American Journal of Preventive Cardiology 9 (2022) 100300

Contents lists available at ScienceDirect

American Journal of Preventive Cardiology

journal homepage: www.journals.elsevier.com/the-american-journal-of-preventive-cardiology

Leveraging structured and unstructured electronic health record data to detect reasons for suboptimal statin therapy use in patients with atherosclerotic cardiovascular disease ☆

Glenn T. Gobbel[a,b], Michael E. Matheny[a,b], Ruth R. Reeves[a,b], Julia M. Akeroyd[c], Alexander Turchin[d,e], Christie M. Ballantyne[f], Laura A. Petersen[c], Salim S. Virani[c,f,*]

[a] Geriatric Research, Education and Clinical Center, Tennessee Valley Healthcare System, Department of Veterans Affairs, Nashville, TN, USA
[b] Department of Biomedical Informatics, School of Medicine, Vanderbilt University, Nashville, TN, USA
[c] Health Policy, Quality & Informatics Program, Michael E. DeBakey VA Medical Center Health Services Research & Development Center for Innovations, Houston, Texas; and Section of Health Services Research, Department of Medicine, Baylor College of Medicine, Houston, TX, USA
[d] Harvard Medical School, Boston, MA, USA
[e] Division of Endocrinology, Brigham and Women's Hospital, Boston, MA, USA
[f] Section of Cardiology and Cardiovascular Research, Department of Medicine, Baylor College of Medicine, Houston, TX, USA

ARTICLE INFO

Keywords:
Clinical decision support
Natural language processing
Clinical informatics
Lipid-lowering therapy
Cholesterol management
Statins
Cardiovascular disease
Atherosclerosis
Clinical inertia
Side effects

ABSTRACT

Objective: To determine whether natural language processing (NLP) of unstructured medical text can improve identification of ASCVD patients not using high-intensity statin therapy (HIST) due to statin-associated side effects (SASEs) and other reasons.

Methods: Reviewers annotated reasons for not prescribing HIST in notes of 1152 randomly selected patients from across the VA healthcare system treated for ASCVD but not receiving HIST. Developers used reviewer annotations to train the Canary NLP tool to detect and extract notes containing one or more of these reasons. Negative predictive value (NPV), sensitivity, specificity and Area Under the Curve (AUC) were used to assess accuracy at detecting documents containing reasons when using structured data, NLP-extracted unstructured data, or both data sources combined.

Results: At least one documented reason for not prescribing HIST occurred in 47% of notes. The most frequent reasons were SASEs (41%) and general intolerance (20%). When identifying notes containing any documented reason for not using HIST, adding NLP-extracted, unstructured data significantly (p<0.05) increased sensitivity (0.69 (95% confidence interval [CI] 0.60–0.76) to 0.89 (95% CI 0.81–0.93)), NPV (0.90 (95% CI 0.87 to 0.93) to 0.96 (95% CI 0.93–0.98)), and AUC (0.84 (95% confidence interval [CI] 0.81–0.88) to 0.91 (95% CI 0.90–0.93)) compared to structured data alone.

Conclusions: NLP extraction of data from unstructured text can improve identification of reasons for patients not being on HIST over structured data alone. The additional information provided through NLP of unstructured free text should help in tailoring and implementing system-level interventions to improve HIST use in patients with ASCVD.

Statin therapy, especially high-intensity statin therapy (HIST), significantly reduces all-cause mortality and the incidence of recurrent cardiovascular events in high-risk patients, particularly those with established atherosclerotic cardiovascular disease (ASCVD) [1–3]. Based on multiple randomized clinical trials demonstrating a direct relationship between the magnitude of low-density lipoprotein (LDL-C) reduction and improvement in recurrent ASCVD events, [2, 4-6] the 2018 American Heart Association/ American College of Cardiology (AHA/ACC) multi-society cholesterol guideline strongly recommends treatment of patients with ASCVD with HIST as a Class I recommendation [7].

While treatment guidelines continue to recommend HIST in patients with ASCVD, [8,9] statin utilization and, especially, HIST use continues to lag despite their clear benefits in high-risk patients including those with ASCVD [8–17]. Although statin associated side effects (SASEs), often denoted as intolerance and commonly related to myalgia, may account for a substantial proportion of this treatment gap, studies have

☆ The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

* Corresponding author at: Health Services Research and Development (152), Michael E. DeBakey Veterans Affairs Medical Center, 2002 Holcombe Blvd., Houston, TX 77030.

E-mail address: virani@bcm.edu (S.S. Virani).

https://doi.org/10.1016/j.ajpc.2021.100300
Received 10 August 2021; Received in revised form 19 October 2021; Accepted 27 November 2021

also shown clinical inertia (failure to initiate or intensify therapy when clinically indicated) as another important reason [18,19]. It is important to identify ways to efficiently understand these reasons so specific interventions could then be designed to address each of these gaps.

Some electronic health record (EHR) systems, like that of the Department of Veterans Affairs (VA), provide data regarding medication adverse events in structured form. However, such events are commonly entered only as unstructured free text in clinical notes [20]. In a previous report, we demonstrated that structured data alone can be used to detect SASEs in patients with ASCVD receiving care in the VA health care system with a sensitivity and negative predictive value (NPV) of 63.4% and 85.3%, respectively, compared to manual chart review by human annotators [21]. Although NPV was within an acceptable range for precisely identifying patients without SASEs, the lower sensitivity indicated that over a third of patients with SASEs could be missed using structured data alone. Review and extraction of SASEs from free text entered in the EHR by clinicians, if made practical, might improve sensitivity. Given the volume of unstructured text, manual note review for detecting SASEs and tailoring the information delivered to clinicians is not feasible, but natural language processing (NLP) can automate note review, SASE identification, and system throughput [20,22-24]. Here we describe the augmentation of the Canary NLP tool to identify the SASE-containing notes of patients with ASCVD treated within the VA healthcare system [20, 22-24]. We test the performance of the tool before and after adaptation for use on ASCVD patient notes generated within the VA, and we describe the process of adapting this tool for use. Furthermore, we assessed whether the tool adds information beyond that provided by identification of SASEs using structured data alone.

