Published in final edited form as:

Int J Med Inform. 2009 April ; 78(Suppl 1): S34–S42. doi:10.1016/j.ijmedinf.2008.09.001.

Identifying QT prolongation from ECG impressions using a general-purpose Natural Language Processor

Joshua C. Denny, MD, MS^{a,b}, Randolph A. Miller, MD^a, Lemuel Russell Waitman, PhD^a, Mark Arrieta, MS^a, and Joshua F. Peterson, MD, MPH^{a,b}

^a Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

^b Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

Abstract

Objective—Typically detected via electrocardiograms (ECGs), QT interval prolongation is a known risk factor for sudden cardiac death. Since medications can promote or exacerbate the condition, detection of QT interval prolongation is important for clinical decision support. We investigated the accuracy of natural language processing (NLP) for identifying QT prolongation from cardiologist-generated, free-text ECG impressions compared to corrected QT (QTc) thresholds reported by ECG machines.

Methods—After integrating negation detection to a locally-developed natural language processor, the KnowledgeMap concept identifier, we evaluated NLP-based detection of QT prolongation compared to the calculated QTc on a set of 44,318 ECGs obtained from hospitalized patients. We also created a string query using regular expressions to identify QT prolongation. We calculated sensitivity and specificity of the methods using manual physician review of the cardiologist-generated reports as the gold standard. To investigate causes of “false positive” calculated QTc, we manually reviewed randomly selected ECGs with a long calculated QTc but no mention of QT prolongation. Separately, we validated the performance of the negation detection algorithm on 5,000 manually-categorized ECG phrases for any medical concept (not limited to QT prolongation) prior to developing the NLP query for QT prolongation.

Results—The NLP query for QT prolongation correctly identified 2,364 of 2,373 ECGs with QT prolongation with a sensitivity of 0.996 and a positive predictive value of 1.000. There were no false positives. The regular expression query had a sensitivity of 0.999 and positive predictive value of 0.982. In contrast, the positive predictive value of common QTc thresholds derived from ECG machines was 0.07–0.25 with corresponding sensitivities of 0.994–0.046. The negation detection algorithm had a recall of 0.973 and precision of 0.982 for 10,490 concepts found within ECG impressions.

Conclusions—NLP and regular expression queries of cardiologists’ ECG interpretations can more effectively identify QT prolongation than the automated QTc intervals reported by ECG machines.

Corresponding Author: Joshua Denny, MD, MS, Eskind Biomedical Library, Room 442, 2209 Garland Ave, Nashville, TN 37232, USA, josh.denny@vanderbilt.edu.

Conflict of Interest

There are no financial conflicts of interest. Drs. Denny and Miller helped develop the KnowledgeMap Concept Identifier system and participated in parts of this study.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Future clinical decision support could employ NLP queries to detect QTc prolongation and other reported ECG abnormalities.

Keywords

electrocardiogram; QT prolongation; Unified Medical Language System; natural language processing; concept identification; negation detection; decision support

Introduction

Electrocardiograms (ECG) provide significant medical information that can facilitate new approaches for clinical decision support interventions. ECGs are typically stored in medical record systems as an image, structured calculations derived from the ECG, and semi-structured text. The image contains the waveform tracing while the text contains automated measures generated by the ECG machine and a free-text impression of the tracing by a physician, usually a cardiologist. The ECG impression may communicate clinical findings that indicate both cardiac and extra-cardiac disease, and may help manage the risks of medication use. For example, an ECG impression could be “Normal sinus rhythm, HR 65. Meets criteria for left ventricular hypertrophy and has a prolonged QT interval.” ECG impressions present two major types of information: a morphologic description of the ECG tracing (e.g., “prolonged QT interval”) and interpretations of those findings (e.g., “left ventricular hypertrophy,” myocardial infarction, or atrial fibrillation). While many researchers and ECG manufacturers have developed automated feature extraction programs based on ECG waveforms, these algorithms are imperfect, with accuracies of 42–96%[1–3]. Automated algorithms are generally superior for morphological descriptions than for interpretations[1,2]. Many factors, such as concurrent arrhythmia or ischemia, can alter the accuracy of morphological descriptions. For these reasons, cardiologists’ interpretations of ECGs remain the consensus gold standard[1,4,5].

The QT interval is an ECG measure that describes the time between ventricular depolarization (resulting in the “QRS complex”) and its repolarization (resulting in the “T wave”). The QT interval varies with heart rate, and thus the QT interval is typically reported adjusted for heart rate (QTc) via Bazett’s formula[6]. When the QTc crosses a threshold, often 450–500ms, the QT interval is described as prolonged. However, other features of the ECG tracing can introduce measurement error, including the presence of arrhythmias, intraventricular conduction disturbances, and additional waveforms such as the presence of a U wave. Despite these potential measurement errors with automated QTc calculation, QTc as reported by ECG machines are often used in clinical practice and research studies.[7–9] True prolongation of the QT interval can result from cardiac toxicity of many different medications[10] or represent underlying cardiac conduction disease. QT prolongation is a key risk factor for the development of Torsades de Pointes, a potentially fatal cardiac dysrhythmia. Drugs which prolong the QT interval may be marked with a black-box warning or removed from the market if contraindicated use becomes common[10,11]. A decision support system that alerts providers to potential adverse medication effects or contraindications may improve prescription selection and patient safety.

