



Toward Alignment in the Reporting of Economic Evaluations of Diagnostic Tests and Biomarkers: The AGREEDT Checklist

Michelle M.A. Kip, Maarten J. IJzerman, Martin Henriksson, Tracy Merlin, Milton C. Weinstein, Charles E. Phelps, Ron Kusters, and Hendrik Koffijberg

Objectives. General frameworks for conducting and reporting health economic evaluations are available but not specific enough to cover the intricacies of the evaluation of diagnostic tests and biomarkers. Such evaluations are typically complex and model-based because tests primarily affect health outcomes indirectly and real-world data on health outcomes are often lacking. Moreover, not all aspects relevant to the evaluation of a diagnostic test may be known and explicitly considered for inclusion in the evaluation, leading to a loss of transparency and replicability. To address this challenge, this study aims to develop a comprehensive reporting checklist. **Methods.** This study consisted of 3 main steps: 1) the development of an initial checklist based on a scoping review, 2) review and critical appraisal of the initial checklist by 4 independent experts, and 3) development of a final checklist. Each item from the checklist is illustrated using an example from previous research. **Results.** The scoping review followed by critical review by the 4 experts resulted in a checklist containing 44 items, which ideally should be considered for inclusion in a model-based health economic evaluation. The extent to which these items were included or discussed in the studies identified in the scoping review varied substantially, with 14 items not being mentioned in ≥ 47 (75%) of the included studies. **Conclusions.** The reporting checklist developed in this study may contribute to improved transparency and completeness of model-based health economic evaluations of diagnostic tests and biomarkers. Use of this checklist is therefore encouraged to enhance the interpretation, comparability, and—indirectly—the validity of the results of such evaluations.

Keywords

approval, biomarkers, checklist, diagnostic test, health economic evaluation, reporting

Date received: March 7, 2018; accepted: August 5, 2018

Detailed evaluation of the clinical utility and also health economic impact of new diagnostic tests prior to their implementation in clinical practice is important to limit overuse of tests, ensure benefits to patients, and support efficient use of health care resources.¹ Different frameworks have been developed for the phased evaluation of diagnostic tests.^{2–6} All these frameworks recognize that after evaluating the safety, efficacy, and accuracy of a diagnostic test, the impact of this test on health outcomes and costs should be determined. Evaluating tests in randomized controlled trials (RCTs), however, is often not feasible for ethical, financial, or other reasons,

particularly in early test development stages.^{7–10} Indeed, RCTs evaluating the impact of diagnostic tests on patient outcomes are rare.¹¹ As an alternative, methods to develop decision-analytic models for the health economic evaluation of diagnostic tests, synthesizing all available evidence from different sources, have long been

Corresponding Author

Hendrik Koffijberg, Department of Health Technology and Services Research, Faculty of Behavioural, Management and Social Sciences, Technical Medical Centre, University of Twente, P.O. Box 217, 7500 AE Enschede, the Netherlands (h.koffijberg@utwente.nl).

available.^{6,12–16} It is widely recognized that such models are a useful and valid alternative to evaluate the impact of new health technologies in general^{17,18} and diagnostic methods in particular.^{12,14}

However, the comprehensive evaluation of the impact of new tests is typically much more complex than, for example, evaluation of the impact of new drugs. Among others, this is due to the indirect impact of tests on health outcomes by improved patient management (also referred to as “clinical utility”¹⁹), the use of combinations and sequences of tests in clinical practice (depending on previous test results), and the often complex interpretation of test outcomes. In practice, model-based impact evaluations of tests therefore actually involve the evaluation of diagnostic testing *strategies* (i.e., test-treatment combinations).

Owing to the complexity of these diagnostic testing strategies, many model-based impact evaluations of tests make use of simplified models that do not incorporate all aspects of clinical practice. Simplified models are used because 1) evidence regarding all aspects involved in health economic test evaluations might be lacking, 2) inclusion of all aspects likely increases model complexity, or 3) researchers may not be aware of all aspects of test evaluation. For example, it is often not reported how the incremental effect of a new test, when used in combination with other tests, is determined and how the correlation between the outcomes of these different tests (applied solo or in sequence) is handled.^{20–23} Similarly, the selection of patients in whom the test is performed, the consequences of incidental findings (also referred to as chance findings), and the occurrence of test failures or indeterminate test results are often not reported.^{24–26}

Although simplifications of the decision-analytic models used for such evaluations may sometimes be necessary and can be adequately justified, implicit simplification due to unawareness of all relevant evaluation aspects or without proper justification may lead to nontransparent and incorrect evaluation results.

General frameworks and guidelines regarding which aspects to include in decision-analytic modeling and how to report modeling outcomes are available^{27–29} but not specific enough to cover the complexities of diagnostic test evaluation. Furthermore, previous research into (aspects of) diagnostic test evaluation mostly focused on specific diseases or on specific types or combinations of diagnostic tests.^{23,30–35} A generic and comprehensive overview of all potentially relevant aspects in health economic evaluation of diagnostic tests and biomarkers that may be used to guide such evaluations is currently lacking.

The purpose of this article is, therefore, to provide such an overview as a generic checklist, intended to be applicable to all types of diagnostic tests and not specific to a single disease or condition or subgroup of individuals. Thereby, this checklist aims to allow researchers to explicitly consider all aspects potentially relevant to the health economic evaluation of a specific test, from a societal perspective. Therefore, this checklist is referred to as the “AGREEDT” checklist, which is an acronym of “AliGnment in the Reporting of Economic Evaluations of Diagnostic Tests and biomarkers.” Use of the checklist does not need to complicate such evaluations, as some aspects described may not be relevant to particular evaluations, but rather suggests that choices to exclude certain aspects are adequately justified.

