An Intervention to Tag Findings Suspicious for Lung Cancer on Chest Computed Tomography Has Good Sensitivity and Number Needed to Diagnose

Jennifer R Dusendang, MPH1; Lori C Sakoda, PhD1; Thomas H Urbania, MD2; Sora Ely, MD3; Todd Osinski, MD2; Ashish Patel, MD3; Lisa J Herrinton, PhD1

INTRODUCTION

Each year, more than 200,000 US residents are diagnosed with lung cancer.1 Chest computed tomography (CT) is a key modality for diagnosing lung cancer. However, follow-up of findings suspicious for lung cancer may be hindered if the free-text radiology report is not clear or specific about recommended follow-up.2

In 2015, Kaiser Permanente Northern California implemented a standardized reporting system3 to tag findings suspicious for lung cancer seen on diagnostic chest CT in accordance with Fleischner guidelines.4-6 The intervention included a tagging system to automate forwarding of chest CT scans highly suspicious for lung cancer (#PUL5) combined with a review by a multidisciplinary review group. In a quasi-experimental evaluation, implementation of the tagging system combined with case management and multidisciplinary review of cases suspicious for malignancy was effective for increasing the diagnosis of early-stage lung cancer within 120 days (odds ratio, 1.24; 95% confidence interval [CI] 1.09–1.41).3

In this analysis, we assessed the frequency of lung cancer diagnosis associated with each #PUL category. In addition, for #PUL5 (suspicous for malignancy), we computed the sensitivity, specificity, and other performance characteristics in relation to lung cancer diagnosis within 120 days. Last, we conducted chart reviews to assess why some newly diagnosed lung cancer cases were assigned tags other than #PUL5.

PATIENTS AND METHODS

Kaiser Permanente Northern California is an integrated, community-based health-care system that provides comprehensive care to approximately 4.3 million members and currently performs 75,000 chest CT procedures each year that are read by > 300 diagnostic radiologists. All medical information is documented in an electronic medical record. The chest CT tagging intervention was built into the radiology dictation software. Tags use the prefix “#PUL” followed by a character (0–6 or X) that encodes specificity and predictive values, and number needed to diagnose. We also performed a chart review to assess why some patients diagnosed with lung cancer were not tagged #PUL5.

RESULTS

Of the 39,409 patients with a tagged CT report, 1105 (2.8%) had a new primary lung cancer diagnosis within 120 days. Among the 2255 patients tagged #PUL5, 821 were diagnosed with lung cancer, with a sensitivity of 74% (95% confidence interval, 72%–77%). The positive predictive value was 36% (35%–38%), number needed to diagnosis was 2.7 (2.6–2.9), and specificity and negative predictive values were > 95%. Chart review identified opportunities to improve system defaults and clarify concepts.

CONCLUSION

The intervention performed well but needed improvement. Automating CT reports is simple and generalizable, and enabled reduction of care gaps and system improvement.
whereas for others, the team makes recommendations to the primary care physician, with the care coordinator following up to ensure the plan is implemented.

A separate tag, #PUL6 (known lung cancer or suspected metastasis to the lung), is intended for patients whose lung cancer was diagnosed pathologically before their current chest CT or whose cancer is metastatic to the lung. For lung nodules that are benign, subsolid, or indeterminate, the Fleischner Society guidelines are added to the radiology report along with standard text containing follow-up recommendations for the ordering provider.

As detailed in a separate publication presenting a controlled cohort study that used a step-wedge design, the system resulted in several advantages, with the key advantage being the increased diagnosis of early-stage lung cancer. Other advantages include clearer documentation, review of suspicious cases by a multidisciplinary team, reduction in the burden to primary care providers, and the opportunity to measure performance improvement.