Natural language processing (NLP) and concept-based indexing can encode free text reports to standardized vocabularies such as the Unified Medical Language System (UMLS).[12] Researchers have used NLP systems to identify clinical syndromes and common biomedical concepts from radiology reports[13–15], clinical notes[16–18], problem lists[19], nursing documentation[20], and medical education documents[21]. Among many such systems, some studied NLP systems include the National Library of Medicine’s MetaMap program[22], the MedLEE (Medical Language Extraction and Encoding) system developed at Columbia[16], a system developed by Nadkarni and colleagues[17], the Mayo Vocabulary Server[23],

MEDSYNDIKATE[18], SAPHIRE[24], several systems by Chapman and colleagues[25,26], and the HITEx system[27]. Each of these takes a unique approach to accomplish the goal of mapping natural language text to structured output matched to standardized vocabularies. They differ primarily in their methodology and their degree of syntactic and semantic processing. The MetaMap system effectively identifies UMLS concepts from biomedical text using statistical phrase boundaries, a rigorous score-based approach, and word-variant recognition. It has been combined with negation detection algorithms (e.g., “no ischemia”), such as the NegEx algorithm[25], to enhance its utility. The MedLEE system, perhaps one of the most well-developed NLP systems, also identifies UMLS concepts and detects negation signals. It parses sentences to group semantic modifiers (e.g., “family history of heart failure” indicates the experiencer was not the patient) and identify temporal references (e.g., “colonoscopy in 2005”). The Mayo Vocabulary Server also employs a syntactic parse to detect negation signals identify concepts from standardized vocabularies, such as SNOMED-CT.[28]

Similar to MetaMap and other concept identification systems, the KnowledgeMap concept identifier (KMCI) is a general purpose concept identifier, using score-based algorithms to identify UMLS concepts from natural language text [21]. The KMCI system was designed to accommodate poorly-formatted document types with ad hoc abbreviations and acronyms, using a combination of linguistic, frequency, and concept co-occurrence data to accurately identify unknown abbreviations, acronyms, and underspecified concepts (e.g., the document phrase “1st degree block” maps to the closest UMLS string match “1st degree atrioventricular block”).

Previously, we reported the use of KMCI to identify UMLS concepts from cardiologist-generated ECG impressions[29]. The system identified concepts with an overall recall of 0.90 and precision of 0.94[29]. Underspecified concepts were especially frequent in this dataset, which presents a challenge for accurate concept identification; 27% of the closest UMLS concept matches required inference of a missing word in the target UMLS term. Of interest for the present study, “QT” was not listed as a synonym for “QT interval” in the UMLS and required disambiguation from other, very different concepts such as “long QT” or “short QT” which all have one additional word. The physician reviewers also classified each concept into a predefined category. The KMCI system performed the best for myocardial perfusion changes, ECG rhythms, and extracardiac manifestations (each with a recall and precision in excess of 0.98). Its poorest performance was for nonmedical concepts. This earlier version of KMCI, however, did not have the ability to detect negated or possible findings.

In this paper, we report on the application of a “negation-detection-enhanced” version of KMCI to identify QT prolongation from a four-year collection of ECGs. The ultimate goal is to process ECG impressions in real time to support development of clinical decision support systems that use such inputs to provide advice at the point of care.

Methods

Creation of ECG database

Vanderbilt University Medical Center has developed an anonymized database of all orders, laboratory results, and ECGs for all inpatients admitted for 2–30 days from 1999–2003 as part of an ongoing research study investigating drug effects. Nearly all inpatient medication orders are written by providers using an electronic care provider order entry system.[30,31] The ECGs were imported in an XML format from an ECG management system and anonymized. Every ECG report includes machine-calculated intervals and estimated heart rate as well as a cardiologist-generated free-text impression. Cardiologists create an impression for all ECGs by selecting among personalizable stock phrases (e.g., “normal sinus rhythm”) and editing stock phrases as necessary (e.g., “normal sinus rhythm *with rare PVCs*”), or typing

unconstrained comments *de novo* (e.g., “LA abnormality, PVCs, and inferolateral ST-T changes”). Finally, cardiologists code each ECG with a standard severity: normal, otherwise normal, borderline normal, or abnormal. We extracted all ECG reports from our research repository and loaded them into a relational database. There were 44,318 ECGs in the database with more than 155,000 sentences from 23,080 hospital admissions for 16,821 patients. The Institutional Review Board approved this study.

From the ECG dataset, we randomly selected a test set of about 5,000 sentences for development of the KMCI negation detection and QT prolongation recognition query. Another 5,000 randomly selected sentences were reserved as a validation set for negation detection. All subsequent data analysis (including QT prolongation detection) was performed with the entire dataset of 44,318 ECG reports.

Negation tagging algorithm and evaluation

Negation detection algorithms identify phrases which qualify the presence of concepts before or after the negating phrase. For instance, a cardiologist may state “no PR prolongation,” indicating the absence of PR prolongation, or “myocardial infarction *cannot be ruled out*,” which signifies that “myocardial infarction” is a possible finding in this patient. Many useful negation detection algorithms exist [25,28,32,33]. We applied a modified version of the NegEx algorithm [25] that uses regular expressions to mark concepts as *negated*, *possible*, or *asserted*. The KMCI detection scheme used a total of 205 phrases indicating negation or possibility, including symbols such as “?” and “r/o” as indicators of “possible.” We empirically chose a window of 8 words before or after a negating phrase by reviewing a sample of negation-tagged ECG impressions. In this dataset, we found that most periods marked abbreviations rather than sentence breaks, so our algorithm ignored periods within identified sentences. Semicolons, unmatched closing parentheses, and other negating phrases terminated the current negation window.

