Few prostate cancer guidelines are of high methodologic quality and therefore trustworthy

Farhad Tondroanamag,[1] Daniel A. Gonzalez-Padilla,[2] Caitlin Bakker[3], Philipp Dahm[4,5]
[1]Faculty of Medicine, University of Helsinki, Helsinki, Finland; [2]Department of Urology, Clínica Universidad de Navarra, Madrid, Spain; [3]Dr. John Archer Library, University of Regina, Regina, SK, Canada; [4]Minneapolis Veterans Affairs Health Care System, Urology Section, Minneapolis, MN, United States; [5]University of Minnesota, Department of Urology, Minneapolis, MN, United States

Cite as: Tondroanamag F, Gonzalez-Padilla DA, Bakker C, et al. Few prostate cancer guidelines are of high methodologic quality and therefore trustworthy. *Can Urol Assoc J* 2026 August 20; Epub ahead of print. http://dx.doi.org/10.5489/cuaj.9674

Published online August 20, 2026

Corresponding author: Dr. Philipp Dahm, University of Minnesota, Department of Urology, Minneapolis, MN, United States; pdahm@umn.edu

***

ABSTRACT

KEY MESSAGES

- Most prostate cancer guidelines published in English do not fulfill criteria of the National Academy of Medicine for trustworthy, high-quality guidelines.
- Only six guidelines were rated as high-quality, and over one-third (8/23) of guidelines were categorized as low -uality.
- Year of publication and breadth of guideline scope were not associated with guideline quality.

Introduction: Prostate cancer is the most common malignancy among men worldwide, underscoring the importance of high-quality, evidence-based clinical practice guidelines (CPGs) in this field. We conducted this study to evaluate the methodologic quality of prostate cancer CPGs.

Methods: Following a prespecified protocol, we conducted a cross-sectional study of English-language prostate cancer CPGs published between 2014 and 2024, searching Medline (Ovid), Trip, EMBASE, ECRI, Guidelines International Network, and Guidelines Central databases. We included guidelines covering screening, diagnosis, and/or management of prostate cancer. Two reviewers independently applied the modified National Guideline Clearinghouse Extent of Adherence to Trustworthy Standards (NEATS) instrument, consisting of 10 items and generating summary scores from 10 (lowest) to 50 (highest) points. Guidelines were categorized as low- (≤16), intermediate- (17–31), or high-quality (≥32).

Results: Of 1005 references screened, 23 international guidelines met the inclusion criteria. Only six CPGs (26%) were rated high-quality; over one-third (8/23, 34.7%) were categorized as low-quality. Most CPGs inadequately reported conflicts of interest, multidisciplinary panel composition, or systematic review methods underpinning recommendations. No significant

associations were observed between guideline quality and year of publication or breadth of scope.
Conclusions: Most prostate cancer CPGs published in English demonstrate limited methodologic rigor, not meeting National Academy of Medicine standards for trustworthy guidelines. Guideline users should critically appraise these documents prior to implementation, and developers are encouraged to adopt established frameworks, such as GRADE, to improve quality.

INTRODUCTION

Prostate cancer is the leading cancer in men across 112 countries, representing 15% of all male cancers worldwide. The annual number of new cases is expected to increase from 1.4 million in 2020 to 2.9 million by 2040.[1] Ongoing advances and substantial research continue to drive rapid changes in the diagnosis and management of prostate cancer.

Clinical Practice Guidelines (CPGs) play a central role in healthcare worldwide by promoting evidence-based decision-making, standardizing clinical practice, enhancing patient safety, and improving the overall quality and efficiency of care. Developing a high-quality and trustworthy guideline is a time-consuming and resource-intensive task; it requires rigorous methodologies, including systematic literature reviews, critical appraisal of the evidence, transparent reporting, and the formulation of clear and actionable recommendations, all of this performed by a team with expertise in CPGs.[2, 3] Given the exponential growth in the number of CPGs published globally—often by diverse organizations using different methods—the need to evaluate their methodological quality has become increasingly relevant. Poorly developed CPGs not only risk delivering suboptimal or even harmful recommendations but also undermine trust in the guideline itself.

