

CADTH Health Technology Review

# Lung-RADS Versus Pan-Canadian Early Detection of Lung Cancer Study Screening for Patients at High Risk of Lung Cancer

**Authors:** Angela M. Barbara, Hannah Loshak

**ISSN:** 2563-6596

**Disclaimer:** The information in this document is intended to help Canadian health care decision-makers, health care professionals, health systems leaders, and policy-makers make well-informed decisions and thereby improve the quality of health care services. While patients and others may access this document, the document is made available for informational purposes only and no representations or warranties are made with respect to its fitness for any particular purpose. The information in this document should not be used as a substitute for professional medical advice or as a substitute for the application of clinical judgment in respect of the care of a particular patient or other professional judgment in any decision-making process. The Canadian Agency for Drugs and Technologies in Health (CADTH) does not endorse any information, drugs, therapies, treatments, products, processes, or services.

While care has been taken to ensure that the information prepared by CADTH in this document is accurate, complete, and up to date as at the applicable date the material was first published by CADTH, CADTH does not make any guarantees to that effect. CADTH does not guarantee and is not responsible for the quality, currency, propriety, accuracy, or reasonableness of any statements, information, or conclusions contained in any third-party materials used in preparing this document. The views and opinions of third parties published in this document do not necessarily state or reflect those of CADTH.

CADTH is not responsible for any errors, omissions, injury, loss, or damage arising from or relating to the use (or misuse) of any information, statements, or conclusions contained in or implied by the contents of this document or any of the source materials.

This document may contain links to third-party websites. CADTH does not have control over the content of such sites. Use of third-party sites is governed by the third-party website owners' own terms and conditions set out for such sites. CADTH does not make any guarantee with respect to any information contained on such third-party sites and CADTH is not responsible for any injury, loss, or damage suffered as a result of using such third-party sites. CADTH has no responsibility for the collection, use, and disclosure of personal information by third-party sites.

Subject to the aforementioned limitations, the views expressed herein are those of CADTH and do not necessarily represent the views of Canada's federal, provincial, or territorial governments or any third-party supplier of information.

This document is prepared and intended for use in the context of the Canadian health care system. The use of this document outside of Canada is done so at the user's own risk.

This disclaimer and any questions or matters of any nature arising from or relating to the content or use (or misuse) of this document will be governed by and interpreted in accordance with the laws of the Province of Ontario and the laws of Canada applicable therein, and all proceedings shall be subject to the exclusive jurisdiction of the courts of the Province of Ontario, Canada.

The copyright and other intellectual property rights in this document are owned by CADTH and its licensors. These rights are protected by the Canadian *Copyright Act* and other national and international laws and agreements. Users are permitted to make copies of this document for non-commercial purposes only, provided it is not modified when reproduced and appropriate credit is given to CADTH and its licensors.

**About CADTH:** CADTH is an independent, not-for-profit organization responsible for providing Canada's health care decision-makers with objective evidence to help make informed decisions about the optimal use of drugs, medical devices, diagnostics, and procedures in our health care system.

**Funding:** CADTH receives funding from Canada's federal, provincial, and territorial governments, with the exception of Quebec.

Questions or requests for information about this report can be directed to [Requests@CADTH.ca](mailto:Requests@CADTH.ca)

## Table of Contents

---

|  |           |
|--|-----------|
| <b>Abbreviations</b> .....   | <b>6</b>  |
| <b>Key Messages</b> .....  | <b>7</b>  |
| <b>Context and Policy Issues</b> .....                                   | <b>7</b>  |
| <b>Research Questions</b> .....  | <b>8</b>  |
| <b>Methods</b> .....   | <b>9</b>  |
| Literature Search Methods.....   | 9         |
| Selection Criteria and Methods .....                                     | 9         |
| Exclusion Criteria.....  | 9         |
| Critical Appraisal of Individual Studies .....                           | 9         |
| <b>Summary of Evidence</b> .....   | <b>9</b>  |
| Quantity of Research Available.....                                      | 9         |
| Summary of Study Characteristics.....                                    | 10        |
| <b>Limitations</b> .....   | <b>15</b> |
| <b>Conclusions and Implications for Decision- or Policy-Making</b> ..... | <b>16</b> |
| <b>References</b> .....  | <b>17</b> |
| <b>Appendix 1: Selection of Included Studies</b> .....                   | <b>19</b> |
| <b>Appendix 2: Characteristics of Included Publications</b> .....        | <b>20</b> |
| <b>Appendix 3: Critical Appraisal of Included Publications</b> .....     | <b>28</b> |
| <b>Appendix 4: Main Study Findings and Authors’ Conclusions</b> .....    | <b>36</b> |
| <b>Appendix 5: References of Potential Interest</b> .....                | <b>40</b> |

## List of Tables

---

|  |    |
|--|----|
| Table 1: Selection Criteria.....   | 10 |
| Table 2: Characteristics of Included Primary Clinical Studies .....  | 20 |
| Table 3: Characteristics of Included Economic Evaluations .....  | 24 |
| Table 4: Characteristics of Included Guidelines.....   | 27 |
| Table 5: Strengths and Limitations of Diagnostic Test Accuracy Studies Using the QUADAS-2 Checklist <sup>9</sup> ..... | 28 |
| Table 6: Strengths and Limitations of Economic Evaluations Using the Drummond Checklist <sup>10</sup> .....            | 33 |
| Table 7: Strengths and Limitations of Guideline Using AGREE II <sup>11</sup> .....                                     | 35 |
| Table 8: Summary of Findings by Outcome – Diagnostic Test Accuracy, AUC .....  | 36 |
| Table 9: Summary of Findings by Outcome – Diagnostic Test Accuracy, Diagnostic Parameters .....                        | 36 |
| Table 10: Summary of Findings of Included Economic Evaluation .....  | 38 |
| Table 11: Summary of Recommendations in Included Guideline.....  | 39 |

## List of Figures

---

Figure 1: Selection of Included Studies ..... 19

## Abbreviations

|                  |   |
|------------------|---|
| <b>AUC</b>       | area under the curve                        |
| <b>BTS</b>       | British Thoracic Society                    |
| <b>CHEST</b>     | American College of Chest Physicians        |
| <b>ICER</b>      | incremental cost-effectiveness ratio        |
| <b>IQR</b>       | interquartile range                         |
| <b>LDCT</b>      | low-dose computed tomography                |
| <b>Lung-RADS</b> | Lung Imaging Reporting and Data System      |
| <b>NLST</b>      | National Lung Screening Trial               |
| <b>PanCan</b>    | Pan-Canadian Early Detection of Lung Cancer |
| <b>QALY</b>      | quality-adjusted life-year                  |
| <b>SSN</b>       | subsolid nodule                             |

## Key Messages

- Evidence of variable quality from 6 diagnostic test accuracy studies indicates that the Pan-Canadian Early Detection of Lung Cancer (PanCan) model may perform better at determining which lung nodules identified by low-dose CT are cancerous compared to the Lung Imaging Reporting and Data System. However, evidence from 3 other studies, also of variable quality, suggests that the risk calculators have similar diagnostic test accuracy.
- No studies were identified that compared the clinical utility of PanCan versus the Lung Imaging Reporting and Data System.
- Results from 2 economic evaluations were inconsistent about the cost-effectiveness of the 2 lung cancer risk models. However, each study applied the models to different types of lung nodules.
- One evidence-based guideline recommended that PanCan be used in the UK for initial risk assessment and for the management of lung nodules.

## Context and Policy Issues

In 2021, lung cancer represented 13% of new cancer cases and 25% of cancer deaths in Canada.<sup>1</sup> The 5-year lung cancer-specific survival rate for the 2015 to 2017 time frame was 19% in males and 26% in females.<sup>1</sup> Low-dose CT (LDCT) screening leads to earlier detection of lung cancer and therefore improving survival compared to usual care.<sup>2</sup> Two screening protocols are available to estimate the risk of lung cancer and guide the management of lung nodules identified by first (baseline) LDCT screening:

- the Lung Imaging Reporting and Data System (Lung-RADS)
- the Pan-Canadian Early Detection of Lung Cancer (PanCan).

Lung-RADS was developed by the American College of Radiology<sup>3</sup> to standardize the reporting and management of lung nodules. Modelled after the American College of Radiology Breast Imaging Reporting and Data System, the BI-RADS,<sup>4</sup> Lung-RADS categorizes the risk of cancer based on 3 primary nodule characteristics: size, type, and growth rate. The nodule type is defined by its solidity. Solid nodules are homogeneous and obscure the lung parenchyma, whereas subsolid nodules have sections that are solid and non-solid nodules have no solid parts. Subsolid nodules can be pure ground-glass nodules, which appear opaque or hazy on scans, or part-solid nodules, which contain both solid and ground-glass components.<sup>5</sup> The Lung-RADS categories indicate an increasing likelihood of malignancy, where 1 means negative; 2 means benign; 3 means probably benign; 4A means suspicious, with 5% to 15% probability of clinically active cancer in the next year; 4B means very suspicious, with more than 15% probability of cancer in the next year; and 4X means very suspicious but not otherwise specified. The category classifications and management recommendations associated with those classifications were updated in 2019 to version 1.1 based on empirical evidence and clinical experience.<sup>6</sup>

PanCan was developed by McWilliams et al. (2013)<sup>7</sup> and is also referred to as the Brock University model or the Vancouver risk calculator for the locations of its conception. To generate a probability of having lung cancer on a continuous scale, PanCan utilizes patient characteristics (age, sex, family history of lung cancer, having emphysema) and nodule characteristics (size, type, location, number of nodules, and signs of spiculation). A nodule

risk index of less than 1.5% means normal finding, 1.5% to 5% means low risk of malignancy, 6% to 30% means moderate risk of malignancy, and greater than 30% means high risk of malignancy. The PanCan results will determine if a person should undergo further diagnostic testing, such as annual CT screening.

The parameters of the 2 models were mathematically derived from screening data. PanCan was validated using data from chemoprevention trials of BC Cancer<sup>7</sup> and the Danish Lung Cancer Screening Trial.<sup>8</sup>

Because lung cancer screening is a process and not a diagnostic test, there is no “gold standard” per se. Rather, the results of the screening algorithms will determine the follow-up interval(s) and further diagnostic testing. For example, when the PanCan equation predicts a greater than 10% risk of a nodule being malignant, PET or CT is then used as additional diagnostic testing.<sup>9</sup> Another example is that Lung-RADS recommends follow-up LDCT, chest CT or PET, and/or tissue sampling depending on the probability of malignancy and comorbidities for category 4B nodules.<sup>6</sup>

While both protocols for nodule classification and subsequent management are available and in use in Canada, neither is universally accepted. The objective of the current review was to evaluate the evidence regarding the comparative diagnostic accuracy, clinical utility, and cost-effectiveness of Lung-RADS compared to PanCan for patients at high risk of lung cancer undergoing screening with LDCT to identify malignant lung nodules. Additionally, evidence-based guidelines regarding the use of either Lung-RADS or PanCan were sought.

## Research Questions

1. What is the comparative diagnostic accuracy of the Lung Imaging Reporting and Data System versus the Pan-Canadian Early Detection of Lung Cancer nodule risk calculation for the identification of malignant lung nodules in patients at high risk of lung cancer undergoing screening with low-dose CT?
2. What is the comparative clinical utility of the Lung Imaging Reporting and Data System versus the Pan-Canadian Early Detection of Lung Cancer nodule risk calculation for the identification of malignant lung nodules in patients at high risk of lung cancer undergoing screening with low-dose CT?
3. What is the comparative cost-effectiveness of the Lung Imaging Reporting and Data System versus the Pan-Canadian Early Detection of Lung Cancer nodule risk calculation for the identification of malignant lung nodules in patients at high risk of lung cancer undergoing screening with low-dose CT?
4. What are the evidence-based guidelines describing use of the Lung Imaging Reporting and Data System and/or the Pan-Canadian Early Detection of Lung Cancer nodule risk calculation for the identification of malignant lung nodules in patients at high risk of lung cancer undergoing screening with low-dose CT?



## Methods

### Literature Search Methods

A limited literature search was conducted by an information specialist on key resources including MEDLINE, the Cochrane Database of Systematic Reviews, the international HTA database, the websites of Canadian and major international health technology agencies, as well as a focused internet search. The search strategy comprised both controlled vocabulary, such as the National Library of Medicine's MeSH (Medical Subject Headings), and keywords. The main search concepts were the Lung Imaging Reporting and Data System and the Pan-Canadian Early Detection of Lung Cancer nodule risk calculation. No filters were applied to limit the retrieval by study type. Where possible, retrieval was limited to the human population. The search was also limited to English-language documents published between January 1, 2011 and November 4, 2021. A second search was done for low-dose CT and lung cancer screening, with CADTH-developed search filters applied to limit retrieval to guidelines. The second search was also limited to English-language documents published between January 1, 2016 and November 4, 2021.

