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

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-152083

N.B.: When citing this work, cite the original publication.
https://doi.org/10.1177/0272989X18797590

Original publication available at:
https://doi.org/10.1177/0272989X18797590

Copyright: SAGE Publications (UK and US)
http://www.uk.sagepub.com/home.nav
Title: Towards Alignment in the Reporting of Economic Evaluations of Diagnostic Tests and biomarkers: the AGREEDT checklist

Running title: A checklist for diagnostic test evaluations

Authors:

Michelle MA Kip, MSc.¹
Maarten J IJzerman, PhD¹
Martin Henriksson, PhD²
Tracy Merlin, PhD³
Milton C Weinstein, PhD⁴
Charles E Phelps, PhD⁵
Ron Kusters, PhD¹,⁶
Hendrik Koffijberg, PhD¹

¹Department of Health Technology and Services Research, MIRA institute for Biomedical Technology and Technical Medicine, University of Twente, Enschede, the Netherlands
²Department of Medical and Health Sciences, Linköping University, Linköping, Sweden
³Adelaide Health Technology Assessment (AHTA), School of Population Health, University of Adelaide, Adelaide, South Australia, Australia
⁴Department of Health Policy and Management Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
⁵Departments of Economics and Political Science, University of Rochester, Rochester, NY, USA
⁶Laboratory for Clinical Chemistry and Haematology, Jeroen Bosch Ziekenhuis, Den Bosch, the Netherlands
Corresponding author:
Hendrik Koffijberg, PhD
University of Twente, MIRA institute, department HTSR, P.O. Box 217, 7500 AE Enschede, the Netherlands.
Email: h.koffijberg@utwente.nl
Phone: +31 (0)53 489 5384

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

Word count (excluding abstract, acknowledgements, references, and figure legends): 3414

No funding has been received for the conduct of this study and/or the preparation of this manuscript.
Abstract

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 three main steps: 1) the development of an initial checklist based on a scoping review; 2) review and critical appraisal of the initial checklist by four independent experts; 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 four 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.
Introduction:

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 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. Amongst 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 which 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), as well as 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 non-transparent and incorrect evaluation results.

General frameworks and guidelines regarding which aspects to include in decision-analytic modelling and how to report modelling 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 paper 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.

Methods

This study consisted of three main steps: 1) the development of an initial checklist based on a scoping review; 2) review and critical appraisal of the initial checklist by four experts (CEP, MCW, MH, and TM) not involved in the scoping review; 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 last 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, using the following combinations of search terms in title and abstract: (health economic OR cost-effectiveness) AND diag* AND (model OR Markov OR
tree OR modeling OR modelling) already resulted in a total of 1,844 articles. Besides the very 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 which include the words ‘diagnostic’ or ‘diagnosis’). Because of these challenges, and the fact that different evaluations are very likely to in- and exclude the same aspects, a scoping review was performed instead of a systematic literature review, followed by critical appraisal by four 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], in particular 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 one 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, two 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 last five 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 upon 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 one 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 three 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 four 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, as well as 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 evaluations 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 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, four articles did not specifically
evaluate the cost-effectiveness of a (combination of) diagnostic test(s), two concerned a letter
to the editor, and seven articles focused on methodological aspects of the evaluation of
diagnostic strategies, for example, in the context of single disease, or on specific types or
combinations of diagnostic tests (as mentioned in the introduction). In addition, one article
was excluded because the full text could not be obtained or purchased by the university
library, from online databases, 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 in six 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 may for example 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 (for example in terms of reduced patient visits), or on society (for example measures that aim to prevent widespread antibiotic resistance). Quantifying these aspects may provide a broader view on the potential impact of diagnostic testing.

**Critical appraisal and validation of the reporting checklist**

Besides the critical appraisal of experts the list was updated, with one item being removed, and 15 items being added, one 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, eight 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 to be 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 Appendix 2. An overview of this process, including the scoping review and the 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 onwards, 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 three 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, 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 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 as 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 ‘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 very large number of published economic evaluations of diagnostic tests. Therefore, a scoping review was performed instead, by one reviewer. As the judgement regarding whether an aspect was incorporated in a health economic evaluation was sometimes found to be difficult, it cannot be excluded that these judgements may have differed slightly when performed by a different reviewer. In addition, as the decision to limit the search strategy to the last five years was based on reviewing two 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 healthcare). Despite the abovementioned limitations, the
critical review of the checklist by four 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,
eight of these added items actually involved a further specification (i.e. a sub-item) 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 eight sub-items 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
judgement 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 the majority of
these items are actually sub-items, 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 when 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.