In this paper, we aimed to systematically appraise the most current CPGs on prostate cancer using the modified National Guideline Clearinghouse Extent of Adherence to Trustworthy Standards (NEATS) instrument,[4] which offers improved reproducibility compared to the original instrument,[5] which was developed to assess adherence to the Institute of Medicine (now: National Academy of Medicine) standards for trustworthy guidelines.[3]

METHODS

We conducted a cross-sectional study of published CPGs retrieved after a comprehensive systematic search of the literature including Medline (Ovid), Trip, EMBASE, ECRI, Guidelines International Network, and Guidelines Central databases. A medical information specialist developed the search strategies (see supplemental materials #1). We included CPGs available in the English language only. Two members of the research team (FT and DAGP) independently screened eligible documents using the Rayyan platform at both a title/abstract and full-text stage.[6] We established consensus as to which guidelines to include through discussion and consensus; if needed, through arbitration by a third member (PD) of the

investigative team. We included all documents that self-identified as guidelines that addressed any aspect of prostate cancer management including screening, diagnosis, and treatment of prostate cancer (either localized, advanced, or metastatic). We included only one CPG per developer; in cases in which there were multiple prostate cancer guidelines available from one organization, we chose the latest one. The following documents were excluded: draft-only guidelines, guidelines specifically directed towards technical details on diagnosis or treatment (such as those on nuclear medicine tests or anatomic pathology processing). We also excluded CPGs with a narrow focus on the management of prostate cancer treatment complications (urinary incontinence, sexual dysfunction, bone health, etc.), those developed for emergent purposes like prostate cancer management during the COVID-19 pandemic, and such CPGs directed towards genetic testing related to prostate cancer-associated genes. We also excluded documents that self-identified as "appropriateness criteria" or "consensus documents" without claiming to be a guideline.

We applied the modified NEATS instrument to assess the CPGs,[4] which consists of ten items, of which two are categorical items (yes/no/unknown) and eight are scored on a 5-point Likert scale (from one to five with larger value representing greater methodological rigor). The maximum score possible was 50 points (40 from the eight questions with 5-point Likert points and 10 from the yes or no/unknown categorical questions which were given either 5 or 1 points respectively).

The modified NEATS instrument was applied at a guideline level for those CPGs that contained 20 or fewer recommendations, when guidelines contained more than 20 recommendations, a representative subset was evaluated. This approach was adopted because only two of the ten items (#4c: synthesis of evidence for each recommendation and #5: strength of recommendation for each recommendation) of the modified NEATS instrument assess elements at the recommendation level, and most clinical practice guidelines included had fewer than 30 recommendations. Accordingly, evaluating up to 30 recommendations per guideline was considered sufficient to ensure representativeness. The study protocol was prospectively registered in the Open Science Framework (protocol available at: https://osf.io/xp3cr).

We performed descriptive statistical analyses and non-parametric testing using SPSS Statistics (version 29.0; IBM Corp., Armonk, NY, USA). Post-hoc based on the respective variable distribution, we categorized CPGs by the year of guideline release (2014-2020 versus 2021-2024) and its scope as narrow (addressing screening, clinically localized or advanced prostate cancer) versus broad (covering two or more topics). We also compared original versus updated CPGs as to their methodological rigor. In the absence of pre-established thresholds to help readers interpret the results, we categorized the modified NEATS summary score into three categories as follows: 16 or less points = low, 17 - 31 points = intermediate, 32 or more points = high quality. We used R software (version 4.4.2; R Core Team, 2024; R Foundation for Statistical Computing) to generate figures and calculate Cohen's kappa as a measure of interobserver agreement.

RESULTS

Our search identified a total of 1,005 references; once duplicates were removed, there were 844 references; Figure 1), which we then screened in two stages (title/abstract and full text). We excluded six documents at the full-text stage (see supplementary materials #2) ultimately including a total of 23 international CPGs published between 2014 and 2024 (Table 1). More than half of guidelines (Table 2) were updates of a prior version of the same CPG (n=14; 60.9%); the remainder were original guidelines (n=9; 39.1%). The most common geographic areas of origin were North America and Europe, each with seven CPGs (30.4%). Approximately half had a narrow scope (11; 47.8%). We were only able to clearly discern the actual numbers of recommendations for 18 guidance documents; for these, the median number of recommendations per CPG was 22.5 (interquartile range: 11.3; 47.8) with a range from one to 159.