Department of Health Technology and Services Research, Faculty of Behavioural, Management and Social Sciences, Technical Medical Centre, University of Twente, Enschede, the Netherlands (MMAK, MJJJ, RK, HK); Department of Medical and Health Sciences, Linköping University, Linköping, Sweden (MH); Adelaide Health Technology Assessment (AHTA), School of Public Health, University of Adelaide, Adelaide, South Australia, Australia (TM); Department of Health Policy and Management Harvard T. H. Chan School of Public Health, Boston, MA (MCW); Departments of Economics, Political Science, and Public Health Sciences, University of Rochester, Rochester, NY (CEP); and Laboratory for Clinical Chemistry and Haematology, Jeroen Bosch Ziekenhuis, Den Bosch, the Netherlands (RK). No funding has been received for the conduct of this study and/or the preparation of this manuscript. Prof. Merlin reports that she was previously commissioned by the Australian Government to develop version 5.0 of the “Guidelines for Preparing a Submission to the Pharmaceutical Benefits Advisory Committee.” Some of the content concerning “Product Type 4—Codependent Technologies” influenced the guidance suggested in the current article. Prof. Weinstein reports that he was a consultant to OptumInsight on unrelated topics. All other authors declare that there is no conflict of interest.

Methods

This study consisted of 3 main steps: 1) the development of an initial checklist based on a scoping review, 2) review and critical appraisal of the initial checklist by 4 experts (CEP, MCW, MH, and TM) not involved in the scoping review, and 3) development of a final checklist based on the review by experts. Finally, each item from the checklist is illustrated using an example from previous research.

Scoping Review

In the past decades, hundreds of model-based health economic evaluations of diagnostic tests have been published, across a wide range of medical contexts. A still narrow literature search in PubMed in January 2017 resulted in a total of 1844 articles using the following

combinations of search terms in title and abstract: (*health economic OR cost-effectiveness*) AND *diagn** AND (*model OR Markov OR tree OR modeling OR modelling*). Besides the large number of studies that have been published in this field, systematic identification of health economic evaluations is found to be challenging.³⁶ This is partly caused by the multitude of MeSH terms in PubMed related to diagnostic strategies (over 48 MeSH terms exist that include the word *diagnostic* or *diagnosis*). Because of these challenges and the fact that different evaluations are very likely to include and exclude the same aspects, a scoping review was performed instead of a systematic literature review, followed by critical appraisal by 4 independent experts. A key strength of a scoping review is that it can provide a rigorous and transparent method for mapping areas of research,³⁷ particularly when an area is complex or has not been reviewed comprehensively before.³⁸

This scoping review was performed in PubMed in January 2017, searching for the following combination of search terms in the title of the article: (*health economic OR cost-effectiveness*) AND *diagn** AND (*model OR Markov OR tree OR modeling OR modelling*) NOT *diagnosed*. The term *NOT diagnosed* was added to prevent retrieving many articles including patients who are already diagnosed with a certain condition, instead of focusing on the diagnostic process itself. The search was limited to articles published in English or Dutch. Studies were excluded, based on title and abstract, if they did not concern original research or did not evaluate the cost-effectiveness of the use of 1 or more tests (regardless of the effectiveness measure, for example, additional cost per additional correct diagnosis or per additional quality-adjusted life year). In addition, as guidelines for performing health economic evaluations continue to be updated,³⁹⁻⁴¹ it was expected that the more recent studies would provide the most comprehensive overview of all potentially relevant items that need to be included in the checklist. To check this assumption, the PubMed search was repeated without limiting the search to studies published ≤ 5 years ago, resulting in 128 additional articles. Following this, 2 articles that were published > 5 years ago were randomly selected.^{42,43} A thorough review of both articles did not result in any additional relevant items for inclusion in the checklist. Therefore, the search was limited to articles published in the past 5 years. One author screened studies for exclusion (MMAK) and consulted with a second author (HK) if necessary.

Design of the reporting checklist

All articles resulting from the scoping review were searched for items related to model-based health

economic test evaluation of diagnostic tests that were either included explicitly in the evaluation, or that were only mentioned but not included (mostly in the introduction or discussion sections). Generic items, not specific to diagnostic test evaluation were not included in the new checklist as these are already covered in existing checklists. Examples of such generic items include choosing the time horizon and perspective of the evaluation.²⁷⁻²⁹ However, some overlap remains as the checklist does include items which are considered applicable to diagnostic test evaluation that are only covered partially or at a high level in existing guidelines.

A thorough screening of all articles was performed by MMAK resulting in an initial list of aspects considered to be potentially relevant. As the checklist was intended to provide a comprehensive overview of all potentially relevant aspects, all of these aspects were added to the checklist, unless it was considered to be already included in currently available guidelines (based upon agreement between MMAK and HK). The definition of each aspect was based on agreement between MMAK and HK.

Critical Appraisal and Validation of the Reporting Checklist

As diagnostic tests and imaging are used for a large variety of (suspected) medical conditions, an expert panel with a broad field of experience was required for critical appraisal of the checklist. Therefore, the expert panel was composed in such a way that at least 1 expert was experienced in each of the different areas of interest (i.e., biomarkers or imaging) and in each of the different purposes of diagnostic testing (i.e., diagnosis, screening, monitoring, and prognosis). In addition, to maximize the likelihood that the final checklist is generalizable to different countries and settings, the experts chosen lived on 3 different continents. Four experts were invited (CEP, MCW, MH, and TM) to participate via email, and none of them declined.

