Feasibility of the Partial Automation of Data Abstraction for the Quality and Safety Review System

Prepared by:
Clinovations Government + Health
Washington, DC

Contributing Team:
Clinovations GovHealth
Nicole Kemper, M.S.
Laura McQueen, R.N., M.S.
Fatema Rafiqi, Ph.D.
Anita Samarth
Manan Trivedi, M.D., M.P.P.
MedStar
Katie Adams
Allan Fong, M.S.
Daniel Hoffman
Raj Ratwani, Ph.D.

Contents

Figures

1 Executive Summary

1.1 Background

The Agency for Healthcare Research and Quality (AHRQ) developed and maintains the Quality and Safety Review System (QSRS), a patient safety surveillance system with clinical event data populated by human abstractors who review medical records at the Centers for Medicare & Medicaid Services Clinical Data Abstraction Center. QSRS is being considered as a replacement for the current Medicare Patient Safety Monitoring System, and was developed so that it could also be made available for other users (e.g., hospitals). QSRS is a patient safety surveillance system designed by AHRQ to detect adverse events (AEs) from a sample of hospital and patient records and to provide reports on rates of AEs reviewed.

The rapid adoption of electronic health records (EHRs) meeting industry standards required by the Office of the National Coordinator for Health Information Technology (ONC) Certification Program provides an opportunity to determine the degree to which QSRS may be enhanced with automated download of discrete data values in electronic medical records and the opportunity to apply natural language processing (NLP).

As of 2014, nearly 97 percent of non-Federal acute care hospitals have implemented certified EHR technology (CEHRT), and over 75 percent of hospitals have adopted basic EHR functions. Given this high adoption rate, AHRQ's Center for Quality Improvement and Patient Safety contracted Clinovations Government + Health (Clinovations GovHealth) and its subcontracting partner, MedStar Health National Center for Human Factors in Healthcare (MedStar), to perform a feasibility study of full or partial automation of the QSRS abstraction process using EHR data.

This study consisted of five tasks:

- **Task 1 - Study, Review, and Analyze:** Perform a review of QSRS events, algorithms, and data abstraction guidelines and perform an analysis of certified EHR technology.
- **Task 2 - Environmental Scan:** Conduct an environmental scan of peer-reviewed and grey-literature publications to identify existing capabilities for electronic tools and methods for

detection and surveillance of adverse events in health care settings, and the adverse events specified in QSRS.

- **Task 3 - Analyze EHR and Other Source Information:** Identify and review adverse event detection and surveillance capabilities from EHR vendors, reporting/analytics vendors, and health systems/customers.
- **Task 4 - Feasibility Analyses and Review:** Analyze EHR capabilities to support QSRS automation, identify opportunities for automation leveraging EHRs and health information technology (IT) capabilities, to discuss barriers and opportunities for automation.
- **Task 5 - Final Report and Presentation:** Develop a final presentation and final narrative report covering the findings from Tasks 1, 2, 3, and 4.

1.2 Analysis

The team developed an evaluation framework, applied to the 205 QSRS questions, to assess the feasibility of automation at a question level and overall module level. QSRS abstraction guidelines provide multiple data sources versus a singular, primary data source, as guidance for addressing each question. As EHR documentation practices and culture vary across health systems, hospitals within the health system, and provider types (e.g., gastroenterologist vs. hospitalist; medical/surgical nurse vs. intensive care unit nurse), multiple data sources should be considered to address questions. Certain questions may have a definitive primary data source that can negate the need to research additional sections of the patient chart and are well-positioned for automation.

CEHRT provides only for a small subset of functionality needed to support QSRS algorithms. CEHRT standards and interoperability requirements do not support capture and exchange of the metadata needed to effectively address QSRS algorithms. The detailed clinical information needed to address surveillance according to QSRS algorithms is contained within nursing assessments, flowsheets, and physician/provider progress notes.

Commercial EHR vendors are addressing adverse events (AEs). For a number of AEs that are a part of QSRS, EHRs offer clinical decision support rules/alerts and safety/population health reports that support both prevention and reporting of AEs. Commercial EHRs rely on clinical documentation practices to chart in specified manners or require health systems to setup mappings to address localized workflows.

Figure 1 below depicts the source data for the 205 QSRS questions. This classification was based upon study activities with clinician users of multiple EHRs, stakeholder reviews, and direct analysis with EHR vendors.

Questions that can be addressed using numeric values or structured and coded data are the easiest to automate, and likely no not require advanced analytics, machine learning, or extensive natural language processing (NLP) to analyze the EHR source data to make a QSRS question determination. Information stored in free text or uncoded values that vary across organizations require NLP for automation.

In addressing NLP feasibility, we developed the following framework and classified all questions against this framework (Figure 2).

