 1: Linkage test result to patient care		Question 2: Linkage patient care to clinical outcomes	
	Model input used	Data source	Model input used	Data source
False negatives	All discharged after repeat testing	None specified (assumption)	1-year mortality: Previously published c/e model input	Single-centre study ($n = 2092$, 2008–09, UK) assessing troponin cut-offs for MI diagnosis, including 1-year treatment and outcomes (MI and death) [34]. Used subset with patient characteristics matching model population (as cited in [13])
			> 1-year overall mortality: RR compared to general population based on study	Single-centre study ($n = 760$, 1990–91, US), assessing mortality by type of MI [38] and Canadian life tables
True negatives	All discharged after repeat testing	None specified (assumption)	Life-expectancy of general population	Canadian life tables
False positives	All treated	None specified (assumption)	Life-expectancy of general population	Canadian life tables

c/e cost-effectiveness, *ER* emergency room, *MI* myocardial infarction, *(P)PCI* (primary) percutaneous coronary intervention, *RR* relative risk

converted these into good practice recommendations to support the translation of patient management to clinical outcomes when developing model-based economic evaluations of diagnostic tests (see Box 2). The first aspect around availability of treatment is already captured in the final recommendation in Box 1, and therefore is not included in Box 2.

There were a number of observations from the case study which may be useful to note here. First, with regard to selecting treatment and its effect, there may be a need to trade precision for accuracy when synthesising data on outcomes. For example, Westwood et al. used a patient-level re-analysis, which meant reducing the original sample size from 2092 to 170 [16]. Furthermore, ensure that the outcomes used as model input indeed link to the selected treatment. For example, Vaidya assumed that all patients with a positive test undergo PCI. Mortality was then derived from two PCI registries [17]. Additionally, in all of the models where linked outcomes were assessed in a one-way sensitivity analysis, outcomes had a significant impact on model results [16–18]. We therefore recommend that this is done as standard practice. Another consideration is to assess the recency of evidence for linked outcomes and consider whether practice patterns have changed to the extent that the evidence will be outdated. Typically, neither benefits nor harms are assumed for true negative patients, although this may not be a reasonable assumption, depending on the risks of the diagnostic test itself. If harm of the diagnostic procedure is a concern, this should be considered in the model (e.g., if invasive procedures are needed to obtain samples for testing), especially if different among modelled treatment strategies.

Second, modelling outcomes of wrongly classified patients can be challenging, as data may be lacking. In our case study, Vaidya et al. used the PCI procedure risk for false positive patients [17]; and two publications used outcomes from a clinical study with patients discharged from hospital despite having an MI for false negative patients [16, 17]. If evidence is missing and assumptions need to be made, clinical validation by experts will become important, as will be robust sensitivity analyses of those assumptions.

More generally, when reporting diagnostic models with linked evidence, it would increase transparency if model inputs were presented by test result category (true positives, false positives, etc.), as this would make it easier to assess what evidence has been used for outcomes and whether linkage assumptions are reasonable.

Similar to models investigating therapeutic interventions, model calibration may be required, as using linked evidence may lead to overly optimistic or pessimistic cumulative model outcomes. For example, the standard of care arm model outcomes, such as projected life expectancy, could be compared to available evidence not used for model input, such as from disease-specific national statistics.

Box 1. Proposed Preliminary Good Practice Recommendations for Linking Test Results to Patient Management in Model-Based Economic Evaluations of Diagnostic Tests

Change in Diagnostic Thinking

Recommendation: Consider whether everyone in the target population would get a valid test result

Is it possible that the test may produce an inconclusive test result, or an uninterpretable or missing test result? For example, a urine culture may be contaminated with bacterial overgrowth in a sample where it is delayed in reaching the laboratory. These data should have been reported within diagnostic accuracy studies (according to STARD [40]). What happens to individuals where this occurs? Is the test repeated? Is a different test used? Both of these scenarios have cost and potential outcome implications and therefore should be accounted in the model. Additionally, inconclusive, uninterpretable or missing test results may be more or less likely if the individual has the disease in question.