The initial checklist was critically appraised and validated independently by all 4 experts, who received the checklist via email. They were asked to provide individual, qualitative judgments on whether all items in this list were clear and unambiguous, to indicate any missing or redundant items in this list, and to provide suggestions for further improvement.

Finalization of the Reporting Checklist

Based on the experts' suggestions, several changes were made to the reporting checklist. Those changes involved

the rewording of items, removal of redundant items, and the addition of missing items to the checklist. As this checklist is intended to provide an exhaustive list of all aspects relevant to the health economic evaluation of diagnostic tests and biomarkers, all suggestions for the addition of missing items were adopted. All changes made to the checklist were decided upon agreement between MMAK and HK (for a full description, see online Appendix 1). The revised checklist was again critically appraised by all authors and agreed upon. Finally, the articles included in the scoping review were reread by MMAK to assess whether the final checklist items were included or mentioned.

Funding

This study was not funded.

Results

Results of the Scoping Review

The literature search resulted in 77 articles that were screened for inclusion in the scoping review, of which 14 articles were excluded. Of these, 4 articles did not specifically evaluate the cost-effectiveness of a (combination of) diagnostic test(s), 2 concerned a letter to the editor, and 7 articles focused on methodological aspects of the evaluation of diagnostic strategies (e.g., in the context of single disease, or on specific types or combinations of diagnostic tests, as mentioned earlier). In addition, 1 article was excluded because the full text could not be obtained or purchased by the university library, from online databases, from the website of the publisher, or by contacting the authors. This resulted in a total of 63 studies that were included in the scoping review. An overview of this selection process is provided in Figure 1.

A critical evaluation of the 63 articles resulted in an initial list of 29 items. These items were divided into 6 main topics: 1) time to presentation of the individual to the health professional (i.e., the clinical starting point), 2) use of diagnostic tests, 3) test performance and characteristics, 4) patient management decisions, 5) impact on health outcomes and costs, and 6) wider societal impact, which may accrue to patients, their families, and/or health care professionals. This societal impact, for example, may concern the impact on caregivers (in terms of time spent on hospital visits and caregiving and the accompanying impact on productivity), on the health system or health professional (e.g., in terms of reduced patient visits), or on society (e.g., measures that aim to prevent widespread antibiotic resistance). Quantifying

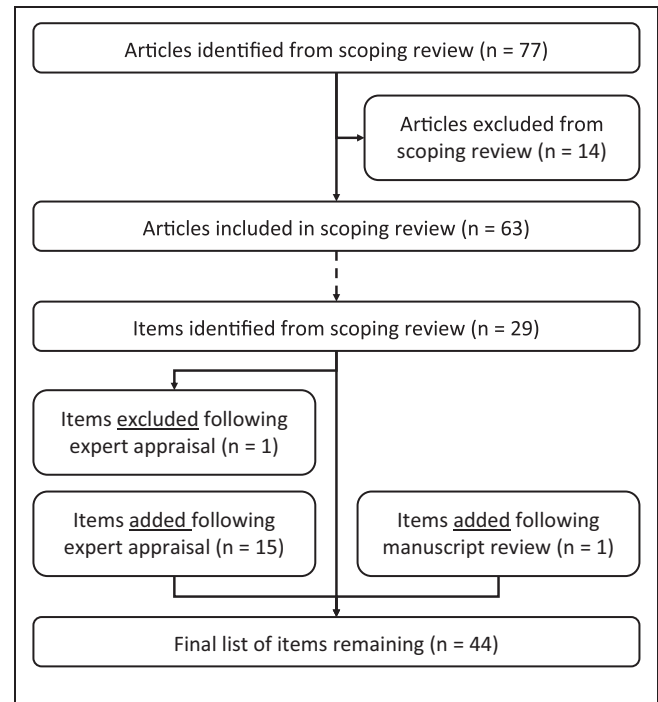


Figure 1 Result of scoping review and checklist design process. This figure first gives an overview of the selection process of articles in the scoping review, as well as the number of checklist items this resulted in, and subsequently shows the results of the expert appraisal on the items included in the final checklist.

these aspects may provide a broader view on the potential impact of diagnostic testing.

Critical Appraisal and Validation of the Reporting Checklist

Following the critical appraisal of experts, the list was updated, with 1 item being removed and 15 items being added; 1 additional item was added based on the suggestion of a reviewer of the manuscript during the submission process. Finally, this resulted in a reporting checklist consisting of 44 items, as shown in Table 1. Of these 16 added items, 8 involved a further specification of the tests' diagnostic performance, as included below item 3.2 in the checklist. The item that was removed concerned the generalizability of the results, which was considered not specific to diagnostic test evaluations. The full reporting checklist, including an overview in which of the studies from the scoping review each of the items was included or considered, as well as an example for each of the items, is provided in online Appendix 2. An overview of this process, including the scoping review and the

Table 1 Reporting Checklist to Indicate Which Items Were Included in the Health Economic Evaluation of Diagnostic Tests and Biomarkers