Figure 1. QSRS Source Data Analysis

Numeric
- YES: 14
- NO: 191

Structured and Coded
- YES: 63
- NO: 142

Structured and Uncoded
- YES: 77
- NO: 128

Structured Free Text
- YES: 64
- NO: 141

Unstructured Free Text
- YES: 132
- NO: 73

Requires NLP for Automation
- YES: 134
- NO: 71

Figure 2. QSRS Question Types in Order of Complexity With Associated Computational Challenges

Increasing Algorithm Complexity →

Type of Questions	Example	Challenges
Presence of a concept/entity	Did the patient have a urinary tract catheter inserted during the stay? [GENERIC25]	Negation and synonyms
Numeric value extraction	During this hospitalization, did the patient have a PTT value greater than 100 seconds? [MEDIQ10]	Part-of-speech tagging and word sense disambiguation
Multiple concept detection	Which secondary morbidities developed? [PUQ7]	Co-referencing
Temporal occurrence of a concept and concept referencing	On or within the first 24 hours of admission, was a history of allergies and/or sensitivities documented? [MEDIQ1]	Defining temporal relations
Contingency	Did bleeding develop more than 24 hours after admission and within 1 day of [('PTT'>100 seconds) OR ('Protamine administration') OR… [MEDIQ35]?	Co-referencing
"Fuzzy" concepts	Did the patient undergo an **unplanned** transfer to a higher level care area within the facility or to another facility? [OTHERQ13]	Reasoning and Subjectivity
Open-ended responses	If at all, describe how the device harmed the patient. [DEVICEQ6]	Summarization

We determined 58 percent of QSRS questions (118 out of 205) could be reasonably automated with the use of currently available software and applications either without NLP or "low complexity" NLP.

Upon adding "medium complexity" NLP for consideration, an additional 40 questions can be automated, resulting in 77 percent of QSRS questions that are categorized as not requiring NLP or those as low and medium relative NLP complexity. In summary, we believe that 58 percent of QSRS questions are relatively easy to automate and 77 percent of QSRS questions are feasible for automation using available capabilities in the market today.

For the remaining 23 percent that require NLP and are classified as "high complexity," in parallel with the approaches and pilots recommended in this section (which best address the 77 percent of questions that are feasible for automation), AHRQ should consider a review of these questions and determine whether AHRQ may identify areas for introduction and engagement with standards development organizations and other standards and specifications bodies, as well as health IT certification.

1.3 Recommendations

Three modules (Pressure Ulcers, HAI-CDI, and Blood/Blood Products) yielded NLP difficulty levels of only low or medium, making them ideal for pilots or proof-of-concept studies for automation. An additional area for AHRQ focus are the Entry Questions, both generic and module specific. Automation of Entry Questions enables the automation of processing the greatest number of charts where only those that are flagged during the Entry Question process would require manual chart review or abstraction.

AHRQ may also want to consider mimicking the eMeasure development process to facilitate automation of data collection of EHRs, if QSRS is to be expanded for widespread use. Health system culture is interested in focusing its health IT interventions on predictive analytics and decision support versus surveillance and leveraging of its EHR investments versus performing manual chart abstraction using limited coding or quality improvement/performance improvement resources and budgets.

We recommend a series of next steps and pilots that support evaluation and testing of partial automation approaches for QSRS. Given that much of the information needed to address QSRS algorithms is text-based, natural language processing (NLP) and advanced analytics capabilities provide promise for automating certain QSRS questions. These recommendations include—

- **Proof-of-Concept Project: Test Rule-Based and NLP Approaches to Populate QSRS** – Identify effective interface designs to support QSRS automation. Determine whether potential automated approaches will work by piloting on a sample of data from different EHRs.
- **Pilot Projects/Competition: Test Partial Automation Interface-Level Guidance** – Use NLP to facilitate automatic identification of the correct information to human-in-the-loop processes like highlighting where the appropriate information might be found.
- **Comparison Study: Automation With One Vendor Versus Manual Abstraction** – Single vendor or competition approach to automate as much of QSRS possible, prioritizing the modules indicated in this report. Conduct a parallel effort using manual abstraction only for comparison.
- **Comparison Study: C-CDA (Consolidated Clinical Document Architecture) and HL7 ADT (Admit-Discharge-Transfer) Interfaces Versus Manual Abstraction** – Conduct a study to determine the efficacy of available electronic data using standards formats such as the C-CDA, HL7 ADT, and billing/claims transactions to determine if the needed information can support QSRS functions and needs. Conduct a parallel portion of the study where manual abstractors are reviewing the same charts that are processed electronically.
- **Additional Research/Pre-Work** – Additional research using data sources external to the EHR: Mini-Automation Feasibility Assessment that considers non-EHR data sources: Scaled down environmental scan to ensure tools used by Performance Improvement, Quality Improvement, and Risk Management teams in hospitals are considered; Risk Management/Insurance Premium

Opportunities for use of QSRS Surveillance Data; and Review of Hospital Resources and Tools for Predictive Versus Retrospective Chart Analysis.

2 Background and Approach

2.1 Background/Need

The Agency for Healthcare Research and Quality (AHRQ) developed and maintains the Quality and Safety Review System (QSRS), a patient safety surveillance system with clinical event data populated by human abstractors who review medical records at the Centers for Medicare & Medicaid Services (CMS) Clinical Data Abstraction Center (CDAC). QSRS is being considered as a replacement for the current Medicare Patient Safety Monitoring System (MPSMS), and was developed so that it could also be made available for other users (e.g., hospitals). QSRS is a patient safety surveillance system AHRQ designed to detect adverse events (AEs) from a sample of hospital and patient records and to provide reports on rates of AEs reviewed.