Conflicts of interest

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 paper. Prof. Weinstein reports that he is a consultant to OptumInsight on unrelated topics. All other authors declare that there is no conflict of interest.
References:


[20] Jethwa PR, Punia V, Patel TD, Duffis EJ, Gandhi CD, Prestigiacomo CJ. Cost-effectiveness of digital subtraction angiography in the setting of computed tomographic...


effectiveness and cost-effectiveness of the PROGENSA(R) prostate cancer antigen 3 assay
and the Prostate Health Index in the diagnosis of prostate cancer: a systematic review and
Analysis of the Diagnosis and Treatment of Primary Aldosteronism in Japan. Hormone and
metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme.
[54] Suen SC, Bendavid E, Goldhaber-Fiebert JD. Cost-effectiveness of improvements in
diagnosis and treatment accessibility for tuberculosis control in India. The international
journal of tuberculosis and lung disease : the official journal of the International Union
against Tuberculosis and Lung Disease. 2015; 19(9):1115-24, i-xv.
point-of-care testing to assist with the diagnosis, management and monitoring of haemostasis:
a systematic review and cost-effectiveness analysis. Health technology assessment. 2015;
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
al. Next-Generation Sequencing Panels for the Diagnosis of Colorectal Cancer and Polyposis
the American Society of Clinical Oncology. 2015; 33(18):2084-91.


Table 1: Reporting checklist to indicate which items were included in the health economic evaluation of diagnostic tests and biomarkers.

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

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

<table>
<thead>
<tr>
<th>ITEMS OF THE EVALUATION OF DIAGNOSTIC TESTS AND BIOMARKERS</th>
<th>INCLUDED IN EVALUATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIME TO PRESENTATION</td>
<td></td>
</tr>
<tr>
<td>ONSET OF COMPLAINTS OR ONSET OF SUSPICION BY PHYSICIAN AND START OF DIAGNOSTIC TRAJECTORY</td>
<td>YES*</td>
</tr>
<tr>
<td>1.1 The study should contain a description of the individuals that 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).</td>
<td></td>
</tr>
<tr>
<td>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). (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).</td>
<td></td>
</tr>
</tbody>
</table>
### DECISION REGARDING WHICH DIAGNOSTIC TEST(S) IS/ARE PERFORMED, IN WHICH PATIENTS, AND IN WHAT ORDER

<p>| 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 two (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 work-up 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 (for example perceived condition or risk, or symptom presentation). |</p>
<table>
<thead>
<tr>
<th>TEST PERFORMANCE AND CHARACTERISTICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIAGNOSTIC TEST PERFORMANCE AND ITEMS RELATED TO THE SAMPLING AND TESTING</td>
</tr>
</tbody>
</table>

| 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. cut-off 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 impacts the estimated test performance |
| 3.2.4 | Describe how uncertainty/variation in the test performance (ROC curve) was handled or explained, for example due to inter-rater and intra-rater reliability, or experience of the clinician. |
| 3.2.5 | Describe the logic, or analysis, applied to choose the cut-off value (i.e. the point on the ROC curve) for the test, for example |

| YES* | NO** |
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 cut-off 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 Receiver Operating Characteristic (ROC) curves for each test).

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 *The impact of incidental findings on performing additional tests is addressed.*

<table>
<thead>
<tr>
<th>PATIENT MANAGEMENT DECISIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IMPACT OF A TEST ON THE DIAGNOSIS AND/OR PATIENT MANAGEMENT STRATEGY (BASED ON THIS DIAGNOSIS)</strong></td>
</tr>
<tr>
<td><strong>YES</strong></td>
</tr>
</tbody>
</table>

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).‡

4.4 Consider the consistency of test results over time (for example: 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.

