


Do clinical practice guidelines consider evidence about diagnostic test consequences on patient-relevant outcomes? A critical document analysis

Mariska K. Tuut MSc^{1,2}  | Jako S. Burgers MD, PhD^{1,3} |
Trudy van der Weijden MD, PhD¹ | Miranda W. Langendam PhD⁴

¹School CAPHRI, Department of Family Medicine, Maastricht University Medical Centre, Maastricht, The Netherlands

²PROVA, Varsseveld, The Netherlands

³Dutch College of General Practitioners, Utrecht, The Netherlands

⁴Department Epidemiology and Data Science, Amsterdam UMC/University of Amsterdam, Amsterdam, The Netherlands

Correspondence

Mariska K. Tuut, Maastricht University, CAPHRI, PO Box 616, 6200 MD Maastricht, The Netherlands.
Email: mariska.tuut@maastrichtuniversity.nl

Abstract

Rationale, aims and objectives: Supporting evidence for diagnostic test recommendations in clinical practice guidelines (CPGs) should not only include diagnostic accuracy, but also downstream consequences of the test result on patient-relevant outcomes. The aim of this study is to assess the extent to which evidence-based CPGs about diagnostic tests cover all relevant test-treatment pathway components.

Methods: We performed a systematic document analysis and quality assessment of publicly accessible CPGs about three common diagnostic tests: C-reactive protein, colonoscopy and fractional exhaled nitric oxide. Evaluation of the impact of the full test-treatment pathway (diagnostic accuracy, burden of the test, natural course of target condition, treatment effectiveness, and link between test result and administration of treatment) on patient relevant outcomes was considered best practice for developing medical test recommendations.

Results: We retrieved 15 recommendations in 15 CPGs. The methodological quality of the CPGs varied from poor to excellent. Ten recommendations considered diagnostic accuracy. Four of these were funded on a systematic review and rating of the certainty in the evidence. None of the CPGs evaluated all steps of the test-treatment pathway. Burden of the test was considered in three CPGs, but without systematically reviewing the evidence. Natural course was considered in two CPGs, without a systematic review of the evidence. In three recommendations, treatment effectiveness was considered, supported with a systematic review and rating of the certainty in the evidence in one CPG. The link between test result and treatment administration was not considered in any CPG.

Conclusions: The included CPGs hardly seem to consider evidence about test consequences on patient-relevant outcomes. This might be explained by reporting issues and challenging methodology. Future research is needed to investigate how to facilitate guideline developers in explicit reliable consideration of all steps of a test-treatment pathway when developing diagnostic test recommendations.

KEYWORDS

clinical guidelines, diagnosis, evidence-based medicine, systematic reviews

1 | INTRODUCTION

Clinicians use medical tests to confirm or exclude a clinical diagnosis (e.g., polymerase chain reaction-test to diagnose COVID-19), to test the likelihood of a certain clinical diagnosis (e.g., prostate-specific antigen-test to screen for risk on prostate cancer) or for follow-up of patients to monitor recovery (e.g., rehabilitation checklists).¹ Test results guide (treatment) decisions. The clinical value of a medical test depends on various elements: the patient population characteristics (e.g., prevalence of the disease), test characteristics (e.g., sensitivity and specificity) and its downstream consequences (e.g., benefits and harms of treatment) on patient-important outcomes.²

Clinical practice guidelines (CPGs) provide recommendations to support professionals and patients in clinical decision-making, with the ultimate goal of improving or maintaining patients' health. In the development of CPGs, the benefits and harms of the interventions of interest are systematically assessed with regard to patient-relevant outcomes. The grading of recommendations, assessment, development and evaluation (GRADE) approach is designed to facilitate this process.³

Diagnostic CPGs provide recommendations about the use of a certain test (or test strategy). Supporting evidence for these recommendations consists of studies about diagnostic accuracy.⁴ However, acceptable test characteristics (sensitivity and specificity) are not enough to improve patients' health. CPG developers should also consider downstream consequences (e.g., burden of the test and the proportion of patients with a certain test result who receive the recommended treatment) on patient-relevant outcomes (e.g., mortality, morbidity and quality of life) (see Figure 1).^{6,7}

The interpretation of evidence about the value of therapeutic interventions is complex, and there is room for improvement.⁸ This applies even more to evidence about diagnostic tests and its translation into CPG recommendations.⁹⁻¹¹ There have been few randomized controlled trials on the value of test-treatment pathways for

patient-relevant outcomes.⁹ Evaluating the value of diagnostic tests on patient-relevant outcomes in CPGs is thus complex since it requires integration of various pieces of evidence for the different links in a chain (see Figure 1).