The study population included adult health plan members who underwent chest CT examination from August 2015 to July 2017 that was assigned a #PUL tag by the reading
RESULTS

The study included 39,409 patients with a tagged chest CT. As reported previously, nearly half were 60 to 80 years old at the time of the index CT, about 40% were nonwhite, 54% were female, and 80% had a preceding chest x-ray.\(^1\)

Six percent of CTs were tagged #PUL5 (suspicous for malignancy) and 2%, #PUL6 (known lung cancer or suspected metastasis (Table 2).\(^2\)) Lung cancer was diagnosed within 120 days of the index CT in 36.4% (CI, 34.4%–38.4%) of #PUL5 patients. It was also diagnosed in 11.7% (CI, 9.6%–13.8%) of #PUL6 patients (with some being metastatic to the lung and some not). In contrast, it was diagnosed in 1% to 2% tagged #PUL2 (infectious/inflammatory), #PUL3 (subsolid), or #PULX (technically limited). It was diagnosed in < 1% tagged #PUL0 (no nodule), #PUL1 (benign), or #PUL4 (indeterminate).

For the #PUL5 tag, the sensitivity for diagnosing lung cancer was 821 of 1105, or 74% (CI, 72%–77%); the positive predictive value was 821 of 2255, or 36% (CI, 35%–38%), and the NND was 2255 of 821 or 2.7 (CI, 2.6–2.9) (Table 3). The specificity for identifying cases without lung cancer and the negative predictive value exceeded 95%, as expected for rare diseases. As detailed in the following paragraph, #PUL6 was sometimes used incorrectly in place of #PUL5. To assess the opportunity to improve sensitivity through further training of radiologists in use of #PUL6, we combined #PUL5 and #PUL6 in a post hoc analysis, noting the sensitivity increased to 926 of 1105, or 84% (CI, 82%–86%), and the NND increased to 3153 of 926, or 3.4 (CI, 3.3–3.5).

Two hundred eighty-four cases were newly diagnosed with lung cancer but given a code other than #PUL5 (suspicous for malignancy). To understand this discordance, we performed 66 chart reviews. Reviews of #PUL0 (no nodule) (10 reviews of 56 cancers) showed that radiologists had not updated their templated reports from the default phrase “No Pulmonary Nodule” and overemphasized nodules (≤ 3 cm), leaving masses (> 3 cm) untagged. In addition, in 1 patient, a nodule was detected on the right lung, but the left lung was described as “no pulmonary nodule,” triggering the wrong code. #PUL1 (benign) (3 reviews of 3 cancers) included 1 case that was diagnosed before the CT exam, 1 judged to be granulomatous disease and 1 with a 13.5-mm nodule that was miscoded. #PUL2 (infectious/inflammatory) (10 reviews of 43 cancers) had findings suspicious for malignancy combined with infectious/inflammatory findings for which a repeat CT was ordered to assess resolution. For #PUL3 (subsolid) (10 reviews of 12 cancers), each report described ground-glass findings, but none mentioned a solid component, although the word “nodule” was sometimes used. Multiple ground-glass nodules were sometimes interpreted as infection. The study radiologist concluded that ground-glass nodules represent various clinical scenarios and noted that no guideline recommends upgrading ground-glass nodules to #PUL5 unless they are part solid or growing quickly, regardless of size. #PUL4 (indeterminate) (10 reviews of 44 cancers) had multiple nodules suggesting infection/inflammation, a large cavitated density, and small

---

radiologist.\(^3\) We examined only the first chest CT, which was defined as the index CT. We excluded patients who had a chest CT during the 2-year period before their index CT or who had a history of lung cancer.