<table>
<thead>
<tr>
<th>IMPACT ON HEALTH OUTCOMES AND COSTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMPACT OF THE PATIENT MANAGEMENT STRATEGY ON DISEASED AND NON-DISEASED INDIVIDUALS, IN TERMS OF HEALTH OUTCOMES AND COSTS</td>
</tr>
<tr>
<td>YES*</td>
</tr>
</tbody>
</table>

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 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.
| 5.2 | 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).‡ |
| 5.3 | Consider patient's adherence to treatment (which includes aspects that may indicate (partial) non-adherence, for example, following only some of the treatment recommendations, as well as aspects that affect the degree of administration of treatment). |
| 5.4 | Consider the occurrence (and consequences) of treatment-related adverse events. |
| 5.5 | 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 whether the patient received a correct or an incorrect diagnosis). |
| 5.6 | Describe the probability of, or time it takes to repeat or extend the diagnostic work-up when the patient management strategy proves to be ineffective, either directly or over time (regardless 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. |
| 5.7 | Describe the impact of ineffective or unnecessary treatment or management on health outcomes and/or costs (including both side
effects and costs, and regardless 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.

5.7.1 Consider the impact of delay in treatment initiation on health outcomes and/or costs

<table>
<thead>
<tr>
<th>SOCIETAL IMPACT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WIDER (SOCIETAL) IMPACT OF THE CHOSEN DIAGNOSTICS AND MANAGEMENT STRATEGY</strong></td>
</tr>
<tr>
<td>YES*</td>
</tr>
</tbody>
</table>

6.1 Consider the psychological impact of diagnostic outcome and management strategy on patients, including value of knowing (in terms of reassurance or anxiety), patient preferences regarding undergoing diagnostic tests, and the (accompanying) impact on caregivers, or relatives, etc.

6.1.1 *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).*

6.2 Consider the additional impact of diagnostic outcome and management strategy on the health system or healthcare professionals.

6.3 Consider the additional impact of diagnostic outcome and management strategy for society.

‡ Existing 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.
**Figure legends:**

**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, 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.

**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.*
APPENDIX 1 – Description of items added to or excluded from the checklist, based on experts’ critical revisions

ITEMS ADDED TO THE CHECKLIST:

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEST PERFORMANCE AND CHARACTERISTICS</td>
<td>Specify the costs of the diagnostic test.</td>
</tr>
<tr>
<td></td>
<td>Describe the evidence base for the estimated test performance.</td>
</tr>
<tr>
<td></td>
<td>Describe the positivity criterion (i.e. cut-off value) applied to the test or testing strategy.</td>
</tr>
<tr>
<td></td>
<td>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.</td>
</tr>
<tr>
<td></td>
<td>Describe how uncertainty in the test performance (ROC curve) was handled or explained, for example due to inter-rater and intra-rater reliability, or experience of the clinician.</td>
</tr>
<tr>
<td></td>
<td>Describe the logic, or analysis, applied to choose the cut-off value (i.e. the point on the ROC curve) for the test.</td>
</tr>
<tr>
<td></td>
<td>Consider whether test performance is dependent on disease prevalence.</td>
</tr>
<tr>
<td></td>
<td>Consider whether test performance is based on a combination of tests (and on a combination of areas under the Receiver Operating Characteristic (ROC) curves for each test).</td>
</tr>
<tr>
<td></td>
<td>Consider how likely/to what extent bias in the available/applied evidence impacts the estimated test performance</td>
</tr>
<tr>
<td></td>
<td>Consider the feasibility of obtaining (sufficient) sample and/or usability of the sample that is obtained.</td>
</tr>
<tr>
<td></td>
<td>Consider costs of retesting (after obtaining insufficient/unusable sample, or after test failure or indeterminate/not assessable result).</td>
</tr>
<tr>
<td></td>
<td>The impact of incidental findings on performing additional tests is addressed.</td>
</tr>
<tr>
<td>PATIENT MANAGEMENT DECISIONS</td>
<td>Clearly specify the impact of the test in selecting the patient management strategy.</td>
</tr>
<tr>
<td></td>
<td>Consider the consistency of test results over time (for example: genetic mutations may be affected by treatment prescribed after the initial diagnosis).</td>
</tr>
<tr>
<td></td>
<td>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.</td>
</tr>
<tr>
<td>SOCIETAL IMPACT</td>
<td>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).</td>
</tr>
</tbody>
</table>