Currently, data abstraction for QSRS is manually performed by clinical abstractors who follow a series of questions in the software, guided by algorithms seeking specific answers to questions determined by medical experts. QSRS provides 19 modules for event descriptions within the hospital setting. CDAC abstractors review and extract data from 20,000 to 40,000 charts identified by CMS each fiscal year for CMS reviews. Today, charts are requested and provided by hospitals as photocopies or in CD-ROM format for patients age 18 and older, not restricted to CMS/Medicare beneficiaries. QSRS abstraction for a set of records in a fiscal year takes approximately 9 months, with an abstractor spending about 1 hour and 15 minutes per chart to perform chart review and complete the abstraction process. QSRS provides information on adverse event incidents that may have occurred but may not have been realized or noted within the medical record. It also has a way for abstractors to note a patient safety event that they saw while reviewing the chart, whether or not it was noted as an event by the software.

The rapid adoption of electronic health records (EHRs) meeting industry standards required by the Office of the National Coordinator for Health Information Technology (ONC) Certification Program provides an opportunity to determine the degree to which QSRS may be enhanced with automated download of discrete data values in electronic medical records and the opportunity to apply natural language processing (NLP).

As of 2014, nearly 97 percent of non-Federal acute care hospitals have implemented certified EHR technology, and over 75 percent of hospitals have adopted basic EHR functions[1]. Given this high adoption

rate, AHRQ's Center for Quality Improvement and Patient Safety contracted Clinovations Government + Health (Clinovations GovHealth) and its subcontracting partner, MedStar Health National Center for Human Factors in Healthcare (MedStar), to perform a feasibility study of full or partial automation of the QSRS abstraction process using EHR data.

2.2 Method/Approach

An in-depth, multiphased approach was developed and implemented to address the complexity of the feasibility assessment and to ensure that all facets of potential feasibility were explored.

The initial task of work involved reviewing all relevant and available information on QSRS and the algorithms, modules and questions that define it. During this task, an initial review of QSRS was completed by applying the algorithms and questions to EHR charts. The second task of work was the completion of an environmental scan and literature review to identify the current state of computational methods available to detect adverse events in hospital settings from EHRs. Building off the first two tasks of work, the third and fourth tasks of work were completed in parallel. The third task of work involved identifying stakeholder groups and conducting interviews with representatives from each group to obtain additional information related to the focus of the environmental scan. The fourth and final task of work involved an extremely detailed analysis of the QSRS system in the context of two identified EHR developers and one NLP/analytics vendor.

To support this multiphased approach, a multidisciplinary team was assembled to provide the necessary skill sets. The team was comprised of practicing clinicians, clinical informaticists, research scientists, health IT system implementers, and interoperability/standards experts. This diverse team ensured that the assessment was robust and inclusive of all relevant areas impacting the ultimate feasibility of automating the QSRS process using EHR data.

2.2.1 QSRS Analysis (Task 1)

To support subsequent tasks of work and the overall feasibility assessment it was essential that the project team obtained a solid foundational understanding of QSRS. This involved reviewing all documents and information provided by AHRQ.

Objectives of this QSRS analysis included obtaining a working knowledge of—

- The QSRS Event Descriptions, Algorithms and Questions
- The complexity and nuances of QSRS
- The background and history of QSRS development
- The current QSRS process and workflows
- Possible challenges of implementing QSRS

The initial task of work involved obtaining baseline information on the QSRS system and ensuring that the project team had all relevant information to support the other tasks. This work involved reviewing documents provided by AHRQ which included, but was not limited to—

- QSRS Manual: Abstraction instructions and guidelines
- QSRS Event Descriptions and Algorithms: Flowcharts depicting decision logic for QSRS questions by module
- QSRS User Guide: Guide for administration and use of QSRS for CMS CDAC use
- Publicly available information regarding QSRS and MPSMS
- Demonstration of QSRS functionality and discussion of current workflow for manual data abstraction at CDAC

The in-depth review of the QSRS documents allowed the project team to have the working knowledge to fully understand the complexity and nuances of the QSRS system and to approach the other tasks of work with the appropriate knowledge and perspective.

2.2.2 EHR Physician User Feedback (Task 1)

Review of QSRS questions was were performed by a practicing hospitalist active in use of three certified EHRs. The goal of this activity was not to perform a comprehensive chart review or in-depth chart abstraction analysis, but to perform a review of QSRS algorithms and chart abstraction guidelines while collecting feedback from a physician who documents within multiple commercial EHRs and has prior experience charting in the Department of Veterans Affairs' VistA EHR and the Department of Defense AHLTA EHR. This activity involved using the QSRS algorithms and associated questions and considering the location of such information in the EHR and obtaining a preliminary sense of where QSRS data is located in these EHRs, who it is charted by, and in what format(s). This initial analysis provided a solid foundational understanding of applying the QSRS system to real-time review of records contained within an EHR, and combined with the results of the QSRS analysis and Environmental Scan, informed the approach and methods for stakeholder interviews and detailed vendor analysis.