One author (JP, an internist), unfamiliar with the NegEx algorithm or this implementation, scored the testing set of 5000 sentences from randomly selected ECGs via a color-coded HTML interface that highlighted the negating phrase and words modified by them. All medical concepts or medical modifiers were considered for scoring; the evaluation was not limited to QT prolongation concepts. Concepts were marked as a correctly identified negated concept (i.e., a true positive, TP), false positives (FP), true negatives (TN), and false negatives (FN). We calculated recall of negated concepts as $TP/(TP+FN)$; precision as $TP/(TP+FP)$; and negative predictive value, the probability that the concept is not negated (i.e., an asserted finding), as $TN/(TN+FN)$. Following the evaluation, three new negation phrases (“replaced <negated concept>”, “<negated concept> replaced by”, and “<negated concept> is/are gone”) were added and validated over several thousand new sentences. After verifying that these additions introduced no false positives, we added them to the application before processing the entire dataset.

Development of concept-based ECG database

We applied KMCI to identify Unified Medical Language System (UMLS) concepts from the free text ECG impressions, using the optimizations identified via our prior study [29]. We added a few synonyms and derivational transformations to KMCI’s lexicon and modified the sentence-identification algorithm to ignore most spaces and periods when determining sentence breaks. We used the 2006AC version of the UMLS [12]; the only restriction on concept matching was that KMCI favored underspecified concepts if they contained words such as “heart” and “electrocardiogram” (see [29] for a full list). Candidate UMLS concepts with these words received higher scores than candidates with other words when the words do not match a document word. In addition, KMCI favors, in a document, ambiguous concepts that occur

frequently as exact-match concepts (much like a prior probability) or that co-occur with exact-match concepts. To utilize this feature, we processed ECGs in bulk so that exact-match concepts over the corpus of ECGs served to help determine ambiguous matches. We applied the negation algorithm following the concept identification step to mark each concept as positive, possible, or negated. The concept-identified ECGs were linked to the original ECG impressions and the calculated intervals from the original ECG reports, forming an identified ECG dataset.

Identification of Electrocardiograms with QT prolongation

Through manual perusal of the ECG dataset, we identified potentially-matching UMLS concepts representing “QT prolongation,” including any text indicating a probable or possible QT or QTc prolongation. To verify that we had found all UMLS concepts representing QT prolongation, we also did text searches for all matching UMLS concepts from the ECG impressions containing the strings “QT” or “QTc.” The NLP query for QT prolongation consisted of these UMLS concepts (see Table 2) that were either asserted or possible in the ECG impressions. We extracted the ECGs reports matching the NLP query along with their automated QT and QTc intervals identified by the ECG management system. We also extracted all ECG reports exceeding common QTc thresholds used by cardiologists to indicate possible QT prolongation: 400ms, 450ms, 500ms, or 550ms. We chose the cardiologist-generated impression as our gold standard. After developing the concept query, two physician authors (JD, JP) simultaneously reviewed all ECG impressions containing the letters “QT” (including phrases such as “short QTc”, “QT unmeasurable”, and “QT interval unchanged” as well as prolonged QT statements) to identify the gold standard set of ECG reports. Since there are no synonyms for the QT interval that do not contain the letters “QT,” this sampling method provided an accurate assessment of possible false negatives. We compared the NLP and string queries to the calculated QT and QTc intervals (which are continuous numbers) using the aforementioned thresholds. For the QT and QTc intervals, we also used the continuous values as measured by the ECG machine to calculate the area under the receiver operator characteristic curves (AUC), using the cardiologist’s impression as the gold standard.

Following the evaluation of the NLP query, we developed a regular expression string query to represent the QT prolongation concepts encountered in the dataset (Table 2). Since this case-insensitive search could match any part of the string, it matches any string containing “long” (e.g., “prolonged,” “prolongation,” or “longer”) or “length” (e.g., “lengthened”) and also the letters “qt” (including “QTc” and “QTU,” which, we determined in the evaluation, was an important word not included in the UMLS). All strings matching this query were evaluated by manual review for accuracy and compared with the NLP query and ECG machine QTc thresholds.

To further assess the negative predictive value of the textual queries for QT prolongation and to evaluate for causes of possible miscalculation of the QT interval by the ECG machine, we reviewed a random sample of 100 ECGs with calculated QTc intervals > 450 ms but no QT prolongation concept identified by the NLP parser. We reviewed these for references to QT prolongation or for potential causes of a miscalculated QTc by the ECG machine. We could not examine the ECG images because they were not stored in our anonymized research database.

Calculations for this study considered a true positive as any ECG correctly identified as representing QT prolongation according to the human review of ECG impressions (our gold standard). Sensitivity (recall), positive predictive value (precision), and negative predictive value were calculated by comparing the query outputs to the gold standard. Specificity was calculated as $TN/(TN + FN)$. The F-measure was calculated as $2 * Recall * Precision / (Recall + Precision)$. [34] Student t-tests were used to compare parametric data. AUC and statistical analyses were calculated using Stata, version 9.2 (StataCorp LP, College Station, TX).

