

Automated evaluation of electronic discharge notes to assess quality of care for cardiovascular diseases using Medical Language Extraction and Encoding System (MedLEE)

Jung-Hsien Chiang,¹ Jou-Wei Lin,^{2,3} Chen-Wei Yang¹

► Supplementary appendices are published online only. To view these files please visit the journal online <http://jamia.bmj.com/content/vol17/issue3>

¹Institute of Medical Informatics and Department of Computer Science, National Cheng Kung University, Tainan, Taiwan

²Cardiovascular Center, National Taiwan University Hospital Yun-Lin Branch, Dou-Liou City, Yun-Lin, Taiwan

³Department of Medicine, College of Medicine, National Taiwan University, Taiwan

Correspondence to

Dr Jung-Hsien Chiang,
Department of Computer
Science, National Cheng Kung
University, Tainan, Taiwan;
jchiang@mail.ncku.edu.tw

J-HC and J-WL contributed
equally.

Received 3 September 2008
Accepted 26 February 2010

ABSTRACT

The objective of this study was to develop and validate an automated acquisition system to assess quality of care (QC) measures for cardiovascular diseases. This system combining searching and retrieval algorithms was designed to extract QC measures from electronic discharge notes and to estimate the attainment rates to the current standards of care. It was developed on the patients with ST-segment elevation myocardial infarction and tested on the patients with unstable angina/non-ST-segment elevation myocardial infarction, both diseases sharing almost the same QC measures. The system was able to reach a reasonable agreement (κ value) with medical experts from 0.65 (early reperfusion rate) to 0.97 (β -blockers and lipid-lowering agents before discharge) for different QC measures in the test set, and then applied to evaluate QC in the patients who underwent coronary artery bypass grafting surgery. The result has validated a new tool to reliably extract QC measures for cardiovascular diseases.

INTRODUCTION

Quantification of quality of care (QC) measures in the process and outcome domains has become one of the most important issues in current medical care.¹ Automatic acquisition of QC measures from medical records and comprehensive databases requires a tremendous amount of information workflow. With advanced information technology, health providers now have more tools to assess QC measurements and even to improve patient outcomes.²

In our previous study, we found that data-mining from electronic discharge notes could provide sufficient information for QC measures for ST-segment elevation myocardial infarction (STEMI).³ This finding has generated an interest in building a “flexible assessment engine” that is able to automatically retrieve essential information from medical narrations. Therefore, the objective of this study was to apply information extraction algorithms to develop an automated acquisition system for QC measures of cardiovascular diseases and to validate its accuracy.

BACKGROUND

Quality of care with the indicator disease

Among the three components for QC assessment (structure, process and outcome), sufficient evidence has supported measuring, reporting and improving the process as quantifiable actions, especially with an emphasis on evidence-based

medicine for promoting practices, can improve patient outcomes.^{3,4}

Cardiovascular diseases, for example acute myocardial infarction (AMI), have been constantly used as indicator diseases in the QC measurement. Recognizing the substantial burden caused by heart diseases, major cardiology societies have summarized evidence-based studies and developed clinical guidelines for STEMI and unstable angina/non-ST-segment elevation myocardial infarction (UA/NSTEMI) for the clinicians all over the world to follow.^{5–7} In addition, the National Registry of Myocardial Infarction 4 study performed quality of healthcare research that evaluated QC for STEMI and UA/NSTEMI in all of the major US hospitals.⁸ The current process of QC measures for STEMI raised by the National Registry of Myocardial Infarction 4 study includes timely reperfusion, medication use at the emergency room and during hospitalization, lipid management and management of complications (eg, shock and heart failure).¹ UA/NSTEMI shares the same set of QC measures, except that early reperfusion is not always required (early vs delayed revascularization).⁹ In addition, the patients undergoing coronary artery bypass grafting surgery (CABG) experience a very high coronary risk and also require life long treatment with antiplatelet agents, β -blockers and lipid-lowering therapy. Therefore, the same set of measures could be applied to evaluate QC across the continuum of coronary heart disease (CHD), that is, stable CHD, CABG, UA/NSTEMI, and STEMI.

Assessing QC measurement by data-mining of the electronic medical databases

In recent years, some hospitals have started to build electronic medical record systems to replace paper medical charts.¹⁰ However, it is still difficult to obtain comprehensive information for QC measures due to unstructured free-text data in electronic medical records (an example is shown in appendix A).

Natural Language Processing (NLP) and Medical Language Extraction and Encoding System (MedLEE)

Natural language processing (NLP) offers an approach for capturing data from narratives and creating structured reports for further computer processing. NLP is an effective method of accurately identifying and reusing data from clinical notes in several domains, such as real-time decision

support for community-acquired pneumonia in radiology findings and adverse event detection in outpatient visit notes as well as quality measurement from discharge notes and laboratory databases.^{11–16}

The difficulty in using NLP in the medical domain is that there is a diversity of lexicons and semantics in different subdomains. A natural language parser MedLEE (Medical Language Extraction and Encoding System) has been used to recognize medical concept entities and has provided promising results in various aspects of medical care.^{17–19} MedLEE was originally designed for decision support applications in the domain of chest x-ray reports and showed high accuracy, sensitivity and specificity in extracting specific clinical information from discharge summaries and x-ray reports.²⁰ Recent studies have expanded the applications of MedLEE to different medical fields, such as pharmacovigilance, analysis of nursing narratives, and representation of biological and phenotypic information.^{21–24} In this study, we applied MedLEE, which was available for research purposes to extract and encode discharge note information. Quantification of QC measures from electronic discharge notes would serve as a new application for MedLEE. User licensure was approved by Columbia University.