2.2.3 Environmental Scan (Task 2)

Our team conducted an environmental scan of peer-reviewed and grey-literature publications to identify existing capabilities of electronic tools and methods for detection and surveillance of adverse events in health care settings, and the adverse events specified in QSRS.

A rigorous environmental scan of computational methods currently available to detect adverse events in hospital settings from electronic health records was conducted. The purpose of the environmental scan is to inform and determine the feasibility of applying computational methods to partially or fully automate the process of accurately populating queries from QSRS. The environmental scan included a search of both the formal peer-reviewed literature and the informal "grey" literature consisting of unpublished studies, white papers, industry documents, and relevant blog posts.

2.2.4 Stakeholder Interviews (Task 3)

The work completed in tasks 1 and 2 guided the strategy for identifying stakeholder groups and selecting the appropriate representatives from each group to reach out to, as well as the areas to address in the interviews. Our team determined that the following three market segments should be represented and included as part of this feasibility study:

- EHR vendors
- NLP/analytics vendors
- Providers/health systems

Semi-structured discussions were held with the representatives from each group. A formal Discussion Guide with topics and questions was used to provide consistency across the interviews and allow for analysis and synthesis of collected findings.

2.2.5 Automation Feasibility Analysis (Task 4)

QSRS Algorithm/Question-by-Question Analysis

The fourth and final task of work was completed in parallel with the third task of work and involved two main activities. The first was an in-depth and detailed analysis of each of the 205 QSRS questions, conducted module by module. This analysis was completed in two steps, the first being completed by the team's practicing physician and clinical informaticists. The analysis included identifying and documenting the following: primary chart location, charted by, format, and NLP analysis needed.

Once the first step of analysis was completed for each module, the module was transitioned to the research scientists who completed the second level of analysis specifically focused on NLP. This analysis included identifying the question type as well the level of complexity using the framework developed for this study.

In-Depth Vendor Analysis

Following the initial question review and analysis, the next activity involved working closely with two EHR vendors and one NLP/analytics vendor. The general findings from the environmental scan and the stakeholder discussions were combined with the findings from the question-by-question, module-by-module analysis of all 205 QSRS questions, and reviewed with the vendors to accurately determine what capabilities exist in current technologies, both EHRs and NLP/analytics, that support the equivalent of electronic abstraction and/or analysis of QSRS algorithm input data from EHRs.

3 QSRS Analysis and EHR Chart Review (Task 1)

3.1 QSRS Analysis

3.1.1 Overview

As noted above, the initial task of work involved obtaining a good understanding of the Quality and Safety Review System (QSRS) in terms of the specific algorithms and questions and the application of both to the chart abstraction process. AHRQ provided numerous QSRS reference documents that the project team reviewed and analyzed in detail. In addition, the project team participated in a demonstration of QSRS to fully understand the current workflow for manual abstraction. As part of this task, the project team included QSRS question review by a clinical user of multiple commercial and government electronic health records (EHRs) to acquire a solid foundational understanding of the complexity and challenges of QSRS.

3.1.2 Findings

QSRS is a patient safety surveillance system designed to detect adverse events in hospital medical records and provide reports on rates of those adverse events for the records reviewed. The scope of QSRS is broad, attempting to provide information on almost all adverse events that occur in hospitals ("all-cause harm"). For many selected events, QSRS provides additional detail beyond the occurrence of the event, e.g., not just fall rates in a specific population or set of hospital medical records over a given time period, but also data on falls that resulted in injury, and on each of several specific types of injury. The current QSRS system relies on human abstractors reviewing paper or electronic EHRs and answering questions that are automatically prompted based on age, sex, coded diagnoses, and procedures in the patient's medical record. The abstractors' answers to these and to previous questions are designed to indicate whether precisely defined adverse events occurred. (A system that the Centers for Medicare & Medicaid services [CMS] has used to produce patient safety data since 2002, the Medicare Patient Safety Monitoring System, operated by AHRQ and CMS, also depends on abstractors reviewing patient charts to produce patient safety data.)

The ability to transfer data from CMS Uniform Billing Form UB-04 into QSRS is already built into the current version of QSRS that is being used at the CMS Clinical Data Abstraction Center (CDAC). Work is also

underway to further develop this functionality in a version of the software that would be used in hospitals or health care systems using admission, discharge, and transfer [ADT] systems outside of the CDAC, and data formats other than the UB-04.

The essence of QSRS is to make definitions of adverse events precise, so that measurements from charts at one hospital will be equivalent to measurements at that same hospital over time, or to charts from a different hospital. The specific questions provided by the QSRS software to an abstractor are designed to drive predetermined definitions and result in a specific, detailed report regarding the occurrence of any adverse events during an inpatient stay. Hence the questions are very precise and need to be answered exactly as asked to populate the specified reports.