Description of identified ECG database

To investigate other potential applications of the ECG NLP parser, we developed UMLS concept queries for other common cardiac diagnoses found in ECG impressions that may be of interest for decision support (see Table 4). Query logic was developed by finding the UMLS concepts representing the topic of interest in the database of matched concepts. For example, the concept query for myocardial infarction involved the tree of concepts related to “myocardial infarction” and “infarct” (since all references to infarcts in this dataset could be assumed to be myocardial).

Results

General characterization of ECG reports

The KMCI algorithm identified 375,838 concepts from the 44,318 ECGs in the database mapping to 23,080 unique admissions. Cardiologists identified 67% percent of the ECGs as “abnormal,” 11% as “borderline,” 18% as “normal” or “otherwise normal,” and 4% as unmarked. Of KMCI-identified concepts in the entire ECG dataset, 339,554 (90.3%) were marked by the NegEx algorithm as asserted; 29,107 (7.7%) concepts were qualified as “possible”; and 7177 (1.9%) concepts were negated. The physician review identified 2,373 ECG impressions containing QT prolongation concepts.

Validation of negation detection algorithm

Table 1 shows the results of the negation analysis. The 5,000 sentences in the negation test set contained a total of 10,480 UMLS concepts. Overall recall was 0.973 and overall precision 0.982. The negative predictive value of finding negation (probability that a statement was positive given it was identified as positive) was 0.998. All false negatives were due to three phrases not present in the regular expression list: “replaced <negated concept>”, “<negated concept> replaced by”, and “<negated concept> is/are gone.” These phrases were added to the list of negating phrases before running the NLP query for QT prolongation. Several of the false positives were instances in which negating phrases were amid multiple concept words (e.g., “ST *no longer* depressed,” in which the negated concept “ST depression” is separated by the negation phrase “no longer”). KMCI typically identified the correct UMLS concept for these phrases. Misspellings also caused some errors.

Natural language processing and regular expression queries for QT prolongation

Table 2 shows the concepts and words used for the QT prolongation queries, along with the frequency of each in the database. There were 254 unique ECG impression sentences with a range of 2–18 words (median 11 words, weighted median 5 words) matching a QT prolongation query; 15 of these strings (e.g., “QT interval long for rate”) accounted for nearly 90% of all matching impressions. Three of these phrases (accounting for 75% of all impressions) appear to be stock phrases from the ECG management software. The overall precision for the concept query was 2450/2451 (1.000) for individual concept matches and 2364/2364 (1.000) for ECGs compared to the gold standard; the one concept error occurred in an ECG that also contained a correctly-matched concept asserting QT prolongation. The precision for regular expression matches was 2495/2539 (98.3%) for individual matches and 2370/2413 (98.2%) for classifying ECGs. Table 3 shows the results of different methods of predicting prolonged QT intervals; 2,364 ECGs (5.3% of all ECGs) were identified as representing QT prolongation by our concept query. The average QTc interval for those with prolonged QT intervals was 487 ms (range 363–716 ms); ECGs without mention of QT prolongation averaged 429 ms ($p < 0.001$, range 46–785 ms). Overall, the calculated QT interval had an AUC of 0.73 for predicting QT prolongation; the QTc interval AUC was 0.91. Ninety-two percent of the statements regarding ECGs containing QT prolongation concepts were asserting presence of the concept; 8% were

qualified as “possible,” with the words “borderline”, “possible”, “question”, or “cannot exclude.”

Failure analysis

There were nine ECGs (with seven unique strings) marked as representing QT prolongation in the gold standard set but were not detected by the NLP query. Three of these strings contained the letters “QTU” instead of QT or QTc, which is not present in the UMLS or KMCI’s synonym tables. Two strings correctly matched QT prolongation concepts we had failed to include in the NLP query (C0023976: “Long QT syndrome” and C0855333: “Electrocardiogram QT corrected interval prolonged”) due to an oversight when creating the QT query. The remaining two string errors were due to NLP parsing errors (e.g., failing to distribute the word “longer” for “longer PR and QT intervals now”). The single concept matching error (false positive) was with the sentence: “Since prior tracing, rate increased 44 bpm, and QT is partially better” in which “QT” was inappropriately overmatched to “QT prolongation” instead of “QT interval” based on its prevalence in the ECG set and the presence of the word “increased” within the sentence. However, this ECG impression also contained a correctly matched QT prolongation concept, meaning that the ECG was correctly classified as containing QT prolongation.

The NegEx algorithm identified 4 ECG impressions as containing negated QT prolongation concepts, and 190 ECGs with QT prolongation concepts marked as “possible.” In three of the four negated concepts, the algorithm’s negation assignment was incorrect (i.e., the concept was actually asserted, making these false positive negations); only one QT prolongation concept was a true positive negation. Each false positive negation resulted when KMCI traversed a comma or conjunction from an adjacent phrase negated with a form of the verb “replace” (e.g., “sinus rhythm has replaced rapid Afib and QT has lengthened”). Despite these negation assignment errors with individual concepts, all the ECG reports with QT prolongation concepts were still correctly classified using the NLP query because their impressions also contained true positive asserted QT prolongation concepts. For example, the full impression of the above example was: “QT prolongation. Sinus rhythm has replaced rapid Afib and QT has lengthened.” The first QT prolongation concept was correctly marked as asserted even though the second (“QT has lengthened”) was not. Without using NegEx, KMCI would detect four additional individual concept hits, one of which would be incorrect, although at the ECG level would still classify all identically as with NegEx. NegEx more accurately detected “possible” signals with a precision of 98% over 192 QT prolongation concepts marked as “possible” by NegEx.