ITEMS REMOVED FROM THE CHECKLIST:

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOCIETAL IMPACT</td>
<td>Likelihood that the results can be generalized to other countries and/or settings (for example different patient groups based on age, severity or other characteristics).</td>
</tr>
</tbody>
</table>
DESCRIPTION OF REASONS FOR THE CHANGES MADE TO THE CHECKLIST:

Items added to the checklist:

Twelve items were added to the category ‘test performance and characteristics’. The item ‘specify the costs of the diagnostic test’ was included as this is specific for and very important in the health economic evaluation of diagnostic tests. The following eight items (as shown on the previous page) were added with the goal to more accurately specify how the diagnostic performance was either estimated or derived from literature, and to incorporate uncertainty or bias of this diagnostic performance in the health economic evaluation.

The item ‘Consider the feasibility of obtaining (sufficient) sample and/or usability of the sample that is obtained’ was added to the checklist, whereas the previous version of the checklist incorporated this issue within the item ‘usefulness or usability of the results, or usability of the sample that is obtained’. Therefore, this item in the checklist was changed to ‘consider costs of retesting (after test failure or indeterminate/not assessable result)’, in order to assess the usability of the sample and the usability of the test result in separate items. In addition, the costs of retesting after an indeterminate/not assessable result, or after obtaining an unusable sample, was added.

In addition, although the impact of diagnostic tests on incidental findings was already included in the checklist, it was decided to explicitly include an item regarding addressing the impact of incidental findings on performing additional tests.

In the category ‘patient management decisions’, three items were added which aim to 1) explicitly specify the aim of the diagnostic test (e.g. is the test used for diagnosis, screening, disease staging), 2) to consider the consistency of test results over time (as outcomes of diagnostic tests may change over time, for example due to spontaneous cure, or by the occurrence of genetic mutations due to treatment), and 3) to not only consider the time the test itself takes, but also the time it takes to perform the test (by the health care professional) and the time it takes to interpret the results.

Finally, in the category ‘societal impact’, the previous version of the checklist already incorporated the impact of a diagnostic test on relatives in terms of reassurance or caregiving, but this item did not specifically address the impact of a test on relatives themselves. Therefore, an item was added to specify the impact of test outcomes on subsequent testing and/or treatment of relatives.

Items removed from the checklist:

The item regarding the generalizability of the study’s findings was excluded as it was agreed upon that this item should be included/considered in any health economic evaluation, and is not specific to diagnostic test evaluation.
**APPENDIX 2 – overview and definitions of the final reporting checklist.**

This item shows the checklist of items that should be considered for inclusion in health economic evaluations of diagnostics tests and biomarkers. Each item is illustrated using an example from a selection of previously published papers.

### TIME TO PRESENTATION

**ONSET OF COMPLAINTS OR ONSET OF SUSPICION BY PHYSICIAN AND START OF DIAGNOSTIC TRAJECTORY**

<table>
<thead>
<tr>
<th>ITEM INCLUDED IN STUDY</th>
<th>ITEM MENTIONED (BUT NOT INCLUDED)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 The study should contain a description of the individuals that 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).</td>
<td>[2-64]</td>
</tr>
<tr>
<td>Example from study: [1] Patients present with chest pain at the general practitioner.</td>
<td></td>
</tr>
</tbody>
</table>

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). *(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).*

<table>
<thead>
<tr>
<th>Item Mentioned</th>
<th>Example from study: [65] In patients with non-traumatic intracerebral hemorrhage, the time since onset of symptoms and presentation at the emergency department influences the clinical condition at presentation.</th>
</tr>
</thead>
<tbody>
<tr>
<td>[2, 4, 13, 19, 20, 33, 34, 37, 43, 57]</td>
<td>[25, 53, 56]</td>
</tr>
</tbody>
</table>