Action: If relevant, adjust either the model structure to allow for inconclusive, uninterpretable or missing test results or the model inputs (e.g., increase overall test costs to account for repeat tests). Where uncertainty exists in the proportion of inconclusive, uninterpretable or missing test results, explore the impact of varying the proportion via one-way sensitivity analyses.

Recommendation: Consider the timing of the test and test result

Consider when the test would be done in routine clinical practice and how long it would take to get test result back. If there is a long wait for a test or test result, then is it possible that other sources of information would be used to inform a diagnostic decision and subsequent management? Would treatment be started without the test result, and then reviewed once the test result is available?

Action: Engage with relevant experts (e.g., clinical, laboratory) to understand how the test would fit within routine practice and the likely time to test and test result. Explore whether this is likely to be within an acceptable time-frame to support patient management decision making. Adjust model structure/inputs accordingly.

Recommendation: Consider whether the test result is likely to be the sole determinant of the diagnosis

What information other than the test results may impact the diagnostic thinking? Consider whether there are other factors such as patient characteristics, symptoms, other tests and examinations that could lead to a different diagnostic conclusion. If so, to what proportion of individuals does this apply?

Action: Understand how the test result would be used in routine practice to inform a diagnostic decision. Is there any other information which could 'override' a positive or negative test result? This is particularly important to consider when thinking about false positive and false negative results. For example, if an individual had a negative test result, but other information suggests that the individual does have the disease in question—which would hold more weight? Real-world data, medical guidelines and clinical expert opinion may be appropriate sources for the required information.

Change in Recommended Management

Recommendation: Consider whether a positive test result is likely to be the sole determinant of the choice of patient management

Is it likely that all patients who test positive would be recommended the same patient management? Are there certain patient characteristics, such as frailty or co-morbidities which mean that some patients would not be managed in the same way? Does the severity of the disease have an impact on the type of treatment given?

Action: Review the literature to identify studies that provide evidence on recommended patient management for those with different test results, i.e., 'change in management' studies. Adjust model structure/inputs to reflect this evidence. Where uncertainty lies, explore via scenario and sensitivity analyses. In the absence of any evidence, engage with clinical experts to explore how those with a positive (true and false) test result would be managed in routine practice. Avoid making blanket assumptions about recommended patient management unless truly reflective of the clinical scenario, e.g., in the case of a companion diagnostic.

Recommendation: Consider whether a negative test result is likely to be the sole determinant of the choice of patient management

Is it likely that all patients who test negative would be recommended the same patient management? Are there certain scenarios where some patients would be managed differently, e.g., if they remain symptomatic? If the patient re-presents for testing, is the severity of the disease likely to be worse? Does this have implications on patient management?

Action: Review the literature to identify studies that provide evidence on recommended patient management for those with different test results, i.e., 'change in management' studies. Adjust model structure/inputs to reflect this evidence. Where uncertainty lies, explore via scenario and sensitivity analyses. In the absence of any evidence, engage with clinical experts to explore how those with a negative (true and false) test result would be managed in routine practice. Avoid making blanket assumptions about recommended patient management unless truly reflective of the clinical scenario, e.g., in the case of a companion diagnostic.

Change in Actual Management

Recommendation: Consider whether all individuals are likely to receive the recommended patient management

Are all patients likely to consent to the recommended patient management? Is the treatment available for all those eligible? Is affordability or access an issue? Are there any patients, clinical and system factors, which may mean that some individuals cannot receive the recommended patient management?

Action: Review the literature to identify studies that provide evidence on actual patient management for those with different test results, i.e., 'change in management' studies. Real-world data could also be useful resources to establish expected versus observed treatment rates and to allow investigation of associated factors. Adjust model structure/inputs to reflect this evidence. Where uncertainty lies, explore via scenario and sensitivity analyses. In the absence of any evidence, engage with clinical experts to explore how actual patient management could differ to recommended patient management. Avoid making blanket assumptions about actual patient management unless truly reflective of the clinical scenario.