### Selection Criteria and Methods

One reviewer screened citations and selected studies. In the first level of screening, titles and abstracts were reviewed and potentially relevant articles were retrieved and assessed for inclusion. The final selection of full-text articles was based on the inclusion criteria presented in Table 1.

### Exclusion Criteria

Articles were excluded if they did not meet the selection criteria outlined in Table 1, if they were duplicate publications, or if they were published before 2011. Guidelines with unclear methodology were also excluded.

### Critical Appraisal of Individual Studies

The included publications were critically appraised by 1 reviewer using the following tools as a guide: the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) checklist<sup>9</sup> for diagnostic test accuracy studies, the Drummond checklist<sup>10</sup> for economic evaluations, and the Appraisal of Guidelines for Research & Evaluation (AGREE) II instrument<sup>11</sup> for guidelines. Summary scores were not calculated for the included studies; rather, the strengths and limitations of each included publication were described narratively.

## Summary of Evidence

### Quantity of Research Available

A total of 258 citations were identified in the literature search. Following screening of titles and abstracts, 217 citations were excluded and 41 potentially relevant reports from the electronic search were retrieved for full-text review. Three potentially relevant publications were retrieved from the grey literature search for full-text review. Of these potentially relevant

articles, 32 publications were excluded for various reasons and 12 publications met the inclusion criteria and were included in this report. These comprised 9 diagnostic test accuracy studies, 2 economic evaluations, and 1 evidence-based guideline. Appendix 1 presents the PRISMA<sup>12</sup> flow chart of the study selection.

Additional references of potential interest are provided in Appendix 5.

## Summary of Study Characteristics

Nine diagnostic test accuracy studies,<sup>13-21</sup> 2 economic evaluations,<sup>22,23</sup> and 1 evidence-based guideline<sup>24</sup> were identified for inclusion in this review. No relevant systematic reviews, health technology assessments, or randomized controlled trials were identified.

Details regarding the characteristics of included publications are provided in Appendix 2.

## Study Design

Seven diagnostic test accuracy studies<sup>13,14,16-18,20,21</sup> retrospectively evaluated pulmonary nodules from participants enrolled in previous studies that used a single-gate approach for patient selection (i.e., patients with unknown lung cancer status). Four studies<sup>13,14,18,21</sup> included patients who were randomly assigned to LDCT screening in the multi-centre National Lung Screening Trial (NSLT).<sup>25</sup> Patients were enrolled from August 2002 through April 2004. One diagnostic test accuracy study included patients who were randomized to LDCT screening

**Table 1: Selection Criteria**

| Criteria                  | Description  |
|---------------------------|--|
| <b>Population</b>         | Q1 to Q4: Patients at high risk of lung cancer undergoing screening with low-dose CT to identify malignant lung nodules  |
| <b>Intervention</b>       | Q1 to Q3: Lung-RADS<br>Q4: Lung-RADS and/or PanCan   |
| <b>Comparator</b>         | Q1 to Q3: PanCan<br>Q4: Not applicable   |
| <b>Reference standard</b> | Q1: Confirmed lung cancer diagnosis; i.e., as determined by biopsy/histology, pathology, surgery, bronchoscopy, or other follow-up diagnostic procedure<br>Q2 to Q4: Not applicable  |
| <b>Outcomes</b>           | Q1: Comparative diagnostic test accuracy; e.g., positive and negative predictive value, sensitivity (effectiveness in identifying all cases of malignant nodules and lung cancer), specificity (effectiveness in accurately identifying malignant nodules and cases of lung cancer)<br>Q2: Comparative clinical utility; e.g., benefits and harms to patients, including time to treatment, impact on quality of life, feasibility of screening test, adverse events from the screening, incidental findings<br>Q3: Comparative cost-effectiveness; e.g., quality-adjusted life-years, costs per unit of health benefit<br>Q4: Recommendations regarding the use of either Lung-RADS and/or PanCan; e.g., which screening method is optimal; guidance as to which intervention is preferable in particular patient populations, settings, contexts; clinical and other considerations when using either screening method |
| <b>Study designs</b>      | Health technology assessments, systematic reviews, randomized controlled trials, non-randomized studies, economic evaluations, evidence-based guidelines and recommendations   |

Lung-RADS = Lung Imaging Reporting and Data System; PanCan = Pan-Canadian Early Detection of Lung Cancer.

from October 2004 to March 2006 in a single-centre trial.<sup>20</sup> Two studies included patients previously enrolled in prospective cohort studies: 1 was a population-based multi-centre study that enrolled patients from April 2017 to December 2018<sup>16</sup> and the other study recruited patients from December 2007 to December 2010 at a single tertiary institution.<sup>17</sup>

Two diagnostic test accuracy studies enrolled new patients and also used a single-gate approach for patient selection: 1 retrospective study screened patients from December 2012 to June 2016 at a single centre<sup>15</sup> and 1 multi-centre prospective study enrolled patients between June 2015 and December 2017.<sup>19</sup>

The 2 economic evaluations were conducted as cost-utility analyses, using lifetime horizons. One study<sup>22</sup> used a Monte Carlo simulation model for subsolid nodules using data from the literature, NLST, and national databases. Major assumptions included that follow-up CT screening led to definitive treatment and a willingness-to-pay threshold of US\$100,000 per quality-adjusted life-year (QALY). The perspective taken was that of the health care system and society. The other study<sup>23</sup> used a predictive logistic regression model for nodules assigned a Lung-RADS category of 4A, 4B, or 4X. Costs were derived from a previous cost-effectiveness study and the perspective of health care payers and policy-makers was taken. Assumptions were made about survival and mortality rates, follow-up detection of nodules originally screened as benign, and growth rates. Model parameters in both studies included nodule properties, patient characteristics, mortality, and treatment.

The evidence-based guideline<sup>24</sup> was developed by the British Thoracic Society (BTS), which included respiratory physicians, radiologists, respiratory specialty trainees, a thoracic surgeon, a pathologist, and a respiratory nurse practitioner. The guideline was informed by systematic reviews of the literature. The recommendations were classified based on Scottish Intercollegiate Guidelines Network criteria. A grade between A (highest) and D (lowest) or Checkmark (no research evidence) was assigned to each recommendation. no research evidence was assigned to each recommendation. Scientific evidence that informed the recommendations was classified between 1++ (highest quality) and 4 (lowest quality). Before publication, the draft guideline was made available online for public consultation and feedback was invited from stakeholder organizations.

## Country of Origin

The diagnostic test accuracy studies were conducted in, and enrolled patients from, Australia,<sup>17</sup> Canada,<sup>19</sup> Denmark,<sup>20</sup> Germany,<sup>20</sup> the Netherlands,<sup>20</sup> South Korea,<sup>16</sup> and the US.<sup>13-15,18,21</sup>

The 2 economic evaluations were conducted by authors in the US.<sup>22,23</sup>

The guideline was intended for use in the UK.<sup>24</sup>

## Patient Population

Four diagnostic test accuracy studies<sup>13,14,18,21</sup> included participants in the LDCT screening arm of NSLT of between 55 years and 74 years of age, had a history of cigarette smoking of at least 30 pack-years, and, if former smokers, had quit within the previous 15 years. Each study included a different subset of NSLT participants: 1 study included a random set of 434 patients with subsolid or part-solid nodules,<sup>13</sup> 1 study included 58 patients with images of pre-cancers and 127 patients with benign nodules,<sup>14</sup> another study included 6,956 patients with solid nodules only,<sup>18</sup> and the fourth study assessed 2,813 patients with all nodule types.<sup>21</sup>

Two other studies included patients using the eligibility criteria of NSLT: the study by Kessler et al. (2020) enrolled 486 patients with a mean age of 63 years<sup>15</sup> and the study by Marshall et al. (2017) included 256 patients aged 60 years or older.<sup>17</sup> The study by Tremblay et al. (2019) enrolled 775 patients who either met the NSLT eligibility criteria or were aged 55 years to 80 years of age with an estimated 6-year lung cancer risk of 1.5% or more.<sup>19</sup>

The diagnostic test accuracy study by van Riel et al. (2017) included 613 current or former smokers (aged 50 years to 75 years) who had any nodule identified by LDCT screening.<sup>20</sup> The study by Kim et al. (2021) included 4,578 patients (median age of 62 years, 54% smokers) with non-calcified nodules determined after LDCT screening.<sup>16</sup>

Patients in 1 economic evaluation included a hypothetical cohort of 10 million current and former smokers ranging from 55 years to 75 years of age and assumed to have subsolid nodules (SSNs) at baseline LDCT.<sup>22</sup> Patients in the other economic evaluation included a simulated cohort of 100,000 patients aged 61 years to 71 years assigned Lung-RADS category 4 nodules.<sup>23</sup>

The target population of the included guideline was adult patients with pulmonary nodules. The intended users of all recommendations included health care professionals such as clinicians (e.g., physicians, general practitioners, radiologists, surgeons) and nurses.<sup>24</sup>

## Interventions and Comparators

The 9 diagnostic test accuracy studies<sup>13-21</sup> assessed their study populations using both Lung-RADS and PanCan. Confirmed lung cancer diagnosis, as determined by follow-up diagnostic procedures, was considered the reference standard for their analyses. The screening algorithms were applied to the baseline scans and the lung cancer diagnosis was assessed over the follow-up periods and ranging from 2 years to 6.5 years.

Hammer et al. (2019) reported that the assessments using both screening algorithms were conducted without knowledge of the results of the reference standard.<sup>13</sup> In the other 8 studies, it was unclear if the screeners were blinded to the final diagnoses of the patients.<sup>14-21</sup>

The economic evaluations examined the cost-effectiveness of Lung-RADS and PanCan.<sup>22,23</sup> In 1 study, Lung-RADS was compared to 2 different guidelines for nodule management using PanCan: the American College of Chest Physicians (CHEST) guidelines, which recommended PET or CT following a high-risk score and non-surgical lung biopsy following intermediate PET/CT results; and the BTS guideline, which recommended PET or CT following an intermediate risk score.<sup>23</sup>

The relevant intervention considered in the guideline was PanCan for initial risk assessment of the probability of malignancy in pulmonary nodules and management of SSNs.<sup>24</sup>

## Outcomes

The diagnostic test accuracy studies calculated various parameters of diagnostic performance. Seven studies reported the area under the curve (AUC).<sup>13,15-20</sup> Six studies reported sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV),<sup>15-19,21</sup> and 3 studies reported accuracy.<sup>14,15,21</sup>

Seven studies calculated diagnostic parameters. Two studies used a PanCan risk score of 5% as the threshold for positivity<sup>19,21</sup> and 2 studies used a PanCan risk score of 10% to indicate a positive result.<sup>14,17</sup> Two studies used Lung-RADS category 3 as the threshold for a

positive result<sup>17,19</sup> and 2 studies used category 4A/4B as the threshold for a positive result.<sup>14,21</sup> Three studies calculated diagnostic test accuracy for different thresholds used by each risk algorithm to determine whether a nodule was positive.<sup>15,16,18</sup>

The economic evaluations calculated the costs and QALYs for each screening strategy and incremental cost-effectiveness ratios (ICERs) expressed as ratios of incremental cost incurred per QALY gained, comparing different screening strategies.<sup>22,23</sup>

The BTS guideline provides recommendations relevant to the current report, which considered the diagnostic test accuracy of PanCan in estimating the probability that a lung nodule would be diagnosed as cancer within a 2- to 4-year follow-up period.<sup>24</sup>

## Summary of Critical Appraisal

### *Diagnostic Test Accuracy Studies*

There were several strengths common to the 9 diagnostic test accuracy studies:<sup>13-21</sup> the screening tests, the fact that their conduct and interpretation matched the review question, and that the thresholds used for the screening tests were pre-specified; the target condition (i.e., lung cancer), as defined by the reference standard (i.e., lung cancer diagnosed during the follow-up period), matched the research question; and the study participants, care providers, and settings appeared to be representative of the population and that they were care settings of interest. Eight diagnostic test accuracy studies<sup>13,15-21</sup> also clearly described objectives, interventions, controls, inclusion criteria, outcomes, and main findings, and avoided the use of a case-control study design. Eight studies reported sources of funding<sup>13,14,16-21</sup> and 7 studies presented characteristics of included patients and lung nodules.<sup>13,15-17,19-21</sup> The authors of 7 studies disclosed no conflicts of interest.<sup>13-18,21</sup>

As for limitations, it was unclear if the screeners who conducted the assessments using Lung-RADS and PanCan were blinded to the results of the reference standard (i.e., having no knowledge of the final diagnosis) in 8 studies.<sup>14-21</sup> The authors of 2 studies had conflicts of interest.<sup>19,20</sup> Three studies had a potential risk of bias because of missing data.<sup>15,20,21</sup> Non-consecutive patients were included in the study by Hammer et al. (2019).<sup>13</sup> In the study by Hawkins et al. (2016), there were additional limitations: a case-control design was used, it was unclear how patients were selected, and patient characteristics were not reported.<sup>14</sup>

### *Economic Evaluations*

The 2 economic evaluations<sup>22,23</sup> shared the following strengths: the research question and its economic importance were stated; sources of effectiveness estimates, primary outcome, details of the simulation models, and methods for the estimation of quantities and unit costs were described; the time horizon of costs and benefits, discount rate, and details of statistical tests and sensitivity analyses were given; the incremental analysis was reported; conclusions were given; and the authors stated that they had no conflicts of interest. One economic evaluation<sup>23</sup> provided confidence intervals for costs and QALYs, but the other study did not.<sup>22</sup> The studies also shared the following limitations: no description of current price adjustments for inflation was provided, no justification for the selected discount rate was provided, and sources of funding were not disclosed.