Of the 100 manually reviewed ECGs with QTc intervals longer than 450 ms but that did not contain QT prolongation by the gold standard determination, 32 had a bundle branch block; 24 had various ST segment or T wave abnormalities; 24 had an arrhythmia, aberrant complexes, or a pacemaker; and 23 had myocardial ischemia or infarct. No ECG impression contained textual comments directly suggesting QT prolongation. Only 4 ECGs had no significant electrocardiographic abnormalities that could not alter calculation of the QT interval.

Table 4 shows concept findings over the entire database for several other queries that are potential targets for decision support or clinical research questions. Asserted concepts are the predominant category across each of these concepts and negated concepts are rare (between 1–7% of each diagnosis).

Discussion

We studied the application of a concept-based, natural language processing system to identify QT prolongation within cardiologist-generated ECG reports. The performance of the NLP system had near perfect precision and recall when compared to a human reviewer. Furthermore,

seven of the nine false negatives were easily fixed by revising the query to include overlooked synonyms and concepts. The regular expression query also performed similarly in identifying QT prolongation, and provides a simple, scalable method to identify QT prolongation from ECGs. However, the dedicated string query does not detect qualifiers such as negation, produced more false positives, and lacks the generalizability to other conditions offered by the NLP method. Both textual queries performed far superior to commonly used ECG machine-reported QTc thresholds for diagnosing QT prolongation, which had positive predictive values ranging from 6–25%. A manual review of ECGs with prolonged QTc values but no “QT Prolongation” concept in the impression suggested that the QTc threshold approach often fails because of other waveform abnormalities such as bundle branch blocks, pacemakers, or arrhythmias. Cardiologists are trained to take coexisting waveform abnormalities into account when reviewing the ECG tracing, which likely accounts for the discrepancy between the NLP method and the QTc method. An automated system employing NLP analysis of ECG impressions was much more accurate at identifying QT prolongation than machine-reported intervals.

The high performance of the NLP system required highly accurate concept identification. Negation within ECG impressions overall was rare, including only one instance of a QT prolongation concept. When applied to all UMLS concepts, NegEx marked only 1.9% of all concepts as negated and 7.7% as possible. Negation detection was not important for correctly classifying ECG impressions with QT prolongation; however, Table 4 demonstrates that other concepts, such as myocardial infarction and ST segment elevation, are often negated or qualified as “possible.” A general purpose NLP tool to parse ECG impressions for some decision report applications would require a highly accurate negation tagger. The NegEx negation algorithm performed well in detecting negation within this dataset with a recall of 97.3% and a precision of 98.2%. The probability of a concept the system identified as positive truly being positive was 0.998. The high recall and precision of negation detection in this dataset is likely due to a constrained vocabulary and the relative simplicity of the ECG impression sentences compared to prior studies evaluating NegEx in other clinical document types, which had recalls of 78–97% and precisions of 85–91%[25,28,32].

Analysis of negation detection for QT prolongation revealed the algorithm incorrectly identified three concepts as negated. Each of these negation failures involved an inappropriate negation assignment to a separate independent clause, indicating the utility of a more complex negation detection algorithm such as presented by Huang and Lowe[35], Elkin et al.[28,33], or Mutalik et al.[32] The NegEx algorithm used in this experiment used a rather simple “negation window” technique that assigns negation status to any word occurring within a certain distance before or after a negation phrase. This simplification caused some errors in our dataset. A more advanced algorithm would use a syntactic parse of the sentence, recognizing the presence of prepositions or coordinating conjunctions to correctly size the negation window.

The handling of “possible” findings, included in our QT prolongation query, would vary depending on application. For the purposes of clinical decision support, inclusion of potential findings may help prevent adverse events. For example, one would prefer to avoid starting a drug known to prolong the QT interval in a patient that had a “borderline long QT,” assuming an alternative was available. In addition, many uncertain ECG findings require further workup. A patient with potential ischemia requires further evaluation, and one would likely discontinue cyclooxygenase-2 inhibitors in this patient. For clinical research, however, one may desire to exclude uncertain diagnoses, as many ECGs indicating “cannot rule out” may represent benign or nonspecific findings. Finally, negating phrases such as “no longer” indicate both the current absence of a finding as well as a prior history of it; the current algorithm only identifies the former. Such information may help determine treatment efficacy.

The developed NLP method to process ECGs could be a valuable resource for clinical decision support and pharmacoepidemiology. Medications that prolong QT interval have been defined in registries[10] and could be linked to “QT prolongation” concepts found in ECG impressions. One clinical decision support component could intercept orders for QT-prolonging medications prescribed when the patient is known to have pre-existing QT prolongation. A second application could investigate the association between medication orders and subsequent QT prolongation in order to define new drug-drug interactions or single-drug causes of QT prolongation. Interactions which may be difficult to discover in clinical trials or *in vitro* studies, may be discernable via such surveillance, such as the addition of a potent cytochrome P450 inhibitor that raises serum concentrations of a known offending agent. Ideally, a medication intervention could not only intercept medications that prolong the QT interval but also those that significantly interact with those already prescribed. Due to the high provider override rates in most medication decision support systems, due in part to poor specificity[36], a medication decision support system for QT prolongation requires use of the cardiologist-generated impression rather than current calculated intervals. The ability to use cardiologist-generated impressions for decision support requires rapid availability of their interpretations in electronic format, which may not be feasible at all institutions.