SYSTEM DESCRIPTION

The goal was to develop an automated system that could understand the contexts in the discharge notes so that QC measures for coronary heart disease could be retrieved with accuracy. An automated information-retrieval engine combining both keyword search and NLP search was designed to receive narrative inputs and then to produce coded data as outputs that were ready for computation and data-mining. Java (JDK version 1.6) was used as the programming language on the Solaris 10.0 operating system on a Sun Blade 2000 workstation. Due to the similarity of QC measures across the spectrum of CHD, discharge notes of STEMI were used to develop the system. Those of UA/NSTEMI were used to evaluate the performance of the system, and those of CABG were input into the system as demonstration for a potential application.

Materials

Our institution, a 2400-bed university-affiliated teaching hospital, is a tertiary referral center for cardiovascular diseases in Taiwan. Residents use Microsoft Word to type discharge notes into 21 structured sections (an example is shown in appendix A). As soon as the resident finishes typing a note, the sections are extracted and stored in individual fields in a Microsoft Access database, along with imported echocardiograph reports and laboratory test results. The text of each section was placed as

a string. The use of clinical information system (CIS) data was approved by the Institutional Review Board.⁵

Patient selection and quality of care measures

We used electronic discharge notes generated between 2002 and 2004 for the 281 cases with STEMI to develop rules for mapping MedLEE's output to the variables of interest in our study. After system development, the 627 cases with UA/NSTEMI were used as the test set. Finally, the system was applied to QC analysis of the 645 cases that underwent CABG within the same time period. The system was designed to extract the eight major QC measures conducted in the National Registry of Myocardial Infarction 4 study (table 1).

The discharge notes of the patients with UA/NSTEMI were input into the system to examine the performance of the system after the development in the STEMI set. The same measures (from #1 to #8, table 1) were applied. There was a minor revision on Measure 1. Unlike STEMI, in which cardiologists have to perform primary coronary angioplasty within the 12 h golden time, the reperfusion strategy can be classified into early invasive approach and early conservative approach in UA/NSTEMI.⁹

Measure 2 identified all the complications during admission for UA/NSTEMI. Measures 3–6 were the identification of medication use, as in the STEMI set. The proportion of the patients who should have used this medication and were indeed prescribed with such medication was defined as the “attainment rate” to a certain drug. Measure 7 (the check-up of LDL and its level) and measure 8 (achieving LDL goal 1 year after discharge) could be obtained from the laboratory information system in a prospective follow-up.

Methodology

The computerized analysis of the discharge notes was divided into the following three steps: (1) preprocessor transformed the texts into the format that the automated system could parse; (2) the main engine, Medical Language Extraction and Encoding System (MedLEE), mapped the texts into structured formats; and (3) a few postprocessing modules were then applied to deal with the structured output and to map MedLEE's output to the measures described in table 1.

Preprocessing

Based on experience with the development set, we modified some medical terms and symbols that were frequently used in this medical subdomain but could not be recognized by MedLEE, for example, changing t-PA to tPA. We eliminated symbols or Chinese fonts that could cause abnormal outputs from MedLEE, and then sent the output of the preprocessor into MedLEE.

Table 1 Summary of the eight measures in process of quality of care for ST-segment elevation myocardial infarction and unstable angina or non-ST-segment elevation myocardial infarction

Measure 1	Proportion of early reperfusion within 12 h after onset of angina for ST-segment elevation myocardial infarction and early invasive approach (vs early conservative approach) for unstable angina or non-ST-segment elevation myocardial infarction
Measure 2	Proportion of cardiovascular and non-cardiovascular complications during admission
Measure 3	Proportion of ACE inhibitor (ACE-I)/angiotensin receptor blocker use before discharge in the patients with depressed left ventricular (LV) function but no contraindications (LV ejection fraction less than 45%)
Measure 4	Proportion of aspirin use before discharge in the patients without contraindications
Measure 5	Proportion of β -blocker use before discharge in the patients without contraindications
Measure 6	Proportion of statin use before discharge.
Measure 7	Proportion of LDL-C measurement during admission
Measure 8	Proportion of LDL-C goal attainment 1 year postdischarge

Application of Medical Language Extraction and Encoding System (MedLEE)

MedLEE mapped the text to UMLS concepts and used semantic and syntactic analysis processes to assign modifiers to the concepts and to determine attribute values such as certainty and temporal information. Figure 1 depicts an example that contains modifiers tagged as <certainty>, <date> and <code>. MedLEE output represented the core information we used to identify the eight QC measures.