### USE OF DIAGNOSTIC TESTS

**DECISION REGARDING WHICH DIAGNOSTIC TEST(S) IS/ARE PERFORMED, IN WHICH PATIENTS, AND IN WHAT ORDER**

<table>
<thead>
<tr>
<th>ITEM INCLUDED IN STUDY</th>
<th>ITEM MENTIONED (BUT NOT INCLUDED)</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
<td>[2-64]</td>
</tr>
<tr>
<td>Example from study: [65] Computed tomography angiography can be used to detect macrovascular causes in patients with non-traumatic intracerebral hemorrhage.</td>
<td></td>
</tr>
</tbody>
</table>

2.1 Consider whether more than two (possible) diagnostic strategies can be compared, each involving a single test or combination of tests.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[15, 20, 22, 29, 33]</td>
<td></td>
</tr>
</tbody>
</table>

2.2 Consider whether the evaluated diagnostic strategies include multiple tests, which can be performed in parallel or in sequence.

<table>
<thead>
<tr>
<th>Example from study: [65] Patients with non-traumatic intracerebral hemorrhage first undergo computed tomography angiography. Patients with a negative result undergo additional testing.</th>
<th>[3-6, 8-14, 16-19, 21, 24-49, 51-54, 56-64]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[2]</td>
<td></td>
</tr>
</tbody>
</table>
2.2.1 Consider whether some tests of the diagnostic work-up are performed conditional on previous test outcomes, leading to a selection of patients undergoing specific tests.

Example from study: [65]
In patients with non-traumatic intracerebral hemorrhage, tests are performed in sequence conditional on previous test outcomes (see 2.2).

[3-6, 8-10, 14, 16-18, 20, 24-28, 30-40, 42-48, 51, 53, 54, 56-62, 64] [13, 49]

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).

Example from study: [1]
The point-of-care troponin test is only performed in patients with at least 4 hours chest pain.

[3, 27, 34, 40, 44, 48, 54] [2, 9, 13, 26, 30, 32, 36, 46, 51]

2.4 Consider whether different tests are applied based on implicit (shared) decision-making (for example perceived condition or risk, or symptom presentation).

Example from study: [1]
In chest pain patients, the general practitioner decides, based on the perceived risk of the patient, whether to use a point-of-care troponin test and/or electrocardiogram, or to immediately refer the patient to the hospital.

[9, 10, 14, 17, 34, 35, 37, 43, 44, 51, 54, 58, 63] [13, 20, 26, 27, 30-32, 40, 46, 55, 57, 61]

TEST PERFORMANCE AND CHARACTERISTICS
DIAGNOSTIC TEST PERFORMANCE AND ITEMS RELATED TO THE SAMPLING AND TESTING

<table>
<thead>
<tr>
<th>ITEM INCLUDED IN STUDY</th>
<th>ITEM MENTIONED (BUT NOT INCLUDED)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specifying the costs of the diagnostic test.</td>
<td>Example from study: [1] The costs of the point-of-care troponin test were estimated at €15. [2-22, 24-64]</td>
</tr>
<tr>
<td>Specifying 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.</td>
<td>Example from study: [1] The accuracy of the point-of-care troponin test is based on previously published literature. [2-22, 24-64]</td>
</tr>
<tr>
<td>Describe the evidence base for the estimated test performance.</td>
<td>Example from study: [1] The sensitivity and specificity of the point-of-care troponin test were derived from Diercks et al. [2-11, 13-22, 24-63]</td>
</tr>
<tr>
<td>Describe the positivity criterion (i.e. cut-off value) applied to the test or testing strategy.</td>
<td>Example from study: [66] Anemia was defined as hemoglobin below 13.7 g/dL (8.5 mmol/L) for males and below 12.1 g/dL (7.5 mmol/L) for females. [2-4, 10, 19, 30, 31, 40, 42, 44, 56, 59] [9, 13]</td>
</tr>
<tr>
<td>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.</td>
<td>Example from study: [29] Significant correlations between the test outcomes for all three soil transmitted helminth species in both infected and uninfected children suggest that the two tests are conditionally dependent, highlighting the inappropriateness of using a combined reference standard for evaluation. [4, 9, 16, 22, 30, 42, 44, 45, 51, 53, 64] [2, 3, 7, 15, 21, 29, 31, 34, 47-50, 55, 59]</td>
</tr>
<tr>
<td>3.2.3.1</td>
<td>Consider how likely/to what extent bias in the available/applied evidence impacts the estimated test performance</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Describe how uncertainty/variation in the test performance (ROC curve) was handled or explained, for example due to inter-rater and intra-rater reliability, or experience of the clinician.</td>
</tr>
<tr>
<td>3.2.5</td>
<td>Describe the logic, or analysis, applied to choose the cut-off 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.</td>
</tr>
<tr>
<td>3.2.6</td>
<td>Describe whether different test performances and cut-off 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).</td>
</tr>
<tr>
<td>3.2.7</td>
<td>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).</td>
</tr>
</tbody>
</table>
Consider whether test performance is based on a combination of tests (and on a combination of areas under the Receiver Operating Characteristic (ROC) curves for each test).