Box 2. Proposed Preliminary Good Practice Recommendations for Translating Patient Management to Clinical Outcomes in Model-Based Economic Evaluations of Diagnostic Tests

Effectiveness of Treatment

Recommendation: Outcomes most relevant for the model target population undergoing testing should be selected, even if estimates for outcomes decrease in precision

Thinking about the studies on which you are basing treatment effectiveness outcomes, is the patient group the same as your target population undergoing testing? Are there any inclusion or exclusion criteria which mean that they are likely to differ, e.g., include a population with a different spectrum of disease severity? Is the setting different (primary vs secondary care) raising concerns about applicability?

Action: When incorporating data on treatment effectiveness, there may be a need to trade precision for accuracy when synthesising data on outcomes. Outcomes most relevant for the model target population undergoing the testing should be selected, even if estimates for outcomes decrease in precision. The impact of lower precision in these estimates can be conveyed through sensitivity analyses. If the low precision in these estimates has a notable impact on cost-effectiveness, consider conducting a value of information analysis to explore the value of reducing the uncertainty in the treatment effectiveness parameters, i.e., explore whether further research is potentially worthwhile.

Recommendation: Always explore the impact that treatment effectiveness has on the overall cost-effectiveness results

Given the typically indirect nature that a diagnostic test impacts on health outcomes, the cost-effectiveness of a test is typically heavily driven by the effectiveness of the patient management or treatment following the test. Thus, it is important to understand the extent to which treatment effectiveness and uncertainty in those parameters impact upon the overall cost-effectiveness of the diagnostic strategy.

Action: Always conduct one-way sensitivity analyses on linked treatment outcome parameters.

Recommendation: Consider when treatment effectiveness studies were conducted and whether the timing of the study is likely to impact upon outcome data

Evidence on treatment effectiveness may come from older studies. Consider whether there are factors which make it likely that the outcome data would be different if the study were conducted in those who would currently receive the test.

Action: Prioritise applicable, high quality, more recent evidence on treatment effectiveness over older studies. Where evidence is only available from older studies, in addition to the sensitivity analyses recommended above, highlight this issue in the write-up of your results and discuss any potential factors that may reduce the applicability of this evidence to current patient populations undergoing testing.

Recommendation: Validate assumptions around outcomes for negative and misclassified patients

What are the health outcomes for those who are misdiagnosed (i.e., those who receive either a false positive or false negative result) and those who test negative? Applicable outcome data are rarely 'directly' available for these subgroups and, in many cases, assumptions have to be made.

Action: Validate any assumptions made around outcomes for negative and wrongly classified patients by seeking out real-world data or through discussion with clinical experts. Any assumptions should be clearly described and their impact should be tested using sensitivity analyses.

Applicability of the evidence for treatment effectiveness

Recommendation: Consider whether there remain (see recommendation 1) minor differences between the target population for testing and those who have participated in the treatment-effectiveness studies used to provide outcome data

It can often be challenging to unpick what evidence has been used to assign health outcomes to different test result outcomes, and the quality and applicability of that evidence. Describing, appraising and citing the evidence sources in a structured table which has a row for each test result outcome (e.g., true positives, false positives, etc.) and using a standardised critical appraisal tool can help overcome this issue.

Action: Use a structured framework or checklist to assess and describe the quality, applicability and certainty of treatment effectiveness studies used to provide outcome data. Report the results for each test outcome (i.e. false positives, true positives etc.).

Recommendation: Consider whether the new diagnostic strategy is likely to lead to a change in the (severity) spectrum of disease diagnosed compared to current practice

In situations where a diagnostic test detects more patients with a disease or it detects disease much earlier than current tests, then it is likely that the spectrum of disease in those correctly diagnosed with the disease is different to those previously diagnosed with the disease. This will directly impact the applicability of any existing treatment effectiveness evidence [42], as the benefit from treatment will differ. A risk of overdiagnosis in the situation also needs to be considered.