Postprocessing

MedLEE does not make inferences from individual concepts and does not reason about relationships among concepts. Therefore, a critical step in using MedLEE was to map MedLEE's XML output to the QC variables. We developed five main modules to automatically extract required information from discharge notes and laboratory databases and, in addition, to map MedLEE's output to the eight QC measures, survival to discharge or referral (figure 2). Next, we describe each of the five post-processing modules.

1. Outcome Extraction

Outcome information could be extracted from either the "Condition" or the "Course" data field in the electronic discharge notes. All possible values were mapped into three mutually exclusive categories: "Death," "CABG," and "Discharge." Rule-based classification algorithms were applied to outcome extraction.

2. Reperfusion Detection

Five steps were applied to identify the certainty and timing of early reperfusion.

1) "Concept searching" used UMLS CUI codes (eg, *UMLS: C0018795_cardiac catheterization procedures*) to match the narration of the records with target concepts.

2) "Certainty and rule determiner" determined whether the certainty attribute was "high certainty" instead of "no certainty" or "moderate certainty." Some exception patterns, such as "neither ... nor..." were filtered by the rule determiner to avoid vague certainty.

3) "Temporal determiner" was designed to obtain precise timing of reperfusion, that is whether coronary angioplasty took place at the current admission and whether reperfusion took place within 12 h. Since most of the sentences did not note the precise procedure time, we used time frame in a contextual

```
<medlee>
<sentence>
  <structured form="xml">
    <procedure v="cardiac catheterization" idref="p2">
      <certainty v="high certainty" idref="p6"/>
      <date v="000205" idref="p12">
        <reltime v="on" idref="p10"/>
      </date>
      <parsemode v="mode1"/>
      <sectname v="impression"/>
      <sid idref="s1"/>
      <code v="UMLS:C0018795_cardiac catheterization procedures" idref="p2"/>
    </procedure>
  </structured>
</tt>
<sent id="s1">
  <phr id="p2">Cardiac catheterization</phr>
  <phr id="p6">was performed</phr>
  <phr id="p10">on</phr>
  <phr id="p12">Feb 5</phr>
</sent>
</tt>
</sentence>
```

Figure 1 Example of XML output from Medical Language Extraction and Encoding System. The original text is 'Cardiac catheterization was performed on Feb 5, 2000.' The major tags used in the analysis were <certainty>, <date> and <code>.

aspect to determine the presence of early reperfusion. To be specific, we marked the target sentence and assigned the time from the prior one, and then compared it with the event date (the *InDate* field).

4) "Pattern filter" was designed to filter out non-relevant reperfusion concepts and applied to aggregate the frequencies of relevant concepts for concept pattern-matching purposes, and the following example will serve to clarify this mechanism. In general "Cardiac catheterization procedure" might co-occur with some certain words or phrases related to coronary anatomy, such as a typical pattern written in the discharge notes would read as "Coronary cath: LM: patent, LAD: 100% stenosis, LCX: patent, RCA: patent, PTCA to LAD." This example indicated the principal function of pattern filter: the more frequent the relevant concepts co-occurred with the idea of reperfusion, the more effective the filtering function was.

5) "Source determiner" integrated four fields ("*InDate*," "*PI*," "*Course*," and "*Lab*") in which information about reperfusion could be documented to gain better matching results, assigned the time from the prior sentence, and then compared it with the event date (the "*InDate*" field) to calculate time between admission and reperfusion.

3. Complication Search

In-hospital complications were usually described in the "Complications" field but sometimes appeared in the "Course" or "PI" field. The previously described "concept searching algorithm" was used to identify and to count the occurrence of these events during admission.

4. Medication Search

The attainment rate was calculated after adjustment of the denominator by removing patients with contraindications to certain medication and those who died before discharge. A "contraindication detection" algorithm was applied to identify contraindications for each medication (appendices B and C). Keyword search and the previously described "concept searching algorithm" were used to extract medication information from the "Discharge" fields and to find the left ventricular ejection fraction (LVEF) from the "Course" field or in the *ECHO* data. Once LVEF was less than 45%, the requirement for ACE-I/ARB use was triggered. In addition, a "pattern matching" algorithm was designed to extract related physiological and laboratory data [heart rate (HR), systolic blood pressure (SBP), renal function (creatinine, CRE), electrolyte (potassium, K⁺)].

The brand names and generic names of a medication, for example, Tapal or Bokey (a brand of aspirin), were not always tagged by MedLEE. A "string matching" algorithm was used to link the unknown strings with the medication lexicon, as shown in figure 3. In the "Medication Search" module, we allowed a more flexible search strategy for misspelling tolerance. We allowed four levels of tolerance, from zero to three, indicating the number of characters that could be misspelled and still match a medication name in our lexicon. Tolerance was zero when the length of the term was less than 5, 1 between 6 and 8, 2 between 9 and 11, and 3 when it was greater than 11. For example, tolerance was set to zero for a short brand name "Tapal" but set to 2 for a longer one "propranolol."