### *Evidence-Based Guideline*

The BTS evidence-based guideline<sup>24</sup> provided a clear description of its scope and purpose, including objectives, the range of clinical questions covered in the guideline, the intended users, and the target population. The final recommendations were easily identifiable and were

written using language that was clear and unambiguous. The guideline development groups appeared to include individuals from all relevant professional groups. Patient preferences were not sought or incorporated as part of the development process; however, a draft guideline document was made available online for public consumption before publication. Systematic literature searches were used to identify evidence for consideration when developing recommendations. Search strategies, databases searched, and the timing of the literature searches were clearly described. The guideline provided a detailed description of the methods used for selecting articles and stated that the screening process was done in duplicate. The guideline included a description of how the recommendations were formulated. Recommendations were externally reviewed by stakeholders before their publication and included a procedure for updating the guideline in the future. Regarding applicability, the facilitators and barriers to the implementation of the recommendations were not addressed. A description of potential conflicts of interest was included; however, sources of funding were not disclosed, making it unclear if the funders' views had any impact on the content of the guideline. Finally, it should be noted that this guideline was developed for use in the UK; therefore, the generalizability of the recommendations to the Canadian context is unclear.

Additional details regarding the strengths and limitations of the included publications are provided in Appendix 3.

## Summary of Findings

Appendix 4 presents the main study findings.

### *Diagnostic Accuracy of Lung-RADS Versus PanCan*

Evidence regarding the diagnostic accuracy of Lung-RADS versus PanCan for the detection of malignant pulmonary nodules was available from 9 diagnostic test accuracy studies.<sup>13-21</sup>

#### Area Under the Curve

The study by van Riel et al. (2017) reported that PanCan performed significantly better than Lung-RADS for discriminating benign nodules from malignant nodules, based on AUCs (0.87 versus 0.81,  $P = 0.003$ ).<sup>20</sup> Three other studies<sup>13,17,19</sup> reported greater, but non-significant, AUCs for PanCan (0.78, 0.90, 0.97) compared to Lung-RADS (0.70, 0.84, 0.93) and 3 other studies<sup>15,16,18</sup> found similar AUCs for both screening tests.

#### Diagnostic Parameters

The study by White et al. (2019)<sup>21</sup> found that PanCan had statistically significantly greater specificity and accuracy (85%, 85%) compared to Lung-RADS (76%, 76%); sensitivity was also greater for PanCan (93% versus 87%), but the difference was not statistically significant. The study by Hawkins et al. (2016)<sup>14</sup> reported that PanCan had greater accuracy compared to Lung-RADS (79% versus 71%; statistical significance was not reported). One study found that both PanCan and Lung-RADS were 100% sensitive, but PanCan had higher specificity (95% versus 80%; statistical significance was not reported).<sup>17</sup> The study by Kessler et al. (2020)<sup>15</sup> found that PanCan had higher sensitivity, although it was not a statistically significant finding (74% versus 58%), but it had a lower specificity than Lung-RADS (94% versus 98%; statistical significance was not reported). Three studies<sup>16,18,19</sup> found similar diagnostic parameters between PanCan and Lung-RADS.

### *Clinical Utility of Lung-RADS Versus PanCan*

No relevant evidence regarding the comparative clinical utility of Lung-RADS versus PanCan was identified; therefore, no summary can be provided.

## ***Economic Evaluations***

One economic evaluation reported that Lung-RADS compared to PanCan was cost-effective under a willingness-to-pay threshold of \$100,000 per QALY, with an ICER of \$52,993 per QALY.<sup>22</sup>

The second economic evaluation found that the BTS guideline using PanCan was associated with more QALYs than Lung-RADS or the CHEST guidelines using PanCan and lower costs compared to the CHEST guidelines using PanCan. The BTS guideline compared to Lung-RADS had an ICER of \$52,643 per QALY gained.<sup>23</sup>

## ***Guidelines***

### **Initial Assessment of the Probability of Malignancy in Pulmonary Nodules**

The BTS guideline<sup>24</sup> recommends the use of PanCan for initial risk assessment of pulmonary nodules in people aged 50 years or older who are smokers or former smokers. This is a grade C recommendation, based on level 2+ evidence from a validation study conducted in the UK.<sup>26</sup>

The BTS guideline<sup>24</sup> recommends consideration of PanCan for the initial risk assessment of pulmonary nodules in all patients. This is a grade D recommendation, based on level 3 evidence.

### **Management of Subsolid Nodules**

The BTS guideline<sup>24</sup> recommends the use of PanCan to calculate the risk of malignancy in SSNs larger than 5 mm that are unchanged at 3 months. This is a grade C recommendation, based on level 2+ evidence.

## **Limitations**

Seven of the 9 diagnostic test accuracy studies were retrospective.<sup>13-15,17,18,20,21</sup> The studies that included patients from NSLT<sup>13,14,18,21</sup> shared limitations related to the original trial. Scanners used in the trial were less technologically advanced than scanners available today. Also, the trial was conducted at a variety of medical institutions in the US, many of which were recognized for their expertise in radiology and in the diagnosis of cancer;<sup>25</sup> therefore, applicability to all Canadian facilities is uncertain.

No studies on the benefits and harms of Lung-RADS versus PanCan were found. It is unclear which of these 2 screening protocols may result in improved clinical outcomes for patients undergoing LDCT screening to identify malignant lung nodules.

Apart from 1 prospective diagnostic test accuracy study,<sup>19</sup> none of the primary studies or economic evaluations<sup>22,23</sup> were conducted in Canada. Similarly, the included guideline was not intended for professionals in Canada.<sup>24</sup> Therefore, the generalizability of the findings from the included literature and the applicability of the recommendations from the included guideline to Canadian settings are unclear.

## Conclusions and Implications for Decision- or Policy-Making

This review comprised 9 diagnostic accuracy studies,<sup>13-21</sup> 2 economic evaluations,<sup>22,23</sup> and 1 evidence-based guideline.<sup>24</sup>

Evidence from 6 retrospective studies suggests that PanCan had superior diagnostic test accuracy compared to Lung-RADS for predicting malignancy.<sup>13-15,17,20,21</sup> However, evidence from 1 prospective study and 1 retrospective analysis suggests that the screening protocols have a similar diagnostic performance.<sup>16,18</sup> Findings from the sole prospective study performed in Canada also found similar diagnostic test accuracy between Lung-RADS and PanCan.<sup>16,18,19</sup>

Results from the 2 economic evaluations were inconsistent about the cost-effectiveness of the 2 lung cancer risk models. One study<sup>22</sup> reported an ICER of \$52,993 per QALY for Lung-RADS compared to PanCan, while the other study reported an ICER of \$52,643 per QALY for the BTS guideline using Pan Can compared to Lung-RADS.<sup>23</sup> However, each study applied the models to different types of lung nodules.

The BTS guideline recommends PanCan for initial risk assessment of the probability of malignancy in pulmonary nodules and management of SSNs.<sup>24</sup> However, there are no Canadian guidelines that recommend the use of Lung-RADS or PanCan for the identification of malignant lung nodules in patients at high risk of lung cancer undergoing LDCT screening.

The limitations of the included literature should be considered when interpreting the findings of this report. Further research investigating the diagnostic accuracy of Lung-RADS versus PanCan in Canadian settings would help confirm if PanCan performs better or is similar to Lung-RADS for distinguishing malignant pulmonary nodules. Clinical utility research is needed to evaluate the benefits and harms to patients using Lung-RADS compared to PanCan. Future economic evaluations conducted from Canadian perspectives and guideline recommendations intended for Canadian settings may be helpful to further inform clinical and policy decisions.