Example from study: [46]
To account for correlations between tests within diagnostic strategies, the sensitivities and specificities of tests are calculated conditional on positive/uncertain results in earlier tests in the strategy.

[4, 5, 16, 25, 29, 30, 34, 42, 46, 52, 53, 56, 58-61]  [2, 9, 13, 49]

Consider the feasibility of obtaining (sufficient) sample and/or usability of the sample that is obtained.

Example from study: [51]
For the base-case analysis, it was assumed that 3% of tests required to be repeated due to processing failure (of the sample) or an indeterminate result.

[3, 10, 29, 30, 51, 60]

Consider the occurrence of test failures or indeterminate/not assessable results.

Example from study: [51]
For the base-case analysis, it was assumed that 3% of tests required to be repeated due to processing failure (of the sample) or an indeterminate result.

[5, 9, 15, 24, 26, 28, 30, 44-47, 51, 52, 59, 61, 62]  [4, 17, 22, 35, 43, 49, 57]

Consider costs of retesting (after obtaining insufficient/unusable sample, or after test failure or indeterminate/not assessable result).

Example from study: [51]
For the base-case analysis, it was assumed that 3% of tests required to be repeated due to processing failure (of the sample) or an indeterminate result.

[9, 26, 28, 30, 44, 51]  [15, 43]

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).

Example from study: [65]
Digital subtraction angiography involves a catheterization, and the most common complications involved a groin hematoma (2%), and (possible) thromboembolic complications (2%).

[4, 5, 8, 9, 26, 28, 30, 39, 41-44, 46, 47, 57, 59, 64]

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).

Example from study: [1]
The duration of a point-of-care troponin test is estimated at 10 minutes.

[11, 13, 19, 20, 22, 38, 41, 42, 45, 48, 51, 53, 54, 57, 61, 62]  [4, 18, 24, 25, 27, 30, 36, 44, 46, 50, 55]

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).

Example from study: [28]
The study includes the indirect costs associated with incidental findings during calcium scoring and cardiac computed tomography angiography.

[28, 30]  [4, 11, 13, 14, 16, 21, 39, 43, 57, 62, 64]

The impact of incidental findings on performing additional tests is addressed.

Example from study: [28]
We assumed that the incidental cases (as mentioned in 3.8) would require non-contrast chest computed tomography a few months later at an additional costs of €80.00.