Items of the Evaluation of Diagnostic Tests and Biomarkers	Included in Evaluation	
	Yes ^a	No ^b
Time to presentation		
<i>Onset of complaints or onset of suspicion by physician and start of diagnostic trajectory</i>		
1.1	_____	_____
The study should contain a description of the individuals who enter the diagnostic pathway (i.e., a patient develops a [new] condition or disease, which may or may not result in symptoms or complaints, or undergoes diagnostic testing as part of screening or genetic testing).		
1.2	_____	_____
Consider the time to start of the diagnostic trajectory or the time until a monitoring or screening test is (repeatedly) performed (initiated by symptoms/complaints or initiated as part of regular screening or monitoring).		
<i>(The time between 1.1 and 1.2 is the time during which individuals are at risk of complications from disease and progression, in the absence of a diagnosis and thus also in the absence of treatment.)</i>		
Use of diagnostic tests		
<i>Decision regarding which diagnostic test(s) is/are performed, in which patients, and in what order</i>		
2	_____	_____
Specify for which purpose(s) the test is used (e.g., screening, diagnosing, monitoring, guide dosage, commencement or cessation of therapy, triaging, staging, prognostic) and define the entire diagnostic and clinical pathway.		
2.1	_____	_____
Consider whether more than 2 (possible) diagnostic strategies can be compared, each involving a single test or combination of tests.		
2.2	_____	_____
Consider whether the evaluated diagnostic strategies include multiple tests, which can be performed in parallel or in sequence.		
2.2.1	_____	_____
Consider whether some tests of the diagnostic workup are performed conditional on previous test outcomes, leading to a selection of patients undergoing specific tests.		
2.3	_____	_____
Consider whether subgroups can be defined based on explicit criteria or patient characteristics, in which different tests would be performed (not solely dependent on previous test outcomes).		
2.4	_____	_____
Consider whether different tests are applied based on implicit (shared) decision making (e.g., perceived condition or risk, or symptom presentation).		
Test performance and characteristics		
<i>Diagnostic test performance and items related to the sampling and testing</i>		
3.1	_____	_____
Specify the costs of the diagnostic test.		
3.2	_____	_____
Specify test performance, in terms of sensitivity, specificity, negative predictive value (and its complement), and/or positive predictive value (and its complement), either or not combined with a decision rule or algorithm.		
3.2.1	_____	_____
Describe the evidence base for the estimated test performance.		
3.2.2	_____	_____
Describe the positivity criterion (i.e., cutoff value) applied to the test or testing strategy.		
3.2.3	_____	_____
Consider whether the estimated test performance may be biased, for example, due to lack of evidence on conditional dependence or independence, lack of a (perfect) gold standard (i.e., classification bias), verification bias, analytic bias, spectrum bias, diagnostic review bias, and incorporation bias.		
3.2.3.1	_____	_____
Consider how likely/to what extent bias in the available/applied evidence affects the estimated test performance.		
3.2.4	_____	_____
Describe how uncertainty/variation in the test performance (receiver operating characteristic [ROC] curve) was handled or explained, for example, due to interrater and intrarater reliability, or experience of the clinician.		
3.2.5	_____	_____
Describe the logic, or analysis, applied to choose the cutoff value (i.e., the point on the ROC curve) for the test, for example, depending on whether the test is used as a single test or part of a sequence of tests.		
3.2.6	_____	_____
Describe whether different test performances and cutoff values were considered for different subgroups of patients and/or environmental characteristics. For example: based on specific subgroup(s) of patients, timing of the test in the diagnostic trajectory, or selection of patients based on previous test outcomes (if any).		
3.2.7	_____	_____
Consider whether test performance is dependent on disease prevalence (which also includes the impact of spectrum bias on disease prevalence and, as a consequence, on test performance) or affected by other patient characteristics or conditions.		
3.2.8	_____	_____
Consider whether test performance is based on a combination of tests (and on a combination of areas under the ROC curves for each test).		

(continued)

Table 1 (continued)