Statistical analysis involved computation of percentages and their CIs. Within groups of patients defined by #PUL tag, we calculated the frequency of lung cancer diagnosis\(^4\) within 120 days of the initial CT examination, with 120 days selected a priori on the basis of the clinical workflow. For #PUL5 (suspicious for malignancy), we computed the sensitivity, specificity, positive predictive value, negative predictive value, number needed to diagnose (NND), and 95% CIs using exact methods and standard logit confidence intervals.\(^8-10\)

A separate tag, #PUL6 (known lung cancer or suspected metastasis to the lung), is intended for patients whose lung cancer was diagnosed pathologically before their current chest CT or whose cancer is metastatic to the lung. However, radiologists may not have complete information about the patient and may not review the patient’s chart to confirm a preceding cancer diagnosis. Because radiologists often coded #PUL6 in place of #PUL5, we conducted a post hoc sensitivity analysis combining the 2 tags to gain insight about how changes in the training of radiologists regarding use of the #PUL6 tag and more specific wording of the #PUL6 description could improve the performance of the reporting system.

For the chart reviews of lung cancers assigned tags other than #PUL5, we obtained random samples of up to 10 patients per #PUL tag. Post hoc, we noted several cases of early-stage lung cancer tagged #PUL6 and conducted additional chart reviews for these patients. Chart reviews were performed from 120 days before to 120 days after the index CT by the first author (JRD), with the senior author (LJH) duplicating review of the CT report. We sought to understand the indication for the index CT, the radiologist’s findings, and subsequent clinical actions. This material was analyzed qualitatively, summarized in a table, and shared with the lead radiologist (TU), who performed additional chart review and articulated learnings.

The Research Determination Committee for the Kaiser Permanente Northern California region determined the project does not meet the regulatory definition of research involving human subject per 45 CFR 46.102(d).

---

**ORIGINAL RESEARCH ARTICLE**

An Intervention to Tag Findings Suspicious for Lung Cancer on Chest Computed Tomography Has Good Sensitivity and Number Needed to Diagnose
nODULES WITH OTHER SUSPICIOUS FINDINGS SUCH AS LYMPHADENOPATHY, WITH 1 PATIENT HAVING LUNG CANCER DIAGNOSED BEFORE THEIR CT SCAN. IN 4 PATIENTS, THESE FINDINGS LED TO THE RECOMMENDATION TO REPEAT CT USING CONTRAST OR TO OBTAIN A PULMONOLOGY CONSULTATION. WE NOTED THAT THE FLEISCHNER SOCIETY GUIDELINES FOR MANAGING INDETERMINATE NODULES ARE COMPLEX, AND RECOMMENDATIONS FOR FOLLOW-UP REQUIRE SYNTHESIS OF INFORMATION OF NODULE TYPE AND SIZE ASSESSED BY RADIOLOGISTS WITH RISK FACTORS ASSESSED BY THE ORDERING PHYSICIAN SUCH AS FAMILY HISTORY, SMOKING HISTORY, AND DIFFERENTIAL DIAGNOSIS.

FOR PATIENTS TAGGED #PUL6 (KNOWN LUNG CANCER OR SUSPECTED METASTASIS TO THE LUNG), WE RANDOMLY SELECTED 10 PATIENTS OF 105 CANCERS, OF WHICH 8 HAD CLEAR EVIDENCE IN THE CT REPORT OF METASTATIC DISEASE AND WERE APPROPRIATELY TAGGED; 2 DID NOT, OF WHOM 1 HAD EARLY-STAGE DISEASE. WE THEN REVIEWED ALL 13 EARLY-STAGE CASES TAGGED #PUL6, OF WHICH 1 HAD A RECENT LUNG CANCER DIAGNOSIS; 2, RECENT DIAGNOSES OF CANCER OTHER THAN LUNG; 4, DISTANT HISTORIES OF CANCER OTHER THAN LUNG; AND 1, WHAT APPEARED TO THE RADIOLOGIST AS A LIVER CANCER WITH LUNG METASTASIS. THE OTHER 6 DID NOT HAVE MENTION OF A RECENT OR DISTANT CANCER DIAGNOSIS AS PART OF THEIR INDICATION. WE NOTED THAT PATIENT HISTORIES WERE HETEROGENEOUS. CHEST CT ALONE DID NOT PROVIDE ADEQUATE INFORMATION TO THE RADIOLOGIST FOR MAKING AN ACCURATE DIAGNOSIS OF METASTATIC DISEASE, AND THE DEFINITION OF #PUL6 COULD BE CLARIFIED.