1. Methods

1.1. Study design and patient cohort

We sought to develop an accurate NLP-based tool that would analyze unstructured text and determine if that text contained one or more documented reasons for not placing a patient with ASCVD on guideline-concordant statin therapy. To accomplish this, we first defined ASCVD patients based on the presence of any ICD-9 code for ischemic heart disease, ischemic cerebrovascular disease, or peripheral artery disease within the patient's record. The methodology for the ascertainment of ASCVD is described in detail in our prior studies and showed a positive predictive value of 95% for the identification of ASCVD patients using our methodology compared with manual chart review [21]. Given our aim to identify the presence or absence of one or more reasons, such as SASEs, for non-guideline concordant statin use in ASCVD patients (i.e. ASCVD patients not on HIST), we restricted the cohort to patients with ASCVD either not on a statin or being treated with low to moderate intensity therapy. In view of our eventual goal of creating a clinical decision support tool to assist clinicians in the outpatient setting to optimizing statin intensity to the highest dose in patients with ASCVD as recommended by the cholesterol guidelines, we further restricted our study to veterans with one or more outpatient encounters in a VA clinic between October 1, 2014 and September 30, 2015. We then randomly selected 1152 of these patients for inclusion in our study. Previous work has shown that the number of notes needed to train an NLP system is dependent on both the variation in phrases used to express a particular concept, like SASE, and the prevalence of the concept in notes. Based on a preliminary review of SASE documentation in our cohort, we estimated that 1152 patients would provide a representative and sufficient number of notes for developing our NLP system and statistically testing its accuracy. To ensure that our NLP system training accounts for both absolute and relative intolerance to statin therapy, our sample included 50% of the patients not on any statin therapy, 25% on low-intensity statin therapy, and 25% of the patients on moderate-intensity statin therapy.

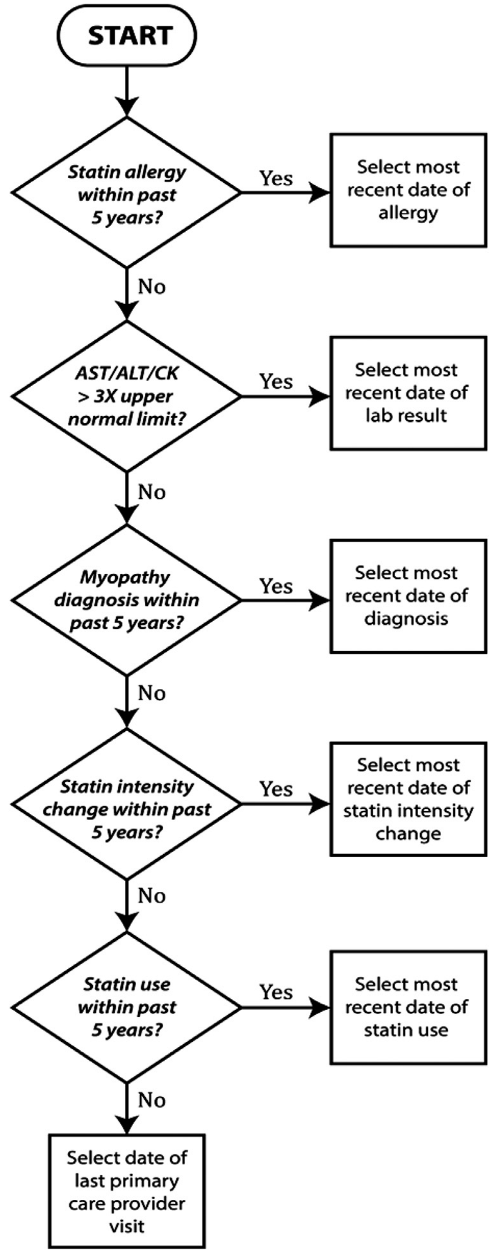


Fig. 1. Flow diagram used to identify dates for selecting notes more likely to include a rationale for statin therapy guideline non-adherence (AST – aspartate transaminase; ALT – alanine transaminase; CK creatine kinase.

1.2. Anchor date based note selection

The Corporate Data Warehouse (CDW) of the Veterans Health Administration was the source for all notes used in this study. Due to concern regarding the potentially low prevalence of notes documenting a reason for prescribing a suboptimal statin intensity, we developed an algorithm to increase that prevalence within the corpus to ensure there would be sufficient notes for developing, training, and testing the NLP tool. Using available patient data, the algorithm proceeded through an ordered, sequential set of criteria to identify an "anchor date," defined as the most likely date for documentation of a reason for suboptimal statin therapy (Fig. 1). Because of the presumption that reasons for changing statin dosage were more likely to be documented around the anchoring event, we used a one-year window from 9 months prior to 3 months after the anchor date for note selection.

1.3. Study sample

We restricted the corpus to notes corresponding to the 1152 patients mentioned in the previous "Study Design and Patient Cohort" subsection and written by clinicians likely to document reasons for non-adherence to statin therapy guidelines, including physicians, nurse practitioners, physician assistants, and pharmacists working in primary care or cardiology clinics. The average number of notes per patient was 79, and we randomly selected a single note from each set of patient notes. We randomly divided the resulting corpus of 1152 notes into 3 batches of 57 notes for training of annotators and 9 batches of 128 notes for NLP system development. Patients within each batch were stratified to include those not on a statin (50%), those on low intensity therapy (25%), and those on moderate intensity therapy (25%).

To ensure that notes used for developing and testing the system contained a wide range of phrases that might be used to express the concepts of interest (such as SASEs), each note in the corpus was selected from a distinct patient. To enhance the inclusion of gender and ethnicity specific language as a part of the NLP model development process and to improve the representativeness and external validity of our findings, we oversampled notes from female and African-American patients (underrepresented among VA population) to ensure that each development and testing batch had 25% female and 25% African-American patients.

1.4. Structured data for SASE identification

The adverse drug event reporting system (VA ADERS) within the VA is a robust source for post-marketing drug surveillance, and VA healthcare providers can enter adverse drug events (ADRs) into the system on a voluntary basis [25]. This information is deposited in a central repository (ADR files), and these files served as the source of the structured data used in our study. Inclusion of any of 7 commonly used statin medications (simvastatin, lovastatin, rosuvastatin, atorvastatin, pravastatin, fluvastatin, and pitavastatin) within the ADR files associated with a given patient was accepted as confirmatory evidence for the presence of one or more SASEs within that patient.