Action: If a diagnostic test identifies a population with a different spectrum of disease compared to existing tests, 'end-to-end' studies are needed.

Recommendation: Prioritise treatment effectiveness studies which report data on final outcomes rather than surrogate outcomes

Data on final (or most important) health outcomes, e.g., mortality, can sometimes be challenging or may take a long time to collect, and surrogate outcomes are used instead. This is only acceptable where there is robust evidence that an observed change in a surrogate outcome is associated with a concomitant change in the final health outcome [42]. By using evidence on surrogate outcomes, another point where evidence must be linked within the model has been introduced and the validity of this linking of evidence must be fully described and appraised.

Action: In circumstances where data are only available on surrogate outcomes, evidence of causality must be presented to support linking evidence of surrogate outcomes to final outcomes. The impact of this assumed relationship on cost-effectiveness must always be explored using sensitivity and scenario analyses.

5 Discussion

Economic models typically utilise evidence from a variety of sources with differing grades of evidence [26]. Our case example illustrates this, with evidence across a spectrum of grades being used, from meta-analysis, to clinical studies, to real-world data studies and clinical expert advice.

While in cost-effectiveness models of new treatments, typical challenges in assessing clinical outcomes are (a) the extrapolation of evidence beyond observed data in order to relate efficacy to real-world effectiveness, and (b) evidence selection and synthesis for comparators, economic models for diagnostic tests have additional specific challenges that need to be addressed. We aimed to identify these challenges and formulate a preliminary set of recommendations to support model development in the field.

Modelling for diagnostic tests with dichotomous outcomes allows representation of an explicit set of outcomes flowing from the different diagnostic test classifications, i.e., true and false positives, true and false negatives. However, while the modeller can assign actions to test results accurately, the health care provider will not know, when presented with a test result, which patients are correctly (true positive and true negative), and which are incorrectly assigned (false negative and false positive). Other information, such as that from repeated testing or other diagnostic tests, will eventually identify incorrectly classified patients. However, when there is no other existing method to confirm or exclude the diagnosis, such as for first-in-class tests, uncertainty of the correct classification of an individual patient will remain.

In our case study, while evidence was available on the outcomes of patients sent home with an MI (false negative) and those who underwent an invasive procedure while not having an MI (false positive), this may become much more complex in chronic diseases when the correction of the wrong course of action may be months to years later and when this period significantly affects outcomes, such as in oncology.

Real-world data studies may be useful to identify the proportion of patients actually receiving treatment based on a test result. "Purpose-built" datasets (in contrast to datasets re-purposed for research, such as claims data), particularly disease registries, are likely to be the most suited. Such purpose-built cohorts may then also be used to determine true and false positives (for an example see Fernandes et al. [27]).

In addition, a common assumption (highlighted in our case study) within economic modelling for diagnostic tests is that all patients receiving a diagnosis will receive the same course of action, and that all patients with a diagnosis ruled out will receive another common course of action. We know this is unlikely to happen in clinical practice, although clinical practice guidelines and local hospital care protocols aim to foster such consistency in clinical decision making and behaviour. For example, prescription of antibiotics in patients with ventilator-associated pneumonia remained high in a randomised, controlled trial comparing a biomarker-guided approach to antibiotic prescription with standard of care, despite the high negative predictive value of the biomarker-based approach [28].

A further consideration when implementing the methodological recommendations in this paper is to ensure that model development and results are transparently reported. The AGREEDT ('AliGnment in the Reporting of Economic Evaluations of Diagnostic Tests and biomarkers') reporting checklist is a comprehensive reporting tool, which encourages explicit reporting of (1) the impact of a test on patient management strategies, and (2) the impact of patient management strategies on health outcomes and costs [41]. Future research should focus on developing a standardised appraisal and reporting tool for cost-effectiveness models of diagnostic tests where a linked-evidence approach is used, incorporating and expanding on our proposed preliminary good practice recommendations.