CHART REVIEW OF #PULX (TECHNICALLY LIMITED) (10 REVIEWS OF 21 CANCERS) REVEALED HETEROGENEOUS CIRCUMSTANCES, ALTHOUGH LUNG CANCER WAS PART OF THE DIFFERENTIAL DIAGNOSIS OR WAS HIGHLY CONCERNING IN MOST. DESPITE NOT BEING CODED #PUL5 (SUSPICIOUS FOR MALIGNANCY), 6 OF THE RADIOLOGY REPORTS DESCRIBED MASSES, AND FINDINGS WERE COMMUNICATED DIRECTLY TO THE ORDERING PHYSICIAN, LEADING TO CLINICAL FOLLOW-UP.

DISCUSSION

WE ASSESSED THE PERFORMANCE CHARACTERISTICS OF A STANDARDIZED REPORTING SYSTEM THAT ADDS TAGS TO CHEST CT REPORTS AND AUTOMATICALLY FORWARDS FINDINGS SUSPICIOUS FOR LUNG CANCER TO AN EXPERT REVIEW COMMITTEE. THE GOALS OF THE REPORTING SYSTEM WERE TO EXPEDITE WORKUP FOR NEW LUNG CANCER CASES THROUGH THE USE OF A MULTIDISCIPLINARY REVIEW GROUP AND TO ASSIST PRIMARY CARE PROVIDERS IN SCHEDULING FOLLOW-UPS FOR NODULE MANAGEMENT IN ACCORDANCE WITH CLINICAL GUIDELINES. WE PREVIOUSLY REPORTED THE SYSTEM’S EFFECTIVENESS FOR DIAGNOSING EARLY-STAGE LUNG CANCER, NOTING THE INTERVENTION WAS ASSOCIATED WITH A 24% GREATER ODDS OF DIAGNOSING EARLY-STAGE LUNG CANCER WITHIN 120 DAYS (ODDS RATIO, 1.24; 95% CI, 1.09–1.41), BUT NOT WITH AN INCREASED RATE OF SURGICAL TREATMENT.3 HERE, WE REPORT #PUL5 (SUSPICIOUS FOR MALIGNANCY) TO HAVE A SENSITIVITY OF 74%

<table>
<thead>
<tr>
<th>#PUL tag</th>
<th>No. of new primary lung cancers</th>
<th>Lung cancer cases, %</th>
<th>Patients with lung cancer, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>#PUL0: None</td>
<td>21,623</td>
<td>55</td>
<td>5</td>
</tr>
<tr>
<td>#PUL1: Benign</td>
<td>2,258</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>#PUL2: Infection/Inflammatory</td>
<td>3,018</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>#PUL3: Subsolid</td>
<td>690</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>#PUL4: Indeterminate</td>
<td>636</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>#PUL5: Suspicious</td>
<td>225</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>#PUL6: Known cancer</td>
<td>89</td>
<td>3</td>
<td>24</td>
</tr>
<tr>
<td>#PULX: Technically limited</td>
<td>205</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>39,409</td>
<td>100</td>
<td>7</td>
</tr>
</tbody>
</table>

CI = confidence interval.
(CI, 72%–77%), with an NND of 2.7 (CI, 2.6–2.9) for tagging new lung cancer. Although this level of sensitivity is acceptable, since completing this work, we have sought to increase the sensitivity through several workflow improvements, detailed in the following paragraph. The NND of 2.7 means that for every case of lung cancer, the multidisciplinary review group reviews 2.7 imaging reports and associated notes, with some cases being expedited easily and other requiring more time. The multidisciplinary group reports this level of NND to be acceptable.