Since this method provides a full concept index using all concepts available in the UMLS, it also supports queries for other key clinical concepts (such as the queries in Table 4). While a complicated string-matching algorithm (in this case designed after the primary analysis) performed similar to the QT prolongation concept query, it would lack flexibility and scalability. By applying highly accurate NLP tools, we can quickly assess multiple queries, enabling a broader range of research and decision support tools. For example, recent studies have implicated commonly used medications in the incidence of sudden death[11,37], myocardial infarction[38,39], and second or third degree atrioventricular block[40], each of which could be targets for clinical decision support interventions. In addition, the parent-child relationships between concepts in the UMLS are useful in grouping more granular concepts into larger groups, such as mapping strings such as “acute anterior infarction” and “ST elevation MI” as types of “myocardial infarction.”

In this study, we used a general purpose concept-identification program with the entire UMLS. We optimized the algorithm to enhance synonymy and favor underspecified matches that match cardiology-related concepts. To further improve performance on ambiguous concept matches, we processed the ECGs in bulk, allowing KMCI to use the frequency of exactly-matched concepts from the set of ECGs to favor common concepts and frequently co-occurring concepts when encountering ambiguous matches. Given the high frequency of underspecified concepts in ECG impressions, other general purpose concept identification algorithms may require similar optimizations.

The study had several limitations. The performance of the negation algorithm and concept identifier may not translate to other repositories of ECG impressions. Other institutions may use additional ad hoc abbreviations, acronyms, or idiosyncratic language that could hinder KMCI’s performance, and our cardiologists may differ from other institutions in how they interpret ECGs for QT prolongation. However, we made no specific optimizations for the specific format of our ECG reports. Second, we considered the cardiologists’ impression as our gold standard since we did not have access to the original ECG images to validate with an independent review for QT prolongation; some cases of QT prolongation may have been missed by the reviewing cardiologists. However, we expect that this would be unlikely to dramatically alter our results since QT prolongation is a well-characterized and potentially lethal ECG finding. Furthermore, the prevalence of QT prolongation identified via the NLP query (5.3%) has greater face validity than the prevalence identified by QTc interval > 450 ms (26.6%). Third, our negation evaluation was performed by review of the algorithm-generated

negation assignments of 5,000 randomly chosen ECG impression sentences, which introduces a potential bias toward the negation status generated by the algorithm. Fourthly, while we have accurately identified concept matches and their negation status, this is not the same as asserting normality. Our algorithm tells the presence or absence of “atrial fibrillation,” for instance, but cannot tell that there were no arrhythmias. These questions may be addressed by classifying concepts by type (e.g., “rhythm” or “perfusion abnormalities”) and defining normal status (e.g., the absence of arrhythmias is the normal state). Finally, our exploratory list of concepts in Table 4 has not been formally assessed for accuracy and provides only a rough prevalence of these findings in set of ECGs.

Conclusion

The use of textual queries through custom searches or NLP techniques allows highly accurate identification of QT prolongation within free-text ECG impressions. We believe this technique could enable large-scale research on drug adverse events and development of new decision support tools to improve cardiovascular medication safety.

Summary Table

What was already known on this topic?

- Electrocardiograms (ECG) provide significant medical information and are available in many electronic medical records. The ECG reports consist of automated intervals provided by an ECG machine and a cardiologist-generated free-text impression describing the findings.
- Often measured via an electrocardiogram, the QT interval is the time between ventricular depolarization to repolarization. The QT interval is affected by the heart rate and thus is often adjusted for rate as the QTc.
- Typically defined as a QTc longer than 450–500ms, QT prolongation is a known risk factor for sudden cardiac death. Many medications are known to promote or exacerbate QT prolongation.

This study adds the following knowledge:

- Natural language processing with negation detection can extract concepts from ECG impressions with high accuracy.
- Natural language processing and regular expression string queries of cardiologist-generated ECG impressions are superior to ECG-machine calculated QTc thresholds for detecting QT prolongation, representing a methodology for clinical decision support applications.

Acknowledgments

This work is supported by two National Library of Medicine grants, T15 LM007450 and R01 LM007995.

Role of the funding source

The funding source had no involvement in the preparation, study design, analysis, writing, or decision to publish this data.