1.5. Annotation schema development

The annotation schema defined and provided examples of specific concepts of interest for reviewers to identify SASEs within the clinical, unstructured text notes. The goal was to use the concept-related phrases captured by the reviewers to evaluate NLP tool accuracy and to identify approaches for optimizing concept recognition performance. To develop the schema, clinical experts and the annotation team first delineated specific concepts that might be used to document why a patient was not on HIST. The annotation team then annotated small subsets of 20–30 trial notes for the various concepts, and the entire group then reviewed the annotations for further schema refinement. This continued iteratively until no further changes to the schema were indicated. Concepts of interest corresponded to sequences of tokens in the text, and we defined tokens as strings of characters corresponding to minimal elements of meaning such as words, numbers, and units of measure. The schema included four broad, mention level concepts, including references to 1) statins and other lipid-lowering pharmaceutical agents, 2) symptoms associated with statin use (subdivided into musculoskeletal, gastrointestinal, neurocognitive, hepatobiliary, allergic, cutaneous, and those associated with other organ systems) and general intolerance (presence of symptoms but no further specification of symptom or organ system), 3) patient refusal, and 4) perception of a lack of statin efficacy. Reviewers also assigned assertion status (positive or negative) to all annotated concepts. Once reviewers completed annotating mention level concepts, they annotated the note at the document level to indicate whether the note contained any reason (i.e., positive assertion of symptom associated with statin use, patient refusal, or lack of statin efficacy) for a patient not taking a high-intensity statin.

1.6. Annotation of unstructured notes

For annotation purposes, two nurse annotators sequentially reviewed batches of 57 notes. Training was stopped once inter-annotator agreement (IAA) exceeded 0.80, which occurred after the third batch. Annotator agreement at the document level was evaluated based on Cohen's Kappa [26]. Both annotators then reviewed the same 9 batches of 128 notes (1152 notes total), and a third reviewer compared the two sets of annotated notes and adjudicated any disagreements. This adjudicated note set constituted the reference standard for training the NLP tool and testing its performance. IAA at the concept level was defined as the F1-measure between the mention level annotations of two reviewers with the annotations of one reviewer serving as a reference. F1-measure is used as the IAA metric at the concept level because, unlike patient level assessments, true negatives (TNs) used in calculating Cohen's Kappa are not present in that analytic frame.

1.7. NLP tool development

We generated the NLP system for SASE extraction for analysis of VA clinical notes by adapting an existing NLP model, which was originally developed for identifying SASEs in provider notes from Mass General Brigham using a publicly available NLP platform, Canary [23,24]. Users of Canary can configure the platform to identify concepts in the text by first defining the grouping of words within "word classes", which are user-generated categories of words with similar semantic meaning. Users can then craft grammatical rules known as "phrase structures" organized into sequential levels or "tiers", which enables Canary to synthesize higher-level concepts based on the recursive combination of lower-level word classes and phrase structures. The parallel processing provided by the Canary tool can support a throughput of 10.4 MB of text per minute, which equates to 1651 notes per minute based on the average note size of our test corpus of 6.3 kB [23]. Because sentence boundaries found in clinical notes are often ill defined, which can interfere with concept detection and reduce performance of NLP systems that rely on accurate boundary delineation, we identified sentence boundaries using the RapTAT NLP tool that had been previously optimized for sentence detection in VA notes [27]. RapTAT removed line wrapping and delineated sentence boundaries as a pre-processing step prior to concept detection by the Canary-based NLP tool.

Using an iterative development process (**Fig. 2**, **Left**), we gradually modified the rules of the original Canary model to adapt to the VA system, employing cycles of system performance testing on batches of 128 notes randomly selected from the full note corpus, which was followed by root error analysis of false positive (FP) and false negative (FN) errors at the mention level and any modifications to Canary word classes or phrase structures indicated by that analysis. After iterating through 5 note batches ($n = 640$), we detected no further increases in performance. We refer to those first 5 batches as the development set. The remaining 4 batches (n = 512) constituted the testing set (**Fig. 2**, **Right**), which was used to assess final tool performance.

1.8. Error evaluation

We evaluated system errors at two levels of granularity. The most granular level was based on the agreement between concept mentions identified by the NLP system and those identified by the reference standard from human annotation, which allowed us to measure tool performance when optimizing the tool to detect the specific phrases corresponding to concepts outlined in the annotation schema. At this level of granularity, we defined a true positive (TP) as a concept mention detected by the tool that overlapped any of the characters of a phrase contained in the reference standard from human annotation and corresponding to the same concept. FPs were phrases detected by the tool but not found in the reference standard, and FNs were phrases in the reference standard but undetected by the tool.