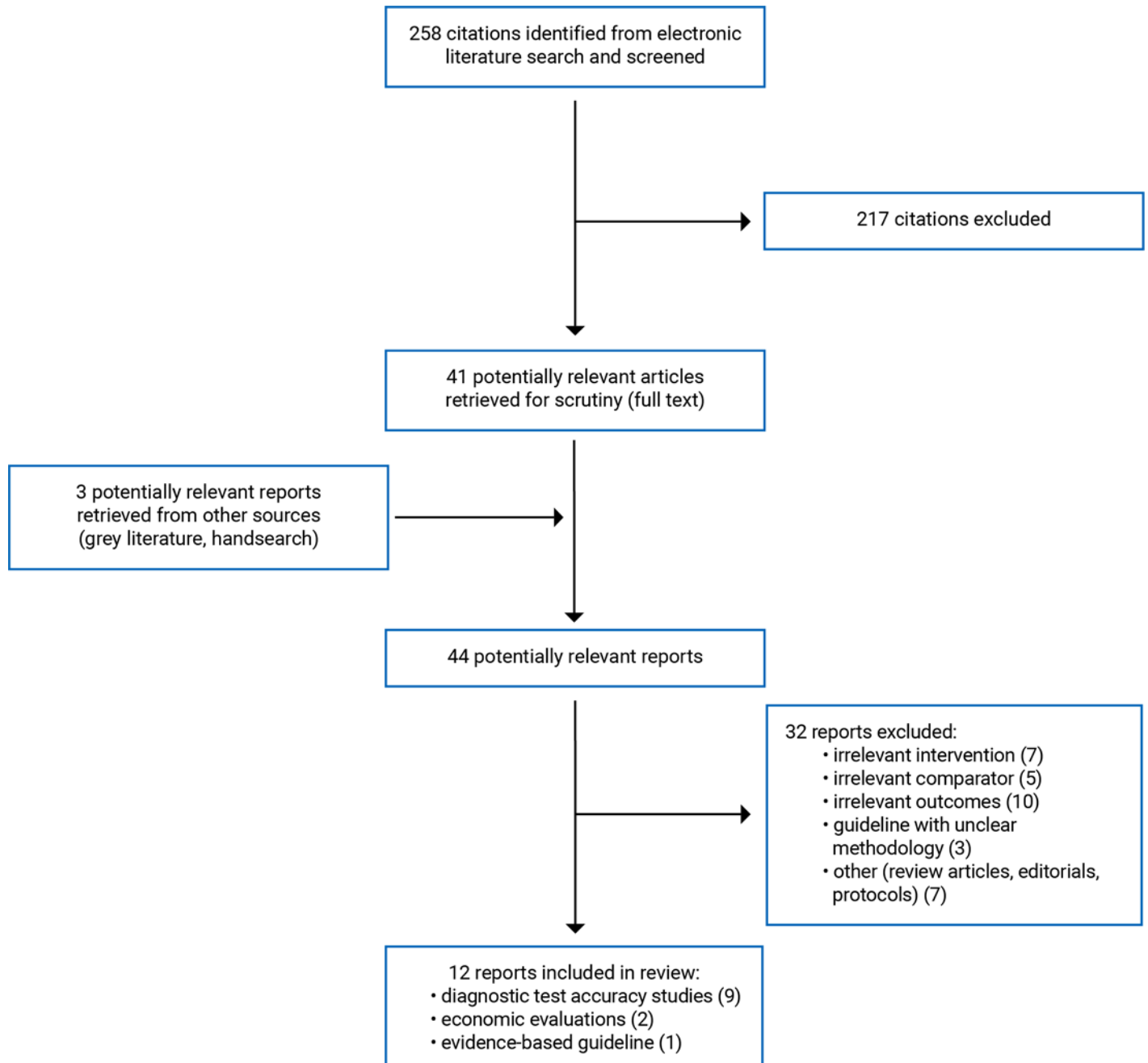
## References

- Canadian Cancer Statistics 2021. Ottawa (ON): Canadian Cancer Statistics Advisory Committee, Statistics Canada, Public Health Agency of Canada; 2021: [https://cdn.cancer.ca/-/media/files/research/cancer-statistics/2021-statistics/2021-pdf-en-final.pdf?rev=2b9d2be7a2d34c1dab6a01c6b0a6a32d&hash=01DE85401DBF0217F8B64F2B7DF43986&\\_ga=2.203979736.995607737.1638540598-408723823.1638540598](https://cdn.cancer.ca/-/media/files/research/cancer-statistics/2021-statistics/2021-pdf-en-final.pdf?rev=2b9d2be7a2d34c1dab6a01c6b0a6a32d&hash=01DE85401DBF0217F8B64F2B7DF43986&_ga=2.203979736.995607737.1638540598-408723823.1638540598). Accessed 2021 Nov 27.
- Huang KL WS, Lu WC, Chang YH, Su J, Lu YT. Effects of low-dose computed tomography on lung cancer screening: a systematic review, meta-analysis, and trial sequential analysis. *BMC Pulm Med*. 2019;19(1):126. [PubMed](#)
- Kazerooni EA, Armstrong MR, Amorosa JK, et al. ACR CT accreditation program and the lung cancer screening program designation. *J Am Coll Radiol*. 2015;12(1):38-42. [PubMed](#)
- An JY, Unsrdorfer KML, Weinreb JC. BI-RADS, C-RADS, CAD-RADS, LI-RADS, Lung-RADS, NI-RADS, O-RADS, PI-RADS, TI-RADS: Reporting and Data Systems. *Radiographics*. 2019;39(5):1435-1436. [PubMed](#)
- Dzadzadziszko K, Szurowska E. Pulmonary nodule radiological diagnostic algorithm in lung cancer screening. *Transl Lung Cancer Res*. 2021;10(2):1124-1135. [PubMed](#)
- Chelala L, Hossain R, Kazerooni EA, Christensen JD, Dyer DS, White CS. Lung-RADS Version 1.1: Challenges and a Look Ahead, From the AJR Special Series on Radiology Reporting and Data Systems. *AJR Am J Roentgenol*. 2021;216(6):1411-1422. [PubMed](#)
- McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med*. 2013;369(10):910-919. [PubMed](#)
- Winkler Wille MM, van Riel SJ, Saghir Z, et al. Predictive Accuracy of the PanCan Lung Cancer Risk Prediction Model -External Validation based on CT from the Danish Lung Cancer Screening Trial. *Eur Radiol*. 2015;25(10):3093-3099. [PubMed](#)
- Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536. [PubMed](#)
- Higgins JPT, Green S, editors. Figure 15.5.a: Drummond checklist (Drummond 1996). *Cochrane handbook for systematic reviews of interventions*. London (GB): The Cochrane Collaboration; 2011: [http://handbook-5-1.cochrane.org/chapter\\_15/figure\\_15\\_5\\_a\\_drummond\\_checklist\\_drummond\\_1996.htm](http://handbook-5-1.cochrane.org/chapter_15/figure_15_5_a_drummond_checklist_drummond_1996.htm). Accessed 2021 Dec 3.
- Agree Next Steps Consortium. The AGREE II Instrument. Hamilton (ON): AGREE Enterprise; 2017: <https://www.agreetrust.org/wp-content/uploads/2017/12/AGREE-II-Users-Manual-and-23-item-Instrument-2009-Update-2017.pdf>. Accessed 2021 Dec 3.
- Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol*. 2009;62(10):e1-e34. [PubMed](#)
- Hammer MM, Palazzo LL, Kong CY, Hunsaker AR. Cancer Risk in Subsolid Nodules in the National Lung Screening Trial. *Radiology*. 2019;293(2):441-448. [PubMed](#)
- Hawkins S, Wang H, Liu Y, et al. Predicting Malignant Nodules from Screening CT Scans. *J Thorac Oncol*. 2016;11(12):2120-2128. [PubMed](#)
- Kessler A, Peng R, Mardakhaev E, Haramati LB, White CS. Performance of the Vancouver Risk Calculator Compared with Lung-RADS in an Urban, Diverse Clinical Lung Cancer Screening Cohort. *Radiol Imaging Cancer*. 2020;2(2):e190021. [PubMed](#)
- Kim H, Kim HY, Goo JM, Kim Y. External validation and comparison of the Brock model and Lung-RADS for the baseline lung cancer CT screening using data from the Korean Lung Cancer Screening Project. *Eur Radiol*. 2021;31(6):4004-4015. [PubMed](#)
- Marshall HM, Zhao H, Bowman RV, et al. The effect of different radiological models on diagnostic accuracy and lung cancer screening performance. *Thorax*. 2017;72(12):1147-1150. [PubMed](#)
- Sundaram V, Gould MK, Nair VS. A Comparison of the PanCan Model and Lung-RADS to Assess Cancer Probability Among People With Screening-Detected, Solid Lung Nodules. *Chest*. 2021;159(3):1273-1282. [PubMed](#)
- Tremblay A, Taghizadeh N, MacGregor JH, et al. Application of Lung-Screening Reporting and Data System Versus Pan-Canadian Early Detection of Lung Cancer Nodule Risk Calculation in the Alberta Lung Cancer Screening Study. *J Am Coll Radiol*. 2019;16(10):1425-1432. [PubMed](#)
- van Riel SJ, Ciompi F, Jacobs C, et al. Malignancy risk estimation of screen-detected nodules at baseline CT: comparison of the PanCan model, Lung-RADS and NCCN guidelines. *Eur Radiol*. 2017;27(10):4019-4029. [PubMed](#)
- White CS, Dharaia E, Dalal S, Chen R, Haramati LB. Vancouver Risk Calculator Compared with ACR Lung-RADS in Predicting Malignancy: Analysis of the National Lung Screening Trial. *Radiology*. 2019;291(1):205-211. [PubMed](#)
- Hammer MM, Eckel AL, Palazzo LL, Kong CY. Cost-Effectiveness of Treatment Thresholds for Subsolid Pulmonary Nodules in CT Lung Cancer Screening. *Radiology*. 2021;300(3):586-593. [PubMed](#)
- Hammer MM, Gupta S, Kong CY. Cost-Effectiveness of Management Algorithms for Lung-RADS Category 4 Nodules. *Radiol Cardiothorac Imaging*. 2021;3(2):e200523. [PubMed](#)
- Callister M, Baldwin D, Akram A, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax*. 2015;70:ii1-ii54. [PubMed](#)
- National Lung Screening Trial Research Team, Aberle D, Adams A, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *N Engl J Med*. 2011;365(5):395-409. [PubMed](#)

26. Al-Ameri A, Malhotra P, Thygesen H, et al. Risk of malignancy in pulmonary nodules: a validation study of four prediction models. *Lung Cancer*. 2015;89(1):27-30. [PubMed](#)
27. Hammer MM, Byrne SC, Kong CY. Factors Influencing the False Positive Rate in CT Lung Cancer Screening. *Acad Radiol*. 2020;03:03.

## Appendix 1: Selection of Included Studies

Figure 1: Selection of Included Studies



## Appendix 2: Characteristics of Included Publications

Note that this appendix has not been copy-edited.

**Table 2: Characteristics of Included Primary Clinical Studies**

| Study citation, country, funding source  | Study design   | Population characteristics   | Intervention and comparator(s)   | Clinical outcomes, length of follow-up  |
|--|--|--|--|---|
| <b>Diagnostic test accuracy</b>  |  |  |  |   |
| <p>Kim et al. (2021)<sup>16</sup><br/>South Korea</p> <p><b>Funding sources:</b> National R&amp;D Program for Cancer Control, Ministry of Health and Welfare; the National Health Promotion Fund, Ministry of Health and Welfare, Republic of Korea.</p> | <p>Secondary analysis of a national, multi-centre, prospective cohort study (Korean Lung Cancer Screening Project)</p> | <p><b>Inclusion criteria:</b><br/>(1) participants whose CT scans were read using a cloud-based thin-client reading system; (2) participants with at least 1 non-calcified nodule</p> <p><b>Exclusion criteria:</b><br/>(1) participants who had undergone baseline chest CT scans before enrolment; (2) Lung-RADS category 1 (i.e., calcified nodules); (3) masses larger than 30 mm, which were not of interest for a prediction model; (4) participants with missing values for the model inputs</p> <p><b>Number of participants:</b> 4,578<br/><b>Median age, IQR (years):</b> 62, 59 to 67<br/><b>Sex:</b> 97% males</p> | <p><b>Intervention:</b> Lung-RADS<br/><b>Comparator:</b> PanCan<br/><b>Reference standard:</b> Lung cancer diagnosis, confirmed by pathology</p>   | <p><b>Outcome:</b><br/>Diagnostic test accuracy (AUC, sensitivity, specificity, PPV, NPV)</p> <p><b>Follow-up:</b> median 664 days, IQR 562 to 794 days</p> |
| <p>Sundaram et al. (2021)<sup>18</sup><br/>US</p> <p><b>Funding source:</b> Cancer Center Support Grant</p>  | <p>Retrospective analysis of a multi-centre randomized controlled trial (NLST)</p>                                     | <p><b>Inclusion criteria:</b> Patients in the LDCT arm of NLST, with a new positive finding (i.e., solid nodules) in any of the 3 screening years</p> <p><b>Exclusion criteria:</b> Participants with masses (i.e., lesions larger than 3 cm in maximum diameter), partly solid nodules or GGNs, a prior lung cancer diagnosis, and/or missing nodule size; and participants for whom age at smoking cessation was inconsistent with age at randomization</p> <p><b>Number of participants:</b> 6,956<br/><b>Age:</b> NR<br/><b>Sex:</b> NR</p>  | <p><b>Intervention:</b> Lung-RADS<br/><b>Comparator:</b> PanCan<br/><b>Reference standard:</b> Lung cancer diagnosis; as determined by needle biopsy, surgery, bronchoscopy, further imaging, or other diagnosis procedure</p> | <p><b>Outcome:</b><br/>Diagnostic test accuracy (AUC, sensitivity, specificity, PPV, NPV)</p> <p><b>Follow-up:</b> 2 years</p>                              |

| Study citation, country, funding source  | Study design  | Population characteristics   | Intervention and comparator(s)  | Clinical outcomes, length of follow-up   |
|--|---|--|---|--|
| Kessler et al. (2020) <sup>15</sup><br>US<br><b>Funding source:</b> NR   | Single-centre, retrospective diagnostic test accuracy study                 | <b>Inclusion criteria:</b> Patients eligible for screening, according to NLST and Centers for Medicare and Medicaid Services eligibility criteria<br><b>Exclusion criteria:</b> NR<br><b>Number of participants:</b> 486<br><b>Mean age, SD (years):</b> 63, 5<br><b>Sex:</b> 54% females  | <b>Intervention:</b> Lung-RADS<br><b>Comparator:</b> PanCan<br><b>Reference standard:</b> Lung cancer diagnosis; as determined by histologic evaluation, follow-up imaging, or clinical evaluation  | <b>Outcome:</b> Diagnostic test accuracy (AUC, sensitivity, specificity, PPV, NPV, accuracy)<br><b>Follow-up:</b> mean 40 months, SD 14 months |
| Hammer et al. (2019) <sup>13</sup><br>US<br><b>Funding source:</b> National Institutes of Health/ National Cancer Institute. | Retrospective analysis of a multi-centre randomized controlled trial (NLST) | <b>Inclusion criteria:</b> Patients in the LDCT arm of NLST, with at least 2 scans (i.e., baseline and at least 1 follow-up, with:<br>(a) subsolid pulmonary nodules comprising GGNs smaller than 10 mm; (b) GGNs measuring 10 mm or larger; and (c) part-solid nodules (“mixed” nodules) measuring 6 mm or larger<br><b>Exclusion criteria:</b> Patients with nodules that were not truly subsolid; patients with part-solid nodules smaller than 6 mm<br><b>Number of participants:</b> 434<br><b>Median age, range (years):</b> 62, 55 to 74<br><b>Sex:</b> 51% females | <b>Intervention:</b> Lung-RADS<br><b>Comparator:</b> PanCan<br><b>Reference standard:</b> Lung cancer diagnosis; as determined by needle biopsy, surgery (thoracotomy, thoracoscopy, or mediastinoscopy) bronchoscopy, chest CT imaging; PET/ chest MRI, or other diagnosis procedure | <b>Outcome:</b> Diagnostic test accuracy (AUC)<br><b>Follow-up:</b> 6.5 years  |
| Tremblay et al. (2019) <sup>19</sup><br>Canada<br><b>Funding source:</b> Alberta Cancer Foundation                           | Prospective, diagnostic test accuracy study                                 | <b>Inclusion criteria:</b> Participants meeting NSLT eligibility criteria (i.e., 55-74 years of age; ≥ 30 pack-years smoking history or quit smoking ≤ 15 years prior) or who were 55 to 80 years of age and had estimated 6-year lung cancer risk ≥ 1.5% using validated model (PLCOm2012)<br><b>Exclusion criteria:</b> Participants with nodules previously detected on off-study clinical scans<br><b>Number of participants:</b> 775<br><b>Mean age, range (years):</b> 63.3, 55 to 80<br><b>Sex:</b> 49.9% women   | <b>Intervention:</b> Lung-RADS<br><b>Comparator:</b> PanCan<br><b>Reference standard:</b> Lung cancer diagnosis; including histopathology   | <b>Outcome:</b> Diagnostic test accuracy (AUC, sensitivity, specificity, PPV, NPV)<br><b>Follow-up:</b> mean 763 days, SD 203 days             |