Figure 2 summarizes the ratings across the ten modified NEATS domains. Of the 23 guidelines, only 60.8% (14/23) reported the funding source for their development. Less than one in three CPGs (30.4%; 7/23) were explicit in stating that none of the panel members had any relevant conflicts of interest (COIs); in three CPGs, potential COI excluded the panel chair and was limited to a third of panelists or less. The remaining 13 guidelines (56.5%) did not address potential COIs. The panel composition was multidisciplinary as defined by including clinicians, methodologists, and at least one patient representative in 26% of CPGs (6/23); in 10 CPGs (43.4%) the panel composition was unclear.

Three questions of the modified NEATS instrument address the search strategy (item 4a), study selection (item 4b) and evidence synthesis (item 4c) of the underlying systematic review: Only 26% (6/23), 17.3% (4/23) and 17.3% (4/23) of CPGs, respectively, fully met methodological expectations set by the modified NEATS instrument. Frequently, it was not discernible whether the search strategy matched the underlying question for a given recommendation (65%; 15/23), unambiguous eligibility criteria were missing (73.9%; 17/23) and/or recommendations lacked a pooled effect estimate for each outcome (or when appropriate, a narrative description) as was the case in two-thirds of CPGs (69.5%; 16/23). Approximately one-third of guidelines issued statements that did not explicitly indicate the strength of recommendation (39.1%; 9/23) and another one-third (34.7%; 8/23) failed to discuss both benefit and harm outcomes relevant to the recommendation. Only one quarter of CPGs included evidence summaries that justified the ‘weak’ (or ‘conditional’) as the strength of recommendation. About one in four guideline documents (43.4%; 10/23) issued recommendations that did not explicitly specify population, intervention, and comparator (unless this was obvious). Only two CPGs (8.6%) used prespecified standardized wording that was specific to the strength of recommendation. Most CPGs provided no information about having undergone external peer review (65.2%; 15/23), and only about half (56.5%; 13/23) indicated that there were plans to update the guideline in the future. Kappa values ranged from 0.64 to 0.90, indicating substantial to almost perfect agreement.

Overall, only six among 23 prostate cancer CPGs (26.1%) were categorized as having a high level of trustworthiness, and eight (34.7%) were rated being of low trustworthiness.

We found no statistical association between time of publication, breadth of scope, or version status of the CPGs with the modified NEATS instrument score.

DISCUSSION

Statement of principal findings

This study is the first to apply the modified NEATS instrument to guidelines on prostate cancer, this instrument by Ghadimi et al.[4] was chosen due to its high-reproducibility and limited room for variability in the interpretation by each rater thanks to its algorithmic approach, contrary to the AGREE II tool, which among its limitations, allows for a higher level of subjectivity on each topic evaluated. Our central finding is that most international prostate cancer CPGs published in English language do not fulfill criteria of the National Academy of Medicine (formerly Institute of Medicine) for what constitutes a trustworthy, high-quality guideline. Approximately one in four guidelines were rated as high quality, and over a third were categorized as low quality. We found no association between time of publication, scope, and version with guideline quality.