5. LDL Measurement

This module calculated the performance of lipid check-up during admission and the LDL-C goal attainment rate 1 year later. An algorithm searched for the coded "*Lab*" table and determined whether LDL-C was examined between admission ("*InDate*") and discharge ("*OutDate*"). This algorithm then linked those records of the same patient in different years so as

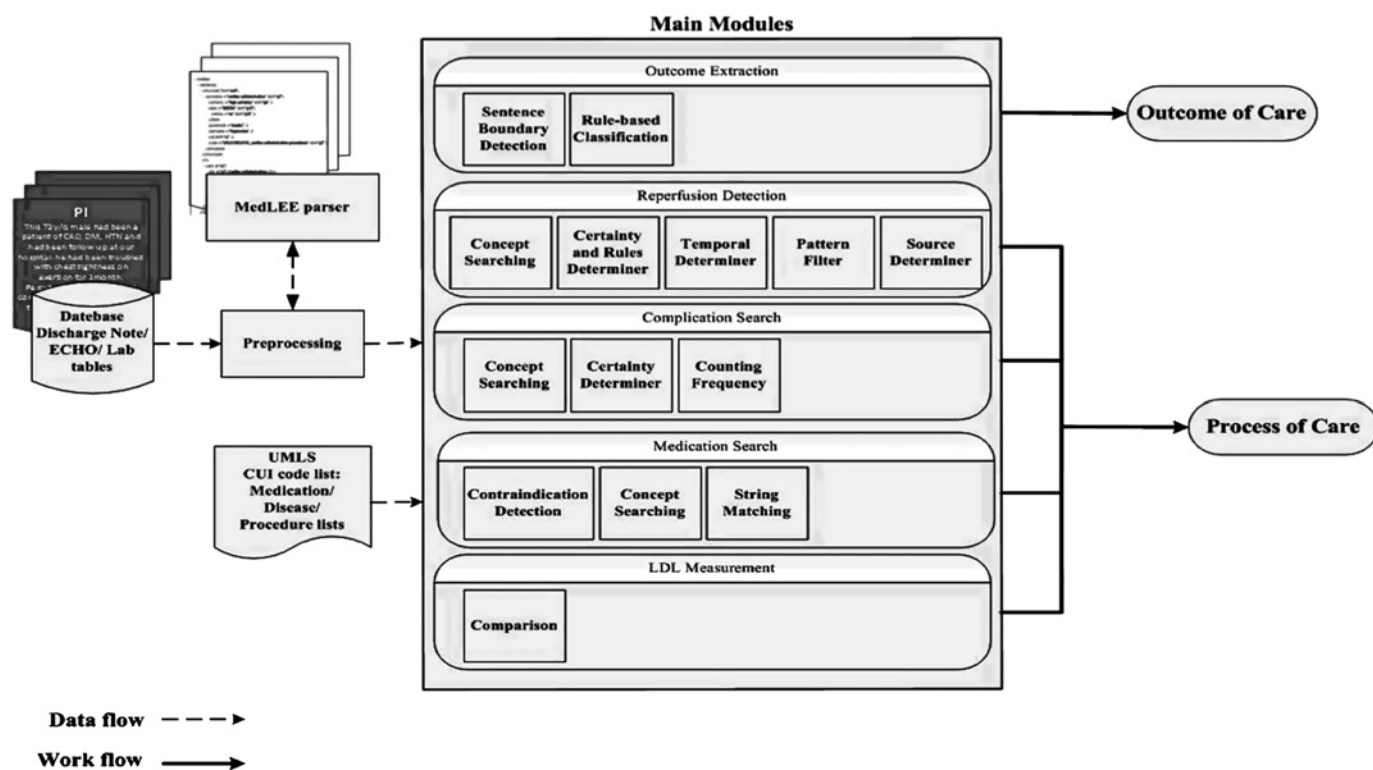


Figure 2 System overview. Medical Language Extraction and Encoding System (MedLEE) processes and encodes unstructured free texts into XML format. The five main modules, namely “Outcome Extraction,” “Reperfusion Detection (Measure 1),” “Complication Search (Measure 2),” “Medication Search (Measures 3, 4, 5, and 6),” and “LDL measurement (Measures 7 and 8),” comprise the assessment of quality of care in the patients with STEMI, UA/NSTEMI, or undergoing coronary artery bypass graft surgery. CUI, Concept Unique Identifier; ECHO, echocardiographic reports; UMLS, Unified Medical Language System.

to extract the LDL-C value 1 year later. Goal attainment was affirmed when LDL-C reached 100 mg/dl.

System evaluation

We evaluated the system’s ability to identify the eight QC measures and made an interpretation of patient outcome (table 1). To evaluate the documentation, the system was tested on the discharge notes from UA/NSTEMI. We report the accuracy in determining early conservative versus early invasive approach (Measure 1), the proportion of patients receiving required medication (Measures 3~6: attainment rate for ACE-I, antiplatelet agents, β -blockers and lipid-lowering agents, respectively) and also the outcomes of these patients. Due to the large amount of information and a variety of disease manifestations, Measure 2 was never validated by the experts but retrieved by the system. No gold standard was provided. Evaluation of the accuracy for Measures 7 and 8 was attempted. However, due to the relatively poor performance of lipid management in our institution and the need to prospectively follow-up laboratory data, it is difficult to derive “gold standards” for these two measures (LDL check-up and follow-up) for all cases for which the system contradicted the cardiologist to show the agreement between system performance and the gold standard.³

To prove the accuracy and efficacy of the automated system, the detailed comparison between the interpretation of the cardiologists and that of the system was drawn as follows. Each case in the test set of 627 cases with UA/NSTEMI was thoroughly reviewed by a cardiologist to establish the gold standard. This cardiologist was asked to read the discharge note and the laboratory results of each patient with UA/STEMI, and to determine the presence or absence of each of the eight QC

measures without knowing the output from the automated acquisition system.