| Study citation, country, funding source   | Study design   | Population characteristics   | Intervention and comparator(s)   | Clinical outcomes, length of follow-up  |
|---|--|--|--|---|
| White et al. (2019) <sup>21</sup><br>US<br><b>Funding source:</b> No funding  | Retrospective analysis of a multi-centre randomized controlled trial (NLST)                  | <b>Inclusion criteria:</b> Patients meeting the NSLT eligibility criteria (i.e., 55-74 years of age; $\geq 30$ years smoking history or quit smoking $\leq 15$ years prior) with $\geq 4$ mm nodules found at initial prevalence screening<br><b>Exclusion criteria:</b> Patients with incomplete information<br><b>Number of participants:</b> 2,813 (4,408 nodules)<br><b>Mean age, SD (years):</b> 64.3, 5.2 (patients with malignant nodules); 62.1, 5.1 (patients with benign nodules)<br><b>Sex:</b> 41% women | <b>Intervention:</b> Lung-RADS<br><b>Comparator:</b> PanCan<br><b>Reference standard:</b> Lung cancer diagnosis; as determined by needle biopsy, surgery, bronchoscopy, further imaging, or other diagnosis procedure  | <b>Outcomes:</b> Diagnostic test accuracy (sensitivity, specificity, accuracy)<br><b>Follow-up:</b> 2 years     |
| Marshall et al. (2017) <sup>17</sup><br>Australia<br><b>Funding sources:</b> National Health and Medical Research Councils; Smart State Project Grant, Queensland Health; National Centre for Asbestos Related Diseases Project Grant; The Prince Charles Hospital Foundation | Retrospective analysis of a prospective cohort study (QLCSS)                                 | <b>Inclusion criteria:</b> Patients aged 60 to 74 years; with minimum lung function (i.e., forced expiratory volume in 1 second $\geq 50\%$ predicted); smokers (i.e., $\geq 30$ pack-years, current or quit within the past 15 years)<br><b>Exclusion criteria:</b> Patients with any medical comorbidity; CT scan within the prior 18 months; poor spirometry; lost to follow-up; or missing data<br><b>Number of participants:</b> 256<br><b>Median age (years):</b> 64.5<br><b>Sex:</b> 67% males                | <b>Intervention:</b> Lung-RADS<br><b>Comparator:</b> PanCan<br><b>Reference standard:</b> Lung cancer diagnosis; as determined by fine need aspirate, bronchoscopy, surgery, or other diagnostic procedure   | <b>Outcome:</b> Diagnostic test accuracy (AUC, sensitivity, specificity, PPV, NPV)<br><b>Follow-up:</b> 5 years |
| van Riel et al. (2017) <sup>20</sup><br>Denmark, Germany, the Netherlands<br><b>Funding source:</b> MeVis Medical Solutions AG  | Retrospective analysis of a randomized controlled trial (Danish Lung Cancer Screening Trial) | <b>Inclusion criteria:</b> Patients aged 50 to 70 years who were current or former smokers with a minimum smoking history of 20 pack-years, normal lung function, and nodules annotated by at least 1 screening radiologist<br><b>Exclusion criteria:</b> NR<br><b>Number of participants:</b> 613<br><b>Mean age, range (years):</b> 58, 50 to 75<br><b>Sex:</b> 53% women  | <b>Intervention:</b> Lung-RADS<br><b>Comparator:</b> PanCan<br><b>Reference standard:</b> Lung cancer diagnosis; as determined by lung cancer mortality, histology/tissue sampling, follow-up LDCT, chest CT or PET, staging, and other diagnostic procedure | <b>Outcome:</b> Diagnostic test accuracy (AUC)<br><b>Follow-up:</b> 9 years                                     |

| Study citation, country, funding source  | Study design  | Population characteristics   | Intervention and comparator(s)  | Clinical outcomes, length of follow-up   |
|--|---|--|---|--|
| Hawkins et al. (2016) <sup>14</sup><br>US<br><b>Funding sources:</b> US Public Health Service; Cancer Center Support Grant; State of Florida Department of Health. | Nested matched case-control study using data from a multi-centre randomized controlled trial (NLST) | <b>Inclusion criteria:</b> Patients in the LDCT arm of NLST with screen-detected lung cancer (cases) or cancer-free findings (controls)<br><b>Exclusion criteria:</b> NR<br><b>Number of participants:</b> 185<br><b>Mean age (years):</b> 64<br><b>Sex:</b> 53% females | <b>Intervention:</b> Lung-RADS<br><b>Comparator:</b> PanCan<br><b>Reference standard:</b> Lung cancer diagnosis; as determined by needle biopsy, surgery, bronchoscopy, further imaging, or other diagnosis procedure | <b>Outcome:</b> Diagnostic test accuracy (sensitivity, specificity, accuracy)<br><b>Follow-up:</b> 2 years |

AUC = area under the curve; GGN = ground-glass nodule; IQR = interquartile range; LDCT = low-dose CT; Lung-RADS = Lung Imaging Reporting and Data System; NLST = National Lung Screening Trial; NPV = negative predictive value; NR = not reported; PanCan = Pan-Canadian Early Detection of Lung Cancer; PPV = positive predictive value; QLCSS = Queensland Lung Cancer Screening Study; SD = standard deviation.

**Table 3: Characteristics of Included Economic Evaluations**

| Study citation country, funding source                                 | Type of analysis, time horizon, perspective   | Population characteristics   | Intervention and comparator(s)                                      | Approach   | Source of clinical, cost, and utility data used in analysis  | Main assumptions  |
|--|---|--|---|--|--|---|
| <p>Hammer et al. (2021)<sup>22</sup><br/>US<br/>Funding source: NR</p> | <p><b>Analysis:</b> Cost-utility analysis.<br/><b>Time horizon:</b> Lifetime<br/><b>Perspective:</b> Health care system and society</p> | <p>A hypothetical cohort of 10 million current and former smokers undergoing LDCT lung cancer screening who are assumed to have a ground-glass nodules at baseline CT. Patient age range was 55 to 75 years at the beginning of screening, and 49% of patients were men.</p> | <p><b>Intervention:</b> Lung-RADS<br/><b>Comparator:</b> PanCan</p> | <p>A state-transition Monte Carlo simulation model with a monthly cycle to investigate the effectiveness of the nodule management guidelines for non-solid nodules. Nodules could grow and develop solid components.</p> | <p>Initial nodule size, nodule growth rates, and the potential for the development of a solid component were determined using data from the literature. Patient characteristics were generated from primary analysis of the NLST data. Age-dependent mortality rates were derived from the National Health Interview Survey, Substance Abuse and Mental Health Services, American Cancer Society cancer prevention studies, and Berkeley Mortality Database. Incidental cancer rates dependent on patient age and smoking status were derived by using the Lung Cancer Policy Model, and the Smoking History Generator was used to model individuals' smoking history. The QALYs and costs were discounted by 3% per year.</p> | <p>Nodules meeting criteria for Lung-RADS category 4B or category 4X at follow-up CT proceed to definitive treatment in the model, as per the assumption that they started out as GGNs and must have grown and/or developed solid components to meet these criteria.</p> <p>Patients are assumed to have a single nodule for purposes of PanCan.</p> <p>A willingness-to-pay threshold of \$100,000 per QALY was used, in keeping with recommendations from the literature for the US health care system.</p> |



| Study citation country, funding source                                 | Type of analysis, time horizon, perspective   | Population characteristics   | Intervention and comparator(s)   | Approach  | Source of clinical, cost, and utility data used in analysis  | Main assumptions  |
|--|---|--|--|---|--|---|
| <p>Hammer et al. (2021)<sup>23</sup><br/>US<br/>Funding source: NR</p> | <p><b>Analysis:</b> Cost-utility analysis.<br/><b>Time horizon:</b> Lifetime.<br/><b>Perspective:</b> Payers and policy-makers within a health care system.</p> | <p>A simulated cohort of 100,000 patients derived from a random subset of 151 patients who underwent LDCT lung cancer screening within the health care network and had been assigned a Lung-RADS category of 4A, 4B, or 4X. The median age was 66 years (range 61 to 71). Males made up 55% of the cohort.</p> | <p><b>Intervention:</b> Lung-RADS (using the following stratification methods: nodule size to determine initial categorization; PET and/or CT for category 4B or 4X nodules, and follow-up for category 4A nodules).<br/><b>Comparators:</b> BTS guideline using PanCan (using the following stratification methods: Brock risk score for initial categorization; PET or CT for high Brock risk score, follow-up for low Brock risk score, biopsy for intermediate PET and/or CT results); or CHEST guidelines using PanCan (using the following stratification methods:</p> | <p>Multivariable logistic regression analysis to predict the results of PET or CT and of follow-up chest CT from patient and nodule characteristics as well as the nodule diagnosis (benign vs. malignant).</p> | <p>Cancer treatment costs were derived from a previous cost-effectiveness study.<sup>27</sup> Costs and QALYs were discounted at 3% per year. Baseline life expectancies by age and sex for smokers were obtained from the State Board of Administration of Florida. Survival of localized lung cancer by cancer stage was derived from the IASLC Lung Cancer Staging Project.</p> | <p>Survival of patients with clinical tumor size 0 cancer was assumed to be 100%, thus median survival was set at 99 years. Survival of patients with metastatic disease was estimated at 5% at 5 years, yielding a median survival of 1.15 years. For a given patient, survival was calculated as the minimum of age-based survival and, if the nodule was malignant, then it was calculated as cancer-based survival (in other words, a patient could not survive past his/her age-based life expectancy). For benign nodules, a surgery-related mortality rate of 2% and a biopsy-related mortality rate of 0.2% were implemented stochastically.<br/><br/>For patients who underwent surgery for a benign nodule, the utility was assumed to be 0.9 for that year, then normal (utility value not provided) thereafter.</p> |

| Study citation country, funding source | Type of analysis, time horizon, perspective | Population characteristics | Intervention and comparator(s)  | Approach | Source of clinical, cost, and utility data used in analysis | Main assumptions   |
|--|---|----------------------------|---|----------|---|--|
|  |   |                            | Brock score for initial stratification; PET or CT for intermediate Brock risk score, follow-up for low Brock risk score). |          |   | Any lung cancers that were assigned a benign diagnosis by the management algorithm were assumed to be detected at a follow-up CT performed as part of the lung cancer screening program 1 year after the work-up. Lung cancers were assumed to grow during this period, but the rate of growth was dependent upon nodule characteristics. The growth rate of a lung cancer was assumed to be 150 days, approximately the average rate of solid lung cancers. |

BTS = British Thoracic Society; CHEST = American College of Chest Physicians; GGN = ground-glass nodule; IALSC = International Association for the Study of Lung Cancer; ICER = incremental cost-effectiveness ratio; LDCT = low-dose CT; Lung-RADS = Lung Imaging Reporting and Data System; NLST = National Lung Screening Trial; NR = not reported; PanCan = Pan-Canadian Early Detection of Lung Cancer; QALY = quality-adjusted life-year; vs. = versus.