Items of the Evaluation of Diagnostic Tests and Biomarkers		Included in Evaluation	
		Yes ^a	No ^b
3.3	Consider the feasibility of obtaining (sufficient) sample and/or usability of the sample that is obtained.	_____	_____
3.4	Consider the occurrence of test failures or indeterminate/not assessable results.	_____	_____
3.5	Consider costs of retesting (after obtaining insufficient/unusable sample or after test failure or indeterminate/not assessable result).	_____	_____
3.6	Consider complications, risks, or other negative/positive aspects directly related to obtaining the sample and/or performing the diagnostic test (either in the intervention or in the control strategy).	_____	_____
3.7	Consider the time taken to perform the test (including waiting time) until the test result is available or until a management decision or treatment is initiated based on this test result (either in the intervention or in the control strategy).	_____	_____
3.8	Consider the impact of additional knowledge gained by performing the diagnostic test (i.e., for a genetic test) or the occurrence and impact of incidental findings (i.e., the unintentional discovery of a previously undiagnosed condition during the evaluation of another condition).	_____	_____
3.8.1	<i>The impact of incidental findings on performing additional tests is addressed.</i>	_____	_____
Patient management decisions		_____	_____
Impact of a test on the diagnosis and/or patient management strategy (based on this diagnosis)		_____	_____
4.1	Clearly specify the impact of the test in selecting the patient management strategy.	_____	_____
4.2	Consider whether other aspects besides test results themselves are part of the decision algorithm (and included in the evaluation). These may involve a shared decision-making process of the physician with patients/relatives or aspects including coverage or physician adherence to treatment guidelines.	_____	_____
4.3	Consider whether the impact of the test result on resulting/selected diagnosis or management strategy varies across subgroups (this difference should not only be caused by differences in diagnostic performance of the test and does not need to include the impact on costs and/or health outcomes within this subgroup). ^c	_____	_____
4.4	Consider the consistency of test results over time (e.g., genetic mutations may be affected by treatment prescribed after the initial diagnosis).	_____	_____
4.5	Consider the impact of performing the test and providing and interpreting the result on the time spend/capacity of the health care professional(s) or the patient.	_____	_____
Impact on health outcomes and costs		_____	_____
Impact of the patient management strategy on diseased and nondiseased individuals, in terms of health outcomes and costs		_____	_____
5	Evaluate the direct impact of the chosen patient management strategy on the number of (in)correctly diagnosed individuals, health outcomes, and/or costs.	_____	_____
5.1	<i>Consider the direct impact of the chosen patient management strategy on health outcomes and/or costs. This concerns the entire period in which patient management may affect a patient's health and/or costs and does not only involve the testing strategy itself.</i>	_____	_____
5.2	<i>Consider whether the direct impact of the chosen patient management strategy on health outcomes and/or costs varies across subgroups. (This does not include only varying the incidence of a certain condition in a sensitivity analysis. The subgroups should be clearly defined and preferably be identifiable based on patient characteristics).^c</i>	_____	_____
5.3	<i>Consider patient's adherence to treatment (which includes aspects that may indicate (partial) nonadherence (e.g., following only some of the treatment recommendations, as well as aspects that affect the degree of administration of treatment).</i>	_____	_____
5.4	<i>Consider the occurrence (and consequences) of treatment-related adverse events.</i>	_____	_____
5.5	<i>Describe the probability or time it takes to observe that the patient management strategy proves to be effective over time or that the patient cures spontaneously (regardless of whether the patient received a correct or an incorrect diagnosis).</i>	_____	_____
5.6	<i>Describe the probability of or time it takes to repeat or extend the diagnostic workup when the patient management strategy proves to be ineffective, either directly or over time (regardless of whether the patient received a correct or an incorrect diagnosis). This also includes the situation in which the patient receives no treatment or unnecessary treatment.</i>	_____	_____
5.7	<i>Describe the impact of ineffective or unnecessary treatment or management on health outcomes and/or costs (including both side effects and costs and regardless of whether the patient received a correct or an incorrect diagnosis). This also includes the situation in which incorrectly no treatment is provided or in which the treatment is delayed.</i>	_____	_____

(continued)

Table 1 (continued)

Items of the Evaluation of Diagnostic Tests and Biomarkers		Included in Evaluation	
		Yes ^a	No ^b
5.7.1	Consider the impact of delay in treatment initiation on health outcomes and/or costs.	_____	_____
Societal impact		_____	_____
<i>Wider (societal) impact of the chosen diagnostics and management strategy</i>		_____	_____
6.1	Consider the psychological impact of the diagnostic outcome and management strategy on patients, including the value of knowing (in terms of reassurance or anxiety), patient preferences regarding undergoing diagnostic tests, the (accompanying) impact on caregivers or relatives, and so on.	_____	_____
6.1.1	<i>Consider the impact of test outcomes on relatives themselves, regarding the value of knowing (spillover knowledge) and regarding subsequent testing and/or treatment in this group (in case of heritable genetic conditions or contagious diseases).</i>	_____	_____
6.2	Consider the additional impact of diagnostic outcome and management strategy on the health system or health care professionals.	_____	_____
6.3	Consider the additional impact of diagnostic outcome and management strategy for society.	_____	_____

^aIf an item is included in the quantitative analysis, indicate the corresponding model parameter(s) and evidence source(s).

^bIf an item is excluded from the quantitative analysis, please explain why the exclusion was necessary.

^cExisting guidelines indicate that subgroup analyses are relevant when different strategies are likely to be (sub)optimal in different subgroups. Subgroup specific analyses can then be performed to address multiple decision problems. Here we consider scenarios where different tests may be used in different subgroups, depending on patient characteristics or previous test outcomes.

critical appraisal by the experts, is shown in Figure 1. The final list of items in this reporting checklist, in chronological order from the start of the diagnostic trajectory and onward, is illustrated in Figure 2.

Results indicate that health economic evaluations of diagnostic tests or biomarkers differ considerably in the items that have been explicitly included (or considered for inclusion) in the corresponding decision-analytic model (Table 1). Some of the items from the checklist were only included (or considered) in a few studies from the scoping review. For example, the impact of incidental findings on performing additional tests, the consistency of test results over time, and the impact of test outcomes on relatives themselves were each only addressed in 3 of the 63 included studies. These items may not have been included in other studies because they were considered not relevant to the specific context, because (scientific) evidence was lacking, or because these items were not considered due to unawareness of their relevance by the authors.

Discussion

Strengths

A strength of this study is that it combines evidence from multiple sources, including a review of literature, as well

as a validation by experts. As the items included in the checklist are defined in general terms and not limited to specific diseases, tests, care providers, or patient management strategies, this reporting checklist can potentially be useful in performing and appraising health economic evaluations worldwide and across a broad spectrum of (novel) diagnostic technologies. In addition, as this checklist specifically focuses on health economic evaluations of diagnostic tests or biomarkers, an area for which no reporting checklists are yet available, it may be a useful extension to existing reporting checklists, such as the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist.²⁷ Finally, use of this checklist can also support *development* of health economic models through increased awareness of all potentially relevant evaluation aspects.