The interpretation from the system was compared with that of the cardiologist. A second cardiologist read the discharge notes and laboratory data for all cases which the system and the cardiologist contradicted. The second cardiologist was unaware of either the interpretation from the first cardiologist or from the output of the automatic acquisition system, so as to provide an independent judgment. The second cardiologist’s judgments were considered the gold standard judgment for those cases.^{25 26}

In the test set, the κ -value was used to show the agreement between system performance and gold standards. In addition, sensitivity, specificity, positive predictive value and negative predictive value were reported for Measure 1 and Measures 3~6 by comparing the system output against the gold standard.

System application

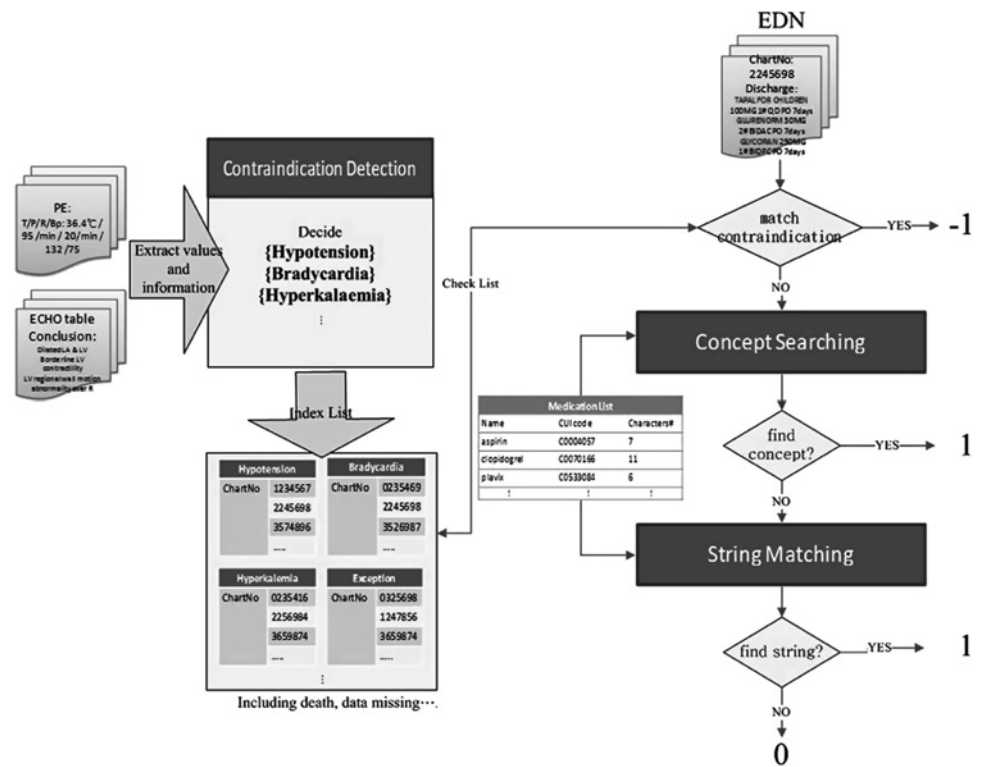
The automated system was then applied to analyze QC measures in the patients status post-CABG, aiming to demonstrate the efficacy of the automation when the medical expert was not involved in reading medical records. We report the percentage of medication prescription (attainment rate to ACE-I, antiplatelet agents, β -blockers and lipid-lowering agents), various complications and case death rate associated with CABG. No gold standard was applied to the application set.

RESULTS

System evaluation results

In the test set, the agreement statistics κ -value between the system and the reference standard could be successfully reported

Figure 3 Workflow of “Medication Search”: “1” represents that the patient needs a medication and the system finds the prescription correctly; “0” represents that the patient needs a medication but the system does not find the medication; “-1” represents that the patient does not need the medication because of contraindications or missing data. EDN, electronic discharge notes.



for five major measures. Measures 3–6 showed good agreement with cardiologists; therefore, we could use these measures to calculate prevalence in the following part of the evaluation. The early reperfusion rate and the attainment rate to medication resulting from manual and automated approaches were very similar, and so the κ -values were high (table 2).

Figure 4 shows relative prevalence of various types of complications (Measure 2) during admission for UA/NSTEMI. Since no expert validation was involved in this measure, we compared the occurrence of various types of complications following STEMI and UA/NSTEMI.