This study provided important information to improve the system, including the need for radiologists to update reports templated #PUL0, the importance of coding all findings suspicious for lung cancer and not just nodules, challenges with measuring the solid component, challenges coding the combination of inflammation with findings suspicious for cancer, separating cases with new and established lung cancer or metastatic disease, and coding suspicious findings together with technical limitations. In response to these findings, we sought to improve the system through design changes and additional radiologist training to clarify that the definition of #PUL5 includes any finding suggestive of malignancy, not just pulmonary nodules, and #PUL6 provides a tag for patients who have received a confirmed cancer diagnosis, not those in whom cancer appears metastatic on CT.

A 2016 survey of American Thoracic Society members reported that the 3 most common systems-level resources to improve the diagnosis of lung cancer included 1) adding Fleischner Society guidelines to radiology reports, 2) adding prompts to the ordering provider on radiology reports describing new nodules, and 3) employing clinical staff to coordinate recommended follow-up.11,12 The system we describe included all 3 of these components. We are not aware of previous evaluations of these components or other efforts such as described here.

A strength of our study was the inclusion of nearly 40,000 community-based patients undergoing diagnostic chest CT for a variety of reasons, many of whom were elderly and therefore underrepresented in past studies. The key limitation was the study’s scope. We did not assess the time required by the radiologist to tag each report or the user-centeredness of the system. Our study allowed for lung cancer diagnoses through 120 days after imaging, thereby allowing time for reimagining growth of features such as ground-glass nodules. However, we did not have adequate follow-up time to assess lung cancer diagnosis more than 120 days after initial CT examination, as needed to evaluate surveillance of subsolid and indeterminate pulmonary nodules. Additional research is needed to examine lung cancer diagnoses beyond 120 days, the impact of greater diagnosis of early-stage lung cancer on lung cancer mortality, and the effectiveness of tagging subsolid and indeterminate pulmonary nodules for improving nodule management. “Infectious/inflammatory” is a broad concept, and for patients with both infectious/inflammatory findings and a pulmonary nodule, it may be possible to define subgroups for whom the nodule evaluation is especially urgent. For ground-glass nodules that have a well-defined solid component, the Fleischner Society guidelines recommend surveillance of the solid component for growth. However, measuring the solid component is challenging because vessels running through these nodules can be mistaken for solid components, and the solid component can appear to change from one scan to the next, with the difference being technical. This results in high interobserver variability. In addition, subsolid nodules grow slowly,13 and the concept of a lung cancer that grows slowly may

<table>
<thead>
<tr>
<th>#PUL tag</th>
<th>Performance characteristic</th>
<th>Numerator</th>
<th>Denominator</th>
<th>Performance, %</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>#PUL5 only</td>
<td>Sensitivity</td>
<td>821</td>
<td>1105</td>
<td>74.3</td>
<td>71.6–76.9</td>
</tr>
<tr>
<td>#PUL5 only</td>
<td>Specificity</td>
<td>36,870</td>
<td>38,304</td>
<td>96.3</td>
<td>96.1–96.4</td>
</tr>
<tr>
<td>#PUL5 only</td>
<td>Positive predictive value</td>
<td>821</td>
<td>2255</td>
<td>36.4</td>
<td>35.0–37.8</td>
</tr>
<tr>
<td>#PUL5 only</td>
<td>Negative predictive value</td>
<td>36,870</td>
<td>37,154</td>
<td>99.2</td>
<td>99.2–99.3</td>
</tr>
<tr>
<td>#PUL5 only</td>
<td>No. needed to diagnosis</td>
<td>2255</td>
<td>821</td>
<td>2.7</td>
<td>2.6–2.9</td>
</tr>
<tr>
<td>#PUL5 + #PUL6</td>
<td>Sensitivity</td>
<td>926</td>
<td>1105</td>
<td>83.8</td>
<td>81.5–85.9</td>
</tr>
<tr>
<td>#PUL5 + #PUL6</td>
<td>Specificity</td>
<td>36,077</td>
<td>38,304</td>
<td>94.2</td>
<td>94.0–94.4</td>
</tr>
<tr>
<td>#PUL5 + #PUL6</td>
<td>Positive predictive value</td>
<td>926</td>
<td>3153</td>
<td>29.4</td>
<td>28.4–30.4</td>
</tr>
<tr>
<td>#PUL5 + #PUL6</td>
<td>Negative predictive value</td>
<td>36,077</td>
<td>36,256</td>
<td>99.5</td>
<td>99.4–99.6</td>
</tr>
<tr>
<td>#PUL5 + #PUL6</td>
<td>No. needed to diagnosis</td>
<td>3153</td>
<td>926</td>
<td>3.4</td>
<td>3.3–3.5</td>
</tr>
</tbody>
</table>

Adult members of Kaiser Permanente Northern California who underwent chest computed tomography (CT) without a history of chest CT in the preceding 24 mo, August 2015 to July 2017. For #PUL only, the number of true positives (TPs) was 821; true negatives (TNs), 36,870; false negatives (FNs), 284; and false positives (FPs), 1434. For #PUL5 and #PUL6, the number of TPs was 926; TNs, 36,077; FNs, 179; and FPs, 2227. The sensitivity is defined as TP/TP + FN; specificity, TN/(FP + TN); positive predictive value, TP/(TP + FP); negative predictive value, TN/(FN + TN); and number needed.
be underappreciated by radiologists. A critical benefit of standardized reporting systems is the opportunity the systems afford, through the creation of data sets, to conduct important research to address unanswered questions regarding lung nodule management and to identify performance improvement opportunities. Artificial intelligence is increasingly being applied to health-care delivery and, theoretically, could be implemented as an extra precaution against underdiagnosing lung cancer from CT scans. However, artificial intelligence faces several barriers to full implementation, including performance, computing power needs, the need to train systems, and differences between high-quality, standardized imaging used for algorithm training and nonstandardized clinical imaging in diverse and complex patients.

In conclusion, radiology tagging systems can categorize patients effectively with abnormal lung findings with good sensitivity and low NND. Systems can be improved by ensuring clear categories, prioritizing the purpose of the system over strict definitions during radiologist training, and utilizing user-centered coding schemes that are easy for wide adoption. Tagging systems provide a simple and useful informatics solution for ensuring appropriate patient care, benchmarking, performance improvement, and research into the effectiveness of guidelines in real-world settings.

Disclosure Statement

The investigators are partners and staff of The Permanente Medical Group and report no other conflicts of interest.

Financial Support

This project was supported by The Permanente Medical Group Delivery Science and Applied Research initiative, and in part by K07 CA188142 (LCK). The funders had no role after making a funding decision.

Author Contributions

Lori C Sakoda, PhD, Thomas H Urbania, MD, Sora Ely, MD, Todd Osinski, MD, Ashish Patel, MD, and Lisa J Herrinton, PhD, participated in study design. Jennifer R Dusendang, MPH, and Lisa J Herrinton, PhD, participated in data collection. Jennifer R Dusendang, MPH, Lori C Sakoda, PhD, Thomas H Urbania, MD, Sora Ely, MD, Todd Osinski, MD, Ashish Patel, MD, and Lisa J Herrinton, PhD, participated in data analysis and manuscript preparation. Lisa J Herrinton, PhD, is the guarantor of this manuscript and its content. Jennifer R Dusendang, MPH, Lori C Sakoda, PhD, Thomas H Urbania, MD, Todd Osinski, MD, Ashish Patel, MD, and Lisa J Herrinton, PhD, contributed substantially to the conception or design of the work; Jennifer R Dusendang, MPH, Lori C Sakoda, PhD, Thomas H Urbania, MD, Sora Ely, MD, Todd Osinski, MD, and Lisa J Herrinton, PhD, contributed substantially to the writing and/or revision of the manuscript; and all authors approved the final version of the manuscript. All authors are accountable for the manuscript’s contents.

References