In addition, use of this checklist does not necessarily require more resources to be allocated to the evaluation or increase the complexity of the resulting decision-analytic model. In general, deliberation on the relevance of all aspects is key, and aspects may be excluded from the evaluation whenever this can be adequately justified. For example, when evaluating the cost-effectiveness of a new point-of-care troponin test used by the general practitioner compared to an existing, older point-of-care troponin test (in the context where the new test would replace the old test), aspects such as

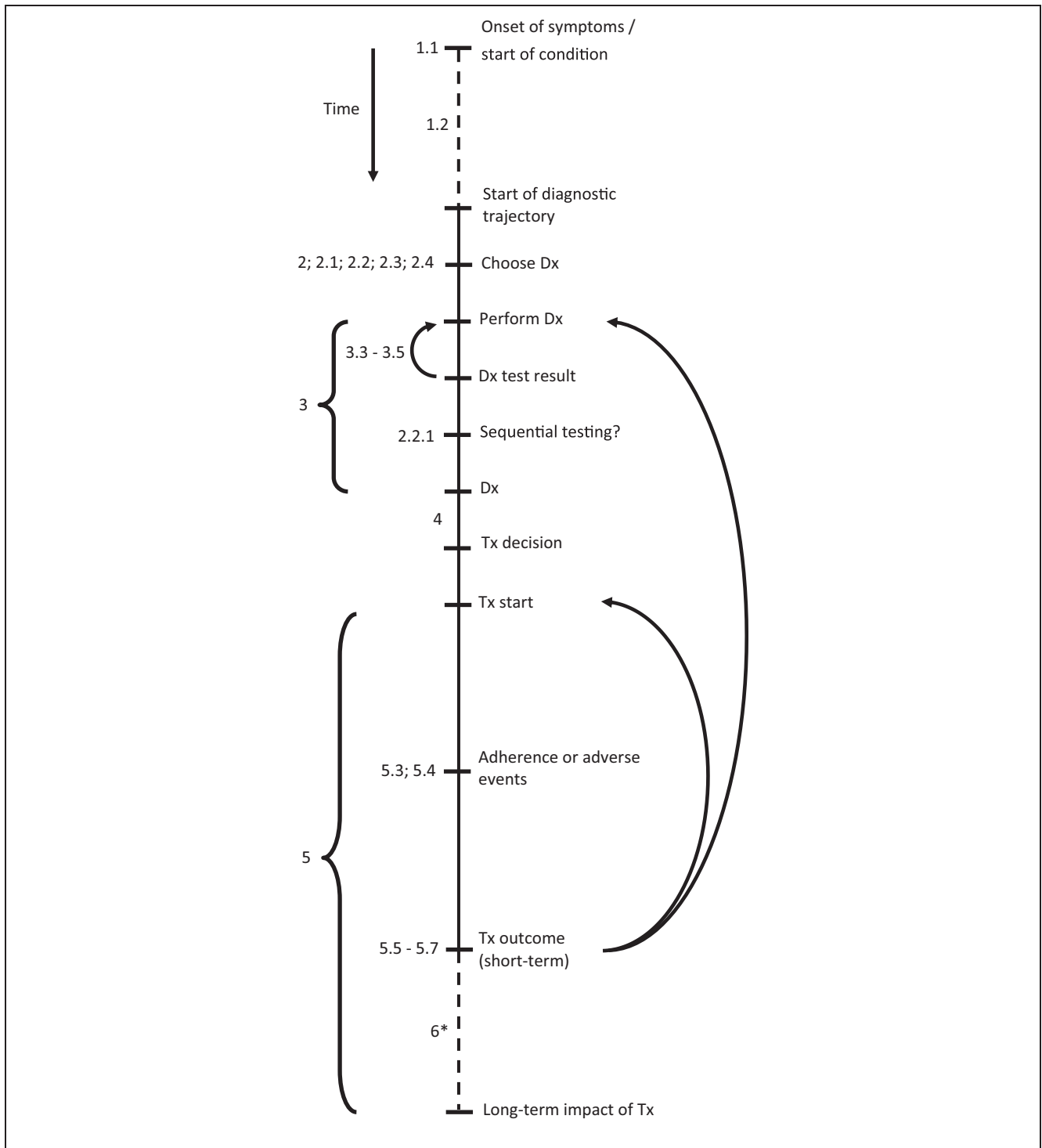


Figure 2 Overview of steps in diagnostic trajectory. This figure gives a conceptual outline of the steps involved in the diagnostic trajectory, in chronological order from top to bottom. The numbers shown at the several steps correspond to the item numbers presented in Table 1. The dashed lines represent steps of which the duration may vary substantially, for example, the time between symptom onset and presentation to a clinician (which may vary from minutes in case of severe symptoms to years for mild and gradually developing conditions). The arrows indicate situations in which either the diagnostic test (result) was not usable or indeterminate (items 3.3–3.5) or situations in which the treatment proves to be ineffective (items 5.5–5.7). As this may be caused by an incorrect diagnosis, the patient may undergo a subsequent round of diagnostic testing and (possibly) treatment. Alternatively, the diagnosis may be correct but the treatment incorrect, in which an alternative treatment may be initiated. *Although the (wider) societal impact of diagnostic testing often involves long-term effects, these effects may sometimes also become apparent in the short term. Dx, diagnostic test; Tx, treatment.

“time to start of the diagnostic trajectory” and “purpose of the test” will not differ between both strategies. In addition, “complication risks” associated with taking the blood sample (in both point-of-care tests) are likely extremely small, which could justify excluding these aspects from the analysis.