Strengths and weaknesses of the study

Strengths of our study include a pre-specified, publicly available a priori protocol, execution by a team or authors with extensive experience in CPG appraisal and development, as well as the involvement of an information specialist who planned and conducted a comprehensive search across multiple databases. We used the modified NEATS instrument that offers clear guidance for scoring CPGs for each criterion and thereby facilitates scoring and improves interobserver agreement.[4] We made sure to include all publicly available supplemental materials and contacted authors if information provided was missing or unclear. A limitation of this study is the focus on English language CPGs only; we recognize that there are likely many other guidelines in major languages such as Mandarin or Spanish that were outside the scope of this investigation. We also faced challenges in retrieving supplemental information for some of the older guidelines due to broken links and the inability to reach the authors; the missing information may have contained additional information that could possibly improve their scoring. We further appreciate that a modified NEATS summary score may be considered a crude assessment of overall guideline quality. However, Figure 2 provides detailed information about the scoring of each domain for each CPG, thereby allowing a more nuanced interpretation. Lastly, we acknowledge the post-hoc nature of the statistical hypothesis testing performed, as well as the small sample size of this study. Another limitation is that not all prostate cancer CPGs published by the included organizations were evaluated; consequently, earlier versions of these guidelines may have achieved higher scores. Therefore, the scores reported in this manuscript do not represent the entirety of CPGs produced by each organization but should instead be interpreted as reflecting a representative sample of the available guidelines. Notwithstanding these limitations, we believe that our findings have meaningful implications.

Strengths and weaknesses in relation to other studies, discussing important differences in results

Given that our study is the first to apply the modified NEATS instrument to guidelines in prostate cancer, it lacks direct comparisons; however, similar studies assessing methodological quality have been conducted. One study has applied this instrument to assess 70 randomly selected CPGs from 2010 and 70 from 2022, calculating a mean score for individual items and across items scored on a 1 – 5-point Likert scale with higher values representing greater adherence. They found that mean scores improved by 0.46 (95% CI: 0.13-0.79) from 2.28 to 2.74. The authors concluded that while there was improvement, the magnitude of this positive change was small. Overall, they judged overall adherence to be moderate and noted that less than one in five CPGs were of high trustworthiness, compared to approximately one of four in our study.

An early formal assessment of prostate cancer guidelines from 2008 identified eight CPGs using the National Guidelines Clearing House and applied the original AGREE instruments for the guideline evaluation.[7] Its main findings related to the considerable heterogeneity as to how different organizations rated the certainty of evidence and developed their recommendations; in its conclusions, it called for greater harmonization of guideline methodology. Subsequent work by the same group of investigators published a similar analysis in 2015 and identified thirteen guidelines using the same resource which they evaluated using the AGREE II tool.[8] Main findings were the variable quality of CPGs in this space with most falling short of established standards for guideline methodology, in particular, regarding applicability and stakeholder involvement. Compared to these two studies, we conducted a more extensive search for eligible guidelines, in part motivated by the fact that the National Guideline Clearing House no longer exists.[9] Like prior studies, we found that only few CPGs meet established standards for methodological rigor and therefore represent guidelines that users, be it individual healthcare providers and their patients, or policy-makers, should trust. Another persistent issue appears to be that of a fragmentation, with guideline developers using a variety of methodological approaches, which can be confusing for users and likely hinders collaboration.[10]

Our findings stand in contrast to a study by a group of investigators affiliated with the European Association of Urology who assessed 16 prostate cancer CPGs using the AGREE II tool. The researchers included members from the guideline producing organizations being evaluated in the evaluation process, which may have introduced bias.[11] Further shortcomings of this study were the lack of an a priori study protocol and the limited information that it provided about the search strategy and the inclusion/exclusion criteria. They found most guidelines (13/16: 81.3%) to be of "high quality", which was defined as a guideline scoring at >60% in at least 5 of 6 AGREE II domains. In our opinion, this is too low of a bar. The label of 'high quality' should require an 80% or greater compliance; in addition, this standard should be met consistently across all six domains. In that study, not a single guideline would have been rated as 'high quality' if such more rigorous standards were applied.

Meaning of the study: Possible explanations and implications for clinicians and policymakers

Our study identified many CPGs developed independently from another by 23 different organizations, using different, and mostly less than rigorous methodology. These findings emphasize the importance of approaching all prostate cancer guideline documents with a healthy degree of skepticism (even when developed by influential organizations and authored by distinguished topic experts) and critically appraising them before implementation. Modified NEATS (and potentially AGREE II) should become a mandatory checklist required for guideline developers. Guideline developers tackling the same or overlapping topics should further harmonize their approach regarding both the certainty of evidence rating and the approach with which recommendations are developed. The grading of recommendations assessment development and evaluation (GRADE) system is the most widely used approach and appears the most natural candidate as a lingua franca for guideline developers that will allow different organizations to share resources and build on each other's work.[10] The development of The International Guideline Training and Certification (INGUIDE) training and credentialing resources should aid in this path forward.[12]