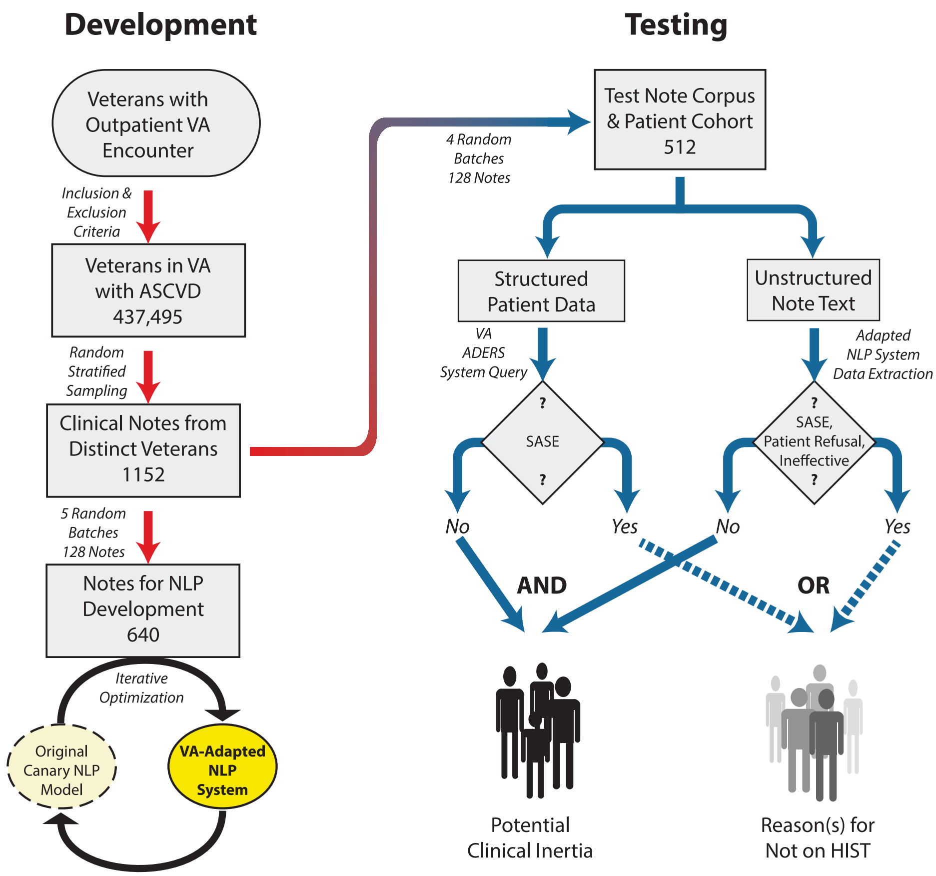


Fig. 2. Flow diagram demonstrating the process used to develop the NLP system (left) responsible for identifying reasons for VA atherosclerotic cardiovascular disease (ASCVD) patients not being on high-intensity statin therapy (HIST). Also shown (right) is the process used to test the accuracy of the system with respect to classifying patients according to whether such a reason existed in a) structured data stored in the VA adverse drug event (ADERS) system, b) unstructured form within the text of clinical notes, or c) neither, indicating potential clinical inertia.

Because we plan to use our tool to generate clinical decision support for improving statin therapy use in ASCVD patients, and because we will base those guidelines on whether there are documented reasons for not using HIST, our primary focus was performance at the less granular patient level. Because all notes used in our study were from distinct patients, performance at the patient level on the unstructured data alone was equivalent to that at the notes level. When assessing system performance based on unstructured data alone, the number of true positive patients equaled the number of notes for which both the annotated reference standard and the NLP-based analysis indicated the presence of at least one positive assertion of a reason for lack of HIST use. The number of true negative patients corresponded to the number of notes for which neither the NLP tool nor the annotated reference standard identified any reason for lack of HIST use. We also assessed system performance at the patient level when using only structured data or both unstructured and structured combined. When assessing system performance using structured data alone, finding that a SASE had been recorded in the ADR file of a patient was accepted as ground truth, and all patients identified as having a history of SASEs based on structured data alone were accepted as TPs. When assessing system performance based on evaluation of both structured and unstructured data from a patient, we considered a finding of statin refusal or perceived lack of efficacy in the unstructured data or an SASE in either dataset as positive evidence of a reason for not using HIST in that patient (Fig. 2, right).

1.9. System performance analysis

Measures used to assess NLP performance with respect to identifying specific mentions of SASEs, patient refusal, and lack of statin efficacy in the text at an individual concept level included positive-predictive value (PPV; equivalent to precision), sensitivity (equivalent to recall), and F1-measure (harmonic mean of PPV and sensitivity) for mentions within all NLP-processed notes. Measures used to assess system performance at the patient level included specificity, NPV, PPV, sensitivity and AUC. For the patient assessments, the area under the receiver-operating characteristic (ROC) curve (AUC) was calculated from the true and false positive rates (TPR and FPR, respectively) for each system evaluated, where TPR was the ratio of TPs to the sum of TPs and FNs (i.e., positive findings of any reason for not being on HIST in the reference set), and FPR was the ratio of FPs to the sum of TNs and FPs (i.e., negative findings in the reference set). When a single point defines the ROC curve as in this study, it can be shown that the AUC is equivalent to (1+TPR-FPR)/2. We used McNemar's test to assess statistical significance ($p<0.05$) of the differences in patient level performance when using different data types (structured, unstructured, and both types combined) to detect the presence or absence of reasons for not adhering to statin guidelines. Bootstrap analysis generated 95% confidence intervals for performance values.

2. Results

The annotators identified one or more reasons for not using the indicated statin therapy in 543 (47.1%) of the 1152 reviewed notes. There was substantial variation in the nature of the reason and the way it was documented, and many notes contained more than one reason. A total of 115 notes documented statin-associated musculoskeletal side effects, 109 documented general statin intolerance, 107 provided evidence of SASEs experienced in non-musculoskeletal organ systems, 49 indicated patient refusal to take statins in the absence of SASEs, 14 suggested lack of statin efficacy in the patient, and 207 of the notes documented the general presence of SASEs in the patient by including one or more statins in a list of medications placed under the heading of "adverse effects" or "allergies."

Table 1
Inter-annotator agreement for mentions of statins, patient refusal of a statin medication, and statin-associated side effects within the 512-note evaluation set.

Concept	Agreements	Disagreements	F1
Statin Mention	373	76	0.91
Statin-Associated Side Effect	399	19	0.91
Statin Refusal	48	19	0.83
Perceived Lack of Efficacy	10	29	0.41
Totals	**830**	**202**	**0.89**

Table 2
Confusion matrix comparing the two annotators, A1 and A2, with respect to detection of the presence (+) or absence (-) of explicit justification for patients not being on guideline-concordant statin therapy within the evaluation set of 512 notes.

		A1		*Totals*
		+	–	
A2	+	165	21	*186*
	–	4	322	*326*
Totals		*169*	*343*	***512***

To calibrate NLP system performance relative to that of the reviewers, we assessed agreement between the annotators specifically within the note set used for final NLP system testing (Table 1). Mention level IAA measured over all 4 concepts and calculated based on F1 was 0.89 (**Table 1; last row**), ranging from a low of 0.41 for mentions implying possible lack of statin efficacy to a high of 0.91 for mentions of medications in the statin drug class and SASEs. The Cohen's Kappa of 0.89 indicated a high level of IAA at the note level, which was the agreement between annotators when assessing whether a given note contained any reason for a clinician not prescribing HIST (**Table 2**).