References

1. Tsai TL, Fridsma DB, Gatti G. Computer decision support as a source of interpretation error: the case of electrocardiograms. *J Am Med Inform Assoc* 2003 Sep-Oct;10(5):478–83. [PubMed: 12807810]

2. Willems JL, Abreu-Lima C, Arnaud P, van Bommel JH, Brohet C, Degani R, et al. The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N Engl J Med* 1991 Dec 19;325(25):1767–73. [PubMed: 1834940]
3. Maglaveras N, Stamkopoulos T, Diamantaras K, Pappas C, Strintzis M. ECG pattern recognition and classification using non-linear transformations and neural networks: a review. *International journal of medical informatics* 1998 Oct-Dec;52(1–3):191–208. [PubMed: 9848416]
4. Olsson SE, Ohlsson M, Ohlin H, Dzaferagic S, Nilsson ML, Sandkull P, et al. Decision support for the initial triage of patients with acute coronary syndromes. *Clin Physiol Funct Imaging* 2006 May; 26(3):151–6. [PubMed: 16640509]
5. Paoletti M, Marchesi C. Discovering dangerous patterns in long-term ambulatory ECG recordings using a fast QRS detection algorithm and explorative data analysis. *Comput Methods Programs Biomed* 2006 Apr;82(1):20–30. [PubMed: 16529841]
6. Bazett H. An analysis of the time relationship of electrocardiograms. *Heart* 1920;7:353–70.
7. Aarnoudse AJ, Newton-Cheh C, de Bakker PI, Straus SM, Kors JA, Hofman A, et al. Common NOS1AP variants are associated with a prolonged QTc interval in the Rotterdam Study. *Circulation* 2007 Jul 3;116(1):10–6. [PubMed: 17576865]
8. Malik M. Errors and misconceptions in ECG measurement used for the detection of drug induced QT interval prolongation. *Journal of electrocardiology* 2004;37(Suppl):25–33. [PubMed: 15534789]
9. Wedam EF, Bigelow GE, Johnson RE, Nuzzo PA, Haigney MC. QT-interval effects of methadone, levomethadyl, and buprenorphine in a randomized trial. *Archives of internal medicine* 2007 Dec 10;167(22):2469–75. [PubMed: 18071169]
10. Drugs that prolong the QT interval and/or induce Torsades de Pointes ventricular arrhythmia. [cited 2006 12/2]; Available from: <http://www.arizonacert.org/medical-pros/drug-lists/drug-lists.htm>
11. Roden DM. Drug-induced prolongation of the QT interval. *N Engl J Med* 2004 Mar 4;350(10):1013–22. [PubMed: 14999113]
12. Unified Medical Language System. [cited 2006 12/2]; 2006AC:[Available from: <http://www.nlm.nih.gov/pubs/factsheets/umls.html>]
13. Fiszman M, Chapman WW, Aronsky D, Evans RS, Haug PJ. Automatic detection of acute bacterial pneumonia from chest X-ray reports. *J Am Med Inform Assoc* 2000 Nov-Dec;7(6):593–604. [PubMed: 11062233]
14. Huang Y, Lowe HJ, Hersh WR. A pilot study of contextual UMLS indexing to improve the precision of concept-based representation in XML-structured clinical radiology reports. *J Am Med Inform Assoc* 2003 Nov-Dec;10(6):580–7. [PubMed: 12925544]
15. Bertaud V, Lasbleiz J, Mougin F, Burgun A, Duvauferrier R. A unified representation of findings in clinical radiology using the UMLS and DICOM. *International journal of medical informatics*. 2008 Jan 4;
16. Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc* 2004 Sep-Oct;11(5):392–402. [PubMed: 15187068]
17. Nadkarni P, Chen R, Brandt C. UMLS concept indexing for production databases: a feasibility study. *J Am Med Inform Assoc* 2001 Jan-Feb;8(1):80–91. [PubMed: 11141514]
18. Hahn U, Romacker M, Schulz S. MEDSYNDIKATE--a natural language system for the extraction of medical information from findings reports. *International journal of medical informatics* 2002 Dec 4;67(1–3):63–74. [PubMed: 12460632]
19. Meystre SM, Haug PJ. Randomized controlled trial of an automated problem list with improved sensitivity. *International journal of medical informatics*. 2008 Feb 14;
20. Bakken S, Hyun S, Friedman C, Johnson SB. ISO reference terminology models for nursing: applicability for natural language processing of nursing narratives. *International journal of medical informatics* 2005 Aug;74(7–8):615–22. [PubMed: 16043086]
21. Denny JC, Smithers JD, Miller RA, Spickard A 3rd. “Understanding” medical school curriculum content using KnowledgeMap. *J Am Med Inform Assoc* 2003 Jul-Aug;10(4):351–62. [PubMed: 12668688]
22. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp* 2001:17–21. [PubMed: 11825149]