Table 4: Characteristics of Included Guidelines

| Intended users, target population   | Intervention and practice considered   | Major outcomes considered  | Evidence collection, selection, and synthesis  | Evidence quality assessment  | Recommendations development and evaluation  | Guideline validation  |
|---|--|--|--|--|---|---|
| <b>British Thoracic Society Guideline (2015)<sup>24</sup></b>   |  |  |  |  |   |   |
| <p><b>Intended users:</b><br/>Practitioners within the UK, including physicians, general practitioners, nurses, radiologists, surgeons and other health care professionals</p> <p><b>Target population:</b><br/>Adults with pulmonary nodules</p> | <p>Risk assessment for malignancy based on clinical and radiological factors; management of subsolid nodules</p> | <p>Risk of cancer; management of subsolid nodules; lung cancer diagnosis; diagnostic test accuracy of risk calculators</p> | <p><b>Evidence collection:</b><br/>Systematic electronic database searches were conducted using Ovid MEDLINE, Ovid Embase and the Cochrane Library to identify potentially relevant studies for inclusion in the guideline. The searches were first run in November 2012 and updated in June 2014.</p> <p><b>Evidence selection:</b><br/>Literature retrieved in the electronic searches was screened for relevance by 2 reviewers</p> <p><b>Evidence synthesis:</b> The body of evidence for each recommendation was summarized into evidence statements and graded using the Scottish Intercollegiate Guidelines Network grading system.</p> | <p>Appraisal was performed to be compliant with the AGREE collaboration. Two guideline reviewers independently appraised each paper using the Scottish Intercollegiate Guidelines Network critical appraisal checklists.</p> | <p>Recommendations were graded according to the strength of the evidence:</p> <p>A: At least 1 meta-analysis, systematic review or RCT rated as 1++ and directly applicable to the target population; or ≥ 1 systematic reviews or RCTs; or a body of evidence consisting principally of studies rated as 1+ directly applicable to the target population and demonstrating overall consistency of results.</p> <p>B: A body of evidence including studies rated as 2++ directly applicable to the target population and demonstrating overall consistency of results; or extrapolated evidence from studies rated as 1++ or 1+.</p> <p>C: A body of evidence including studies rated as 2+ directly applicable to the target population and demonstrating overall consistency of results; or extrapolated evidence from studies rated as 2++.</p> <p>D: Evidence level 3 or 4; or extrapolated evidence from studies rated as 2+.</p> <p>Checkmark: Important practical points for which there is no research evidence, nor is there likely to be any research evidence.</p> | <p>The draft guideline was made available online for public consultation and feedback was invited from stakeholder organizations.</p> |

AGREE = Appraisal of Guidelines for Research & Evaluation; RCT = randomized controlled trial.

## Appendix 3: Critical Appraisal of Included Publications

Note that this appendix has not been copy-edited.

**Table 5: Strengths and Limitations of Diagnostic Test Accuracy Studies Using the QUADAS-2 Checklist<sup>9</sup>**

| Strengths   | Limitations  |
|---|--|
| <b>Kim et al. (2021)<sup>16</sup></b>   |  |
| <ul style="list-style-type: none"> <li>• The objectives, intervention, comparator, and outcomes were clearly described</li> <li>• Patients from the original RCT were reviewed for eligibility into the study</li> <li>• A case-control study design was avoided</li> <li>• Inclusion and exclusion criteria were included</li> <li>• Inappropriate exclusion criteria were avoided</li> <li>• The 2 screening tests, their conduct and interpretation matched the review question</li> <li>• The thresholds used for the screening tests were pre-specified</li> <li>• The target condition as defined by the reference standard matches the question</li> <li>• Patient and nodule characteristics were clearly described</li> <li>• All eligible patients were included in the analysis</li> <li>• Study participants, care providers, and setting appeared to be representative of the population and care setting of interest</li> <li>• The authors declared that they had no potential conflicts of interest</li> <li>• Source of funding was disclosed</li> </ul> | <ul style="list-style-type: none"> <li>• It was unclear if the thoracic radiologists conducting the screening were blinded to the patient’s final diagnosis</li> <li>• Some patients had a follow-up period (&lt; 2 years) that was shorter than the minimum follow-up period required to assess the presence or absence of lung cancer</li> <li>• The findings of this Korean-based study may not be generalizable to the Canadian health system (e.g., incidence of lung cancer in South Korea is lower than in the US)</li> </ul> |

| Strengths   | Limitations   |
|---|---|
| <b>Sundaram et al. (2021)<sup>18</sup></b>  |   |
| <ul style="list-style-type: none"> <li>• The objectives, intervention, comparison, and outcomes were clearly described</li> <li>• Patients from the original RCT were reviewed for eligibility into the study</li> <li>• A case-control study design was avoided</li> <li>• Inclusion and exclusion criteria were included</li> <li>• The screening tests, their conduct and interpretation matched the review question</li> <li>• The thresholds used for the screening tests were pre-specified</li> <li>• The target condition as defined by the reference standard matches the question</li> <li>• All eligible patients were included in the analysis</li> <li>• Study participants, care providers, and setting appeared to be representative of the population and care setting of interest</li> <li>• The authors declared that they had no potential conflicts of interest</li> <li>• Source of funding was disclosed</li> </ul> | <ul style="list-style-type: none"> <li>• The analysis was limited to solid nodules; patients with non-solid and ground-glass nodules were excluded</li> <li>• It was unclear if the screening tests were conducted without knowledge of the reference standard</li> <li>• Patient characteristics were not reported</li> <li>• Limitations of the original trial:               <ul style="list-style-type: none"> <li>◦ Scanners used in the trial were less technologically advanced than scanners currently available</li> <li>◦ The trial was conducted at a variety of medical institutions, many of which were recognized for their expertise in radiology and in the diagnosis of cancer; applicability to community facilities is uncertain</li> <li>◦ The reference standard results were assigned to a lobe rather than a nodule and it is uncertain if a malignancy identified within the lobe was the result of the most suspicious nodule seen at that time or of a new (incident) nodule</li> </ul> </li> </ul> |
| <b>Kessler et al. (2020)<sup>15</sup></b>   |   |
| <ul style="list-style-type: none"> <li>• The objectives, intervention, comparison, and main outcomes were clearly described</li> <li>• A case-control study design was avoided</li> <li>• Inclusion criteria for screening were reported</li> <li>• The screening tests, their conduct and interpretation matched the review question</li> <li>• The thresholds used for the screening tests were pre-specified</li> <li>• The target condition as defined by the reference standard matched the question</li> <li>• Patient and nodule characteristics were clearly described</li> <li>• All eligible patients were included in the analysis</li> <li>• Study participants, care providers, and setting appeared to be representative of the population and care setting of interest</li> <li>• The authors declared that they had no potential conflicts of interest</li> </ul>   | <ul style="list-style-type: none"> <li>• Exclusion criteria were not reported</li> <li>• It was unclear if the radiologists were blinded to the results of the reference standard</li> <li>• There were missing data points in the study population (e.g., family history), which were required for the PanCan score, but not for Lung-RADS</li> <li>• The PanCan was used to assess cancer probability on a per-patient basis, rather than a per-nodule basis</li> <li>• There were lower rates of follow-up in this clinical population, compared to trial populations</li> <li>• The source of funding was not disclosed</li> <li>• Single-centre study in Bronx, New York; the generalizability to the Canadian setting was unclear</li> </ul>  |

| Strengths  | Limitations  |
|--|--|
| <b>Hammer et al. (2019)<sup>13</sup></b>   |  |
| <ul style="list-style-type: none"> <li>• The objectives, intervention, comparator, and outcomes were clearly described</li> <li>• A case-control study design was avoided</li> <li>• Inclusion and exclusion criteria were included</li> <li>• Patient and nodule characteristics were clearly described</li> <li>• The radiologists conducting the screening algorithms were blinded to the patient's final diagnosis</li> <li>• The screening tests, their conduct and interpretation matched the review question</li> <li>• The thresholds used for the screening tests were pre-specified</li> <li>• The target condition (lung cancer) as defined by the reference standard matched the question</li> <li>• Study participants, care providers, and setting appeared to be representative of the population and care setting of interest</li> <li>• The authors declared that they had no potential conflicts of interest</li> <li>• Source of funding was disclosed</li> </ul> | <ul style="list-style-type: none"> <li>• Rather than enrolling a random sample of eligible patients; a random sample of patients from the original trial was reviewed for eligibility into the study</li> <li>• The analysis was limited to subsolid nodules and part-solid; patients with solid nodules were excluded</li> <li>• Limitations of the original trial:               <ul style="list-style-type: none"> <li>◦ Scanners used in the trial were less technologically advanced than scanners currently available</li> <li>◦ The trial was conducted at a variety of medical institutions, many of which were recognized for their expertise in radiology and in the diagnosis of cancer; applicability to community facilities is uncertain</li> <li>◦ The reference standard results were assigned to a lobe rather than a nodule and it is uncertain if a malignancy identified within the lobe was the result of the most suspicious nodule seen at that time or of a new (incident) nodule</li> </ul> </li> </ul> |
| <b>Tremblay et al. (2019)<sup>19</sup></b>   |  |
| <ul style="list-style-type: none"> <li>• The objectives, intervention, comparison, and main outcomes were clearly described</li> <li>• Patients were enrolled consecutively into the study</li> <li>• A case-control study design was avoided</li> <li>• Inclusion criteria were included</li> <li>• Inappropriate exclusion criteria were avoided</li> <li>• The screening tests, their conduct and interpretation matched the review question</li> <li>• The thresholds used for the screening tests were pre-specified</li> <li>• The target condition as defined by the reference standard matches the question</li> <li>• Patient and nodule characteristics were reported</li> <li>• All eligible patients were included in the analysis</li> <li>• Study participants, care providers, and setting appeared to be representative of the population and care setting of interest</li> <li>• The source of funding was disclosed</li> </ul>                                     | <ul style="list-style-type: none"> <li>• It was unclear if the screening tests were conducted without knowledge of the patient's final diagnosis</li> <li>• Two authors have a copyright for 1 of the screening tests (PanCan)</li> </ul>  |

| Strengths  | Limitations  |
|--|--|
| <b>White et al. (2019)<sup>21</sup></b>  |  |
| <ul style="list-style-type: none"> <li>• The objectives, intervention, comparison, and main outcomes were clearly described</li> <li>• Patients from the original RCT were reviewed for eligibility</li> <li>• A case-control study design was avoided</li> <li>• Inclusion and exclusion criteria were included</li> <li>• The screening tests, their conduct and interpretation matched the review question</li> <li>• The thresholds used for the screening tests were pre-specified</li> <li>• The target condition as defined by the reference standard matches the question</li> <li>• Patient and nodule characteristics were clearly described</li> <li>• Study participants, care providers, and setting appeared to be representative of the population and care setting of interest</li> <li>• The authors declared that they had no potential conflicts of interest</li> <li>• The authors disclosed that there was no funding for the study</li> </ul>  | <ul style="list-style-type: none"> <li>• Several patients were excluded because of missing non-nodule characteristics required for screening with PanCan (e.g., family history) but not required for Lung-RADS</li> <li>• Several patients were excluded because there was ambiguity about which 1 of multiple nodules were malignant</li> <li>• It was unclear if the screening tests were conducted without knowledge of the results of the reference standard</li> <li>• Limitations of the original trial:               <ul style="list-style-type: none"> <li>◦ Scanners used in the trial were less technologically advanced than scanners currently available</li> <li>◦ The trial was conducted at a variety of medical institutions, many of which were recognized for their expertise in radiology and in the diagnosis of cancer; applicability to community facilities is uncertain</li> <li>◦ The reference standard results were assigned to a lobe rather than a nodule and it is uncertain if a malignancy identified within the lobe was the result of the most suspicious nodule seen at that time or of a new (incident) nodule</li> </ul> </li> </ul> |
| <b>Marshall et al. (2017)<sup>17</sup></b>   |  |
| <ul style="list-style-type: none"> <li>• The objectives, intervention, comparison, and main outcomes were clearly described</li> <li>• A consecutive sample of patients were enrolled in the study</li> <li>• A case-control study design was avoided</li> <li>• Inclusion and exclusion criteria were included</li> <li>• Inappropriate exclusion criteria were avoided</li> <li>• The screening tests, their conduct and interpretation matched the review question</li> <li>• The thresholds used for the screening tests were pre-specified</li> <li>• The target condition as defined by the reference standard matches the question</li> <li>• Patient and nodule characteristics were clearly described</li> <li>• All eligible patients were included in the analysis</li> <li>• Study participants, care providers, and setting appeared to be representative of the population and care setting of interest</li> <li>• The authors declared that they had no potential conflicts of interest</li> <li>• Sources of funding were disclosed</li> </ul> | <ul style="list-style-type: none"> <li>• It was unclear if the screening tests were conducted without knowledge of the patient's final diagnosis</li> <li>• The generalizability of this Australian study to the Canadian setting were unclear</li> </ul>  |