When the original, unmodified Canary NLP model was run on the first development set of 128 notes, PPV, sensitivity and AUC with regard to detecting notes with one or more reasons for not adhering to statin guidelines were 0.82, 0.17, and 0.28, respectively. The low sensitivity was due to the tool missing concept level mentions of reasons for SASE-based and other reasons noted by the annotators. There were 102 reasons documented in the first batch of notes in the development set, including 78 SASEs, 19 patient refusals, and 1 lack of perceived efficacy; 4 were uncategorized (e.g., anticipated liver transplant). The original Canary model detected 14 of the reasons, all of which were SASEs. Failure analyses revealed that undetected reasons arose from three sources. The first was that the Canary tool often split text on non-sentence boundaries. Because we used and optimized the RapTAT NLP system for detection of sentence boundaries specifically in notes from the VA EHR [27], we decided to pre-process each note and explicitly delineate sentences using RapTAT before further Canary processing. Other errors were associated with the inability of the existing Canary model to identify phrases that may be particular to VA providers when documenting SASEs. Lastly, the original Canary model was not designed to capture two of the reasons for not prescribing statins, patient refusal and perceived lack of efficacy. We addressed these last two error sources by augmenting the original Canary model with additional word classes, reducing the number of "phrase structures" in the Canary model from 1246 to 103 and increasing the number of "tiers" from 3 to 8.

Once iterative error analysis and model modifications led to no further improvements in NLP system performance when detecting reasons for not prescribing high intensity statins, we tested the system on 4, randomly selected note batches (i.e., notes from 512 distinct patients) with respect to its ability to detect the presence or absence of at least one reason (i.e., SASE, refusal, or lack of efficacy) for a patient not being on HIST (Fig. 2, **Right**). Within these notes, there were 132 notes containing one or more reasons for an ASCVD patient not being on HIST. Using structured data alone, we detected only 91 of these patients for a sensitivity of 0.69 (**Table 3, Column 1**). The original Canary model detected just 47 of the patients **(Table 3, Column 2)**, but iterative model modifications improved detection to 108 patients **(Table 3, Column 3).** Combining unique patients identified using structured data with those identified using the optimized Canary model detected the highest number of patients, 117 **(Table 3, Column 4)**, which corresponded to a significant increase in sensitivity over structured data alone, from 0.69 to 0.89. NPV also increased from 0.90 to 0.96 as the number of false negatives decreased from 41 for structured data alone to 15 when structured and unstructured data were combined. Any SASEs found within the structured data were accepted as representative of truth, and, accordingly, PPV and specificity were 1.00 for such data. The incorrect identification of 22 patients as having a rationale for not being on high-intensity statins (false positives) by the use of unstructured data provided by the optimized NLP system corresponded to a significant reduction in PPV and specificity to 0.84 and 0.94, respectively. The highest AUC, which accounts for both sensitivity and specificity, was 0.91, was achieved by combining both structured and unstructured data in the assessment and was significantly higher than the AUC of 0.84 for structured data alone.

System performance varied depending on the intensity of statin therapy (**Supplemental Table 1**). In the low-intensity statin therapy group, combining structured and unstructured data increased AUC only slightly, from 0.77 to 0.79. In this particular group, the improvement in sensitivity afforded by the inclusion of unstructured data was partially offset by a larger decrease in PPV than observed in the other statin intensity groups. In patients not on statins or on moderate-intensity statin therapy, adding unstructured to the structured data increased AUC from 0.83 to 0.89 and from 0.82 to 0.87, respectively.

3. Discussion

Despite the demonstrated benefit of HIST on reducing recurrent cardiovascular event risk and mortality in ASCVD patients, many such patients are treated at doses below those indicated based on current treatment guidelines. Identifying the barriers hindering the use of optimal statin doses could help ascertain the type of information and assistance needed by clinicians to improve guideline-concordant statin use in high-risk patients with ASCVD. The results of the present study demonstrate that NLP can improve detection clinical notes containing documented reasons that might account for a less than optimal statin dose, including SASE, patient refusal, and perceived lack of drug efficacy **(Central Illustration)**. Additionally, there were benefits to using NLP-based analysis of unstructured data in combination with structured data. Adding NLP-based analysis increased both the sensitivity for detecting patients for which there are documented reasons for not using HIST from 0.69 to 0.89 and the AUC from 0.84 to 0.91 (Table 3) over the use of structured data alone.

The VA EHR, like many others, allows for entry of structured elements when documenting patient-specific adverse events associated with particular drugs or other therapeutic interventions, but many clinicians choose to only use free text when entering these events. Although reasons for this are not known, those could include lack of time, clinician's belief that these are not true side effects or adverse events and clinician aversion to entering these in the structured datasets as allergies. Although we did not attempt to determine why a clinician might choose not to create a structured entry, EHR design can influence structured data entry by influencing the effort taken to enter a drug-related adverse event [20]. Another potential contributing factor is the limited expressivity inherent in structured entries [28], which may prevent inclusion of contextual information and exclude entry of other factors contributing to patient not being on HIST. Using only structured data could therefore lead to an incomplete determination of the factors contributing to guideline non-adherence, which could interfere with providing

Table 3
Impact of data type (structured, unstructured, or both types combined) and NLP model optimization on detecting one or more reasons for a patient with cardiovascular disease not being on a high-intensity statin.

	Structured Data Only	Unstructured Data Only (OriginalCanary Model)	Unstructured Data Only (Optimized Canary Model)	Structured & Unstructured (OptimizedCanary Model)
True Positives	91	47	108	117
False Positives	0	7	22	22
True Negatives	380	367	358	358
False Negatives	41	91	24	15
Sensitivity	0.69 *(0.60–0.76)*	[b]0.34 *(0.25–0.41)*	0.82 *(0.74–0.87)*	[a]0.89 *(0.81–0.93)*
Specificity	1.00 *(1.00–1.00)*	[b] 0.98 *(0.97–0.99)*	[b] 0.94 *(0.92–0.96)*	[b] 0.94 *(0.92–0.96)*
PPV	1.00 *(1.00–1.00)*	[b]0.87 *(0.76–0.96)*	[b] 0.83 *(0.76–0.88)*	[b]0.84 *(0.69–0.90)*
NPV	0.90 *(0.87–0.93)*	[b] 0.80 *(0.76–0.83)*	0.94 *(0.91–0.96)*	[a]0.96 *(0.93–0.98)*
AUC	0.84 *(0.81–0.88)*	[b]0.66 *(0.63–0.69)*	0.88 *(0.85–0.91)*	[a]0.91 *(0.90–0.93)*

The bottom half of the table provides performance values and their bootstrap-generated 95% confidence intervals.
PPV – Positive Predictive Value; NPV – Negative Predictive Value; a,b – significant ($p<0.05$) increase (a) or decrease (b) in performance value relative to structured data alone.

clinicians with information most useful in overcoming the true, patient-specific barriers to HIST. Based on the sensitivity for detection of statin-related adverse events (Table 3), our current results suggest that using unstructured data alone may also miss approximately one -third (31%) of the adverse events and other reasons for lack of HIST use. In our analyses, structured data helped identify clinical anchors around which we pulled the notes for processing by our NLP algorithm. Therefore, an informatics-based approach leveraging both structured and unstructured data seems to be the most effective strategy in identifying why patients with ASCVD are not on guideline concordant statin therapy.