Unanswered questions and future research

We have found the modified NEATS instrument to be an easy-to-use tool that should be applied to guidelines in other topic areas. Given the findings of our study on prostate cancer guidelines, there would be value in using this tool to formally assess CPGs in other disease entities. Future studies should also reassess the guidelines on prostate cancer to establish whether there has been interval improvement over time. An even greater priority would be to explore the barriers that prevent guideline developers from adopting higher and ideally unified standards for guideline development and creating mechanisms and incentives to facilitate greater collaboration, improving quality of CPGs, and ultimately, benefiting the society.

CONCLUSIONS

We found that most clinical practice guidelines on prostate cancer published in English language are of low methodological quality and therefore limited trustworthiness. We found no correlation between the year of publication, the number of recommendations or the number of topics covered within the guideline and its overall modified NEATS instrument score. While no guideline was perfect, six guidelines demonstrated a high degree of adherence, demonstrating that it is possible to reach high standards. We hope that findings of this study raise greater awareness for guideline trustworthiness and spurs efforts towards greater methodological rigor.

REFERENCES

1. James ND, Tannock I, N'Dow J, et al. The Lancet Commission on prostate cancer: Planning for the surge in cases. *Lancet* 2024;403:1683-722. https://doi.org/10.1016/S0140-6736(24)00651-2
2. Schunemann HJ, Wiercioch W, Etxeandia I, et al. Guidelines 2.0: Systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ* 2014;186:E123-42. https://doi.org/10.1503/cmaj.131237
3. Institute of Medicine. Clinical practice guidelines we can trust. National Academies Press; 2011.
4. Ghadimi M, Guyatt G, Brignardello-Petersen R. Modifications to the NEATS instrument for more appropriate and reproducible assessment of guidelines trustworthiness. *Clin Public Health Guid* 2024;1:e12025. https://doi.org/10.1002/gin2.12025
5. Jue JJ, Cunningham S, Lohr K, et al. Developing and testing the Agency for Healthcare Research and Quality's National Guideline Clearinghouse Extent of Adherence to Trustworthy Standards (NEATS) instrument. *Ann Intern Med* 2019;170:480-7. https://doi.org/10.7326/M18-2950
6. Ouzzani M, Hammady H, Fedorowicz Z, et al. Rayyan - a web and mobile app for systematic reviews. *Syst Rev* 2016;5:210. https://doi.org/10.1186/s13643-016-0384-4
7. Dahm P, Yeung LL, Chang SS, et al. A critical review of clinical practice guidelines for the management of clinically localized prostate cancer. *J Urol* 2008;180:451-9. https://doi.org/10.1016/j.juro.2008.04.004
8. Gupta M, McCauley J, Farkas A, et al. Clinical practice guidelines on prostate cancer: A critical appraisal. *J Urol* 2015;193:1153-8. https://doi.org/10.1016/j.juro.2014.10.105
9. Munn Z, Qaseem A. Disappearance of the National Guideline Clearinghouse: A huge loss for evidence-based health care. *Ann Intern Med* 2018;169:648-9. https://doi.org/10.7326/M18-2216
10. Dahm P, Chapple CR, Konety BR, et al. The future of clinical practice guidelines in urology. *Eur Urol* 2011;60:72-4. https://doi.org/10.1016/j.eururo.2011.04.007
11. Sakalis V, Bhattacharya Y, Beyer K, et al. AGREE II quality assessment of national and international clinical practice guidelines on prostate cancer management by the OPTIMA consortium. *Eur Urol Open Sci* 2024;70:183-93. https://doi.org/10.1016/j.euros.2024.10.020
12. Schünemann HJ, Nieuwlaat R. The INGUIDE International Guideline Training and Certification Programme. *Clin Public Health Guid* 2024;1:e12008. https://doi.org/10.1002/gin2.12008