Our paper has several limitations. The case study serving as an example to derive recommendations came from one clinical decision problem (identification of MI with a blood test) and the models reviewed to see how our research questions have been addressed have not been systematically and critically appraised by us. Potentially more case studies could have identified additional issues, such as dealing with complex test strategies in parallel and serial testing. Despite this, the major issues were identified, given that we worked from a general model structure (see Fig. 1) and a defined sequence of linkage (from test to treatment to outcomes). Second, the HTA guidelines and the case study came from only a few countries (Australia, Canada, Netherlands, and the UK) with established cost-effectiveness hurdles and hence more advanced methods guidance. However, there is no reason to believe that the challenges and recommendations would not be applicable to other countries. Last, our case study did not investigate the appropriateness of the framing of the decision problem and whether the test should be used in this context; rather, this was assumed to be a given. In practice, however, this may pose an additional set of challenges, such as mentioned in the introduction and outlined by the Australian guidelines [13], whether the available evidence (i.e., generated with current tests) is likely applicable to the population selected with the new test. As with any innovation, by design, this evidence may only become available with coverage of the new test and as real-world evidence is generated as a consequence of its use.

In conclusion, there exist several unique challenges for cost-effectiveness modelling of diagnostic tests which need forethought in the design of an economic evaluation. Selected evidence and assumptions need to be justified, particularly for the link of the test results to the treatment. Upfront consideration of how a test and its results will likely be incorporated into patient diagnostic pathways is key to exploring the optimal design of such models. We propose several preliminary good practice recommendations to aid in these tasks.

Acknowledgements We would like to thank Professor Chris Hyde, University of Exeter for reading and providing useful comments back on a draft of this paper. We would also like to thank Janet Bouttell, Tracy Merlin and Alexander Thompson for their insightful feedback when peer-reviewing our paper, which has helped to shape and improve this manuscript.

Declarations

Funding None.

Financial Disclosures BM, AJA and JK are full-time employees of Roche Diagnostics and hold Roche stock options. BS was Associate Director and part-funded by the NIHR Leeds In Vitro Diagnostics MIC. LG has other research supported by Roche Molecular Systems.

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication (from patients/participants) Not applicable.

Availability of Data and Material All extracted data are included in the manuscript.

Code Availability Not applicable.

Authors' Contributions BM came up with the idea for the paper. BM, JK and AJA conducted the review of the HTA bodies guidelines and the review of the models for the case study. BS led the drafting and redrafting of the manuscript and the development of the good practice recommendations, with input throughout from all authors.