LDL measurement was found in 75.6% of the patients with UA/NSTEMI (Measure 7). The LDL level ranged above 130 mg/dl in one-third of the patients, between 100 and 130 mg/dl in one-third, and below 100 mg/dl in the other one-third. The agreement statistics κ -value was 0.68 in those patients with valid test results. One year later, the attainment rate to the LDL goal was unknown because only three cases were found to have LDL measurement 1 year later (Measure 8). The outcome of the UA/NSTEMI patients was classified into three categories: death, CABG or discharge. The overall accuracy was 95.4%, and the κ -value was 0.82 (table 3).

Application results

The results of the automated analysis of the application set (patients receiving CABG) are shown in table 4. The medication prescription rate was much lower in patients receiving CABG than those with UA/NSTEMI, implying that cardiovascular surgeons obeyed treatment guidelines less rigorously than the cardiologists. Complications after CABG were also reported. The overall case death rate, including both urgent and elective CABG, was 8.7%. Although the complication rates might be clinically feasible, the overall case death rate was much higher than the internationally accepted standards (mortality after elective CABG around 2%).

DISCUSSION

We have successfully implemented an acquisition system that is able to extract parameters for QC measures automatically using natural language processing combined with postprocessing algorithms. We have also validated the accuracy of the system (UA/NSTEMI test set) and demonstrated its use in a related application (patients with CABG). Overall, the automated system could identify the information about reperfusion strategy and discharge medication in a high agreement with that extracted by cardiologists. The information extraction about lipid management (LDL check-up and follow-up) was unsatisfactory because the doctors in our institution did not check and follow-up lipid profiles as rigorously as the guidelines suggest. Also, we illustrated a particular QC pattern for UA/NSTEMI, that is, a relatively high early invasive approach, adequate use of antiplatelet agents and ACE-I/ARB, and poor performance in β -blocker usage and lipid management. The automated system retrieved the medication information and calculated the attainment rate with accuracy.

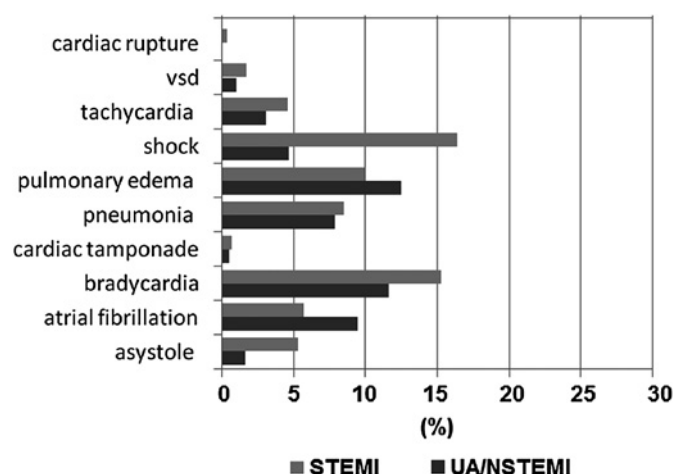
The methodology using textual reports as the information source has eliminated the needs of complicated integration during databases operation. More importantly, this approach has taken into account clinical conditions (eg, shock, hypotension, bradycardia, hyperkalemia, azotemia, etc) in which physicians would decline the use of certain medication. The attainment rate to QC standards might be underestimated if contraindicated cases were not excluded from analysis during a medication search. In addition, the acquisition system is able to find the occurrence of various kinds of cardiovascular events for which it was almost impossible to be retrieved unless a tremendous amount of human time and expertise was involved.

One problem related to system evaluation was overtraining. The gold standards concerning the patients with STEMI had been setup before the beginning of the design of the automated

Table 2 System performance and expert evaluation on the test set

Unstable angina/non-ST-segment elevation myocardial infarction		True positive	False positive	False negative	True negative	Total	Overall accuracy	κ Value	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Attainment rate by system	Attainment rate by gold standards
Measure 1	Early invasive approach by coronary angioplasty	264	33	78	252	627	0.823	0.65	0.77	0.88	0.889	0.76	0.48	0.55
Measure 3	ACE inhibitors/angiotensin receptor blockers before discharge (poor left ventricular ejection fraction)	48	1	2	22	73	0.958	0.91	0.96	0.96	0.980	0.92	0.67	0.68
Measure 4	Aspirin before discharge	474	2	5	24	505	0.986	0.87	0.99	0.92	0.996	0.83	0.94	0.95
Measure 5	β -blockers before discharge	296	1	6	185	488	0.986	0.97	0.98	0.99	0.997	0.97	0.61	0.62
Measure 6	Statins before discharge	167	0	7	340	514	0.986	0.97	0.96	1.00	1.000	0.98	0.32	0.34

Measures 7 and 8 are reported in the text.

**Figure 4** Cardiovascular and non-cardiovascular complications (Measure 2) in both ST-segment elevation myocardial infarction (STEMI) and unstable angina or non-ST-segment elevation myocardial infarction (UA/NSTEMI). vsd, ventricular septal defect.

acquisition system.³ However, this kind of evaluation might have tuned the system to perform well on the patients with STEMI. It was expected that there would be a degradation of κ -values when applying what was learnt on the patients with UA/NSTEMI. However, since there were about threefold cases in the test set, and the calculation of κ -values was sensitive to the case number, the degradation of performance due to over-training was not seen.

Error analysis

The comparison between automated and manually derived attainment rates shows similar numbers except on Measure 1. In the test results (UA/NSTEMI), the distinction between early conservative approach and early invasive approach was found to have the lowest κ -value. Unlike patients with STEMI who usually presented to the hospital for the first time, the patients with UA/NSTEMI might have visited the hospital several times, and coronary angiography and coronary angioplasty might be performed either on previous occasions or in the current admission. Thus, most false-positive and false-negative interpretations resulted from the lack of capacity of the automated system in finding the fine details about the temporal information. One of the solutions was to apply a “temporal tag” to each event (timing of previous catheterization, present symptom onset and timing of the catheterization in this admission) to calculate the time gap in between. However, the lack of the temporal resolution in the reperfusion time was mostly attributed to insufficient documentation in electronic

Table 3 Outcome extraction (three categories: Death, Coronary artery bypass graft and Discharge) for the patients in the unstable angina/non-ST-segment elevation myocardial infarction test set

Categories by experts	By system	Coronary artery bypass graft		Discharge	Total
		Death			
Death		19	1	0	20
Coronary artery bypass graft		0	55	15	70
Discharge		0	12	505	517
Total		19	68	520	607

Each column represents the outcome instances in a predicted category provided by the acquisition system, whereas each row represents the outcome instances in an actual category provided by human experts. Discharge information was missing in 20 cases, leaving a total effective case number of 607.

Table 4 System performance on the application set in which the patients received coronary artery bypass graft surgery

Coronary artery bypass grafting		Positive	Total	Percentage (by system)
Measure 3	ACE inhibitors/angiotensin receptor blockers before discharge (poor left ventricular ejection fraction)	4	26	15.4
Measure 4	Aspirin before discharge	426	589	72.3
Measure 5	β -blockers before discharge	141	562	25.1
Measure 6	Statin before discharge	27	589	4.6
Complications	Types			Percentage
	Hypotension/shock			19.0
	Atrial fibrillation			10.7
	Pulmonary edema			8.0
	Bradycardia			6.4
	Pneumonia			4.9
	Ventricular septal defect			3.4
	Asystole			1.0
	Cardiac tamponade			0.7
Case death	Elective+emergent surgery			8.7

discharge notes rather than MedLEE's errors. The cardiologists would have demonstrated an instinctive capacity in exploring the sequence, but the automated acquisition has not obtained such a function yet.

In terms of a medication search, there were more false-negative than false-positive cases. The system searched for medication information in the "Discharge" field, whereas the residents might have written the discharge prescription in the final paragraph of the "Course" field. Also, the disagreement in outcome extraction in UA/NSTEMI set was related to the misclassification between CABG and DISCHARGE. Some patients were discharged and planned to readmit for CABG, and others actually underwent CABG and were then discharged (table 3). Expanding the searching fields of the automated system is expected to reduce the occurrences of misclassification. Occasionally, medication information could be retrieved from other fields than "Discharge." We did not attempt to parse "Course" or other fields for fear of increasing the complexity and potentially increasing false negatives for medication search.

The errors by MedLEE or errors from the above mapping algorithms were relatively rare. However, due to the similar QC measures in the development set, the test set and the application set, the accuracy could be highly reproducible. On the other hand, there may be doubts as to whether the same accuracy could be maintained if the system were applied to other hospitals or to other medical subdomains.

Limitations

The automated system is dependent on residents documenting the important facets of the hospitalization in a fairly comprehensive electronic discharge note. This may not be applicable in many parts of the world. Therefore, we are conservative in expanding the generalizability of the system to a wider scale.

Generalization of QC evaluation to other medical applications requires a flexible interface that enables clinical researchers to apply the automated system to develop the strategy for information retrieval and extraction. Currently, our system was designed to extract QC measures for cardiovascular diseases, and its application cannot be generalized to other medical domains yet.

Concerning system evaluation, the two cardiologists should have independently reviewed the discharge notes to provide information about inter-rater reliability. However, it was difficult to perform an item-by-item consensus-forming process

between the two independent experts due to multiple measures in hundreds of patients. Instead, the second expert who was independent of the opinions from the first cardiologist and unaware of the system outputs was involved whenever it was necessary. In addition, the simultaneous occurrence of interpretation errors from the first cardiologist and from the system might have slightly decreased the overall accuracy in estimating the attainment rates, but from reviewing the output, we believe errors of that type were relatively rare.

CONCLUSION

In conclusion, our study has provided an innovative tool for assessing QC for cardiovascular diseases that involves in applications of NLP. MedLEE performed well on the task, and our mapping rules successfully mapped MedLEE's output to QC measures and interpretations.

This system will be applied to retrieve quality of care measures for the patients across the continuum of cardiovascular diseases in our institution to evaluate the quality of care for CHD over a decade. Also, this automated approach can provide a standardized and objective evaluation method for comparing the quality of care across different hospitals in our country. In terms of information innovation, future work will target designing a multipurpose information retrieval and extraction device that is able to explore quality of care in various aspects of medical care.

Acknowledgements The authors would like to thank C Friedman, S Johnson, G Hripcsak, L Shagina, and the anonymous reviewers. The proofreading assistance by P Chiang is much appreciated.

Funding This study was supported in part by research grants from National Science Council, Taiwan (NSC 96-2221-E-002-241 and NSC 97-2221-E-006-233).

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Spertus JA, Radford MJ, Every NR, *et al*. Challenges and opportunities in quantifying the quality of care for acute myocardial infarction: summary from the acute myocardial infarction working group of the American Heart Association/American College of Cardiology first scientific forum on quality of care and outcomes research in cardiovascular disease and stroke. *Circulation* 2003;**107**:1681–91.
2. Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. *J Am Med Inform Assoc* 2005;**12**:448–57.
3. Chang SN, Lin JW, Liu SC, *et al*. Measuring the process of quality of care for ST-segment elevation acute myocardial infarction through data-mining

- of the electronic discharge notes. *J Eval Clin Pract* 2008;**14**: 116–20.
4. **Donabedian A.** The quality of care: how can it be assessed? *JAMA* 1988;**260**:1743–8.
 5. **American Heart Association.** Heart and stroke statistical update. <http://www.americanheart.org/statistics/index.html> [online article].
 6. **Antman EM, Anbe DT, Armstrong PW, et al.** ACC/AHA guidelines for the management of patients with st-elevation myocardial infarction. *J Am Coll Cardiol* 2004;**44**:671–719.
 7. **Van de Werf F, Ardissino D, Betriu A, et al.** Management of acute myocardial infarction in patients presenting with ST-segment elevation. The task force on the management of acute myocardial infarction of the european society of cardiology. *Eur Heart J* 2003;**24**:28–66.
 8. **Roe MT, Parsons LS, Pollack CV Jr, et al.** Quality of care by classification of myocardial infarction. *Arch Intern Med* 2005;**165**:1630–6.
 9. **Anderson JL, Adams CD, Antman EM, et al.** ACC/AHA 2007 guidelines for the management of patients with unstable angina/non ST-elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. *Circulation* 2007;**116**:e148–304.
 10. **Rubenfeld GD.** Using computerized medical database to measure and to improve the quality of intensive care. *J Crit Care* 2004;**19**:248–56.
 11. **Hyun S, Johnson SB, Bakken S.** Exploring the ability of natural language processing to extract data from nursing narratives. *Comput Inform Nurs* 2009;**27**:215–23.
 12. **Fiszman M, Haug P.** Using medical language processing to support real-time evaluation of pneumonia guideline. *Proc AMIA Symp* 2000:235–9.
 13. **Honigman B, Light P, Pulling RM, et al.** A computerized method for identifying incidents associated with adverse drug events in outpatients. *Int J Med Inform* 2001;**61**:21–32.
 14. **Pakhomov SS, Hemingway H, Weston SA, et al.** Epidemiology of angina pectoris: role of natural language processing of the medical record. *Am Heart J* 2007;**153**:666–73.
 15. **Pakhomov S, Weston SA, Jacobsen SJ, et al.** Electronic medical records for clinical research: application to the identification of heart failure. *Am J Manag Care* 2007;**13**:281–8.
 16. **Pakhomov SV, Jacobsen SJ, Chute CG, et al.** Agreement between patient-reported symptoms and their documentation in the medical record. *Am J Manag Care* 2008;**14**:530–9.
 17. **Jain NL, Knirsch CA, Friedman C, et al.** Identification of suspected tuberculosis patients based on natural language processing of chest radiograph reports. *Proc AMIA Annu Fall Symp* 1996:542–6.
 18. **Jain NL, Friedman C.** Identification of findings suspicious for breast cancer based on natural language processing of mammogram reports. *Proc AMIA Annu Fall Symp* 1997:829–33.
 19. **Mendonca EA, Haas J, Shagina L, et al.** Extracting information on pneumonia in infants using natural language processing of radiology reports. *J Biomed Inform* 2005;**38**:314–21.
 20. **Friedman C, Knirsch CA, Shagina L, et al.** Automating a severity score guideline for community-acquired pneumonia employing medical language processing of discharge summaries. *Proc AMIA Symp* 1999:256–60.
 21. **Wang X, Hripcsak G, Markatou M, et al.** Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study. *J Am Med Inform Assoc* 2009;**16**:328–37.
 22. **Melton GB, Hripcsak G.** Automated detection of adverse events using natural language processing of discharge summaries. *J Am Med Inform Assoc* 2005;**12**:448–57.
 23. **Bakken S, Hyun S, Friedman C, et al.** ISO reference terminology models for nursing: applicability for natural language processing of nursing narratives. *Int J Med Inform* 2005;**74**:615–22.
 24. **Chen L, Friedman C.** Extracting phenotypic information from the literature via natural language processing. *Stud Health Technol Inform* 2004;**107**:758–62.
 25. **Hripcsak G, Wilcox A.** Reference standards, judges, and comparison subjects: roles for experts in evaluating system performance. *J Am Med Inform Assoc* 2002;**9**:1–15.
 26. **Hripcsak G, Kuperman GJ, Friedman C, et al.** A reliability study for evaluating information extraction from radiology reports. *J Am Med Inform Assoc* 1999;**6**:143–50.