3.2 EHR Chart Review

3.2.1 Approach

To get a better understanding of the questions in QSRS and how the process of finding the answers in a chart would proceed, one of our initial undertakings was to consider the documentation approach in electronic format from various EHR systems and interrogate them with the same QSRS question process that is currently executed at the CDAC. In this approach, led by practicing physician hospitalist, at total of nine simulated scenarios of discharged patients from three different health care systems were analyzed. As the provider actively practiced at locations with three different EHRs, the physician reviewed the process that would be utilized for documenting as well as searching for information to support QSRS questions.

The QSRS algorithm was followed to simulate the chart review process that currently occurs. The main intent of this process was to get an initial understanding of what the QSRS questions were truly looking for, how long the process would take in general, and a preliminary assessment of where the answers to the questions could be reliably found using simulated or test patients.

Annotation of the preferred or primary chart locations of the likely answers and whether the locations were in structured or unstructured formats, and lab or radiology formats was undertaken. Lastly, special notes for certain specific questions were also recorded.

3.2.2 Findings

In general, an overall review of the chart prior to going through any of the specific QSRS questions was found to be useful. The familiarity achieved from an initial chart scan allows a reviewer to focus in on the most likely locations for answers and to answer some QSRS questions without requiring further review of the chart.

Nursing notes both in the initial assessment format and the subsequent "flowsheet" format were determined to be a useful resource for finding answers to many of the QSRS questions. The nursing note entries were also noted to almost always be in a structured format.

As expected, a majority of the question modules were never entered for any given chart because the entry questions were negative thus precluding further questions in that module. And lastly, as questions got more detailed and further down the algorithmic branch, the answers to those questions were less likely to be in a structured format and more likely to be found in free text portions of the chart — namely, the provider progress notes.

3.3 Certified EHR Technology Review

3.3.1 Approach

Upon completing the review of QSRS event descriptions and abstraction guidelines described in Section 3.3 and completing the EHR chart review described in Section 3.4, our team developed an initial understanding of both the most likely as well as potential sources of data for QSRS. The potential sources of data were reviewed alongside EHR functionality, structured data, coded data, document types, and information

exchanges available within EHRs achieving Office of the National Coordinator for Health Information Technology (ONC) 2014 Edition Certification and ONC 2015 Edition Certification. The goal of this review was to identify the ability of certified EHR technology to provide standardized datasets, documents, or interfaces to enable partial automation of the QSRS process.

3.3.2 Findings

EHRs certified to ONC's 2011, 2014, and 2015 Edition certification have the capability to complete a wide range of clinical functions in support of achieving meaningful use of EHR technology and participation in the CMS EHR incentive program. EHR certification and meaningful use functionality largely focus on the ability of EHRs to capture and exchange key clinical information, implement clinical decision support, perform computerized order entry, and support a number of interoperability requirements for health information exchange. EHR certification does not include requirements for electronic nursing documentation or physician clinical documentation of progress notes.

Given the number of QSRS algorithms with detailed chart abstraction guidelines that require data from nursing assessments, nursing flowsheets, and progress notes, it was determined that data available for reporting, data portability, and information exchange from certified EHRs would not be a sufficient data source for the QSRS algorithms.

One of the standardized datasets available using certified EHRs, the Common Clinical Data Set, provides summary information in a structured – and mostly standards-based and coded – format (Figure 3). However, the metadata needed to support QSRS algorithms such as the time medications were administered or the temporal or sequenced based nature of when a medication was administered (e.g., after a specific event) is not required within certified EHRs. The 2014 EHR certification includes a criterion for electronic medication administration record (eMAR), whereas 2011 and 2015 EHR certification do not. Standards-based documents such as the Continuity of Care Document (C-CDA), include a medication list with medications coded using the RxNorm terminology standard, but do not require medication order time or medication administered time, which is critical information to address QSRS algorithms and abstraction guidelines.

Figure 3. Certified EHR Technology Common Clinical Data Set

Patient Name	Vital Signs
Sex	Procedures
Date of Birth	Care Team Members
Race and Ethnicity	Immunizations
Preferred Language	Unique Device Identifiers
Smoking Status	Assessment & Treatment Plan
Problems	Goals
Medications	Health Concerns
Medication Allergies	Assessment & Treatment Plan
Lab Tests & Results	Goals
Vital Signs	Health Concerns

4 Environmental Scan (Task 2)

4.1 Methods

A rigorous environmental scan of computational methods currently available to detect adverse events in hospital settings from electronic health records was conducted. The purpose of the environmental scan is to inform and determine the feasibility of applying computational methods to partially or fully automate the process of accurately populating queries from the Quality and Safety Review System (QSRS). The environmental scan included a search of both the formal peer-reviewed literature as well as the informal "grey" literature consisting of unpublished studies, white papers, industry documents, and relevant blog posts.

The team conducting the environmental scan included expertise in the area of data science, with a focus on the application of computational methods to extract information from clinical text, as well as a background in healthcare safety. This section describes the methods that were used to conduct the searches of both the formal and informal literature sources, the process by which the relevant literature was reviewed, and the results of the environmental scan. The results focus on how current computational methods align with partial or full automation to populate QSRS. These results informed the subsequent tasks of the study which included direct review and analysis of widely used EHR technology and other health information technology (IT) computational solutions across multiple hospital organizations.