References

1. NICE health technology evaluations: the manual (PMG36), 31 January 2022.
2. European Society of Cardiology (ESC): Guidelines for Management of Atrial Fibrillation. August 2020. https://www.escardio.org/Guidelines/Clinical-Practice-Guidelines/Atrial-Fibrillation-Management. Accessed 26 Sep 2022.
3. AHA/ACC/HRS Focused Update of the 2014 AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation. January 2019. https://doi.org/10.1161/CIR.0000000000000665. Accessed 26 Sep 2022.
4. Faria R, Soares MO, Spackman E, et al. Optimising the diagnosis of prostate cancer in the era of multiparametric magnetic resonance imaging: a cost-effectiveness analysis based on the Prostate MR Imaging Study (PROMIS). Eur Urol. 2018;73(1):23–30.
5. Bossuyt PM, Reitsma JB, Linnet K, Moons KG. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. Clin Chem. 2012;58(12):1636–43.
6. Snowsill T. Modelling the cost-effectiveness of diagnostic tests. Pharmacoeconomics. 2023;41:339–51.
7. Shinkins B, Yang Y, Abel L, Fanshawe TR. Evidence synthesis to inform model-based cost-effectiveness evaluations of diagnostic tests: a methodological review of health technology assessments. BMC Med Res Methodol. 2017;17(1):56.
8. Ferrante di Ruffano L, Harris I, Zhelev Z, et al. Health technology assessment of diagnostic tests: a state of the art review of methods guidance from international organizations. Int J Technol Assess Health Care. 2023;39(1):E14.
9. Yang Y, Abel L, Buchanan J, et al. Use of decision modelling in economic evaluations of diagnostic tests: an appraisal and review of health technology assessments in the UK. PharmacoEconomics Open. 2019;3:281–91.
10. Staub LP, Dyer S, Lord SJ, Simes RJ. Linking the evidence: intermediate outcomes in medical test assessments. Int J Technol Assess Health Care. 2012;28(1):52–8.
11. European Society of Cardiology (ESC): Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. 2020. https://www.escardio.org/Guidelines/Clinical-Practice-Guidelines/Acute-Coronary-Syndromes-ACS-in-patients-presenting-without-persistent-ST-segm. Accessed 26 Sep 2022.
12. Hicks KA, Mahaffey KW, Mehran R, et al. Standardized data collection for cardiovascular trials initiative (SCTI). 2017 Cardiovascular and stroke endpoint definitions for clinical trials. Circulation. 2018;137(9):961–72.
13. Guidelines for preparing assessments for the Medical Services Advisory Committee. Australia. Version 1.0, May 2021.
14. CADTH Methods and Guidelines: Guidelines for the Economic Evaluation of Health Technologies: Canada (4th Edition), March 2017.
15. Thokala P, Goodacre SW, Collinson PO, et al. Cost-effectiveness of presentation versus delayed troponin testing for acute myocardial infarction. Heart. 2012;98(20):1498–503.
16. Westwood M, van Asselt T, Ramaekers B, et al. High-sensitivity troponin assays for the early rule-out or diagnosis of acute myocardial infarction in people with acute chest pain: a systematic review and cost-effectiveness analysis. Health Technol Assess. 2015;19(44):1–234.
17. Vaidya A, Severens JL, Bongaerts BW, et al. High-sensitive troponin T assay for the diagnosis of acute myocardial infarction: an economic evaluation. BMC Cardiovasc Disord. 2014;14:77.
18. CADTH. Optimal Use Report. High-sensitivity cardiac troponin for the rapid diagnosis of acute coronary syndrome in