FIGURES AND TABLES

Figure 1. Identification of studies via databases and registers

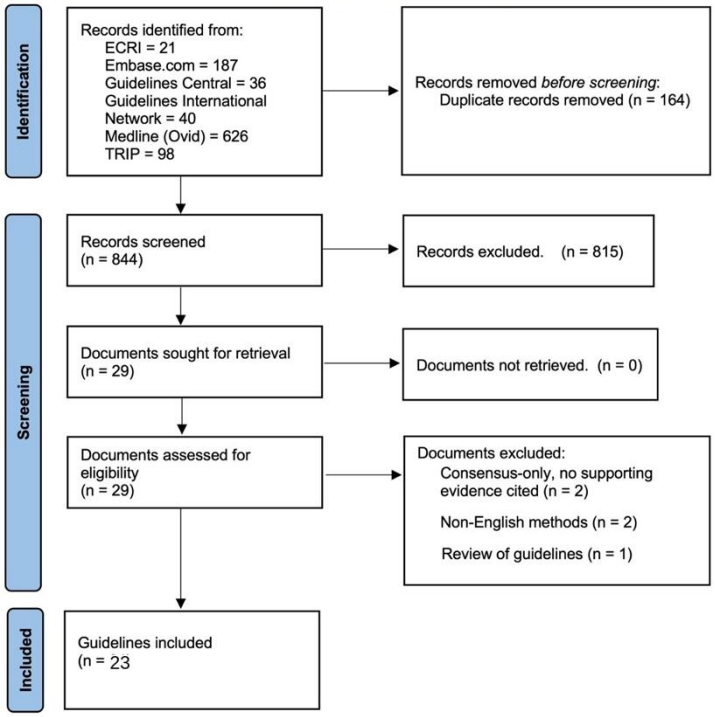

Figure 2. Radar figure depicting the scoring of each clinical practice guideline evaluated, each color represents one item of the modified NEATS instrument, the larger the color the higher the score on each item, the longer the column, the higher the overall scoring.

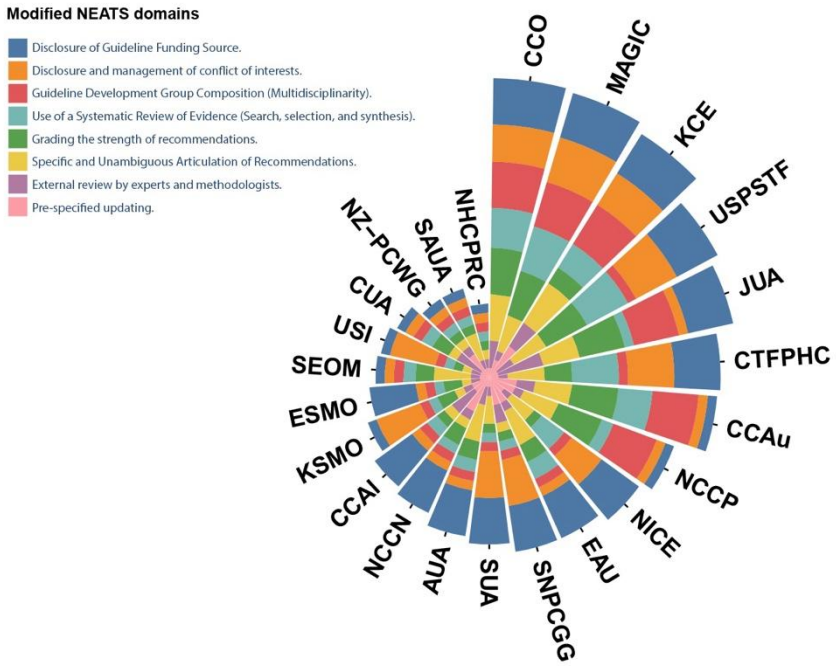

Table 1. Prostate cancer guidelines included in this study (n=23) sorted by modified NEAT score.					
Guideline	**Organization**	**Year**	**No. of recommendations**	**Scope**	**Modified NEATS score**
Chinese guidelines for diagnosis and treatment of prostate cancer 2018 (English version)	National Health Commission of the People's Republic of China	2018	NA	Broad	10
Prostate Cancer Management and Referral Guidance	New Zealand Prostate Cancer Working Group and Ministry of Health	2015	NA	Broad	12
The South African Prostate Cancer Screening Guidelines	South African Urology Association and the Prostate Cancer	2024	14	Narrow	12