Limitations

Performing a systematic literature review was considered not possible given the large number of published economic evaluations of diagnostic tests. Therefore, a scoping review was performed instead by 1 reviewer. As the judgment regarding whether an aspect was incorporated in a health economic evaluation was sometimes found to be difficult, it cannot be excluded that these judgments may have differed slightly when performed by a different reviewer. In addition, as the decision to limit the search strategy to the past 5 years was based on reviewing 2 studies published >5 years ago, this small sample (i.e., 1.6% of studies published >5 years ago) cannot rule out the possibility that items have been missed by excluding all older studies. Also, the scoping review may have been subject to publication bias, as it may have omitted potentially relevant aspects from unpublished studies, as well as from method manuals (including those focusing on economic evaluations of other interventions or technologies in health care). Despite the abovementioned limitations, the critical review of the checklist by 4 independent experts from different countries makes it unlikely that important items have been missed.

In addition, the expert appraisal resulted in the addition of 16 items to the checklist. Although this may seem to be a large extension to the items already identified in the scoping review, 8 of these added items actually involved a further specification (i.e., a subitem) of the test’s diagnostic performance. It was found useful to further specify “test performance” (i.e., item 3.2, which initially integrated several performance measures) into 8 subitems to further increase the transparency and comparability of health economic test evaluations.

Implications for Practice

This study was intended to design a reporting checklist without formulating a quality judgment of the studies included in the scoping review, based on which items of the checklist they did or did not incorporate. Furthermore, some items may have been included *implicitly* in the health economic evaluations identified in the

scoping review, which could thus not be identified by the reviewer. As scientific articles are often restricted in their length, there may often be insufficient space to mention the inclusion (or justified exclusion) of each of the items from this checklist. In these situations, authors are recommended to describe their use of this checklist in an appendix. More specifically, authors are recommended to describe which items from the checklist they included in their evaluation and what evidence was used to inform them. Furthermore, they are recommended to explicitly state the reason(s) for excluding checklist items from their evaluation. Although it may be considered time-consuming to consider all 44 items of this checklist, it should be noted that most of these items are actually subitems, which do not need to be considered if the overarching (higher-level) item is (justifiably) excluded from the evaluation.

In addition, it should be noted that not all items in this checklist can be considered of equal importance. For example, diagnostic performance will typically have a larger impact on health outcomes and costs compared to considering the occurrence of test failures or the consistency of test results over time. However, this checklist is designed to provide an exhaustive overview of all potentially relevant items, regardless of importance. Therefore, use of this checklist will likely increase the chance that all relevant aspects will be included in health economic evaluations of diagnostic tests and biomarkers. Ultimately, it is up to the researchers to make a justifiable decision on which items to incorporate and which to exclude.

Finally, experiences regarding the use of this reporting checklist in practice may be valuable to further enhance its completeness and usability. Furthermore, given the rapid methodological developments in the field of health economic evaluation of diagnostic tests, regular updating of this checklist may be warranted.

Conclusion

Given the complexity and dependencies related to the use of diagnostic tests or biomarkers, researchers may not always be fully aware of all the different aspects potentially influencing the result of a model-based health economic evaluation. The use of the reporting checklist developed in this study may remedy this by increasing awareness of all potentially relevant aspects involved in such model-based health economic evaluations of diagnostic tests and biomarkers and thereby also increase the transparency, comparability, and—indirectly—the validity of the results of such evaluations.

Authors' Note

This research was conducted at the department of Health Technology and Services Research, University of Twente, Enschede, the Netherlands.