[28, 30]  [40]
### 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. | Example from study: [28] The diagnostic procedures either ends when a test is negative, or invasive coronary angiography is performed when the last test is positive or inconclusive. | [2-22, 24-26, 28, 30-36, 38, 39, 41-53, 56-59, 61-64, 67] | [27, 29, 37, 40, 55, 60] |
| 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. | Example from study: [65] Among patients with non-traumatic intracerebral hemorrhage, not all patients with negative results in the initial strategy underwent digital subtraction angiography, owing to reluctance of patients and their treating clinicians because of its (small) risk of complications. | [2, 4, 5, 9, 11, 17, 20, 30-34, 36-38, 43-45, 48, 51, 54, 56, 57, 62, 64] | [3, 6, 7, 10, 13, 21, 24, 31, 42, 52, 53, 58, 61, 63] |
| 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). | Example from study: [68] When an aneurysm is diagnosed, the strategy of whether or not to (surgically) treat a patient, may be influenced by the patient’s age. | [2, 4, 12, 13, 15, 21, 24, 30, 31, 33-35, 39, 40, 42, 48, 50, 52, 54, 58] | [6, 19, 37, 41, 60] |
| 4.4 | Consider the consistency of test results over time (for example: genetic mutations may be affected by treatment prescribed after the initial diagnosis). | Example from study: [54] Among patients with tuberculosis, smear-negative cases may progress to smear-positive, and all individuals with active disease may spontaneously self-cure. | [9, 53, 54] |
| 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. | Example from study: [50] The modified transesophageal echocardiography imaging procedure takes, on average, only 4.5 minutes to perform. | [15, 19, 20, 24, 31, 41, 42, 51, 52] | [29, 38, 50] |

### IMPACT ON HEALTH OUTCOMES AND COSTS

**IMPACT OF THE PATIENT MANAGEMENT STRATEGY ON DISEASED AND NON-DISEASED INDIVIDUALS, IN TERMS OF HEALTH OUTCOMES AND COSTS**

<p>| 5 | Evaluate the direct impact of the chosen patient management strategy on the number of (in)correctly diagnosed individuals, health outcomes, and/or costs. | Example from study: [69] The overall impact of an extensive laboratory work-up in anemia patients on the (correct) diagnosis, on treatment decisions and on accompanying costs was quantified. | [2-64] |
| 5.1 | Consider the direct impact of the chosen patient management strategy, on health outcomes and/or costs. This concerns the entire period in which patient well-being over his or her entire lifetime, the impact on quality of | Example from study: [1] As acute coronary syndrome may impact the patient’s well-being over his or her entire lifetime, the impact on quality of | [4-7, 11-15, 20, 24, 25, 35, 41-44, 46, 48, 49, 58] | [2, 8-10, 16-19, 21-23, 26-34, 36-39,] |</p>
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Example from study</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>5.2</strong></td>
<td>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).</td>
<td>Example from study: [69] Among patients presenting with anemia in primary care, the results in terms of costs and health outcomes (i.e. correct underlying causes) were presented separately for the different underlying causes.</td>
<td>[2, 8, 10, 12-15, 21-24, 28, 30, 33, 38-40, 42, 44, 50, 54, 64] [6, 35, 48, 53, 58]</td>
</tr>
<tr>
<td><strong>5.3</strong></td>
<td>Consider patient’s adherence to treatment (which includes aspects that may indicate (partial) non-adherence, for example, following only some of the treatment recommendations, as well as aspects that affect the degree of administration of treatment).</td>
<td>Example from study: [54] Individuals on TB treatment may successfully complete treatment, fail, default (become lost to follow-up) or die.</td>
<td>[8, 11, 23, 30, 33, 38, 42, 51, 53, 54, 56, 58, 62, 63] [61]</td>
</tr>
<tr>
<td><strong>5.4</strong></td>
<td>Consider the occurrence (and consequences) of treatment-related adverse events.</td>
<td>Example from study: [54] Individuals who fail or default from treatment may acquire resistance to the drugs they have received.</td>
<td>[3, 5, 8, 12, 15, 26, 30, 33, 41, 42, 44, 49-52, 54, 56] [6, 9, 21, 31, 34, 35, 58, 63]</td>
</tr>
<tr>
<td><strong>5.5</strong></td>
<td>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 whether the patient received a correct or an incorrect diagnosis).</td>
<td>Example from study: [69] Although a patient may be diagnosed with the incorrect underlying cause of anemia, this same patient may (by coincidence) receive a management or treatment strategy that is considered effective.</td>
<td>[2, 9-11, 20-24, 26, 31-34, 42, 44, 45, 48, 52-54, 56, 58, 62-64] [61]</td>
</tr>
<tr>
<td><strong>5.6</strong></td>
<td>Describe the probability of, or time it takes to repeat or extend the diagnostic work-up when the patient management strategy proves to be ineffective, either directly or over time (regardless 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.</td>
<td>Example from study: [69] Although a patient may be diagnosed with the correct underlying cause of anemia, the management or treatment strategy (as chosen by the general practitioner) may be considered ineffective.</td>
<td>[2, 4, 6-12, 16-18, 21-24, 26, 28, 30-36, 40, 42-46, 48, 52-54, 58, 62-64] [57, 61]</td>
</tr>
<tr>
<td><strong>5.7</strong></td>
<td>Describe the impact of ineffective or unnecessary treatment or management on health outcomes and/or costs (including both side effects and costs, and regardless 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.</td>
<td>Example from study: [69] The impact of ineffective treatment was quantified (in terms of repeat consultations and additional healthcare costs).</td>
<td>[2-5, 7, 8, 11, 13, 16, 17, 20-22, 24-26, 28, 30-36, 38, 39, 41-43, 45-48, 50-54, 56, 59, 61, 62, 64] [6, 9, 12, 14, 15, 18, 19, 23, 27, 37, 40, 44, 57, 60, 63]</td>
</tr>
<tr>
<td><strong>5.7.1</strong></td>
<td>Consider the impact of delay in treatment initiation on health outcomes and/or costs</td>
<td>Example from study: [1]</td>
<td>[4, 5, 7, 11, 13, 21, 24-26, 30, 33, 35] [6, 8, 9, 12, 14, 15, 17-19, 23, 27, 34,]</td>
</tr>
</tbody>
</table>
In patients with acute coronary syndrome, a delayed referral to the emergency care (and consequently a delay in treatment initiation), is associated with worse health outcomes.