| Strengths   | Limitations   |
|---|---|
| <b>van Riel et al. (2017)<sup>20</sup></b>  |   |
| <ul style="list-style-type: none"> <li>• The objectives, intervention, comparison, and main outcomes were clearly described</li> <li>• Patients from the original RCT were reviewed for eligibility into the study</li> <li>• A case-control study design was avoided</li> <li>• Inclusion criteria were included</li> <li>• The screening tests, their conduct and interpretation matched the review question</li> <li>• The thresholds used for the screening tests were pre-specified</li> <li>• The target condition as defined by the reference standard matches the question</li> <li>• Patient and nodule characteristics were reported</li> <li>• All eligible patients were included in the analysis</li> <li>• Study participants, care providers, and setting appeared to be representative of the population and care setting of interest</li> <li>• The source of funding was disclosed</li> </ul> | <ul style="list-style-type: none"> <li>• It was unclear if the screening tests were conducted without knowledge of the results of the reference standard</li> <li>• Due to lack of follow-information, nodule growth size was excluded; because growth size is a criterion in Lung-RADS (unlike PanCan), risk calculation may have been restricted</li> <li>• The authors disclosed conflicts of interest</li> <li>• The original trial was conducted in a single hospital in Denmark; generalizability to the Canadian setting is unclear</li> </ul> |
| <b>Hawkins et al. (2016)<sup>14</sup></b>   |   |
| <ul style="list-style-type: none"> <li>• The screening tests, their conduct and interpretation matched the review question</li> <li>• The thresholds used for the screening tests were pre-specified</li> <li>• The target condition as defined by the reference standard matches the question</li> <li>• The authors declared that they had no potential conflicts of interest</li> <li>• Sources of funding were disclosed</li> </ul>   | <ul style="list-style-type: none"> <li>• A case-control study design was used</li> <li>• Comparison of the 2 screening tests was not described in the objectives or methods; rather, it was described in a subsection (Risk Score) of the results section</li> <li>• It was unclear how patients (images) were selected for assessment of diagnostic test accuracy</li> <li>• Patient characteristics were not reported</li> <li>• It was unclear if the screening algorithms were conducted without knowledge of diagnostic results</li> </ul>       |

QUADAS-2 = Quality Assessment of Diagnostic Accuracy Studies 2; Lung-RADS = Lung Imaging Reporting and Data System; NLST = National Lung Screening Trial; PanCan = Pan-Canadian Early Detection of Lung Cancer nodule risk calculation; RCT = randomized controlled trial.



**Table 6: Strengths and Limitations of Economic Evaluations Using the Drummond Checklist<sup>10</sup>**

| Strengths   | Limitations  |
|---|--|
| <b>Hammer et al. (2021)<sup>22</sup></b>  |  |
| <p><b>Study design:</b></p> <ul style="list-style-type: none"> <li>• The research question was stated</li> <li>• The economic importance of the research question was stated</li> <li>• The viewpoint of the analysis was clearly stated and justified</li> <li>• The choice of form of economic evaluation was justified in relation to the questions addressed</li> </ul> <p><b>Data collection:</b></p> <ul style="list-style-type: none"> <li>• The sources of effectiveness estimates used were stated</li> <li>• The primary outcome measures for the economic evaluation were clearly stated</li> <li>• Details of the subjects from whom valuations were obtained were given</li> <li>• Methods for the estimation of quantities and unit costs were described</li> <li>• Currency and price data were recorded</li> <li>• Details of the simulation model were given</li> </ul> <p><b>Analysis and interpretation:</b></p> <ul style="list-style-type: none"> <li>• Time horizon of costs and benefits were stated</li> <li>• The discount rate was stated</li> <li>• Details of statistical tests were given</li> <li>• The approach to sensitivity analysis was given</li> <li>• The choice of variables for sensitivity analysis were justified</li> <li>• The ranges over which the variables were varied were justified</li> <li>• Incremental analysis was reported</li> <li>• The answer to the study question was given</li> <li>• Conclusions following from the data were reported</li> <li>• Conclusions were accompanied by the appropriate caveats</li> </ul> <p><b>Miscellaneous:</b></p> <ul style="list-style-type: none"> <li>• Authors stated that they had no conflicts of interest related to the study</li> </ul> | <ul style="list-style-type: none"> <li>• No description of current price adjustments for inflation was provided</li> <li>• No justification for the selected discount rate was provided</li> <li>• Confidence intervals for costs and QALY were not reported</li> <li>• Source of funding was not disclosed</li> </ul> |

| Strengths   | Limitations   |
|---|---|
| <b>Hammer et al. (2021)<sup>23</sup></b>  |   |
| <p><b>Study design:</b></p> <ul style="list-style-type: none"> <li>• The research question was stated</li> <li>• The economic importance of the research question was stated</li> <li>• The viewpoint of the analysis was clearly stated and justified</li> <li>• The choice of form of economic evaluation was justified in relation to the questions addressed</li> </ul> <p><b>Data collection:</b></p> <ul style="list-style-type: none"> <li>• The sources of effectiveness estimates used were stated</li> <li>• The primary outcome measures for the economic evaluation were clearly stated</li> <li>• Details of the subjects from whom valuations were obtained were given</li> <li>• Methods for the estimation of quantities and unit costs were described</li> <li>• Currency and price data were recorded</li> <li>• Details of the predictive logistic regression models were given</li> </ul> <p><b>Analysis and interpretation:</b></p> <ul style="list-style-type: none"> <li>• Time horizon of costs and benefits was stated</li> <li>• The discount rate was stated</li> <li>• Details of statistical tests and confidence intervals were given</li> <li>• The approach to sensitivity analysis was given</li> <li>• The choice of variables for sensitivity analysis was justified</li> <li>• The ranges over which the variables are varied were justified</li> <li>• Incremental analysis was reported</li> <li>• The answer to the study question was given</li> <li>• Conclusions following from the data reported</li> <li>• Conclusions were accompanied by the appropriate caveats</li> </ul> <p><b>Miscellaneous:</b></p> <ul style="list-style-type: none"> <li>• Authors stated that they had no conflicts of interest related to the study</li> </ul> | <ul style="list-style-type: none"> <li>• Model inputs were taken from single studies, rather than a synthesis or meta-analysis of estimates from multiple sources</li> <li>• No description of current price adjustments for inflation was provided</li> <li>• No justification for the selected discount rate was provided</li> <li>• Sources of funding were not disclosed</li> </ul> |

QALY = quality-adjusted life-year.

**Table 7: Strengths and Limitations of Guideline Using AGREE II<sup>11</sup>**

| Item  | BTS Guideline, 2015 <sup>24</sup> |
|---|-----------------------------------|
| <b>Domain 1: Scope and Purpose</b>  |                                   |
| 1. The overall objective(s) of the guideline is (are) specifically described.                                 | Yes                               |
| 2. The health question(s) covered by the guideline is (are) specifically described.                           | Yes                               |
| 3. The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described. | Yes                               |
| <b>Domain 2: Stakeholder Involvement</b>  |                                   |
| 4. The guideline development group includes individuals from all relevant professional groups.                | Yes                               |
| 5. The views and preferences of the target population (patients, public, etc.) have been sought.              | Yes                               |
| 6. The target users of the guideline are clearly defined.   | Yes                               |
| <b>Domain 3: Rigour of Development</b>  |                                   |
| 7. Systematic methods were used to search for evidence.   | Yes                               |
| 8. The criteria for selecting the evidence are clearly described.   | Yes                               |
| 9. The strengths and limitations of the body of evidence are clearly described.                               | Yes                               |
| 10. The methods for formulating the recommendations are clearly described.                                    | Yes                               |
| 11. The health benefits, side effects, and risks have been considered in formulating the recommendations.     | Yes                               |
| 12. There is an explicit link between the recommendations and the supporting evidence.                        | Yes                               |
| 13. The guideline has been externally reviewed by experts before its publication.                             | Yes                               |
| 14. A procedure for updating the guideline is provided.   | Yes                               |
| <b>Domain 4: Clarity of Presentation</b>  |                                   |
| 15. The recommendations are specific and unambiguous.   | Yes                               |
| 16. The different options for management of the condition or health issue are clearly presented.              | Yes                               |
| 17. Key recommendations are easily identifiable.  | Yes                               |
| <b>Domain 5: Applicability</b>  |                                   |
| 18. The guideline describes facilitators and barriers to its application.                                     | Unclear                           |
| 19. The guideline provides advice and/or tools on how the recommendations can be put into practice.           | Yes                               |
| 20. The potential resource implications of applying the recommendations have been considered.                 | Yes                               |
| 21. The guideline presents monitoring and/or auditing criteria.   | Yes                               |
| <b>Domain 6: Editorial Independence</b>   |                                   |
| 22. The views of the funding body have not influenced the content of the guideline.                           | Unclear                           |
| 23. Competing interests of guideline development group members have been recorded and addressed.              | Yes                               |

AGREE II = Appraisal of Guidelines for Research & Evaluation II.

## Appendix 4: Main Study Findings and Authors' Conclusions

Note that this appendix has not been copy-edited.

**Table 8: Summary of Findings by Outcome – Diagnostic Test Accuracy, AUC**

| Strategy  | AUC (95% CI)                    |                                      |                                     |                                    |                                      |                                      |                                      |
|-----------|---------------------------------|--------------------------------------|-------------------------------------|------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|
|           | Kim et al. (2021) <sup>16</sup> | Sundaram et al. (2021) <sup>18</sup> | Kessler et al. (2020) <sup>15</sup> | Hammer et al. (2019) <sup>13</sup> | Tremblay et al. (2019) <sup>19</sup> | Marshall et al. (2017) <sup>17</sup> | van Riel et al. (2017) <sup>20</sup> |
| Lung-RADS | 0.95 (0.91, 0.99)               | 0.84 (0.81, 0.86)                    | 0.87 (0.84, 0.90)                   | 0.70 (0.60, 0.80)                  | 0.93 (0.89, 0.98)                    | 0.84 (IQR 0.69-0.98)                 | 0.81 (NR)                            |
| PanCan    | 0.96 (0.92, 0.99)               | 0.85 (0.82, 0.87)                    | 0.88 (0.85, 0.91)                   | 0.78 (0.67, 0.85)                  | 0.97 (0.95, 1.0)                     | 0.90 (IQR 0.75-1.0)                  | 0.87 (NR)                            |
| P value   | 0.34                            | 0.17                                 | NR                                  | 0.09                               | ns                                   | 0.25                                 | 0.003                                |

AUC = area under the curve; CI = confidence interval; IQE = interquartile range; Lung-RADS = Lung Imaging Reporting and Data System; NR = not reported; PanCan = Pan-Canadian Early Detection of Lung Cancer.

**Table 9: Summary of Findings by Outcome – Diagnostic Test Accuracy, Diagnostic Parameters**

| Strategy                        | Diagnostic parameters (95% CI)   |  |   |  |   |  |  |
|---------------------------------|--|--|---|--|---|--|--|
|                                 | Kim et al. (2021) <sup>16</sup>  | Sundaram et al. (2021) <sup>18</sup>   | Kessler et al. (2020) <sup>15</sup>   | Tremblay et al. (2019) <sup>19</sup>   | White et al. (2019) <sup>21</sup>                               | Marshall et al. (2017) <sup>17</sup>   | Hawkins et al. (2016) <sup>14</sup>        |
| Lung-RADS, 3 and 4A/4B positive | Sens: 95.0 (88.2, 100)<br>Spec: 76.7 (75.8, 77.5)<br>PPV: 1.6 (1.1, 2.1)<br>NPV: 100 (99.9, 100)   | Sens: 0.81 (0.76, 0.85)<br>Spec: 0.78 (0.77, 0.79)<br>PPV: 0.16 (0.15, 0.17)<br>NPV: 0.99 (0.98, 0.99) | Sens: 84.2 (68.1, 93.4)<br>Spec: 79.2 (75.1, 82.8)<br>PPV: 25.5 (18.4, 34.3)<br>NPV: 98.3 (96.2, 99.3)<br>Accuracy: 79.6 (75.8, 83.1) | Sens: 76.2 (52.8, 91.8)<br>Spec: 92.6 (90.5, 94.3)<br>PPV: 22.2 (16.8, 28.8)<br>NPV: 99.3 (98.5, 99.7) | NR  | Sens: 100 (47.8, 100)<br>Spec: 79.6 (74.1, 84.4)<br>PPV: 8.9 (3.0, 19.6)<br>NPV: 100 (98.2, 100) | NR   |
| Lung-RADS, 4A/4B positive       | Sens: 87.5 (77.3, 97.7)<br>Spec: 93.3 (92.8, 93.7)<br>PPV: 5.0 (3.4, 6.6)<br>NPV: 99.9 (99.9, 100) | Sens: 0.48 (0.42, 0.53)<br>Spec: 0.94 (0.93, 0.94)<br>PPV: 0.29 (0.26, 0.32)<br>NPV: 0.97 (0.97, 0.97) | Sens: 58.0 (40.8, 73.7)<br>Spec: 98.0 (96.2, 99.1)<br>PPV: 71.0 (54.8, 83.1)<br>NPV: 96.5 (95.0, 97.6)<br>Accuracy: 94.9 (92.5, 96.6) | NR   | Sens: 87 (80, 93)<br>Spec: 83 (82, 84)<br>Accuracy: 76 (75, 78) | NR   | Sens: 22.4<br>Spec: 93.7<br>Accuracy: 71.4 |