The literature search of formal, peer-reviewed publications was conducted using the PubMed and Medline databases and searching these databases using a set of predetermined keywords (see Task 2 Environmental Scan Report). The search was further constrained in two ways. First, the search was limited to articles published within the last eight years (2009–16). Second, the search was constrained to three patient safety journals and 23 high-impact–factor informatics journals (see Appendix). Each search term was used as a "keyword" search and "exact phrase" search. The keyword search returned articles that included any part of the search term in the title, abstract, or keyword section of the article. The search was then limited using the exact phrase search, which returned articles that included all parts of the search term in exact order in the title, abstract or keyword portions of the article.

From the keyword search, there were 1,354 matching articles from PubMed and 244 from Medline. The abstracts for each of these articles were reviewed by the research team, and the articles were coded as being potentially relevant or not relevant based on the content of the abstract. There were 109 articles coded as potentially relevant, and the full papers were retrieved. Upon further review of these articles it was determined that 9 of the articles were not actually relevant, leaving 100 articles for full review. These 100 articles were reviewed and coded based on the framework described below.

4.2 Coding Framework

A coding framework was developed to extract relevant aspects from each of the 100 articles. The coding scheme was based on our team's data science expertise and the general process by which computational algorithms are developed and tested. This process and the aspects of the articles that were coded are illustrated in Figure 4.. Each computational method described in the article was coded according to this framework.

From each article, we documented the *data source* that was used to develop the computational method. The sources include structured electronic health record (EHR) data, unstructured EHR data, and/or other sources of data (e.g., safety event reporting systems, FDA device reporting systems etc.). The data from the EHR then have to be *processed* in a manner that accounts for potential inconsistencies with the data, such as use of synonyms and misspelled words, and to ensure the data is in a format for efficient algorithms to be applied. We coded based on whether standard data processing approaches were applied such as tokenization, spelling, synonyms, and normalization or whether other more unique data processing methods were applied. Each computational method relies on a *feature extraction* process whereby the particular variables of interest are identified. We coded whether the feature extraction process was manual, which entails the research team selecting the particular variables, or an automatic approach using statistical analysis, or whether a hybrid approach was used. Finally, the actual algorithmic approach was coded, and the overall performance of the algorithms was documented. The algorithmic approach was coded based on whether a machine learning approach or a rule-based approach was used. In addition, it was also documented if the algorithm relies on an ontology, lexicon, or language system. The final stage is to implement the algorithm in a system like QSRS. While this component was not covered by the environmental scan, it will be addressed in subsequent tasks under the contract.

Figure 4. The General Algorithm Development Process and Resulting Coding Framework

14
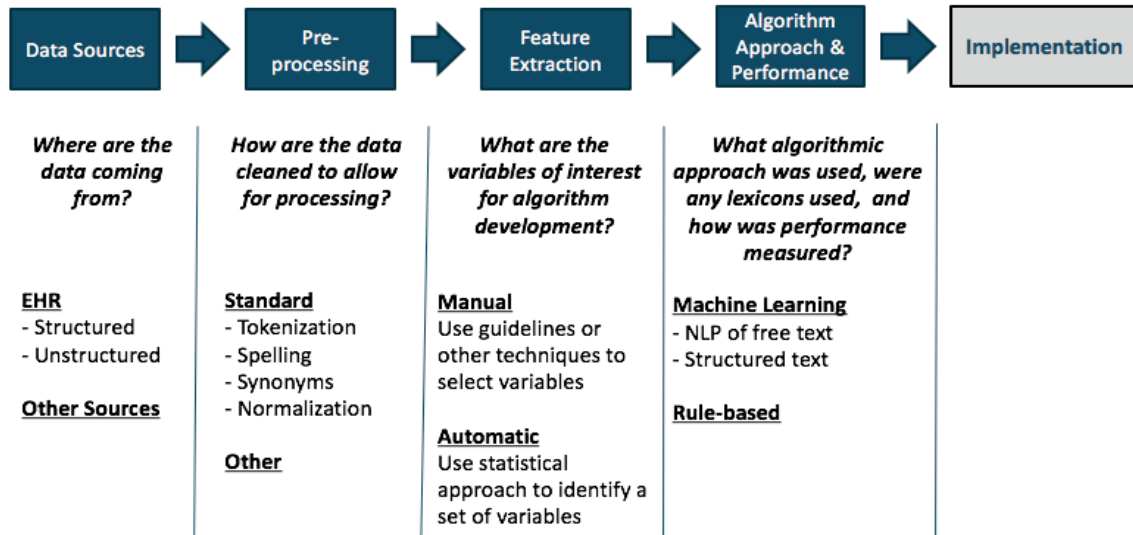
4.3 Grey Literature Search

The grey literature review was conducted using the Google Internet search engine to search the Web for resources using the predetermined set of keywords used for the peer-reviewed literature search. The grey literature search was constrained to results and studies published within the last 5 years. The search was limited to unpublished studies, white papers, industry documents, conference proceedings, relevant blog posts, online discussions, and trade/specialty specific journals.