Our results are important in several ways. First, these results demonstrate that we can very accurately identify the proportion of patients with ASCVD whose clinical records contain documented justification for not prescribing guideline-concordant statin therapy. Our results show that the overall yield of identifying patients with documented justification is highest when using both structured and unstructured data. Implementation of these results at the level of a facility or a health care system could identify what proportion of ASCVD patients are not on guideline-concordant statin therapy due to reasons documented in the clinical record versus no documented reasons. Lack of documented reasons in such cases could indicate presence of clinical inertia (i.e. failure to initiate or intensify therapy when clinically indicated).

Even with detailed clinical note review by experienced annotators, documented reasons for ASCVD patients not being on HIST could only be ascertained in ~47% of those patients. Although failure to document reasons for a patient not being on HIST could partially explain this, our results also show that clinical inertia does play a significant role in low use of HIST among patients with ASCVD.

We have recently initiated a clustered clinical trial that will use the NLP system described herein in a subset of VA medical care centers with a goal of improving HIST prescribing among patients with ASCVD. In this trial, once clinical inertia or the reason for lack of HIST use is identified by the NLP system, appropriate informatics supported tools are being deployed in the EHR to provide point-of-care cognitive support to clinicians to address each of the reasons associated with a lack of guideline- concordant HIST use in ASCVD patients. By leveraging unstructured data from clinician text notes, our analyses also show that this tool can also identify patient refusal to statin therapy in the absence of SASEs, which was identified in 49 patients. This is important when clinicians, facilities, and health care systems are evaluating their performance against national benchmarks.

Optimizing performance of NLP systems commonly requires creating language models carefully molded to the targeted environment and use case, which can limit generalizability of such systems to other healthcare systems and tasks. The NLP tool used in our study, Canary, had been previously developed and used for detection of statin-related adverse events [22]. Given the similarity of that task to our own and that both tasks involved similar note types (progress notes) and environments (large medical centers), we were hopeful that adapting it for VA notes would require only minimal changes. When the previous Canary-based NLP model was used on VA notes to detect potential causes of statin guideline non-adherence, limited sensitivity at the phrase (0.17) and patient (0.36; Table 3) levels indicated that attaining adequate performance would require changes to both the tool and model. Because VA notes 1) can originate from over 130 different medical facilities, each with their own institutional note titles, and 2) may contain facility-specific text automatically inserted for documentation purposes, they likely have greater structural and language variations than notes from a single institution such as the one where the tool was originally developed. Generalizing NLP systems across use cases and institutions is often challenging, and our results indicate that even seemingly minor differences in domain or use case can have a substantial impact on performance of a previously developed NLP system. Our results highlight the need to first adapt each NLP tool to a health care system before its system wide implementation for quality assessment and quality improvement. Given the considerable inter-institutional variation in the language and forms used to document VA patient encounters, we anticipate that the additional Canary NLP system development effected through this study should improve its generalizability and performance when used in other health care organizations. However, optimal performance would likely require further refinement with respect to Canary word classes and phrase structure definition and sentence boundary detection.

When encouraging providers to change from a current therapy plan, tailored guidance would likely have the greatest impact because the optimal therapeutic options may vary depending on the barrier to guideline adherence. For example, clinical inertia would best be addressed by providing different forms of cognitive assistance than those used when helping to circumvent particular SASEs. For the former, identifying ASCVD patients not on guideline-concordant doses of statin therapy, benefits of high-intensity statin therapy, and what constitutes various intensities of statin therapy would suffice. On the other hand, management of patients with SASE may require cognitive support for clinicians on how to discuss risk-benefits of statin therapy in such patients and treatment algorithms geared towards maximizing the tolerated dose of statin therapy followed by the use of non-statin lipid lowering therapies shown to be of benefit in clinical trials. Given the much greater speed of analysis by NLP compared to manual review together with the high NPV (0.94; Table 3) of the system when structured and unstructured data are combined, automatic provisioning of appropriate cognitive assistance tools, such as providing reminders of ASCVD patients not on high-intensity statins and guided decision trees in modifying statin therapy given the underlying barrier, should be possible. Preliminary testing suggests that the system can be engineered to handle even the daily stream of documents produced in a large healthcare system like the VA. Initial analyses indicate a rate of 50 thousand documents per day running as a single process using modest computational resources (4-core virtual CPU with 6 GB of random-access memory). With parallelization, the Canary NLP system can reportedly process over 2 million documents of the size used in our study [23].

One of the limitations of using NLP and unstructured data noted in our analyses is that it can reduce PPV and specificity relative to data entered in structured form or extracted through manual review. One contributing factor to this reduction is that we assume these data sources are correct and that PPV and specificity are perfect (Table 3) even though inaccuracies in structured and manually extracted data do occur. Also contributing are the ambiguities introduced into unstructured text through use of non-standard and highly variable acronyms and abbreviations and syntactic constructs implemented to efficiently document clinical conditions and treatments of patients. Such ambiguities can be difficult to resolve consistently even with manual review, so measures to improve sensitivity using NLP have to be carefully weighed against potential reductions in precision. When using this system for clinical decision support, we will reduce the reliance on precision with respect to identifying specific reasons (e.g., musculoskeletal versus hepatobiliary dysfunction) for lack of HIST use and instead only assess presence or absence of any documented reason for not being on guideline-concordant statin therapy. Our testing of final NLP system performance was at the patient note and not the concept mention level, but further refinements of the Canary model and addition of alternative NLP approaches could make it possible to identify specific reasons for non-adherence and where these are expressed in clinical text. Such a capability would allow for the system to reveal the exact SASEs and other factors that may impact statin therapy and show how these factors are documented in the historical record.