In the GRADE approach for diagnostic tests and test strategies, the first step is to formulate the clinical question, including definition of patient-important outcomes and description of the aim of the test (add-on, replacement or triage). The next step is to assess diagnostic accuracy and downstream consequences of testing. These include the burden of the test, clinical management, natural course of the target condition (to estimate the outcomes of patients with a false negative test result), and the link between test result and management (proportion of patients with a certain test result who receive the recommended treatment). Ideally, each evidence component is based on a systematic review of the literature and the certainty in the evidence for each component is determined separately.⁹ Finally, the evidence components are integrated and the overall certainty in the evidence is assessed.^{12,13} To move from evidence to recommendation, guideline developers use the GRADE evidence-to-decision framework.¹²

The aim of this study is to assess the extent to which evidence-based CPGs about diagnostic tests cover all relevant test-treatment pathway components.

Specific objectives are to assess the types of supporting evidence used for CPG recommendations about diagnostic tests, and to explore determinants of best practices. In the context of CPG development about the value of a diagnostic test, we formulated the following research questions:

1. Which types of evidence (diagnostic accuracy, burden of the test, natural course, treatment effectiveness, link between test result and administration of treatment) are used to support the recommendations?

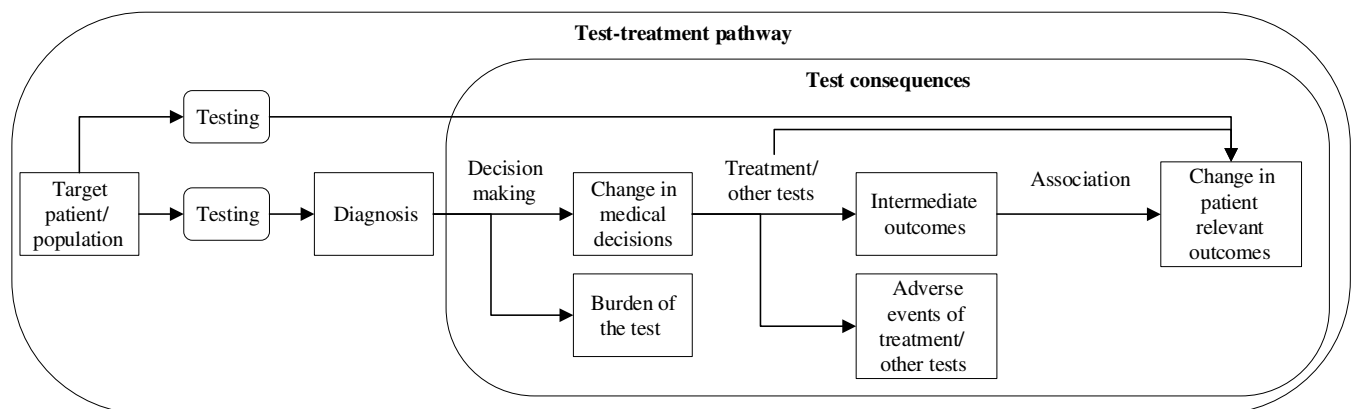


FIGURE 1 Test-treatment pathway (adapted from Harris et al, 2001)⁵

2. Which factors (e.g., composition of the guideline panel, use of the GRADE approach, methodological quality according to AGREE II's domain methodology) contribute to completeness of the evidence?
3. To what extent can differences between CPG recommendations be explained by including different types of evidence?

Answers to these questions elucidate gaps in the implementation of good CPG development methods when developing recommendations about diagnostic tests and test strategies and can help guideline methodologists in developing strategies to facilitate this process.

2 | METHODS

2.1 | Design

In order to assess the types of supporting evidence used for CPG recommendations about diagnostic tests, and to identify factors related to the extent of the supporting evidence, we performed a systematic document analysis of recent versions of publicly accessible CPGs concerning three diagnostic topics.

2.2 | Topics

We chose tests that are frequently used to diagnose three common diseases. We considered tests with different characteristics (primary vs. secondary care, negligible vs. reasonable risk of serious burden of the test, low vs. high costs) to identify possible factors related to differences in methodological approach in the development of the CPGs. We therefore selected the following tests:

- C-reactive protein (CRP) test to increase the likelihood of pneumonia (annual incidence estimated at 0.5%–1.1%) in primary care patients with cough (excluding diagnostic procedures in patients suspected of having a COVID-19 infection),¹⁴
- Colonoscopy to detect colon cancer (annual incidence 1 148 515 new cases) in secondary care patients suspected of having (primary) colon cancer (excluding screening and tests in patients at risk of hereditary types of colon cancer),¹⁵
- Fractional exhaled nitric oxide (FeNO) to diagnose (severe) asthma (prevalence 3.6%) in children and adults in primary and secondary care (excluding monitoring of asthma).¹⁶

2.3 | Search and selection of relevant CPGs

Current, publicly accessible, recent (publication date 2016–2020) CPGs were eligible if they included recommendations about the tests mentioned above, were CPGs at a national or international level, and were published in English, German or Dutch.