As in the literature search of formal, peer-reviewed publications, each search term was used as a "keyword" search and as an "exact phrase" search. The keyword search within the search engine returned various resources that included any part of the search term in the web link, title, or within the text of the document and/or resource.

Each keyword search generated thousands of hits, with a search for "EHR Surveillance" in Google generating about 339,000 results. The top 50 results for each query, listed within the first 5 pages, were filtered for publication date, potential relevance of the title, relevance of the abstract sentence for the resource listed after the Web link, as well as the origins of the resource. Upon further reading and review of the most relevant of the top results, it was determined that approximately 40 of the sources would provide information and/or references relevant to the project.

4.4 Peer-Reviewed Literature Search

4.4.1 Overview

Of the 100 articles, 75 articles were coded as generally relevant and 25 articles were coded as specifically relevant to the topics addressed by QSRS, as shown in Figure 5. Thirty articles (30%) that were reviewed applied computational methods to EHR structured data, 39 articles (39%) used unstructured data, 21 articles (21%) use both structured and unstructured data, 7 articles (7%) used EHR data with data from other sources, and 3 articles (3%) did not state the data sources. When data sources other than EHR data were used, the most common sources were patient safety event report data, vaccine adverse event report (VAERS) databases and the FDA Adverse Event Reporting System.

Figure 5. Categories of Coded Articles

Articles	Count
Generally Relevant	75
Adverse Drug Events (ADEs)[2–16]	15
Hospital-Acquired Infections (HAIs)[17–20]	4
Surgery[21,22]	2
Pressure Ulcers[23]	1
Pneumonia[24]	1
Lab[25]	1
VTE[26]	1
Total	**100**

The coding of data processing revealed that 52 articles (52%) used standard data processing techniques, and the remaining 48 articles (48%) did not report their data processing methods or did not conduct data processing. None of the articles described atypical data processing techniques.

Feature extraction techniques were mostly a manual process (n = 46 articles, 46%) which often involve the research team developing the algorithm identifying the particular features of interest. Eleven articles (11%) used an automated process that uses algorithms to identify the features of interest. Twenty-six articles (26%) used both a manual and automated process, generally in an iterative fashion where algorithms are used to identify a set of features and these features are then manually reviewed. Fifteen articles (17%) did not stipulate their feature extraction process.

The most common algorithmic approach was rule-based (n = 48 articles, 48%). Twenty-seven articles (27%) used a machine learning approach, 22 articles (22%) used both machine learning and rule-based, and 3 articles (3%) did not state the algorithmic approach. Common algorithmic approaches included support vector machine, k-means, random forest classifier, conditional random field, logistic regression, Bayesian network, artificial neural networks, and k-nearest neighbor algorithms.

Forty-four of the articles used an ontology or language system to support the computational approach. Ontologies serve as a method to standardize language by providing a classification of terms in a standardized format. The most common ontology (n = 18 articles) was the Unified Medical Language System (UMLS), followed by seven articles using the systematized nomenclature of medicine (SNOMED).

Several articles used a language system as well. The language systems generally provide a framework for processing clinical text and include the ontology and the algorithms embedded in the system. Five articles used the Medical Language Extraction and Encoding System (MEDLEE), four articles using RxNorm (part of UMLS), four articles using the clinical Text Analysis and Knowledge Extraction System (cTakes), and three articles using MetaMap.

The comparison of performance measures across articles could not be completed given the tremendous differences in the types of performance metrics used and the methods by which performance measurement was conducted. However, in the section below, we describe general results regarding performance in the context of the types of questions in QSRS.

4.4.2 Literature Search Results in the Context of QSRS Queries

Within QSRS there are several different types of questions, and the type of question has an impact on the computational methods that would need to be applied for partial or complete automation of QSRS. Each of the different types of questions presents a different level of complexity to current computational approaches resulting in the likelihood that some questions may be successfully answered with a higher degree of certainty whereas other more complex questions may not have a strong computational solution at this time.

Our research team reviewed the QSRS questions and ranked the questions by the type of information that needs to be extracted to successfully answer the questions. Certain QSRS questions require the recognition of a single concept or the presence of a particular condition while other, more complex questions, require the extraction of multiple concepts, the determination of temporal sequence, contingencies, and "fuzzy" concepts, and open-ended responses. Figure 6, below, summarizes the different types of information and processing required for the various questions in QSRS and describes the computational challenges of automatically extracting the information. The questions are listed in order of increasing complexity in terms of computational approaches to answer the question, with the least challenging question at the top of the table.