23. Elkin PL, Ruggieri AP, Brown SH, Buntrock J, Bauer BA, Wahner-Roedler D, et al. A randomized controlled trial of the accuracy of clinical record retrieval using SNOMED-RT as compared with ICD9-CM. *Proc AMIA Symp* 2001:159–63. [PubMed: 11825173]
24. Hersh WR, Donohoe LC. SAPHIRE International: a tool for cross-language information retrieval. *Proc AMIA Symp* 1998:673–7. [PubMed: 9929304]
25. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of biomedical informatics* 2001 Oct;34(5):301–10. [PubMed: 12123149]
26. Chapman WW, Cooper GF, Hanbury P, Chapman BE, Harrison LH, Wagner MM. Creating a text classifier to detect radiology reports describing mediastinal findings associated with inhalational anthrax and other disorders. *J Am Med Inform Assoc* 2003 Sep-Oct;10(5):494–503. [PubMed: 12807805]
27. Zeng QT, Goryachev S, Weiss S, Sordo M, Murphy SN, Lazarus R. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. *BMC medical informatics and decision making* [electronic resource] 2006;6:30.
28. Elkin PL, Brown SH, Bauer BA, Husser CS, Carruth W, Bergstrom LR, et al. A controlled trial of automated classification of negation from clinical notes. *BMC medical informatics and decision making* [electronic resource] 2005;5(1):13.
29. Denny JC, Spickard A, Miller RA, Schildcrout J, Darbar D, Rosenbloom ST, et al. Identifying UMLS concepts from ECG Impressions using KnowledgeMap. *AMIA Annu Symp Proc* 2005:196–200. [PubMed: 16779029]
30. Neilson EG, Johnson KB, Rosenbloom ST, Dupont WD, Talbert D, Giuse DA, et al. The impact of peer management on test-ordering behavior. *Annals of internal medicine* 2004 Aug 3;141(3):196–204. [PubMed: 15289216]
31. Geissbuhler A, Miller RA. A new approach to the implementation of direct care-provider order entry. *Proc AMIA Annu Fall Symp* 1996:689–93. [PubMed: 8947753]
32. Mutalik PG, Deshpande A, Nadkarni PM. Use of general-purpose negation detection to augment concept indexing of medical documents: a quantitative study using the UMLS. *J Am Med Inform Assoc* 2001 Nov-Dec;8(6):598–609. [PubMed: 11687566]
33. Ceusters W, Elkin P, Smith B. Negative findings in electronic health records and biomedical ontologies: a realist approach. *International journal of medical informatics* 2007 Dec;76(Suppl 3):S326–33. [PubMed: 17369081]
34. Hripcsak G, Rothschild AS. Agreement, the f-measure, and reliability in information retrieval. *J Am Med Inform Assoc* 2005 May-Jun;12(3):296–8. [PubMed: 15684123]
35. Huang Y, Lowe HJ. A novel hybrid approach to automated negation detection in clinical radiology reports. *J Am Med Inform Assoc* 2007 May-Jun;14(3):304–11. [PubMed: 17329723]
36. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. *J Am Med Inform Assoc* 2006 Mar-Apr;13(2):138–47. [PubMed: 16357358]
37. Ray WA, Murray KT, Meredith S, Narasimhulu SS, Hall K, Stein CM. Oral erythromycin and the risk of sudden death from cardiac causes. *N Engl J Med* 2004 Sep 9;351(11):1089–96. [PubMed: 15356306]
38. Bresalier RS, Sandler RS, Quan H, Bolognese JA, Oxenius B, Horgan K, et al. Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. *N Engl J Med* 2005 Mar 17;352(11):1092–102. [PubMed: 15713943]
39. Solomon SD, McMurray JJ, Pfeffer MA, Wittes J, Fowler R, Finn P, et al. Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. *N Engl J Med* 2005 Mar 17;352(11):1071–80. [PubMed: 15713944]
40. Zeltser D, Justo D, Halkin A, Rosso R, Ish-Shalom M, Hohenberg M, et al. Drug-induced atrioventricular block: prognosis after discontinuation of the culprit drug. *J Am Coll Cardiol* 2004 Jul 7;44(1):105–8. [PubMed: 15234417]

Table 1

Comparison of KnowledgeMap Concept Identifier (KMCI) Negation Detection to Gold Standard Physician Review

NegEx classification	<i>Gold Standard Classification</i>			Total Concepts
	Negated	Possible	Asserted	
Identified as negated	129 (1.2%)	0 (0.0%)	3 (0.0%)	132 (1.3%)
Identified as possible	0 (0.0%)	593 (5.6%)	10 (0.1%)	603 (5.8%)
Identified as asserted	18 (0.2%)	2 (0.0%)	9,725 (92.8%)	9,745 (93.0%)
Total Concepts	147 (1.4%)	595 (5.7%)	9,738 (92.9%)	10,480

Table 2

employed to identify QT prolongation within ECG impressions in the NLP query

	ECG matches	True Positives
	2,294	2,293†
	55	55
	2	2

True Positives

ECG matches

100

100

C0205166 Long
C0439590 Prolonged
C0392744 Lengthened

with any of

	ECG matches	True Positives
	2,364 [‡]	2,364 [‡]
	411	407
	2,128	2,088
	2,413 [‡]	2,370 [‡]

Int J Med Inform. Author manuscript; available in PMC 2010 April 1.

list must occur within the same sentence to be considered a match.

the one concept query or for query 1 also contained a correct concept match for QT prolongation such that the ECG was correctly classified.

ore than one query (e.g., QT prolongation. Compared with prior ECG, the QT has lengthened.”)

Table 4

Number of ECGs expressing potential targets for decision support

Concept	Total (%)	Asserted	Possible	Negated
QT interval prolongation	2,373 (5.4)	2,182	190	1
Myocardial infarction (MI)	6,355 (14)	3,381	2,919	55
Acute MI	208 (0.5)	167	36	5
Myocardial ischemia	2,312 (5.2)	1,855	444	13
ST segment elevation	1,015 (2.3)	895	92	28
Wolff-Parkinson-White bypass tract	107 (0.2)	93	7	7
Atrioventricular block, 1 st degree	2,461 (5.5)	1,876	571	14
Atrioventricular block, 2 nd degree	61 (0.1)	58	1	2
Atrioventricular block, 3 rd degree	24 (0.05)	23	1	0
Pericarditis	105 (0.2)	43	62	0
Atrial fibrillation	1,748 (4.0)	1,719	16	13
Atrial flutter	397 (0.9)	373	20	4
Total number of ECGs	44,318			