References

- Price CP. Evidence-based laboratory medicine: is it working in practice? *Clin Biochem Rev.* 2012;33(1):13–9.
- Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ.* 2012;344:e686.
- Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. *CMAJ.* 1986;134(6):587–94.
- Silverstein MD, Boland BJ. Conceptual framework for evaluating laboratory tests: case-finding in ambulatory patients. *Clin Chem.* 1994;40(8):1621–7.
- Anonychuk A, Beasall G, Shorter S, Kloss-Wolf R, Neumann P. A framework for assessing the value of laboratory diagnostics. *Healthcare Management Forum.* 2012;25(3):S4–S11.
- Phelps CE, Mushlin AI. Focusing technology assessment using medical decision theory. *Med Decis Making.* 1988;8(4):279–89.
- Schunemann HJ, Oxman AD, Brozek J, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ.* 2008;336(7653):1106–10.
- Moons KG. Criteria for scientific evaluation of novel markers: a perspective. *Clin Chem.* 2010;56(4):537–41.
- Bossuyt PM, Lijmer JG, Mol BW. Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet.* 2000;356(9244):1844–7.
- Biesheuvel CJ, Grobbee DE, Moons KG. Distraction from randomization in diagnostic research. *Ann Epidemiol.* 2006;16(7):540–4.
- Ferrante di Ruffano L, Davenport C, Eisinga A, Hyde C, Deeks JJ. A capture-recapture analysis demonstrated that randomized controlled trials evaluating the impact of diagnostic tests on patient outcomes are rare. *J Clin Epidemiol.* 2012;65(3):282–7.
- Schaafsma JD, van der Graaf Y, Rinkel GJ, Buskens E. Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness. *J Clin Epidemiol.* 2009;62(12):1248–52.
- Bossuyt PM, McCaffery K. *Additional Patient Outcomes and Pathways in Evaluations of Testing.* Medical Tests—White Paper Series. Rockville, MD: 2009.
- Koffijberg H, van Zaane B, Moons KG. From accuracy to patient outcome and cost-effectiveness evaluations of diagnostic tests and biomarkers: an exemplary modelling study. *BMC Med Res Methodol.* 2013;13:12.
- Merlin T. The use of the ‘linked evidence approach’ to guide policy on the reimbursement of personalized medicines. *Personal Med.* 2014;11(4):435–48.
- Merlin T, Lehman S, Hiller JE, Ryan P. The “linked evidence approach” to assess medical tests: a critical analysis. *Int J Technol Assess Health Care.* 2013;29(3):343–50.
- Briggs A, Claxton K. *Decision Modelling for Health Economic Evaluation.* New York: Oxford University Press; 2006.
- Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? *Health Econ.* 2006;15(7):677–87.
- Bossuyt PM, Reitsma JB, Linnet K, Moons KG. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clin Chem.* 2012;58(12):1636–43.
- Jethwa PR, Punia V, Patel TD, Duffis EJ, Gandhi CD, Prestigiacomo CJ. Cost-effectiveness of digital subtraction angiography in the setting of computed tomographic angiography negative subarachnoid hemorrhage. *Neurosurgery.* 2013;72(4):511–9; discussion 9.
- Koffijberg H. Cost-effectiveness analysis of diagnostic tests. *Neurosurgery.* 2013;73(3):E558–60.
- Novielli N, Sutton AJ, Cooper NJ. Meta-analysis of the accuracy of two diagnostic tests used in combination: application to the d-dimer test and the wells score for the diagnosis of deep vein thrombosis. *Value Health.* 2013;16(4):619–28.
- Novielli N, Cooper NJ, Sutton AJ. Evaluating the cost-effectiveness of diagnostic tests in combination: is it important to allow for performance dependency? *Value Health.* 2013;16(4):536–41.
- Otero HJ, Fang CH, Sekar M, Ward RJ, Neumann PJ. Accuracy, risk and the intrinsic value of diagnostic imaging: a review of the cost-utility literature. *Acad Radiol.* 2012;19(5):599–606.
- Ding A, Eisenberg JD, Pandharipande PV. The economic burden of incidentally detected findings. *Radiol Clin North Am.* 2011;49(2):257–65.
- Mueller M, Rosales R, Steck H, Krishnan S, Rao B, Kramer S. Subgroup discovery for test selection: a novel approach and its application to breast cancer diagnosis. In: Adams NM, Robardet C, Siebes A, Boulicaut J-F, eds. *Advances in Intelligent Data Analysis VIII: 8th International Symposium on Intelligent Data Analysis, IDA 2009, Lyon, France, August 31–September 2, 2009 Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2009:119–30.
- Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *Value Health.* 2013;16(2):e1–5.
- Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *Pharmacoeconomics.* 2006;24(4):355–71.

29. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. *JAMA*. 2016;316(10):1093–103.
30. Dowdy DW, Houben R, Cohen T, et al. Impact and cost-effectiveness of current and future tuberculosis diagnostics: the contribution of modelling. *Int J Tuberculosis Lung Dis*. 2014;18(9):1012–8.
31. Wimo A, Ballard C, Brayne C, et al. Health economic evaluation of treatments for Alzheimer's disease: impact of new diagnostic criteria. *J Intern Med*. 2014;275(3):304–16.
32. Raymakers AJ, Mayo J, Marra CA, FitzGerald M. Diagnostic strategies incorporating computed tomography angiography for pulmonary embolism: a systematic review of cost-effectiveness analyses. *J Thorac Imaging*. 2014;29(4):209–16.
33. Oosterhoff M, van der Maas ME, Steuten LM. A systematic review of health economic evaluations of diagnostic biomarkers. *Appl Health Econ Health Policy*. 2016;14(1):51–65.
34. Sailer AM, van Zwam WH, Wildberger JE, Grutters JP. Cost-effectiveness modelling in diagnostic imaging: a stepwise approach. *Eur Radiol*. 2015;25(12):3629–37.
35. Lee J, Tollefson E, Daly M, Kielb E. A generalized health economic and outcomes research model for the evaluation of companion diagnostics and targeted therapies. *Exp Rev Pharmacoecon Outcomes Res*. 2013;13(3):361–70.
36. Glanville J, Paisley S. Identifying economic evaluations for health technology assessment. *Int J Technol Assess Health Care*. 2010;26(4):436–40.
37. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. 2005;8(1):19–32.
38. Mays N, Roberts E, Popay J. Synthesising research evidence. In: Fulop N, Allen P, Clarke A, Black N, ed. *Methods for Studying the Delivery and Organisation of Health Services*. London, UK: Routledge; 2001.
39. Zorginstituut Nederland. *Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg*. Dieman, Holland; 2015.
40. National Institute for Health and Care Excellence (NICE). *Developing NICE Guidelines: The Manual*. London, UK: NICE; 2014.
41. *Guidelines for the economic evaluation of health technologies: Canada*. 4th ed. Ottawa, ON: Canadian Agency for Drugs and Technologies in Health; 2017 Mar.
42. Tiel-van Buul MM, Broekhuizen TH, van Beek EJ, Bosuyt PM. Choosing a strategy for the diagnostic management of suspected scaphoid fracture: a cost-effectiveness analysis. *J Nucl Med*. 1995;36(1):45–8.
43. Shillcutt S, Morel C, Goodman C, et al. Cost-effectiveness of malaria diagnostic methods in sub-Saharan Africa in an era of combination therapy. *Bull WHO*. 2008;86(2):101–10.