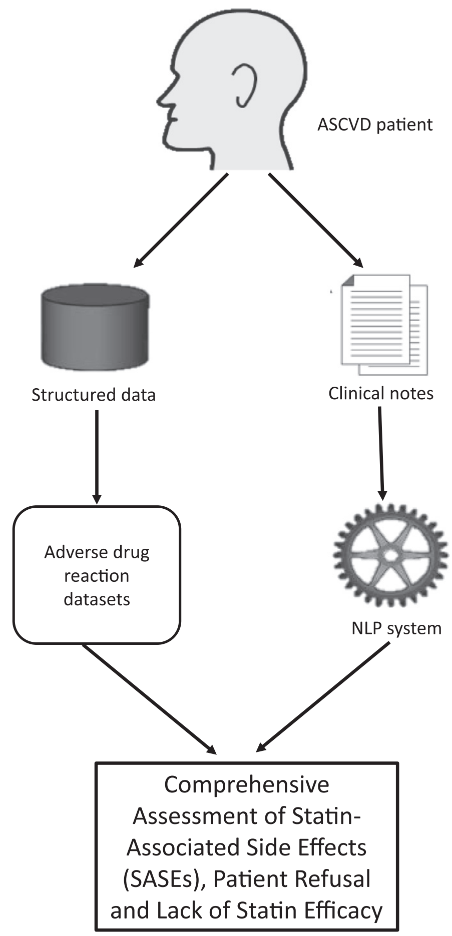
Data obtained through NLP regarding statin therapy of ASCVD patients can augment that found in structured form. When searching for documented reasons for not treating a patient in accordance with treatment guidelines, combining structured with unstructured data can significantly improve sensitivity and NPV over the use of structured data alone while continuing to maintain high specificity and an acceptable level of precision. The additional information provided through NLP of unstructured free text should help in tailoring and implementing system-level interventions to improve HIST use due to SASEs or clinical inertia in patients with ASCVD

Authors' contributions

All authors have contributed to the conceptualization and design of the study and/or the data curation, formal analysis, drafting or reviewing and editing the manuscript. Each author has approved the final article.

Funding

This work was supported by research funding from the Department of Veterans Affairs Health Services Research & Development (IIR 16-072). This work was also supported by the Houston VA HSR&D Center for Innovations grant CIN 13-413.



Conflicts of interest/Competing interests

Dr. Virani receives research support from the Department of Veterans Affairs, World Heart Federation, Tahir and Jooma Family and honorarium from the American College of Cardiology (Associate Editor for Innovations, acc.org). **Dr. Ballantyne**: Grant/Research Support- All significant. (All paid to institution, not individual): Abbott Diagnostic, Akcea, Amgen, Esperion, Ionis, Novartis, Regeneron, Roche Diagnostic, NIH, AHA, ADA. Consultant- Abbott Diagnostics, Althera, Amarin*, Amgen, Arrowhead, Astra Zeneca, Corvidia, Denka Seiken*, Esperion, Genentech, Gilead, Matinas BioPharma Inc, New Amsterdam*, Novartis, Novo Nordisk, Pfizer, Regeneron, Roche Diagnostic, Sanofi-Synthelabo* **Dr. Turchin**: Grant/Research support: Astra Zeneca, Edwards, Eli Lilli, Novo Nordisk, Sanofi. **Others**: None

Ethics approval and consent to participate

This study was performed in line with the principles of the Declaration of Helsinki. The Institutional Review Boards at Baylor College of Medicine and the Michael E. DeBakey VA Medical Center approved the protocol. Informed consent from individual patients was waived due to the observational design of this study.

Acknowledgments

The authors thank Mark Kuebeler, MS, for his efforts and expertise in identification and acquisition of the data used in this project.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ajpc.2021.100300.