the emergency department: a clinical and cost-effectiveness evaluation. 2013; 2(1A).
19. Collinson PO, Gaze DC, Thokala P, Goodacre S. Randomised assessment of treatment using panel assay of cardiac markers--contemporary biomarker evaluation (RATPAC CBE). Health Technol Assess. 2013;17(15):v-vi.
20. Goodacre S, Thokala P, Carroll C, et al. Systematic review, meta-analysis and economic modelling of diagnostic strategies for suspected acute coronary syndrome. Health Technol Assess. 2013;17(1):v–vi (1–188).
21. Pope JH, Aufderheide TP, Ruthazer R, et al. Missed diagnoses of acute cardiac ischemia in the emergency department. N Engl J Med. 2000;342(16):1163–70.
22. Smolina K, Wright FL, Rayner M, Goldacre MJ. Long-term survival and recurrence after acute myocardial infarction in England, 2004 to 2010. Circ Cardiovasc Qual Outcomes. 2012;5:532–40.
23. Westwood M, Ramaekers B, Grimm S, et al. High-sensitivity troponin assays for early rule-out of acute myocardial infarction in people with acute chest pain: a systematic review and economic evaluation. Health Technol Assess. 2021;25(33):1–276.
24. Rathore SS, Curtis JP, Chen J, et al. Association of door-to-balloon time and mortality in patients admitted to hospital with ST elevation myocardial infarction: national cohort study. BMJ. 2009;338: b1807.
25. Anderson HV, Shaw RE, Brindis RG, et al. Risk-adjusted mortality analysis of percutaneous coronary interventions by American College of Cardiology/American Heart Association guidelines recommendations. Am J Cardiol. 2007;99(2):189–96.
26. Mullins DC, Onwudiwe NC, Branco de Araújo GT, et al. guidance document: global pharmacoeconomic model adaption strategies. Value Health Reg Issues. 2014;5:7–13.
27. Fernandes LE, Epstein CG, Bobe AM, et al. Real-world evidence of diagnostic testing and treatment patterns in US patients with breast cancer with implications for treatment biomarkers from RNA sequencing data. Clin Breast Cancer. 2021;21(4):e340–61.
28. Hellyer TP, McAuley DF, Walsh TS, et al. Biomarker-guided antibiotic stewardship in suspected ventilator-associated pneumonia (VAPrapid2): a randomised controlled trial and process evaluation. Lancet Respir Med. 2020;8(2):182–91.
29. Goodacre S, Nicholl J, Dixon S, et al. Randomised controlled trial and economic evaluation of a chest pain observation unit compared with routine care. BMJ. 2004;328(7434):254.
30. Ward S, Lloyd Jones M, Pandor A, et al. A systematic review and economic evaluation of statins for the prevention of coronary events. Health Technol Assess. 2007;11(14):1–160.
31. Sullivan PW, Slejko JF, Sculpher MJ, Ghushchyan V. Catalogue of EQ-5D scores for the United Kingdom. Med Decis Making. 2011;31(6):800–4.
32. Kind P, Hardman G, Macran S. UK population norms for EQ-5D. York (UK): Centre for Health Economics, University of York; 1999. (Discussion paper 172). Available from: http://www.york.ac.uk/che/publications/in-house/archive/1990s/. Accessed 14 Jan 2023.
33. Schweikert B, Hunger M, Meisinger C, König HH, Gapp O, Holle R. Quality of life several years after myocardial infarction: comparing the MONICA/KORA registry to the general population. Eur Heart J. 2009;30(4):436–43.
34. Mills NL, Churchhouse AM, Lee KK, et al. Implementation of a sensitive troponin I assay and risk of recurrent myocardial infarction and death in patients with suspected acute coronary syndrome. JAMA. 2011;305(12):1210–6.
35. Polanczyk CA, Kuntz KM, Sacks DB, Johnson PA, Lee TH. Emergency department triage strategies for acute chest pain using creatine kinase-MB and troponin I assays: a cost-effectiveness analysis. Ann Intern Med. 1999;131(12):909–18.
36. Selmer R, Halvorsen S, Myhre KI, Wisløff TF, Kristiansen IS. Cost-effectiveness of primary percutaneous coronary intervention versus thrombolytic therapy for acute myocardial infarction. Scand Cardiovasc J. 2005;39(5):276–85.
37. Abbott JD, Ahmed HN, Vlachos HA, Selzer F, Williams DO. Comparison of outcome in patients with ST-elevation versus non-ST-elevation acute myocardial infarction treated with percutaneous coronary intervention (from the National Heart, Lung, and Blood Institute Dynamic Registry). Am J Cardiol. 2007;100(2):190–5.
38. Allen LA, O'Donnell CJ, Camargo CA Jr, Giugliano RP, Lloyd-Jones DM. Comparison of long-term mortality across the spectrum of acute coronary syndromes. Am Heart J. 2006;151(5):1065–71.
39. Nikus KC, Eskola MJ, Virtanen VK, et al. Mortality of patients with acute coronary syndromes still remains high: a follow-up study of 1188 consecutive patients admitted to a university hospital. Ann Med. 2007;39(1):63–71.
40. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open. 2016;6(11):e012799.
41. Kip MMA, IJzerman MJ, Henriksson M, et al. Toward alignment in the reporting of economic evaluations of diagnostic tests and biomarkers: the AGREEDT checklist. Med Decis Making. 2018;38(7):778–88.
42. Merlin T, Lehman S, Hiller J, Ryan P. The, "linked evidence approach" to assess medical tests: a critical analysis. Int J Technol Assess Health Care. 2013;29(3):343–50.