| Strategy                      | Diagnostic parameters (95% CI)   |  |   |  |   |   |  |
|-------------------------------|--|--|---|--|---|---|--|
|                               | Kim et al. (2021) <sup>16</sup>  | Sundaram et al. (2021) <sup>18</sup>   | Kessler et al. (2020) <sup>15</sup>   | Tremblay et al. (2019) <sup>19</sup>   | White et al. (2019) <sup>21</sup>                               | Marshall et al. (2017) <sup>17</sup>  | Hawkins et al. (2016) <sup>14</sup>        |
| <b>PanCan, 1.5% threshold</b> | NR   | Sens: 0.78 (0.73, 0.82)<br>Spec: 0.82 (0.81, 0.83)<br>PPV: 0.19 (0.17, 0.20)<br>NPV: 0.99 (0.98, 0.99) | Sens: 79.0 (62.7, 90.4)<br>Spec: 83.9 (80.2, 87.2)<br>PPV: 29.4 (24.2, 35.3)<br>NPV: 97.9 (96.2, 98.9)<br>Accuracy: 83.5 (79.9, 86.7) | NR   | NR  | NR  | NR   |
| <b>PanCan, 5% threshold</b>   | Sens: 87.5 (77.3, 97.7)<br>Spec: 92.3 (91.8, 92.9)<br>PPV: 4.4 (3.0, 5.8)<br>NPV: 99.9 (99.9, 100) | Sens: 0.51 (0.46, 0.57)<br>Spec: 0.93 (0.92, 0.94)<br>PPV: 0.29 (0.26, 0.43)<br>NPV: 0.97 (0.97, 0.98) | Sens: 73.7 (56.9, 86.6)<br>Spec: 93.5 (90.8, 95.6)<br>PPV: 49.1 (39.3, 59.0)<br>NPV: 97.7 (96.1, 98.6)<br>Accuracy: 92.0 (89.7, 94.6) | Sens: 90.5 (69.6, 98.8)<br>Spec: 93.1 (91.1, 94.8)<br>PPV: 26.8 (21.4, 33.0)<br>NPV: 99.7 (99.0, 99.9) | Sens: 93 (86, 97)<br>Spec: 90 (89, 90)<br>Accuracy: 85 (84, 86) | NR  | NR   |
| <b>PanCan, 10% threshold</b>  | Sens: 82.5 (70.7, 94.3)<br>Spec: 95.9 (95.5, 96.2)<br>PPV: 7.4 (5.0, 9.9)<br>NPV: 99.9 (99.9, 100) | NR   | Sens: 65.8 (48.7, 80.4)<br>Spec: 95.8 (93.5, 97.5)<br>PPV: 56.8 (44.5, 68.3)<br>NPV: 97.1 (95.5, 98.1)<br>Accuracy: 93.4 (90.8, 97.3) | NR   | NR  | Sens: 100 (47.8, 100)<br>Spec: 94.8 (91.3, 97.2)<br>PPV: 27.8 (9.7, 53.5)<br>NPV: 100 (98.5, 100) | Sens: 46.5<br>Spec: 93.7<br>Accuracy: 78.9 |

CI = confidence interval; IQR = interquartile range; Lung-RADS = Lung Imaging Reporting and Data System; NPV = negative predictive value; PanCan = Pan-Canadian Early Detection of Lung Cancer nodule risk calculation; PPV = positive predictive value; Sens = sensitivity; Spec = specificity.

**Table 10: Summary of Findings of Included Economic Evaluation**

| Main study findings  | Authors' conclusion   |
|--|---|
| <b>Hammer et al. (2021)<sup>22</sup></b>   |   |
| <p>Costs</p> <ul style="list-style-type: none"> <li>• Lung-RADS: USD 10,150</li> <li>• PanCan 5%: USD 19,116</li> <li>• PanCan 10% USD 16,469</li> </ul> <p>QALY</p> <ul style="list-style-type: none"> <li>• Lung-RADS: 10.53</li> <li>• PanCan 5%: 10.48</li> <li>• PanCan 10%: 10.50</li> </ul> <p>ICER (95% CI)</p> <ul style="list-style-type: none"> <li>• Lung-RADS vs. PanCan: \$52,993 per QALY gained (\$44,407 - \$64,372)</li> </ul> <p>Sensitivity analyses</p> <ul style="list-style-type: none"> <li>• “sensitivity analyses showed similar ICERs as we varied multiple parameters.” (p. 591)<sup>22</sup></li> </ul>   | <p>“Lung CT Screening Reporting and Data System–based strategies perform better than strategies using the Brock risk calculator, with a 4-mm solid component size threshold yielding the greatest quality-adjusted life years (QALYs) at a willingness-to-pay threshold of \$100 000 per QALY.” (p. 593)<sup>22</sup></p>   |
| <b>Hammer et al. (2021)<sup>23</sup></b>   |   |
| <p>Costs (95% CI)</p> <ul style="list-style-type: none"> <li>• Lung-RADS: USD 81,329 (80,798 - 81,819)</li> <li>• BTS using PanCan: USD 82,362 (81,853 – 82,887)</li> <li>• CHEST using PanCan: USD 83,599 (83,107 – 84,101)</li> </ul> <p>QALY (95% CI)</p> <ul style="list-style-type: none"> <li>• Lung-RADS: 10.021 (10.007 – 10.057)</li> <li>• BTS using PanCan: 10.041 (10.025 - 10.065)</li> <li>• CHEST using PanCan: 10.035 (10.019 - 10.058)</li> </ul> <p>ICER (95% CI)</p> <ul style="list-style-type: none"> <li>• BTS using PanCan vs. Lung-RADS: \$52,634 per QALY gained (\$45,122 - \$60,619)</li> </ul> <p>Sensitivity analyses</p> <ul style="list-style-type: none"> <li>• “Under nearly all conditions, the only algorithms on the efficient frontier were BTS and Lung-RADS. The ICERs for BTS versus Lung-RADS were under \$100 000 for all scenarios except an increased life expectancy in patients without cancer, in which case the ICER was \$109 273. Under one condition, an increase in the growth rate of fast-growing lung cancers, the CHEST algorithm was on the efficient frontier and yielded higher QALY and cost than BTS; however, the ICER was very high at \$1 384 951.” (p. 4)<sup>23</sup></li> </ul> | <p>“We found that the two management algorithms on the efficient frontier were Lung-RADS and BTS, with BTS yielding the greatest QALYs. The advantage of the BTS algorithm was seen by its ICER compared with Lung-RADS in statistical analysis by confidence interval and all except one sensitivity analysis we performed; in a condition where the growth rate of faster growing nodules was increased, the ACCP algorithm yielded higher QALYs but at a substantial cost (ICER of over \$1 million).” (p. 4)<sup>23</sup></p> <p>“In conclusion, the BTS algorithm was the cost-effective option with the best outcomes for managing high-risk (Lung-RADS 4) pulmonary nodules. This finding held true under multiple sensitivity analyses, suggesting that it may be generalizable, at least within the United States health care system.” (p. 5)<sup>23</sup></p> |

ACCP/CHEST = American College of Chest Physicians; BTS = British Thoracic Society; CI = confidence interval; ICER = incremental cost-effectiveness ratio; Lung-RADS = Lung Imaging Reporting and Data System; PanCan = Pan-Canadian Early Detection of Lung Cancer; QALY = quality-adjusted life-year; USD = US dollar; vs. = versus.

**Table 11: Summary of Recommendations in Included Guideline**

| Guideline                          | Recommendations and supporting evidence   | Quality of evidence and strength of recommendations  |
|------------------------------------|---|--|
| BTS Guideline (2015) <sup>24</sup> | <p><b>Recommendation:</b> "Use the Brock model (full, with spiculation) for initial risk assessment of pulmonary nodules (<math>\geq 8</math> mm or <math>\geq 300</math> mm<sup>3</sup>) at presentation in people aged <math>\geq 50</math> who are smokers or former smokers." (p. ii17)<sup>24</sup></p> <p><b>Evidence statement:</b> Prediction models for pulmonary nodules based on clinical and radiological parameters have been externally validated. In the only validation study performed in a UK population, the Herder model (incorporating nodule FDG avidity) performed significantly better than other models (Mayo, Brock, Veterans Administration). In sub-centimetre nodules, the Brock score had the highest accuracy (AUC value).</p> | <p><b>Strength of recommendation:</b> Grade C</p> <p><b>Quality of evidence:</b> Evidence level 2+</p> |
|                                    | <p><b>Recommendation:</b> "Consider the Brock model (full, with spiculation) for initial risk assessment of pulmonary nodules (<math>\geq 8</math> mm or <math>\geq 300</math> mm<sup>3</sup>) in all patients at presentation." (p. ii17)<sup>24</sup></p> <p><b>Evidence statement:</b> The use of clinical prediction models is more accurate than clinicians' individual clinical judgment in estimating the probability of malignancy in patients with pulmonary nodules.</p>  | <p><b>Strength of recommendation:</b> Grade D</p> <p><b>Quality of evidence:</b> Evidence level 3</p>  |
|                                    | <p><b>Recommendation:</b> "Use the Brock risk prediction tool to calculate risk of malignancy in SSNs <math>\geq 5</math> mm that are unchanged at 3 months." (p. ii24)<sup>24</sup></p> <p><b>Evidence statement:</b> One prospective study that validated the Brock model included 1,672 SSNs. The guideline authors reported that the Brock model may underestimate risk of malignancy in SSN that persist at 3 months.</p>  | <p><b>Strength of recommendation:</b> Grade C</p> <p><b>Quality of evidence:</b> Evidence level 2+</p> |

AUC = area under the curve; BTS = British Thoracic Society; FDG = fluorodeoxyglucose; SSN = subsolid nodule; SSN = subsolid nodule.

## Appendix 5: References of Potential Interest

Note that this appendix has not been copy-edited.

### Diagnostic Test Accuracy Studies

#### *Alternative Intervention*

Gupta S, Jacobson FL, Kong CY, Hammer MM. Performance of Lung Nodule Management Algorithms for Lung-RADS Category 4 Lesions. *Acad Radiol.* 2021;28(8):1037-1042. [PubMed](#)

Gonzalez Maldonado S, Delorme S, Husing A, et al. Evaluation of Prediction Models for Identifying Malignancy in Pulmonary Nodules Detected via Low-Dose Computed Tomography. *JAMA Netw.* 2020;3(2):e1921221. [PubMed](#)

### Review Articles

Dziadziuszko K, Szurowska E. Pulmonary nodule radiological diagnostic algorithm in lung cancer screening. *Transl Lung Cancer Res.* 2021;10(2):1124-1135. [PubMed](#)

Rzyman W, Didkowska J, Dziedzic R, et al. Consensus statement on a screening programme for the detection of early lung cancer in Poland. *Adv Respir Med.* 2018;86(1):53-74. [PubMed](#)

Field JK, Marcus MW, Oudkerk M. Risk assessment in relation to the detection of small pulmonary nodules. *Transl Lung Cancer Res.* 2017;6(1):35-41. [PubMed](#)

### Guideline Documents: Methodology Not Reported

Veronesi G, Baldwin DR, Henschke CI, et al. Recommendations for Implementing Lung Cancer Screening with Low-Dose Computed Tomography in Europe. *Cancers (Basel).* 2020;12(6):24. [PubMed](#)

Wormanns D, Kauczor HU, Antoch G, et al. Joint Statement of the German Radiological Society and the German Respiratory Society on a Quality-Assured Early Detection Program for Lung Cancer with Low-Dose CT. *ROFO Fortschr Geb Rontgenstr Nuklearmed.* 2019;191(11):993-997. [PubMed](#)

### Additional References

Jonas DE, Reuland DS, Reddy SM, et al. Screening for Lung Cancer with Low-Dose Computed Tomography Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA.* 2021;325(10):971-987. [PubMed](#)

US Preventive Services Task Force, Krist AH, Davidson KW, et al. Screening for Lung Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA.* 2021;325(10):962-970. [PubMed](#)