	Foundation of South Africa				
USI Cancer Prostate Guidelines	Urological Society of India	2022	67	Broad	14
2022 UPDATE: Canadian Urological Association-Canadian Urologic Oncology Group guideline: Metastatic castration-naive and castration-sensitive prostate cancer	The Canadian Uro-Oncology Group (CUOG)/Canadian Urological Association (CUA)	2022	19	Narrow	14
SEOM clinical guidelines for the treatment of advanced prostate cancer (2020)	SEOM - Sociedad Española de Oncología Médica	2020	23	Narrow	15
Prostate cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up	European Society for Medical Oncology (ESMO)	2023	50	Broad	15
2020 Korean guidelines for the management of metastatic prostate cancer	Korean Society of Medical Oncology	2020	22	Narrow	16
NCCN - Prostate Cancer	National Cancer Care Network (NCCN)	2024	NA	Broad	18
Local Prostate Cancer (Cancer Care Alberta)	Cancer Care Alberta	2024	NA	Narrow	18
Saudi Oncology Society and Saudi	Saudi Oncology Society and Saudi	2017	47	Broad	20

Urology Association combined clinical management guidelines for prostate cancer 2017	Urology Association				
Clinically Localized Prostate Cancer: AUA/ASTRO Guideline 2022	American Urological Association (AUA), ASTRO, SUO	2022	44	Broad	20
The Swedish national guidelines on prostate cancer, part 1: early detection, diagnostics, staging, patient support and primary management of non-metastatic disease	Swedish National Prostate Cancer Guidelines Group	2022	NA	Broad	21
EAU - EANM - ESTRO -ESUR - ISUP - SIOG Guidelines on prostate cancer	European Association of Urology (EAU) - EANM - ESTRO - ESUR - ISUP - SIOG	2025	159	Broad	23
NCCP - Diagnosis and staging of patients with prostate cancer	National Cancer Control Programme (Ireland)	2024	28	Narrow	24
Prostate cancer: diagnosis and management	National Institute for Health and Care Excellence (NICE)	2021	131	Broad	26
Japanese clinical practice guidelines for prostate cancer 2023	Japanese Urological Association	2023	14	Broad	29

PSA testing	Cancer Council Australia	2015	43	Narrow	32
Recommendations on screening for prostate cancer with the prostate-specific antigen test (Canadian Task Force)	Canadian Task Force on Preventive Health Care	2014	3	Narrow	35
National Practice Guideline on the Treatment of localized prostate cancer – part 2	College voor Oncologie – Collège de médecins d'Oncologie-Belgian Health Care Knowledge Centre (KCE)	2014	16	Broad	37
Screening for Prostate Cancer - US Preventive Services Task Force Recommendation Statement	USPSTF - US Preventive Services Task Force Recommendation Statement	2018	2	Narrow	38
Multiparametric Magnetic Resonance Imaging in the Diagnosis of Clinically Significant Prostate Cancer	Cancer care Ontario	2021	3	Narrow	41
Prostate cancer screening with prostate-specific antigen (PSA) test: a clinical practice guideline	MAGIC and BMJ	2018	1	Narrow	42

Table 2. Characteristics of included guidance documents (n=23)	
Characteristic	Number (%)
Number of recommendations	
<15	6 (26.1)
16 −43	6 (26.1)
44−159	6 (26.1)
Unclear/not extractable	5 (21.7)
Publication year	
2014–2020	10 (43.5)
2021–2024	13 (56.5)
Region of Origin	
North America	7 (30.4)
Europe	7 (30.4)
Middle East and Africa	3 (13.0)
Asia and Oceania	5 (21.7)
Other	1 (4.3)
Version	
Original	9 (39.1)
Updated	14 (60.9)
Scope	
Narrow	11 (47.8%)
Broad	12 (52.2%)