To identify relevant CPGs, one author (Mariska K Tuut) performed the search and selected the CPGs. The selection was

checked for accuracy by a second author (Jako S Burgers). In February 2021, we searched the international guideline library from guidelines international network (GIN, [<https://guidelines.ebmportal.com/>], including around 3000 CPGs, mostly developed by organizational GIN members), databases from organizational GIN members active in CPG development ($n = 103$), the TRIP database (Turning Research Into Practice [<https://www.tripdatabase.com/>], containing around 10,000 English-language CPGs) and Medline (see Appendix 1 for full search details).

2.4 | Identification of recommendations

We analysed the content of the selected CPGs to identify relevant recommendations, including supporting evidence available online (e.g., tables with study characteristics, evidence documents, GRADE Evidence Profiles), as well as information about the methods of CPG development of the developing organization (e.g., methodology manuals).

2.5 | Data extraction

In the preparatory phase of this study, we piloted data extraction on two recommendations with four authors (Mariska K Tuut, Miranda W Langendam, Jako S Burgers and Trudy van der Weijden) to refine the data extraction form and define the variables for which we needed data extraction in duplicate. One author (Mariska K Tuu) extracted the initial characteristics of each recommendation and CPG (CPG title (including English translation if relevant), initial developing organization, country, publication year, recommendation text (including English translation if relevant)).

Detailed information about each recommendation and CPG was extracted by one author (Mariska K Tuu) and critically reviewed by another author (Miranda W Langendam, Jako S Burgers or Trudy van der Weijden) using a predefined and piloted data extraction form (see Appendix 2 for the data extraction form and the categorisation of the variables). The form consisted of questions about scope and target audience of the CPG and composition of the CPG panel, involvement of methodologist(s), methodological quality of the CPG (using AGREE II, domain methodology, items 7–12),^{17,18} patient involvement (using AGREE II item 5),^{17,18} the types and extent of supporting evidence for the recommendation (consideration and inclusion of systematic evaluation with assessment of the certainty in the evidence about diagnostic accuracy, burden of the test, natural course, treatment effectiveness and link between test result and administration of treatment), grading of the recommendation and use of the GRADE approach, direction of the recommendation, and characteristics of the test and target condition. Disagreements between the reviewers were discussed until consensus was reached.

2.6 | Analysis

We tabulated basic and detailed characteristics of the included recommendations and CPGs. Systematic evaluation (with a systematic

review of the literature and assessment of the certainty in the evidence) of the impact of the full test-treatment pathway (diagnostic accuracy, burden of the test, natural course of target condition, treatment effectiveness, and link between test result and administration of treatment) on patient relevant outcomes was considered best practice for developing medical test recommendations.

We planned to analyse a possible relation between differences in evidence base and methodological factors (e.g., composition of the CPG panel, involvement of patients and methodologists, development approach). However, because the data about the evidence base were quite homogenous we were not able to perform these analyses.

3 | RESULTS

3.1 | Search and selection of relevant CPGs

Full details of the search and selection process are described in Appendix 1. In short, the search identified 15 unique relevant recommendations in 15 CPGs: four about CRP related to the diagnosis pneumonia in primary care,¹⁹⁻²² five about colonoscopy in secondary care patients suspected of having colon cancer,²³⁻²⁷ and six about the use of FeNO to diagnose (severe) asthma.²⁸⁻³² The search and selection process is illustrated in Figure 2.

In Table 1, we present the included CPGs with information about the developing organization, the country of publication and the publication year. All guidelines originated from high-income countries.

3.2 | Quality of the guidelines and use of the GRADE approach

Table 2 presents detailed information about the composition of the CPG panel, the methodological quality of the included CPGs, the

direction and grading of the recommendation and the reported and actual use of the GRADE approach. Nine out of 15 CPGs included a methodologist in the development process, in the CPG panel and/or at bureau level.^{23,24,27-32} In all CPGs about FeNO a methodologist was involved, and in none of the CPGs about CRP. Patient involvement and inclusion of patient perspective varied a lot between the CPGs. AGREE II methodology domain scores varied from 8 to 42 (possible range from worst to best: 6-42), with the highest scores for the CPGs about FeNO. Thirteen of the included recommendations were in favour of the test of interest, only one recommendation about CRP,²² and one recommendation about FeNO,²⁸ advised against the use of the test. Eleven recommendations were graded, which included all recommendations about CRP,¹⁹⁻²² two out of five recommendations about colonoscopy,^{25,26} and five out of six recommendations about FeNO.²⁸⁻³² Seven CPGs reported to have used the GRADE approach^{20,21,24,28-30}; in four of these elements of the GRADE approach (such as a GRADE evidence profile) were recognized.^{24,28-30} No clear differences between the topics were identified in the (reported) use of the GRADE approach.

3.3 | Supporting evidence for the recommendations

Detailed information about the supporting evidence for the included recommendations is presented in Table 3. Ten CPGs out of 15 considered diagnostic accuracy,^{20-22,24,26-30,32} of which four underpinned these considerations with a systematic review of the literature and a judgement of the certainty in the evidence.^{21,28-30} Burden of the test was considered in three CPGs,^{24,27,29} and two CPGs considered the natural course of the disease,^{19,32} all without systematically reviewing the literature. Three CPGs considered treatment effectiveness,^{19,25,28} of which one performed a systematic review of the literature with

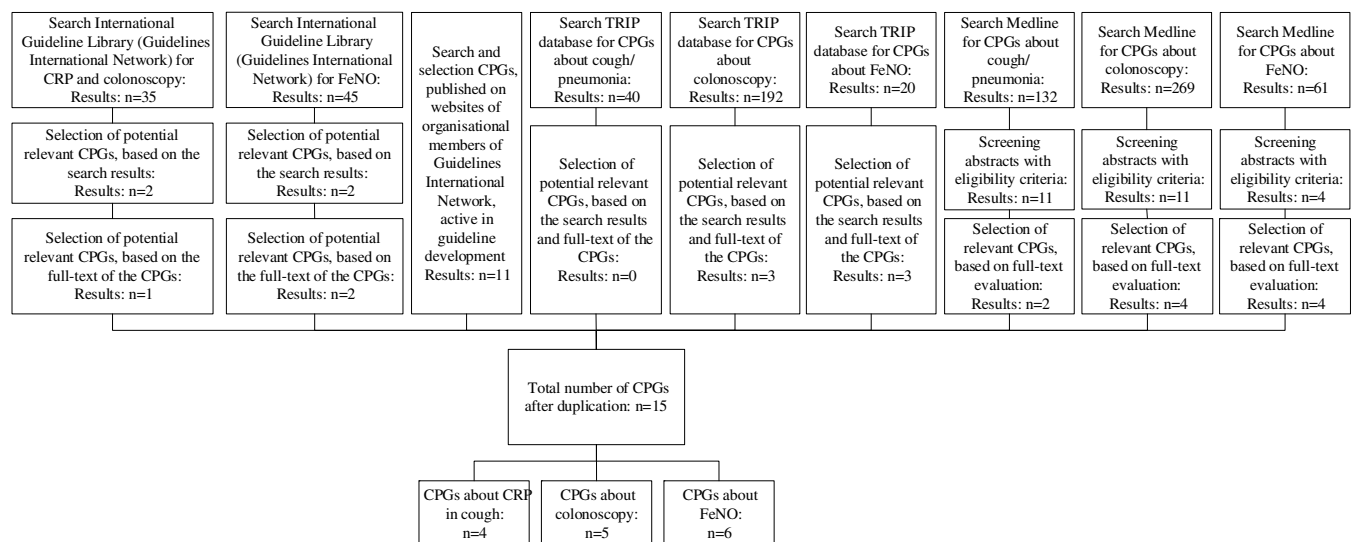


FIGURE 2 Search and selection of relevant CPGs

TABLE 1 Basic characteristics of the included CPGs

Organization	Year	Country	Title (original language)	English-translated title in case of non-English language CPG
CRP				
Deutschen Gesellschaft für Pneumologie und Beatmungsmedizin (DGPB) ²⁰	2016	Germany	Behandlung von erwachsenen Patienten mit ambulant erworbener Pneumonie und Prävention	Prevention and management of adult patients with community acquired pneumonia
American College of Chest Physicians (ACP) ²¹	2019	United States	Adult Outpatients with acute cough due to suspected pneumonia or influenza	
Ministry of Public Health, Qatar (MoPH) ¹⁹	2019	Qatar	The diagnosis and management of community acquired pneumonia	
Deutschen Gesellschaft für Pädiatrische Infektiologie (DGPI) ²²	2017	Germany	Management der ambulant erworbenen Pneumonie bei Kindern und Jugendlichen (pädiatrische ambulant erworbene Pneumonie, pCAP)	Management of community acquired pneumonia in children and adolescents
Colonoscopy				
European Society for Medical Oncology (ESMO) ²⁵	2020	Europe	Localized colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up	
Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF) ²³	2019	Germany	Kolorektales Karzinom	Colorectal cancer
Association of Coloproctology of Great Britain and Ireland (ACPGBI) ²⁶	2017	Great Britain and Ireland	Guidelines for the Management of Cancer of the Colon, Rectum and Anus (2017) - Diagnosis, Investigations and Screening	
Federatie Medisch Specialisten (FMS) ²⁴	2019	The Netherlands	Colorectaal carcinoom	Colorectal cancer
Nederlands Huisartsen Genootschap (NHG) ²⁷	2017	The Netherlands	Rectaal bloedverlies	Rectal bleeding
FeNO				
Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF) ²⁸	2020	Germany	Nationale VersorgungsLeitlinie Asthma	National Guideline on asthma
Ministry of Public Health (MoPH_A) ³¹	2019	Qatar	The diagnosis and management of asthma in adults	
Ministry of Public Health (MoPH_C) ³²	2019	Qatar	The diagnosis and management of asthma in children	
National Asthma Education and Prevention Program (NAEPP) ²⁹	2020	USA	Managing Asthma in Adolescents and Adults	
National Institute for Health and Care Excellence (NICE) ³⁰	2020	UK	Asthma: diagnosis, monitoring and chronic asthma management	
Scottish Intercollegiate Guidelines Network (SIGN) ³³	2019	UK	British guideline on the management of asthma	

judgement of the certainty in the evidence.²⁸ Not any CPG considered the link between the test result and administration of treatment. As a consequence, there were no CPGs that considered all test consequences of the test-treatment pathway.

Since no CPG systematically evaluated all steps of the test-treatment pathway, we were not able to identify a best practice, nor could we study possible relationships between clarifying factors and supporting evidence for a recommendation.

TABLE 2 Detailed characteristics of the CPG quality, the recommendation and the (reported) use of GRADE

Topic	Guideline	CPG panel				Methodology (AGREE II scores)							Recommendation		GRADE	Elements of GRADE approach in CPG
		Methodologist involvement	Patient involvement (AGREE II score)	Systematic evidence search methods	Clear criteria for evidence selection	Clear description of the strengths and limitations of the body of evidence	Clear description methods for formulating recommendations	Health benefits, side effects and risks have been considered	Explicit link between recommendation and supporting evidence	SUM score methodology domain [†]	Direction	Graded	Reported use of GRADE approach			
CRP	DGPB, 2016 ²⁰	-	2	4	2	3	6	3	5	23	+	+	+	+	-	
	ACP, 2019 ²¹	-	5	6	5	2	6	2	6	27	+	+	+	+	-	
	MoPH, 2019 ¹⁹	-	1	2	1	2	1	1	1	8	+	+	+	-	-	
	DGPI, 2017 ²²	-	2	3	1	1	7	5	2	19	-	+	+	-	-	
Colonoscopy	ESMO, 2020 ²⁵	-	1	1	1	1	3	1	2	9	+	+	+	+	-	
	AWMF, 2019 ²³	+	3	7	7	5	7	3	6	35	+	-	-	-	-	
	ACPGI, 2017 ²⁶	-	1	2	1	3	2	1	4	13	+	+	+	-	-	
	FMS, 2019 ²⁴	+	5	2	1	2	4	6	7	22	+	-	-	+	+	
	NHG, 2017 ²⁷	+	1	5	1	1	5	6	7	25	+	+	-	-	-	
	AWMF, 2020 ²⁸	+	7	7	7	7	7	6	7	41	-	+	+	+	+	
FeNO	MoPH_A, 2019 ³¹	+	3	4	4	2	2	2	2	16	+	-	-	-	-	
	MoPH_C, 2019 ³¹	+	3	3	3	2	2	2	2	14	+	+	+	-	-	
	NAEPP, 2020 ²⁹	+	7	7	7	7	7	7	7	42	+	+	+	+	+	
	NICE, 2020 ^{30†}	+	7	7	7	7	6	7	7	41	+	+	+	+	+	
	SIGN, 2019 ³²	+	7	7	3	3	4	5	7	29	+	+	+	+	-	

Note: +, yes; +/-, unclear; -, no; [†]possible range: 6-42; [‡]This CPG contains two separate recommendations concerning the use of FeNO in the diagnosis of childhood respectively adult-onset asthma; the scores are identical.

TABLE 3 Detailed information about the supporting evidence for the recommendations

Topic	Guideline	Diagnostic accuracy		Burden of the test		Natural course		Treatment effectiveness		Link between test result and administration of treatment
		Considered with systematic review of the literature	Considered with systematic review of the literature and certainty in the evidence	Considered with systematic review of the literature	Considered with systematic review of the literature and certainty in the evidence	Considered with systematic review of the literature	Considered with systematic review of the literature and certainty in the evidence	Considered with systematic review of the literature	Considered with systematic review of the literature and certainty in the evidence	
CRP	DGPB, 2016 ²⁰	-	-	-	-	-	-	-	-	-
	ACP, 2019 ²¹	+	+	-	-	-	-	-	-	-
	MoPH, 2019 ¹⁹	-	-	-	+	-	-	+	-	-
	DGPI, 2017 ²²	+	-	-	-	-	-	-	-	-
	ESMO, 2020 ²⁵	-	-	-	-	-	-	+	-	-
Colonoscopy	AWMF, 2019 ²³	-	-	-	-	-	-	-	-	-
	ACPGI, 2017 ²⁶	+	-	-	-	-	-	-	-	-
	FMS, 2019 ²⁴	+	-	+	-	-	-	-	-	-
	NHG, 2017 ²⁷	+	-	+	-	-	-	-	-	-
	AWMF, 2020 ²⁸	+	+	-	-	-	-	+	+	-
FeNO	MoPH_A, 2019 ³¹	-	-	-	-	-	-	-	-	-
	MoPH_C, 2019 ³¹	-	-	-	-	-	-	-	-	-
	NAEPP, 2020 ²⁹	+	+	+	-	-	-	-	-	-
	NICE, 2020 ^{30†}	+	+	-	-	-	-	-	-	-
	SIGN, 2019 ³²	+	-	-	-	+	-	-	-	-

Note: +, yes; +/-, unclear; -, no; †This CPG contains two separate recommendations concerning the use of FeNO in the diagnosis of childhood respectively adult onset asthma; the scores are identical.



4 | DISCUSSION

Our document analysis on a sample of 15 CPGs about CRP, colonoscopy and FeNO diagnostic tests revealed that none of these CPGs reported evidence on all components of the test-treatment pathway. Consideration of any test consequences on patient-relevant outcomes was described in only six CPGs (three CPGs considered burden of the test, two considered natural course of the disease of interest, and four considered treatment effectiveness). Systematic review of the literature, including a judgement of the certainty in the supporting evidence was only reported for four recommendations and covered diagnostic accuracy in all four cases and treatment effectiveness in one case.

The importance of systematically evaluating test consequences for the purpose of developing CPGs has been recognized.^{7,12,13} For instance, one could imagine that a certain diagnostic test might have limited value when it has no treatment consequences (e.g., no treatment available). Or when comparing two tests with the same diagnostic accuracy to ascertain the same disease, one could recognize that differences in test burden may play an important role.

This study suggests that implementation of the systematic evaluation of the value of a test is lagging behind. This also applies to CPGs that claim to use the GRADE approach. There seems to be a gap between following a methodologically robust approach and developing CPGs in practice.

Two issues may explain that gap. First, guideline developers may have considered the downstream consequences of a diagnostic test but did not explicitly report these. It may not be strictly necessary to systematically evaluate all evidence components. However, we still recommend transparent documentation of choices made in the guideline development process. A guideline user should be able to read which elements of a test-treatment pathway were considered and how, and which were not considered and why.

Second, performing systematic literature reviews of the complete test-treatment pathway—including assessment of the certainty in the evidence of test accuracy and downstream consequences—is complex and time-consuming. The use of the GRADE approach for the evaluation of diagnostic tests and test strategies is considered challenging.^{10,11} Strategies to facilitate the use of this approach, such as training of CPG panel members, may improve the application. Unfortunately, we could not determine factors that contribute to successful use of the GRADE approach, because we could not identify a “best practice.”

A lack of transparency in combination with the use of state-of-the-art methods was also described by Arevalo-Rodríguez and colleagues, who studied the methods and reports of 191 rapid reviews of medical tests.³⁴ In the majority of those reviews, the study selection method was not reported. Although almost 20% of the reviews claimed to have applied the GRADE approach, few actually reported the data extraction and quality appraisal methods.

This finding is consistent with a recent report on the application of GRADE in US guidelines.³⁵ Although guideline developers indicated that they used the GRADE approach, only 10% of the included CPGs

reported on all eight criteria for assessing the certainty in the evidence (e.g., indirectness and dose-response gradient), and around half of these included an evidence profile or summary of findings table.

Gopalakrishna et al. studied barriers in the development of recommendations about medical tests in a qualitative study among European CPG developers.³⁶ They also reported challenges in the development of recommendations about medical tests, for example, in the definition of key questions, the types of evidence and outcomes included in the CPG, and synthesizing and appraising the evidence. Awareness and education were reported as the most important ways to solve these challenges.

Our study emphasizes the need for more knowledge and expertise among CPG developers when evaluating diagnostic tests. Currently available competency-based frameworks for CPG developers do not include a special focus on diagnostic test evaluation.^{37,38} This also applies for current training programs of CPG panel members, for example, INGUIDE.³⁹ Facilitating the implementation of GRADE for diagnosis by defining competencies and training needs may improve the quality of CPGs about diagnostic tests.

4.1 | Strengths and limitations

This study evaluated the supporting evidence of recommendations in CPGs on three medical tests. The selection of only three topics is a limitation of this study. However, we chose three diagnostic tests with divergent characteristics (e.g., invasiveness, possible burden of the test, disease of interest, costs) allowing comparison of many CPGs. The homogenous results in all three clusters of CPGs strengthens the external validity of our findings. Additionally, we found large variance in methodological quality of the included CPGs. However, high-scoring CPGs on the AGREE II domain methodology did not reflect a better or more transparent underpinning of the recommendations than lower scoring CPGs.

Due to the document analysis design, we could not retrieve information about the dynamics in the CPG panels that could explain their decisions and reasons for lack of transparency in the CPG documents. We did not contact the CPG developers, since in our opinion CPG users should be able to find the considerations of the panel beyond the recommendations in the published documents of the CPG.

4.2 | Implications for practice

We suggest that developers of CPGs about diagnostic tests clearly describe which elements of a test-treatment pathway were or were not considered and why. In addition, CPG developers should indicate the presence or absence of systematic reviews of the evidence, including determination of the certainty in that evidence, for all evaluated parts of the test-treatment pathway, which is also usual in recommendations about therapy. Facilitating the implementation of GRADE for diagnosis will be useful to improve the clinical content of CPGs.

4.3 | Implications for research

This study highlighted the lack of (transparency about) supporting evidence for diagnostic test recommendations in CPGs. A next step could be to study why CPG developers do not report all elements of the test-treatment pathway, including a review of the evidence and its quality. Furthermore, it is worthwhile to research how to facilitate CPG developers in explicitly and reliably considering all relevant steps of a test-treatment pathway when developing diagnostic test recommendations.

5 | CONCLUSION

Diagnostic test recommendations in the included CPGs are mainly based on evidence and considerations on diagnostic accuracy. Other steps of the test-treatment strategy, such as burden of the test, natural course of the disease of interest, effectiveness of treatment of the disease of interest and the link between the test result and the administration of treatment should receive more attention in CPGs in order to consider evidence about test consequences on patient-relevant outcomes.

CONFLICT OF INTEREST

Mariska Tuut and Miranda Langendam are members of the GRADE Working Group. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHORS' CONTRIBUTIONS

Mariska K Tuut conceptualized the study, including methodology, extracted, analysed, and interpreted the data, and wrote the original draft of the manuscript. Jako S Burgers, Trudy van der Weijden and Miranda W Langendam were involved in the conceptualization of the study, including methodology, performed data checks, reviewed and edited the manuscript and had a supervising role. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available (see Reference list).

ORCID

Mariska K. Tuut  <https://orcid.org/0000-0002-0880-6215>

REFERENCES

- Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ*. 2001;323(7305):157-162.
- Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-based medicine working group. *Jama*. 1994;271(5):389-391.
- Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64(4):383-394.
- Sackett DL, Haynes RB. The architecture of diagnostic research. *BMJ*. 2002;324(7336):539-541.
- Harris RP, Helfand M, Woolf SH, et al. Current methods of the US preventive services task force: a review of the process. *Am J Prev Med*. 2001;20(3 Suppl):21-35.
- Schünemann HJ, Oxman AD, Brozek J, et al. GRADE: assessing the quality of evidence for diagnostic recommendations. *Evid Based Med*. 2008;13(6):162-163.
- Schünemann HJ, Oxman AD, Brozek J, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ*. 2008;336(7653):1106-1110.
- Florez ID, Brouwers MC, Kerkvliet K, et al. Assessment of the quality of recommendations from 161 clinical practice guidelines using the appraisal of guidelines for research and evaluation-recommendations excellence (AGREE-REX) instrument shows there is room for improvement. *Implement Sci*. 2020;15(1):79.
- 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):282-287.
- Gopalakrishna G, Mustafa RA, Davenport C, et al. Applying grading of recommendations assessment, development and evaluation (GRADE) to diagnostic tests was challenging but doable. *J Clin Epidemiol*. 2014;67(7):760-768.
- Tuut M, de Beer H, Burgers J, van de Griendt EJ, van der Weijden T, Langendam M. Applying GRADE for diagnosis revealed methodological challenges: an illustrative example for guideline developers. *J Clin Epidemiol*. 2020;131:123-132.
- Schünemann HJ, Mustafa R, Brozek J, et al. GRADE guidelines: 16. GRADE evidence to decision frameworks for tests in clinical practice and public health. *J Clin Epidemiol*. 2016;76:89-98.
- Schünemann HJ, Mustafa RA, Brozek J, et al. GRADE guidelines: 22. The GRADE approach for tests and strategies - from test accuracy to patient important outcomes and recommendations. *J Clin Epidemiol*. 2019;111:69-82.
- Lim WS, Baudouin SV, George C, et al. Pneumonia guidelines committee of the BTS standards of care committee. BTS guidelines for the management of community acquired pneumonia in adults: update 2009. *Thorax*. 2009;64(3):iii1-iii55.
- International Agency for Research on Cancer. World Health Organisation. Cancer Today: Colon. 2021. <https://gco.iarc.fr/today/data/factsheets/cancers/8-Colon-fact-sheet.pdf>.
- Soriano JB, Kendrick PJ, Paulson KR, et al. Prevalence and attributable health burden of chronic respiratory diseases, 1990-2017: a systematic analysis for the global burden of disease study 2017. *Lancet Respir Med*. 2020;8(6):585-596.
- Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ*. 2010;182(18):E839-E842.
- AGREE Next Steps Consortium. *The AGREE II Instrument [Electronic Version]*, 2017. <http://www.agreetrust.org>.
- Ministry of Public Health, Doha, Qatar. The diagnosis & management of community acquired pneumonia, 2019. <https://www.moph.gov.qa/english/OurServices/eservices/Pages/Clinical-Guidelines.aspx>.
- Ewig S, Höffken G, Kern WV, Rohde G, Flick H, Krause R, et al. Behandlung von erwachsenen Patienten mit ambulant erworbener Pneumonie und Prävention. Bochum: Deutschen Gesellschaft für Pneumologie und Beatmungsmedizin, 2016 <https://www.awmf.org/leitlinien/aktuelle-leitlinien/II-liste/deutsche-gesellschaft-fuer-pneumologie-und-beatmungsmedizin-dgp.html>.
- Hill AT, Gold PM, El Solh AA, et al. Adult outpatients with acute cough due to suspected pneumonia or influenza: CHEST guideline and expert panel report. *Chest*. 2019;155(1):155-167.
- Rose MA, Barker M, Liese J, et al. S2k-Leitlinie Management der ambulant erworbenen Pneumonie bei Kindern und Jugendlichen

- (pädiatrische ambulant erworbene Pneumonie, pCAP) (Guidelines for the management of community acquired pneumonia in children and adolescents [pediatric community acquired pneumonia, pCAP] - Issued under the responsibility of the German Society for Pediatric Infectious Diseases [DGPI] and the German Society for Pediatric Pulmonology [GPP]). *Pneumologie*. 2020;74(8):515-544.
23. Leitlinienprogramm Onkologie der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V., Deutschen Krebsgesellschaft e.V. und Deutschen Krebshilfe. S3-Leitlinie Kolorektales Karzinom, 2019. https://www.awmf.org/uploads/tx_szleitlinien/021-007OLI_S3_Kolorektales-Karzinom-KRK_2019-01.pdf.
 24. Federatie Medisch Specialisten. Colorectal carcinoom, 2019. https://richtlijndatabase.nl/richtlijn/colorectal_carcinoom_crc/startpagina_-_crc.html.
 25. Argiles G, Taberner J, Labianca R, et al. Localised colon cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2020;31(10):1291-1305.
 26. Cunningham C, Leong K, Clark S, et al. Association of coloproctology of Great Britain & Ireland (ACPGBI): guidelines for the management of cancer of the colon, rectum and anus (2017) - diagnosis, investigations and screening. Practice guideline. *Colorectal Dis*. 2017;19(1):9-17.
 27. Nederlands Huisartsen Genootschap. NHG Standaard Rectaal bloedverlies (M89), 2017. <https://www.nhg.org/standaarden/samenvatting/rectaal-bloedverlies>.
 28. Bundesärztekammer, Kassenärztlicher Bundesvereinigung, Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. Nationale VersorgungsLeitlinie (NVL) Asthma, 2020. <https://www.leitlinien.de/themen/asthma>.
 29. Cloutier MM, Dixon AE, Krishnan JA, Lemanske RF Jr, Pace W, Schatz M. Managing asthma in adolescents and adults: 2020 asthma guideline update from the national asthma education and prevention program. *Jama*. 2020;324(22):2301-2317.
 30. National Institute for Health and Care Excellence Asthma: diagnosis and monitoring of asthma in adults, children and young people, 2017. <https://www.nice.org.uk/guidance/ng80>.
 31. Ministry of Public Health, Doha, Qatar National clinical guideline: The diagnosis and management of asthma in adults; 2019. <https://www.moph.gov.qa/english/OurServices/eservices/Pages/Clinical-Guidelines.aspx>
 32. Ministry of Public Health, Doha, Qatar. National clinical guideline: The diagnosis and management of asthma in children; 2019. Accessed March 13, 2021. <https://www.moph.gov.qa/english/OurServices/eservices/Pages/Clinical-Guidelines.aspx>.
 33. Scottish Intercollegiate Guidelines Network SIGN 158. British guideline on the management of asthma, 2019. <https://www.sign.ac.uk/sign-158-british-guideline-on-the-management-of-asthma>.
 34. Arevalo-Rodriguez I, Moreno-Nunez P, Nussbaumer-Streit B, et al. Rapid reviews of medical tests used many similar methods to systematic reviews but key items were rarely reported: a scoping review. *J Clin Epidemiol*. 2019;116:98-105.
 35. Dixon C, Dixon PE, Sultan S, et al. Guideline developers in the United States were inconsistent in applying criteria for appropriate GRADE use. *J Clin Epidemiol*. 2020;124:193-199.
 36. Gopalakrishna G, Leeflang MM, Davenport C, et al. Barriers to making recommendations about medical tests: a qualitative study of European guideline developers. *BMJ Open*. 2016;6(9):e010549.
 37. Norris SL, Meerpohl JJ, Akl EA, et al. The skills and experience of GRADE methodologists can be assessed with a simple tool. *J Clin Epidemiol*. 2016;79:150-158.
 38. Sultan S, Morgan RL, Murad MH, et al. A theoretical framework and competency-based approach to training in guideline development. *J Gen Intern Med*. 2020;35(2):561-567.
 39. Guidelines International Network and McMaster University INGUIDE. International Guideline Development Credentialing & Certification Program. 2021. <https://inguide.org/>.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Tuut MK, Burgers JS, van der Weijden T, Langendam MW. Do clinical practice guidelines consider evidence about diagnostic test consequences on patient-relevant outcomes? A critical document analysis. *J Eval Clin Pract*. 2022;28:278-287. <https://doi.org/10.1111/jep.13619>