References

[1] Rodriguez F, Maron DJ, Heidenreich PA. Intensity of statin treatment and mortality-reply. JAMA Cardiol 2017;2:928–9. doi:10.1001/jamacardio.2017.0549.
[2] Rodriguez F, Maron DJ, Knowles JW, Virani SS, Lin S, Heidenreich PA. Association between intensity of statin therapy and mortality in patients with atherosclerotic cardiovascular disease. JAMA Cardiol 2017;2:47–54. doi:10.1001/jamacardio.2016.4052.
[3] Serban MC, Colantonio LD, Manthripragada AD, Monda KL, Bittner VA, Banach M, Chen L, Huang L, Dent R, Kent ST, Muntner P, Rosenson RS. Statin intolerance and risk of coronary heart events and all-cause mortality following myocardial infarction. J Am Coll Cardiol 2017;69:1386–95. doi:10.1016/j.jacc.2016.12.036.
[4] Cannon CP, Steinberg BA, Murphy SA, Mega JL, Braunwald E. Meta-analysis of cardiovascular outcomes trials comparing intensive versus moderate statin therapy. J Am Coll Cardiol 2006;48:438–45. doi:10.1016/j.jacc.2006.04.070.
[5] Cannon CP, Braunwald E, McCabe CH, Rader DJ, Rouleau JL, Belder R, Joyal SV, Hill KA, Pfeffer MA, Skene AM. Pravastatin or atorvastatin evaluation and infection therapy-thrombolysis in myocardial infarction 22 investigators. Intensive versus moderate lipid lowering with statins after acute coronary syndromes. N Engl J Med 2004;350. doi:10.1056/NEJMoa040583.
[6] LaRosa JC, Grundy SM, Waters DD, Shear C, Barter P, Fruchart JC, Gotto AM, Greten H, Kastelein JJ, Shepherd J, Wenger NK. Treating to New Targets (TNT) Investigators. Intensive lipid lowering with atorvastatin in patients with stable coronary disease. N Engl J Med 2005;352:1425–35. doi:10.1056/NEJMoa050461.
[7] Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, Braun LT, de Ferranti S, Faiella-Tommasino J, Forman DE, Goldberg R, Heidenreich PA, Hlatky MA, Jones DW, Lloyd-Jones D, Lopez-Pajares N, Ndumele CE, Orringer CE, Peralta CA, Saseen JJ, Smith SC Jr, Sperling L, Virani SS, Yeboah J. AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 2018;139:e1082–143 2019. doi:10.1161/CIR.0000000000000625.
[8] Nguyen V, deGoma EM, Hossain E, Jacoby DS. Updated cholesterol guidelines and intensity of statin therapy. J Clin Lipidol 2015;9:357–9. doi:10.1016/j.jacl.2014.12.009.
[9] Valentino M, Al Danaf J, Panakos A, Ragupathi L, Duffy D, Whellan D. Impact of the 2013 American College of Cardiology/American Heart Association cholesterol guidelines on the prescription of high-intensity statins in patients hospitalized for acute coronary syndrome or stroke. Am Heart J 2016;181:130–6. doi:10.1016/j.ahj.2016.07.024.
[10] Nanna MG, Navar AM, Wang TY, Li S, Virani SS, Li Z, Robinson JG, Roger VL, Wilson PWF, Goldberg AC, Koren A, Louie MJ, Peterson ED. Practice-level variation in statin use and low-density lipoprotein cholesterol control in the United States: results from the Patient and Provider Assessment of Lipid Management (PALM) registry. Am Heart J 2019;214:113–24. doi:10.1016/j.ahj.2019.05.009.
[11] Rosenson RS, Kent ST, Brown TM, Farkouh ME, Levitan EB, Yun H, Sharma P, Safford MM, Kilgore M, Muntner P, Bittner V. Underutilization of high-intensity statin therapy after hospitalization for coronary heart disease. J Am Coll Cardiol 2015;65:270–7. doi:10.1016/j.jacc.2014.09.088.
[12] Mahtta D, Ramsey DJ, Al Rifai M, Nasir K, Samad Z, Aguilar D, Jneid H, Ballantyne CM, Petersen LA, Virani SS. Evaluation of aspirin and statin therapy use and adherence in patients with premature atherosclerotic cardiovascular disease. JAMA Netw Open 2020;3:e2011051. doi:10.1001/jamanetworkopen.2020.11051.
[13] Virani SS, Woodard LD, Ramsey DJ, Urech TH, Akeroyd JM, Shah T, Deswal A, Bozkurt B, Ballantyne CM, Petersen LA. Gender disparities in evidence-based statin therapy in patients with cardiovascular disease. Am J Cardiol 2015;115:21–6. doi:10.1016/j.amjcard.2014.09.041.
[14] Hira RS, Cowart JB, Akeroyd JM, Ramsey DJ, Pokharel Y, Nambi V, Jneid H, Deswal A, Denktas A, Taylor A, Nasir K, Ballantyne CM, Petersen LA, Virani SS. Risk factor optimization and guideline-directed medical therapy in US veterans with peripheral arterial and ischemic cerebrovascular disease compared to veterans with coronary heart disease. Am J Cardiol 2016;118:1144–9. doi:10.1016/j.amjcard.2016.07.027.
[15] Rodriguez F, Lin S, Maron DJ, Knowles JW, Virani SS, Heidenreich PA. Use of high-intensity statins for patients with atherosclerotic cardiovascular disease in the Veterans Affairs Health System: practice impact of the new cholesterol guidelines. Am Heart J 2016;182:97–102. doi:10.1016/j.ahj.2016.09.007.
[16] Mahtta D, Ahmed ST, Ramsey DJ, Akeroyd JM, Lee MT, Rodriguez F, Michos ED, Itchhaporia D, Nasir K, Alam M, Jneid H, Ballantyne CM, Petersen LA, Virani SS. Statin prescription rates, adherence, and associated clinical outcomes among women with PAD and ICVD. Cardiovasc Drugs Ther 2020;34:745–54. doi:10.1007/s10557-020-07057-y.
[17] Virani SS, Kennedy KF, Akeroyd JM, Morris PB, Bittner VA, Masoudi FA, Stone NJ, Petersen LA, Ballantyne CM. Variation in lipid-lowering therapy use in patients with low-density lipoprotein cholesterol >/=190mg/dL: insights From the National Cardiovascular Data Registry-Practice Innovation and Clinical Excellence Registry. Circ Cardiovasc Qual Outcomes 2018;11:e004652. doi:10.1161/CIRCOUTCOMES.118.004652.
[18] Dixon DL, Sharma G, Sandesara PB, Yang E, Braun LT, Mensah GA, Sperling LS, Deedwania PC, Virani SS. Therapeutic inertia in cardiovascular disease prevention: time to move the bar. J Am Coll Cardiol 2019;74:1728–31. doi:10.1016/j.jacc.2019.08.014.
[19] Virani SS, Woodard LD, Chitwood SS, Landrum CR, Urech TH, Wang D, Murawsky J, Ballantyne CM, Petersen LA. Frequency and correlates of treatment intensification for elevated cholesterol levels in patients with cardiovascular disease. Am Heart J 2011;162:725–32 e721. doi:10.1016/j.ahj.2011.07.013.
[20] Skentzos S, Shubina M, Plutzky J, Turchin A. Structured vs. unstructured: factors affecting adverse drug reaction documentation in an EMR repository. AMIA Annu Symp Proc 2011;2011:1270–9.
[21] Virani SS, Akeroyd JM, Ahmed ST, Krittanawong C, Martin LA, Slagle J, Gobbel GT, Matheny ME, Ballantyne CM, Petersen LA. The use of structured data elements to identify ASCVD patients with statin-associated side effects: insights from the Department of Veterans Affairs. J Clin Lipidol 2019;13 797–803 e791. doi:10.1016/j.jacl.2019.08.002.
[22] Malmasi S, Hosomura N, Chang LS, Brown CJ, Skentzos S, Turchin A. Extracting healthcare quality information from unstructured data. AMIA Annu Symp Proc 2017;2017:1243–52.
[23] Malmasi S, Sandor NL, Hosomura N, Goldberg M, Skentzos S, Turchin A. Canary: an NLP platform for clinicians and researchers. Appl Clin Inform 2017;8:447–53. doi:10.4338/ACI-2017-01-IE-0018.
[24] Canary A user-friendly information extraction tool; 2021. Available at http://canary.bwh.harvard.edu Accessed February 16.
[25] VA Center for Medication Safety (VA MedSAFE) Adverse events tracking tools; 2021. Available at https://www.pbm.va.gov/vacenterformedicationsafety/vacenterformedicationsafetyadverseeventtrackingtools.asp Accessed May 5.
[26] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74 doi:10.2307/2529310. JSTOR 2529310. PMID 843571.
[27] Gobbel GT, Reeves RM, South BR, Chapman WW, Jarman J, Lay S, Matheny ME. RapTAT: a tool for assisted annotation and reviewer training via online machine learning. *AMIA Annu Symp Proc* San Francisco 2021:CA2015.
[28] Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. J Am Med Inform Assoc 2011;18:181–6. doi:10.1136/jamia.2010.007237.