38, 39, 42, 45, 46, 48, 52-54, 59, 61, 64

37, 40, 44, 51, 57, 60, 62

### SOCIETAL IMPACT

**WIDER (SOCIETAL) IMPACT OF THE CHOSEN DIAGNOSTICS AND MANAGEMENT STRATEGY**

<table>
<thead>
<tr>
<th>ITEM INCLUDED IN STUDY</th>
<th>ITEM MENTIONED (BUT NOT INCLUDED)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Consider the psychological impact of diagnostic outcome and management strategy on patients, including value of knowing (in terms of reassurance or anxiety), patient preferences regarding undergoing diagnostic tests, and the (accompanying) impact on caregivers, or relatives, etc.</td>
<td>[4, 14, 42, 48, 56, 58]</td>
</tr>
<tr>
<td>6.1.1 <strong>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).</strong></td>
<td>[2, 6, 8, 9, 15, 21, 25-27, 30-32, 34, 35, 43, 57]</td>
</tr>
<tr>
<td>Example from study: [14] Among patients with hereditary colorectal cancer and polyposis syndromes, it was (in the base-case analysis) assumed that first-, second-, and third-degree relatives would be contacted.</td>
<td>[14, 56]</td>
</tr>
<tr>
<td>Example from study: [14] If the relative (of the patient with hereditary colorectal cancer and polyposis syndromes) was found to carry the pathogenic variant, he or she was presumably offered intensive colorectal surveillance with colonoscopy at recommended intervals.</td>
<td>[27]</td>
</tr>
<tr>
<td>6.2 Consider the additional impact of diagnostic outcome and management strategy on the health system or healthcare professionals.</td>
<td>[37, 52, 54]</td>
</tr>
<tr>
<td>Example from study: [69] If an anemia patient receives a correct treatment immediately, this lowers the number of follow-up appointments at the general practitioner.</td>
<td>[4, 5, 9, 15, 17, 19-21, 25, 27, 29, 30, 32, 35, 40, 42, 45, 48, 51, 60, 61]</td>
</tr>
<tr>
<td>6.3 Consider the additional impact of diagnostic outcome and management strategy for society.</td>
<td>[16, 29, 54]</td>
</tr>
<tr>
<td>Example from study: [70] Use of the procalcitonin test in intensive care unit patients with sepsis limits the duration of antibiotic therapy, which in turn may reduce antibiotic resistance.</td>
<td>[2, 3, 6, 8, 9, 11, 19, 20, 24, 33, 34, 36, 37, 42, 53, 55, 61-63]</td>
</tr>
</tbody>
</table>
References:


