8-26-2016

An Automated Method for Identifying Individuals with a Lung Nodule Can Be Feasibly Implemented Across Health Systems

Farhood Farjah
University of Washington, Seattle, WA, ffarjah@uw.edu

Scott Halgrim
Optum360, San Diego, CA

Diana S.M. Buist
Group Health Research Institute, Seattle, WA

Michael K. Gould
Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena, CA

See next pages for additional authors

Follow this and additional works at: http://repository.edm-forum.org/egems

Part of the Analytical, Diagnostic and Therapeutic Techniques and Equipment Commons

An Automated Method for Identifying Individuals with a Lung Nodule Can Be Feasibly Implemented Across Health Systems

Abstract

Introduction: The incidence of incidentally detected lung nodules is rapidly rising, but little is known about their management or associated patient outcomes. One barrier to studying lung nodule care is the inability to efficiently and reliably identify the cohort of interest (i.e. cases). Investigators at Kaiser Permanente Southern California (KPSC) recently developed an automated method to identify individuals with an incidentally discovered lung nodule, but the feasibility of implementing this method across other health systems is unknown.

Methods: A random sample of Group Health (GH) members who had a computed tomography in 2012 underwent chart review to determine if a lung nodule was documented in the radiology report. A previously developed natural language processing (NLP) algorithm was implemented at our site using only knowledge of the key words, qualifiers, excluding terms, and the logic linking these parameters.

Results: Among 499 subjects, 156 (31%, 95% confidence interval [CI] 27-36%) had an incidentally detected lung nodule. NLP identified 189 (38%, 95% CI 33-42%) individuals with a nodule. The accuracy of NLP at GH was similar to its accuracy at KPSC: sensitivity 90% (95% CI 85-95%) and specificity 86% (95% CI 82-89%) versus sensitivity 96% (95% CI 88-100%) and specificity 86% (95% CI 75-94%).

Conclusion: Automated methods designed to identify individuals with an incidentally detected lung nodule can feasibly and independently be implemented across health systems. Use of these methods will likely facilitate the efficient conduct of multi-site studies evaluating practice patterns and associated outcomes.

Authors

Farhood Farjah, *University of Washington, Seattle, WA*; Scott Halgrim, *Optum360, San Diego, CA*; Diana S Buist, *Group Health Research Institute, Seattle, WA*; Michael K Gould, *Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena, CA*; Steven B Zeliadt, *Department of Veterans Affairs, VA Puget Sound Health Care System, Health Services Research & Development Center of Innovation*; Elizabeth T Loggers, *Seattle Cancer Care Alliance*; David S Carrell, *Group Health Research Institute, Seattle, WA*.

eGEMs
Generating Evidence & Methods
to improve patient outcomes

An Automated Method for Identifying Individuals with a Lung Nodule Can Be Feasibly Implemented Across Health Systems

Farhood Farjah, MD;[i] Scott Halgrim;[ii] Diana S.M. Buist, PhD, MPH;[iii] Michael K. Gould, MD;[iv] Steven B. Zeliadt, PhD, MPH;[v] Elizabeth T. Loggers, MD, PhD;[vi] David S. Carrell, PhD[iii]

ABSTRACT

Introduction: The incidence of incidentally detected lung nodules is rapidly rising, but little is known about their management or associated patient outcomes. One barrier to studying lung nodule care is the inability to efficiently and reliably identify the cohort of interest (i.e. cases). Investigators at Kaiser Permanente Southern California (KPSC) recently developed an automated method to identify individuals with an incidentally discovered lung nodule, but the feasibility of implementing this method across other health systems is unknown.

Methods: A random sample of Group Health (GH) members who had a computed tomography in 2012 underwent chart review to determine if a lung nodule was documented in the radiology report. A previously developed natural language processing (NLP) algorithm was implemented at our site using only knowledge of the key words, qualifiers, excluding terms, and the logic linking these parameters.

Results: Among 499 subjects, 156 (31%, 95% confidence interval [CI] 27-36%) had an incidentally detected lung nodule. NLP identified 189 (38%, 95% CI 33-42%) individuals with a nodule. The accuracy of NLP at GH was similar to its accuracy at KPSC: sensitivity 90% (95% CI 85-95%) and specificity 86% (95% CI 82-89%) versus sensitivity 96% (95% CI 88-100%) and specificity 86% (95% CI 75-94%).

Conclusion: Automated methods designed to identify individuals with an incidentally detected lung nodule can feasibly and independently be implemented across health systems. Use of these methods will likely facilitate the efficient conduct of multi-site studies evaluating practice patterns and associated outcomes.

[i]University of Washington, [ii]Optum360, [iii]Group Health Research Institute, [iv]Department of Research and Evaluation, Kaiser Permanente Southern California, [v]Department of Veterans Affairs, VA Puget Sound Health Care System, Health Services Research & Development Center of Innovation, [vi]Seattle Cancer Care Alliance

Introduction

A lung nodule is defined by a radiographic abnormality 30 mm or smaller and is most commonly detected by computed tomography (CT). The incidence of lung nodules appears to be increasing over time coincident with a rise in the use of CT.[1] Recent estimates suggest up to 1.5 million new diagnoses per year.[2] Because a lung nodule may represent lung cancer, clinicians must balance the risks of diagnostic evaluation with the benefits of early detection and treatment of lung cancer.[3] While practice guidelines have been available for over a decade to help providers strike this balance,[4,5,6,7] the level of evidence supporting practice guideline recommendations is low.[7]

One barrier to investigating lung nodule care delivery and associated outcomes is the inability to reliably identify a cohort of individuals with an incidentally detected lung nodule.[8] The International Classification of Disease, Ninth Revision (ICD-9) diagnostic codes are inaccurate,[9] and manual abstraction of data from free-text radiology reports is time intensive and expensive. Researchers from Kaiser Permanente Southern California (KPSC) developed an automated method of identifying individuals with a lung nodule using natural language processing (NLP), electronic radiology reports, and administrative data.[9] Their methodology had a high sensitivity (96 percent) and specificity (86 percent) for identifying individuals with a lung nodule. Automated approaches could be used to identify a cohort for a multisite study, if they can feasibly be implemented in other health systems.

Another barrier to studying lung nodule care and outcomes is the inability to efficiently measure documented lung cancer risk factors such as smoking status and nodule size. Practice guideline recommendations regarding the intensity of diagnostic evaluation vary based on an individual's risk of lung cancer.[4,6,7] In order to assess adherence with practice guidelines, one needs to be able to measure the presence or absence of documented lung cancer risk factors. If lung cancer risk factors are frequently documented, then automated methods could be modified to identify persons with a lung nodule *and* ascertain their documented lung cancer risk factors. However, the frequency of documentation is unknown.

Therefore the primary goal of this study was to evaluate the feasibility of implementing an automated method for identifying individuals with a lung nodule in a health system other than the one for which the method was first developed. A secondary goal was to determine the frequency of provider documentation of lung cancer risk factors among individuals with an incidentally detected lung nodule.

Methods

We conducted a retrospective cross-sectional study of in-network members enrolled in the Group Health Cooperative (GHC) Plan who underwent a chest CT in 2012. Common procedural terminology (CPT) codes were used to identify individuals who had undergone a CT (71250, 71260, 71270, 71275). Subjects eligible for study included adults ages 18–89 years who had continuous enrollment from 2010 to 2012, had no claims for a CT between 2010 and 2011, and no prior history of lung cancer based on ICD-9 diagnostic codes found within administrative data or entries within the Western Washington Surveillance Epidemiology and End Results (SEER) tumor registry. Only in-network health plan members were included in this study to ensure access to free-text radiology reports. We excluded individuals with a CT performed between 2010 and 2011 to avoid including subjects who may have been under active surveillance for a previously diagnosed lung nodule. A two-year period was chosen because practice guidelines recommend up to two years of follow-

up.[4,6] Persons with a history of lung cancer were excluded because these patients are recommended to have routine surveillance imaging according to lung cancer practice guidelines.[4,6,10] The Institutional Review Board at GHC approved this study and waived the need for consent.

We selected a random sample of 500 subjects from the larger sample of members who underwent CT in 2012 for medical record review by a trained chart abstractor. In order to evaluate the comparability of the random sample to the overall population, we compared demographic characteristics, the Charlson comorbidity index, frequency and type of prior malignancies, and smoking status. Demographic information was available through enrollment data. The Charlson comorbidity index was determined using all available administrative data in the year prior to the CT.[11] Cancer data were available through the Western Washington SEER registry. Finally, information about smoking status was obtained from the electronic health record (EHR). We used this sample population to develop and evaluate the performance of NLP against this gold standard chart abstraction, and to describe the frequency of documented lung cancer risk factors among individuals with a confirmed lung nodule.

We developed an NLP program based on a previously published algorithm developed by investigators at KPSC. Our algorithm, consistent with the original KPSC algorithm, was rule-based and was developed iteratively using a combination of key words (e.g., nodule, opacity, lesion, mass, tumor), qualifiers (e.g., nodule size, small, tiny), and excluding terms (e.g., adrenal, liver, kidney, thyroid, etc.).[9] Our algorithm was implemented according to standard methods at the Group Health Research Institute and was subsequently applied to the free-text radiology report associated with the CT that qualified the study subject for inclusion in this investigation. While the original algorithm output included only a binary

indicator of a documented lung nodule, our modified version also reported the documented size of the lung nodule (see Appendix).

"True" nodule status was based only on information available in the free-text radiology report collected from standardized abstraction. Any uncertainty about documentation in the radiology report was adjudicated by a board-certified thoracic surgeon with expertise in both lung nodule and lung cancer care. The accuracy of NLP was measured in terms of sensitivity, specificity, and positive and negative predictive values (PPV and NPV, respectively). Associated 95 percent exact binomial confidence intervals (CI) were also reported to allow for comparisons to previously published estimates of accuracy. Feasibility was defined by successful implementation of the automated method and accuracy in identifying individuals with a lung nodule comparable to that reported by KPSC.

In order to estimate the frequency of documented lung cancer risk factors, chart abstractors reviewed all available EHRs 30 days prior to and 30 days after the CT. Risk factors were selected based on variables identified by lung nodule[4,5,6,7] and lung cancer screening practice guidelines;[12] variables used in studies predicting the risk of lung cancer;[13,14,15,16,17,18,19,20,21] and variables recognized and used by clinical experts—in pulmonary medicine (author M.G.), in thoracic surgery (author F.F.), and in medical oncology (author E.T.L.). Risk factors were divided into two broad categories. Radiographic characteristics refer to imaging findings documented in the free-text radiology report that increase the likelihood of a lung cancer diagnosis, such as nodule size, border characteristics (e.g., spiculation versus no spiculation), and location (e.g., right upper lobe). Clinical risk factors refer to demographic and behavioral variables and occupation and environmental exposures documented anywhere within the EHR. The frequency of documentation

was determined by the proportion of patients with any documentation indicating that the specific risk factor was present or absent. Any uncertainty about documentation of a risk factor within the EHR was also adjudicated by a clinician (author F.F.).

Results

Figure 1 shows the planned sequential exclusions resulting in the final study cohort consisting of a random sample of 499 eligible study subjects with available free-text radiology reports. We used a combination of administrative, tumor registry, and structured EHR data to compare the characteristics of the randomly selected study cohort to the overall population of eligible study enrollees who underwent a CT in 2012 (Table 1). Demographic characteristics, the comorbidity index, the frequency and type of prior malignancies, and the distribution of current, former, and never smokers were similar between the two groups.

Table 2 describes the performance and accuracy of various implementations of NLP as well as the performance of the original NLP algorithm. NLP tended to overestimate the prevalence of individuals with lung nodules compared to chart abstraction. The sensitivity, specificity, and NPV of NLP did not vary by implementation across different integrated health systems (i.e., GHC and KPSC). However, the PPV was lower when NLP was implemented at GHC, but so was the prevalence of individuals with lung nodules. Modifications to the NLP to generate two output parameters (binary indicator of nodule and continuous variable for nodule size) did not markedly change the accuracy whether compared to the original KPSC algorithm or its nonmodified implementation at GHC. Additionally, the accuracy of NLP did not vary significantly when it was further modified to classify subjects with a prespecified, discrete nodule size range (e.g., any size nodule 1–30mm, ≤8mm, or >8–30mm).

Among 156 subjects with confirmed documentation of a lung nodule, the frequency of documented radiographic characteristics and clinical lung cancer risk factors was assessed within the 30 days prior to and after the date of CT (Table 3). Commonly documented radiographic characteristics included size, lobar location, and the number of lung nodules. Border characteristics (i.e., spiculation), nodule density, and the presence of calcifications were uncommonly documented. Smoking status, intensity of tobacco exposure (i.e., pack-years, which is a unit for measuring the amount a person has smoked over a long period), number of years quit among former smokers, and body mass index were the most commonly documented clinical risk factors. Underlying pulmonary disease, a personal history of radiation therapy, extrathoracic malignancy, family history of lung cancer, or exposure to second-hand smoke were documented in a minority of individuals. Thirteen clinical risk factors—all of them environmental and occupational exposures—were rarely or never documented.

A post hoc analysis was performed to evaluate the accuracy of using automated methods to ascertain two key, commonly documented lung cancer risk factors—tobacco status (smoker versus never a smoker) and nodule size.[4] Among 156 subjects with a confirmed lung nodule, structured data elements within the EHR revealed that smoking status was documented within a median of 19 days (range 1–1,203 days) of undergoing the CT that identified the lung nodule. Of those 156 subjects, 95 had their smoking status documented within 30 days of their CT. In these 95 patients we were able to compare ascertainment of smoking status via structured elements against gold standard chart abstraction reviewing both structured and unstructured data elements. The sensitivity and specificity of structured data elements for ascertaining smoking status was 95 percent and 89 percent, respectively.

Figure 1. Nodule Size Ascertained by NLP Versus Chart Abstraction Among Individuals with a Confirmed Lung Nodule
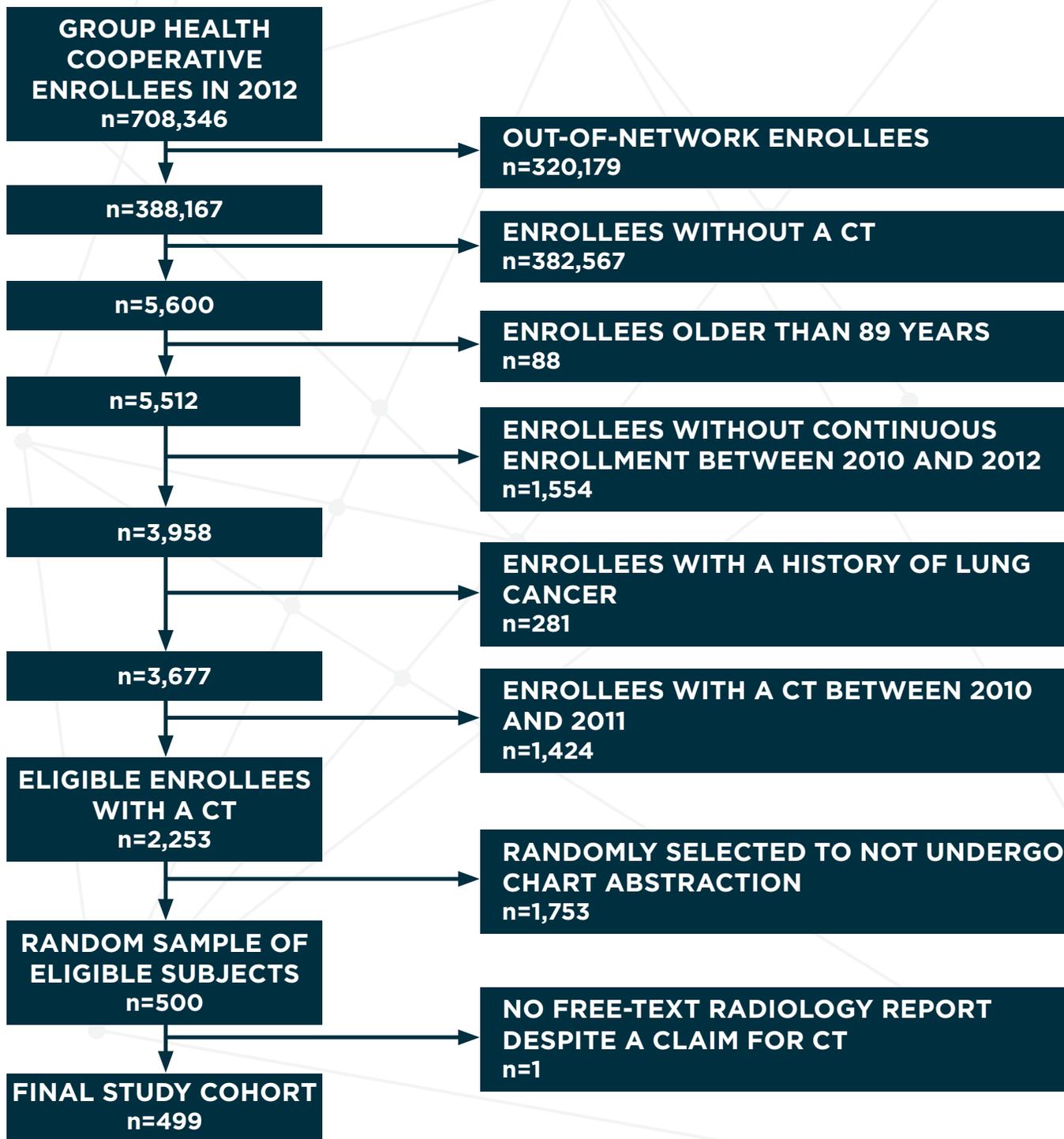
Table 1. Population and Sample Characteristics*

	ALL SUBJECTS WHO UNDERWENT A CT IN 2012	RANDOM SAMPLE OF SUBJECTS WHO UNDERWENT A CT IN 2012
	n=2,253	**n=499**
Median age (range), years	65 (18–89)	66 (20–89)
Men	43%	43%
RACE		
White	83%	83%
Black	4.6%	3.8%
Asian	5.8%	6.4%
AI/AN	1.5%	****
NH or other PI	0.9%	1.2%
Multiple	0.3%	****
Unknown	4.0%	4.4%
Hispanic	4.1%	5.2%
COMORBIDITY INDEX		
0	86%	82%
1	3.7%	4.4%
2+	11%	14%
HISTORY OF MALIGNANCY **		
Head and neck	6.0%	6.8%
Breast	8.0%	8.6%
Prostate	2.9%	2.2%
Colorectal	3.2%	3.2%
Other	13%	13%
Any/multiple	27%	28%
SMOKING STATUS*		
Never	48%	50%
Former	41%	37%
Current	11%	12%
Missing	0.6%	****
MEDIAN PACK-YEARS** (RANGE)**	**23 (11–43)**	**30 (12–46)**
Missing	31%	28%

Notes: Alaskan Native (AN), American Indian (AI), Native Hawaiian (NH), Pacific Islander (PI)

*Information obtained from administrative data, tumor registry data, and structured data elements within the electronic health record (e.g., smoking status). **Columns do not necessarily add up to 100 percent. Data for esophageal cancer and melanoma are not shown because there were five or fewer subjects within one of the samples. ***Ascertained using structured data elements within the electronic health record. These data were obtained within a median of 18 days (range 1–2,593 days) prior to the date of the CT. ****Calculated among former and current smokers.

Table 2. Accuracy (95% Confidence Interval) of NLP in Identifying Individuals with a Lung Nodule

	ORIGINALLY PUBLISHED[8] NLP PERFORMANCE FOR LUNG NODULES 5-30MM n=116	APPLIED ALGORITHM FOR LUNG NODULES 5-30MM n=499	MODIFIED ALGORITHM FOR ANY SIZE LUNG NODULE (≤30 MM) n=499	MODIFIED ALGORITHM FOR ANY SIZE LUNG NODULE ≤8MM n=499	MODIFIED ALGORITHM FOR ANY SIZE LUNG NODULE >8MM-30MM n=499
Prevalence (chart abstraction)	49% (40–56%)	18% (15–22%)	31% (27–36%)	22% (18–26%)	9.2% (6.8–12%)
Prevalence (NLP)	54% (45–64%)	24% (20–28%)	38% (33–42%)	25% (21–28%)	13% (10–16%)
Sensitivity	96% (88–100%)	88% (79–94%)	90% (95% CI 85–95%)	86% (79–92%)	87% (74–95%)
Specificity	86% (75%–94%)	90% (87–93%)	86% (82–89%)	93% (90–95%)	94% (92–96%)
PPV	87% (77–94%)	66% (57–75%)	75% (68–81%)	77% (69–84%)	61% (48–72%)
NPV	96% (87–100%)	97% (95–99%)	95% (92–97%)	96% (94–98%)	99% (97–99%)

Notes: Natural language processing (NLP), positive predictive value (PPV), negative predictive value (NPV), confidence interval (CI).

Table 3. Provider Documentation of Radiologic Characteristics and Lung Cancer Risk Factors among Individuals with a Lung Nodule

	FREQUENCY (95% CONFIDENCE INTERVAL) (n=156)
RADIOGRAPHIC CHARACTERISTICS	
Nodule size	100% (98–100%)
Lobar location of lung nodule	99% (96–100%)
Number of nodules	99% (96–100%)
Calcified nodule	31% (24–39%)
Nodule density	7.7% (4.0–13%)
Spiculated border	5.1% (2.2–10%)
CLINICAL RISK FACTORS*	
Smoking status**	95% (90–98%)
Body mass index	89% (83–94%)
Pack-years tobacco	86% (79–91%)
Years quit among former smokers***	78% (67–86%)
Personal history of extrathoracic malignancy	35% (27–43%)
Underlying pulmonary disease	32% (25–40%)
Family history of lung cancer	10% (6.0–16%)
Personal history of radiation therapy	5.1% (2.2–9.9%)
Exposure to second-hand smoke	3.2% (1.0–7.3%)

Notes: *Data for the following variables were suppressed because five or fewer subjects had documentation about the risk factor: radon, silica, cadmium, asbestos, arsenic, beryllium, chromium, diesel fuel, nickel, coal smoke, soot, and dust. **Ascertained by chart abstraction of structured and unstructured data within the electronic health record (EHR) 30 days prior to the date of the CT. ***Calculated among former smokers.

As no structured data existed for nodule size we investigated the use of our modified NLP output (i.e., a binary indicator of a nodule *and* nodule size) to generate data on lung nodule size as documented in the radiology report. Figure 2 shows a high level of statistical correlation between NLP and chart abstracted lung nodule size (r=0.9).

To better understand the potential for clinically relevant misclassifications that could result from the use of automated methods, we considered eight risk categories identified by the Fleischner Society Guidelines and defined by tobacco status and nodule size.[4] The Fleischner Society Guidelines recommend varying intensities of diagnostic

evaluation depending on an individual's risk category, and misclassification of risk category is therefore a clinically significant error. The frequency of misclassification was 25 percent with 14 percent of subjects' risk underestimated and 11 percent of subjects' risk overestimated. Of the 24 subjects with a misclassification, the error was the result of misclassified smoking status only, nodule size only, or both smoking status and nodule size in 13 percent, 67 percent, and 21 percent, respectively.

Discussion

A higher level of evidence is greatly needed to guide the care of patients with an incidentally detected lung nodule. However, the inability to identify patients with lung nodules and measure their documented lung cancer risk factors have been major barriers

to conducting multicenter observational studies of lung nodule care and outcomes. Our investigation sought to evaluate the feasibility of implementing an automated method for identifying individuals with lung nodules, and to describe the frequency of lung cancer risk factor documentation.

One important finding from this study is that an automated method of identifying individuals with an incidentally detected lung nodule can be feasibly implemented across health systems. We used the published diagrammatic logic underlying an NLP algorithm[9] to implement NLP within a health care system completely independent of the original setting in which it was developed. Not only were we able to successfully implement NLP, its accuracy in our cohort was similar to that reported in the

Figure 2. Correlation Between Chart Abstracted and NLP Abstracted Nodule Size
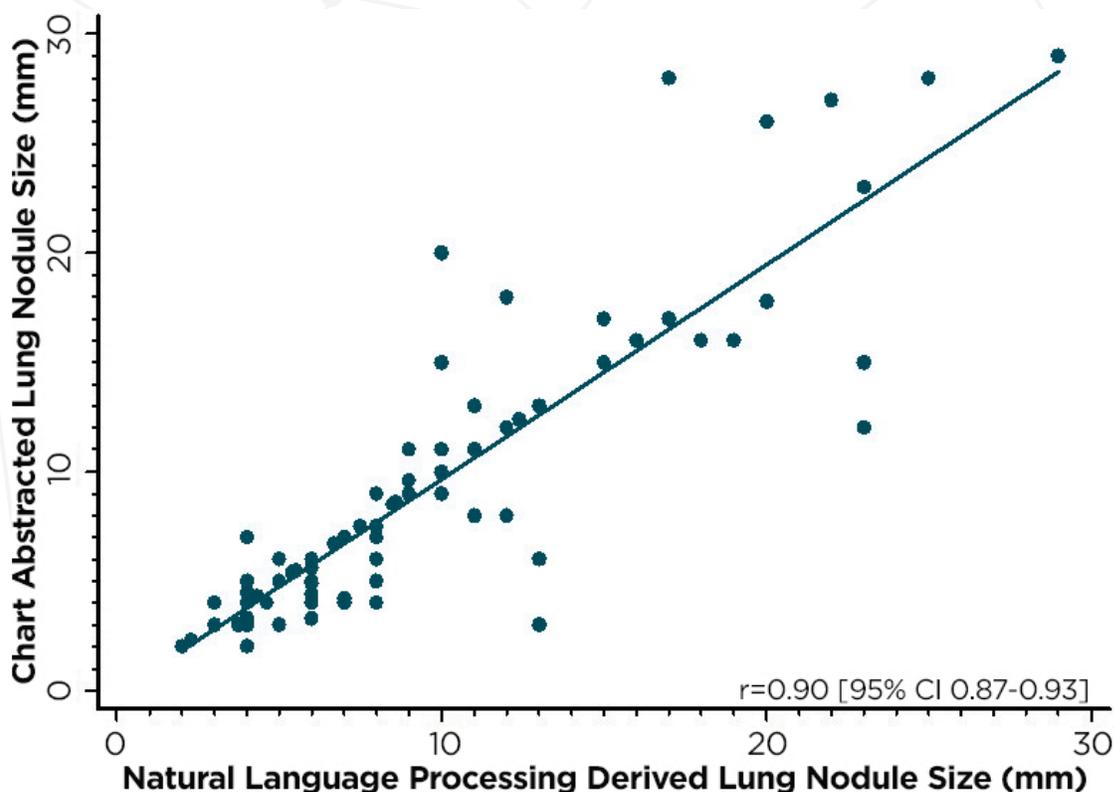
development cohort, with the exception of a lower PPV.[9] The lower PPV is likely explained, at least in part, by the lower prevalence of individuals with incidentally detected lung nodules in our cohort. This lower prevalence of disease is most likely explained by differences in study designs. The original development and validation study of NLP at KPSC oversampled cases with lung nodules.[9] This design difference is the most likely explanation for a lower prevalence of disease and lower PPV in our study. When the KPSC authors implemented this automated method in a cohort without sampling (i.e., a study design similar to ours), the prevalence of lung nodules was similar to ours (28 percent).[2] Other evidence supporting the feasibility of implementing NLP across health systems finds that its accuracy did not change even when we modified the algorithm. Our modification of their algorithm was to add size as an output parameter provided by the NLP (i.e., in addition to a binary indicator of a documented lung nodule). NLP–determined size allows investigators flexibility in assessing clinically relevant subgroups of patients. For instance, the approach to evaluating lung nodules ≤ 8 mm is largely based on smoking status and nodule size,[4] whereas the approach to evaluating nodules > 8 mm is largely dependent on the predicted risk of lung cancer based on age, smoking status, nodule size, nodule lobar location, history of extrathoracic malignancy, and border characteristics of the nodule (i.e., spiculation).[6] The significance of our findings is that the use of automated methods will likely facilitate the efficient conduct of multisite observational studies of lung nodule care and associated outcomes.

Another important finding from our investigation is that clinicians appear to document many important lung cancer risk factors infrequently. These risk factors determine, in part, eligibility for guideline recommended nodule care. For instance, persons with a calcified nodule do not need further follow-up according to practice guidelines.[4,6] Likewise, patients with a history of extrathoracic malignancy do not qualify for guideline recommended nodule care, and are instead recommended to undergo surveillance as determined by specialists based on the underlying cancer, its risk of recurrence, and the risk of lung cancer.[4] Another reason to better understand the frequency of documented lung cancer risk factors is that practice guidelines vary the intensity of the subsequent diagnostic evaluation based on an individual's risk of having lung cancer. For example, for nodules ≤8mm, the Fleischner Society Guidelines recommend a range of possibilities from "no further work-up" to a surgical biopsy, depending largely on nodule size and a crude stratification of lung cancer risk.[4] The American College of Chest Physicians recommends a range of options for individuals with a nodule greater than 8mm, including an early follow-up CT, positron emission tomography, nonsurgical lung biopsy, and surgical resection.[6] The appropriate next step is based on the estimated probability of lung cancer using one of several available risk calculators. These risk calculators use variables such as age, smoking status, border characteristics (i.e., spiculation), nodule size, lobar location of the nodule, and history of extrathoracic malignancy.[16,21,22] These specific aspects of guideline recommendations highlight the importance of being able to measure lung cancer risk factors, a key determinant of the intensity of the recommended evaluation. Our findings have two important implications for future research. First, the feasibility of developing automated methods to ascertain documented lung cancer risk factors is limited by the availability of information within the EHR. Second, the ability to study nodule care by any means— automated or manual abstraction—is also limited by the availability of documented lung cancer risk information. This limitation highlights an opportunity to improve provider documentation with templates and structured data entry.

Our post hoc analysis allows us to speculate on the limitations of a fully automated approach to studying lung nodule care across health systems. Researchers are interested in investigating the frequency of guideline adherence and outcomes associated with guideline concordant care. Given a sensitivity of 90 percent, fully automated methods will falsely exclude 10 percent of all individuals with a "true" nodule —effectively reducing the sample size and power of a study. A specificity of 86 percent means that 14 percent of subjects included in the study will not have a documented nodule (i.e., a false-positive classification of nodule status will erroneously include individuals without a lung nodule in the study). Individuals without a true lung nodule will not undergo subsequent evaluation. However, because of the false classification of having a lung nodule, investigators will observe a lower than expected rate of guideline adherent care. The impact of this bias will depend on the study-specific aims. Finally, our investigation also shows that use of automated methods will lead to clinically significant misclassification of lung cancer risk (as defined by the Fleischner Society Guidelines) in a quarter of patients. Although there was no systematic tendency to over- or underestimated risk, the impact of such bias again depends on the study-specific aims. Overall, automated methods appear to result in some error and potential bias. However, investigators may be able to use this information to predict the impact (conservative versus anticonservative) of these sources of bias on their study. Additionally, this information should allow investigators to weigh the benefit of substantially reducing the need for human chart abstraction—an expensive and time consuming approach to conducting research— against potential bias.

An alternative to purely automated or human data abstraction is a hybrid approach. The high and invariant NPV of NLP in identifying individuals with lung nodules reveals an opportunity to utilize a two-step hybrid approach. NLP could be used to screen for individuals *without* a lung nodule with a very low error rate (1–5 percent). Because the "NLP determined prevalence" of lung nodules is approximately 38 percent, a majority of radiology reports reviewed using automated methods would be eliminated from further review, resulting in a substantial reduction in the need for manual abstraction. Chart abstractors could then manually review "NLP positive" individuals to establish true nodule status and to ascertain size and tobacco status. This approach would likely mitigate most sources of errors mentioned earlier. For this study, it took our chart abstractor 160 hours and 300 subjects to achieve proficiency in reliably ascertaining variables at a rate of 10 minutes per chart. The time required to train and abstract data would likely be less if abstractors were ascertaining only two variables from the EHR (i.e., nodule status and size and smoking status). To place these estimates into perspective, if this hybrid approach were applied to the entire sample of 2,253 individuals who underwent a CT in 2012 at GHC, 1,411 would be identified by NLP as not having a nodule. The remaining 842 subjects would have their EHR reviewed by a chart abstractor. It would take approximately 250 hours to review these 842 charts compared to 486 hours to review the entire sample. The improved accuracy of a hybrid approach must be weighed against the added resources necessary to manually review a subset of subjects flagged as "NLP positive."

This study has several limitations. Our evaluation of the feasibility of independently implementing automated methods at our health system does not guarantee its performance in other systems. One reason for automated methods to perform less well in other settings is if the documentation style of radiologists varies substantially from site to site. Despite this strongly suspected variability

across sites, NLP performed just as well at GHC as it did at KPSC. Nonetheless, investigators seeking to conduct a multicenter observational study of lung nodule care should independently verify the accuracy of NLP in a subset of study subjects. The frequency of documented lung cancer risk factors may also not be generalizable to other settings. In some instances, there may be better documentation at certain sites, whereas in other cases there may be worse documentation at other centers. Finally, this study was not designed or funded to improve the performance of an existing NLP algorithm or to compare performance among all available NLP algorithms. Gershanik and associates used a publicly available toolkit called Information from Searching Content with an Ontology-Utilizing Toolkit (iSCOUT) and reported the precision and recall to be 96 percent and 80 percent, respectively.[23] These performance metrics are equivalent to PPV and sensitivity, respectively. Zeliadt and colleagues used a Lucene-based text search tool to implement NLP and reported a sensitivity, specificity, PPV, and NPV of 77 percent, 93 percent, 71 percent, and 95 percent, respectively.[24] Reasons for differences in performance remain unknown at this time.

Future studies may focus on improving the performance of automated methods and applying the current technology for the purposes of conducting efficient comparative-effectiveness studies. For instance, future investigations may wish to compare all available NLP algorithms with an eye toward developing a new, or modifying an existing, algorithm to achieve even higher accuracy rates for identifying individuals with a lung nodule. Alternatively, researchers may choose to study and leverage the impact of guidelines or quality improvement initiatives that encourage the use of structured reporting for incidentally detected lung nodules. For example, the American College of Radiology has developed a standardized method

of reporting positive findings (i.e., lung nodules) in the context of lung cancer screening. Not only has standardized reporting resulted in better patient care (e.g., fewer false positive screen findings),[25,26] it facilitates the use of automated methods for identifying individuals with lung nodules. Although no such recommendations for the use of structured reporting currently exist for incidentally (not screen) detected nodules, the use of structured reporting would likely enhance the accuracy of automated methods. Finally, given the high accuracy and efficiency of a hybrid approach (i.e., NLP and manual human abstraction) in identifying individuals with a lung nodule, the currently available technology could be applied in the observational setting to study the comparative effectiveness of guideline concordant nodule care. Recent studies have demonstrated low rates of adherence to guidelines in various settings (e.g., community clinics, academic centers, integrated health systems),[27–33] but no study to date has evaluated the relationship between guideline adherence and outcomes. Members of our research team are currently pursuing a multicenter study of the effectiveness, harms, and costs associated with guideline concordant nodule care among individuals enrolled in integrated health systems participating in the Cancer Research Network.[3]

Conclusion

A previously published automated method[9] of identifying individuals with a lung nodule was feasibly implemented in a completely different health system. Providers frequently document smoking status and nodule size, but infrequently document many other important radiographic and clinical lung cancer risk factors. Automated methods may be used to identify persons with a lung nodule and ascertain their documented smoking status and nodule size; however, this approach is associated with misclassification errors that may bias and reduce the statistical power of studies evaluating

the relationship between nodule care and outcomes. Alternatively, a hybrid approach utilizing NLP and chart abstraction may decrease misclassification errors and be substantially less costly and resource intensive than a completely chart abstraction-based approach to studying lung nodule care and associated outcomes.

Acknowledgements

References

1. Smith-Bindman R, Miglioretti DL, Johnson E, Lee C, Feigelson HS, Flynn M, et al. Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996-2010. JAMA 2012;307(22):2400-2409.
2. Gould MK, Tang T, Liu IA, Lee J, Zheng C, Danforth KN, et al. Recent Trends in the Identification of Incidental Pulmonary Nodules. Am J Respir Crit Care Med 2015;192(10):1208-1214.
3. National Lung Screening Trial Research Team. Results of initial low-dose computed tomographic screening for lung cancer. N Engl J Med 2013;368(21):1980-1991.
4. MacMahon H, Austin JHM, Gamsu G, Herold CJ, Jett JR, Naidich DP, et al. Guidelines for Management of Small Pulmonary Nodules Detected on CT Scans: A Statement from the Fleischner Society1. Radiology 2005;237(2):395-400.
5. Tan BB, Flaherty KR, Kazerooni EA, Iannettoni MD, American College of Chest Physicians. The solitary pulmonary nodule. Chest 2003;123(1 Suppl):89S-96S.
6. Gould MK, Fletcher J, Iannettoni MD, Lynch WR, Midthun DE, Naidich DP, Ost DE; American College of Chest Physicians. Evaluation of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines (2nd edition). Chest 2007;132(3 Suppl):108S-130S.
7. Gould MK, Donington J, Lynch WR, Mazzone PJ, Midthun DE, Naidich DP, et al. Evaluation of individuals with pulmonary nodules: when is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2013;143(5 Suppl):e93S-120S.
8. Slatore CG, Horeweg N, Jett JR, Midthun DE, Powell CA, Wiener RS, Wisnivesky JP, Gould MK; ATS Ad Hoc Committee on Setting a Research Framework for Pulmonary Nodule Evaluation. An official American Thoracic Society research statement: a research framework for pulmonary nodule evaluation and management. Am J Respir Crit Care Med 2015;192(4):500-514.
9. Danforth KN, Early MI, Ngan S, Kosco AE, Zheng C, Gould MK. Automated Identification of Patients With Pulmonary Nodules in an Integrated Health System Using Administrative Health Plan Data, Radiology Reports, and Natural Language Processing. J Thorac Oncol 2012;7(8):1257-1262.
10. Ettinger DS, Akerley W, Borghaei H, Chang AC, Cheney RT, Chirieac LR, et al. Non-small cell lung cancer. J Natl Compr Canc Netw 2012;10(10):1236-1271
11. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-83.
12. Wood DE, Baum SL, Dransfield MT, Eapen GA, Ettinger DS, Hou L, et al. Lung cancer screening, version 1.2013. J Natl Compr Canc Netw 2013.
13. Naidich DP, Bankier AA, MacMahon H, Schaefer-Prokop CM, Pistolesi M, Goo JM, et al. Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. Radiology 2013;266(1):304-317.
14. Bach PB, Kattan MW, Thornquist MD, Kris MG, Tate RC, Barnett MJ, et al. Variations in lung cancer risk among smokers. J Natl Cancer Inst 2003;95(6):470-478.
15. Cassidy A, Myles JP, van Tongeren M, Page RD, Liloglou T, Duffy SW, et al. The LLP risk model: an individual risk prediction model for lung cancer. Br J Cancer 2008;98(2):270-276.
16. Gould MK, Ananth L, Barnett PG, Veterans Affairs SNAP Cooperative Study Group. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. Chest 2007 Feb;131(2):383-388.
17. Ost DE, Gould MK. Decision making in patients with pulmonary nodules. Am J Respir Crit Care Med 2012;185(4):363-372.
18. Schultz EM, Sanders GD, Trotter PR, Patz EF, Silvestri GA, Owens DK, et al. Validation of two models to estimate the probability of malignancy in patients with solitary pulmonary nodules. Thorax 2008;63(4):335-341.
19. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, et al. A risk model for prediction of lung cancer. J Natl Cancer Inst 2007;99(9):715-726.
20. Tammemägi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection Criteria for Lung-Cancer Screening. N Engl J Med 2013;368(8):728-736.
21. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The probability of malignancy in solitary pulmonary nodules: application to small radiologically indeterminate nodules. Arch Intern Med 1997;157(8):849.
22. McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K, et al. Probability of cancer in pulmonary nodules detected on first screening CT. N Engl J Med 2013;369(10):910-919.

23. Gershanik EF, Lacson R, Khorasani R. Critical finding capture in the impression section of radiology reports. AMIA Annu Symp Proc 2011;2011:465-469.

24. Zeliadt S, Feemster L, Hammond K, Au D, Takasugi J. Development of a natural language processing tool to use electronic text reports to categorize lung nodules. Poster session presented at: The Crossroads of Radiology. Annual Conference of the American College of Radiology; 2015 May 17-21; virtual meeting.

25. Pinsky PF, Gierada DS, Black W, Munden R, Nath H, Aberle D, Kazerooni E. Performance of Lung-RADS in the National Lung Screening Trial: a retrospective assessment. Ann Intern Med. 2015 Apr 7;162(7):485-91.

26. McKee BJ, Regis SM, McKee AB, Flacke S, Wald C. Performance of ACR Lung-RADS in a clinical CT lung screening program. J Am Coll Radiol. 2015 Mar;12(3):273-6.

27. Wiener RS, Gould MK, Slatore CG, Fincke BG, Schwartz LM, Woloshin S. Resource use and guideline concordance in evaluation of pulmonary nodules for cancer: too much and too little care. JAMA Intern Med 2014;174:871-80.

28. Tanner NT, Aggarwal J, Gould MK, Kearney P, Diette G, Vachani A, Fang KC, Silvestri GA. Management of pulmonary nodules by community pulmonologists: A multicenter observational study. Chest. 2015 Jun 18.

29. Masciocchi M, Wagner B, Lloyd B. Quality review: Fleischner criteria adherence by radiologists in a large community hospital. J Am Coll Radiol. 2012 May; 9(5): 336-339.

30. Lacson R, Prevedello LM, Andriole KP, Gill R, Lenoci-Edwards J, Roy C, Gandhi TK, Khorasani R, Fleischner Society. Factors associated with radiologists' adherence to fleischner society guidelines for

31. management of pulmonary nodules. J Am Coll Radiol. 2012 Jul; 9(7): 468-473.

32. Eisenberg RL, Bankier AA, Boiselle PM. Compliance with fleischner society guidelines for management of small lung nodules: A survey of 834 radiologists. Radiology. 2010 Apr; 255(1): 218-224.

33. Esmaili A, Munden RF, Mohammed TL. Small pulmonary nodule management: A survey of the members of the society of thoracic radiology with comparison to the fleischner society guidelines. J Thorac Imaging. 2011 Feb; 26(1): 27-31.

34. Cancer research network. Healthcare Delivery Research Program, Division of Cancer Control & Population Sciences. (Accessed April 24, 2016, at http://crn.cancer.gov/).