Figure 6. QSRS Question Types in Order of Complexity With Associated Computational Challenges

Type of Questions	Example	Challenges
Presence of a concept/entity	Did the patient have a urinary tract catheter inserted during the stay? [GENERIC25]	Negation and synonyms
Numeric value extraction	During this hospitalization, did the patient have a PTT value greater than 100 seconds? [MEDIQ10]	Part-of-speech tagging and word sense disambiguation
Multiple concept detection	Which secondary morbidities developed? [PUQ7]	Co-referencing
Temporal occurrence of a concept and concept referencing	On or within the first 24 hours of admission, was a history of allergies and/or sensitivities documented? [MEDIQ1]	Defining temporal relations
Contingency	Did bleeding develop more than 24 hours after admission and within 1 day of [('PTT'>100 seconds) OR ('Protamine administration') OR… [MEDIQ35]?	Co-referencing
"Fuzzy" concepts	Did the patient undergo an **<u>unplanned</u>** transfer to a higher level care area within the facility or to another facility? [OTHERQ13]	Reasoning and Subjectivity
Open-ended responses	If at all, describe how the device harmed the patient. [DEVICEQ6]	Summarization

(left margin vertical label: Increasing Algorithm Complexity)

The challenges include identifying negations and synonyms, part of speech tagging, co-referencing, defining time periods, reasoning with information, subjectivity, and summarization. Each of these is described below:

- **Negation**: Negation is the ability for algorithms to distinguish between negated and positive concepts. Negation detection is a necessary step to correctly identify the presence of a concept or condition and involves identifying modifying words or phrases (such as "was not seen," "no") around the condition of interest. For example, an algorithm to identify the placement of a urinary tract catheter will have to be able to categorize sentences such as "urinary tract catheter *was not placed*" and "patient *previously had* urinary tract catheter infection" as negative occurrences.
- **Synonyms**: It is important for algorithms to identify when synonyms for the same concept are being used. Ontologies and dictionaries are often used to identify synonym words or concepts in text. This is particularly relevant for medication names where drugs can be described by both generic and brand names. While several systems have been built to assist in the identification and matching of concepts, this process still requires careful consideration particularly to incorporate colloquial or short hand expressions of concepts.[27,28] Use of synonyms, shorthand expressions, and other colloquial concepts differ between hospitals and provider types (e.g., hospitalists vs. gastroenterologists).
- **Part-of-speech tagging**: The ambiguous use of words is a particularly challenging problem when words in sentences take different functions. For example, the sentence "Sam saw the man with the telescope" can be interpreted either as Sam saw the man by using a telescope or that Sam saw a man carrying a telescope. Additional context in the text is often required to disambiguate these types of sentences.[29] Part-of-speech tagging, or word sense disambiguation, marks words or word phrases by their part of speech based on context in which the word is used.

- **Co-referencing**: Identifying co-references is the process of determining if two or more concepts are referring to the same or equivalent concept. For example in the sentence "She was scheduled to receive a *temporal artery biopsy*, but she never followed up on *that testing*," the "temporal artery biopsy" and "that testing" are both referring to the same test.[30] Correctly identifying co-references and equivalent relationships is important particularly for questions that require extracting and relating multiple concepts and conditions.
- **Defining temporal relations and reasoning**: Correctly answering questions that rely on temporal sequences require the proper extraction and modeling of text that can span various clinical narratives. Answering temporal related questions about concepts and conditions require the ability to first model and define clinical events or concepts ("admission," "transfer," "surgery," "tests," "treatment," etc.), which are referred to as the "EVENT," and temporal expressions (dates, time, duration, and frequencies), which is referred to as "TIMEX." Next, temporal relations, or temporal links, TLINK, are used to define how two or more combinations of EVENTs and TIMEXs are related in clinical text. EVENTs and TIMEXs can be connected by different TLINKs such as "before" (patient was given medication prior to surgery), "simultaneous" (patient's blood pressure was high on admission), "overlap" (patient had fevers and chills), "ended by" (his IV was disconnected on 07-17-13).[31]
- **Reasoning/subjectivity**: Questions that require "additional reasoning" will be difficult for algorithms to process. For example, a question that asks if an event is "unplanned," without specifically defining "unplanned," requires the coder or algorithm to determine what an "unplanned" event entails. Without the usage of specific words ("unplanned," "unexpected," etc.) in the description or specific algorithm rules, automating the response to these questions is complex since the algorithms have to determine what the "fuzzy" concept represents.
- **Summarization**: Summarization is a difficult and unsolved problem in natural language processing research. There are two primary ways to summarize text: text extraction or language generation. Text extraction involves identifying key words or sentences that best represent a body of text. Language generation uses the structure of text to generate new sentences. While practical uses of the first method are more common than the latter, both require significant training and testing and are often context specific.[32]

To address many of these challenges a National Institutes of Health (NIH)-funded research center called the Informatics for Integrating Biology and the Bedside (I2B2) has conducted numerous research competitions with the goal of developing possible solutions. The competitions have led to many promising outcomes from several different research groups and these solutions can be leveraged for purposes of partially automating QSRS.

Topics addressed by I2B2 include medication extraction, identifying relations, co-referencing, and temporal relations. Although some topics may be more complex than others, I2B2 and their research competitions show that potential solutions exist to these topics and can be addressed in the current environment. These areas being addressed by I2B2 are important to this project as it highlights the current complex state of automation. In addition, I2B2 has developed a suite of open source software products as a result of their research and competitions. Many of these open source software suites review specific areas such as natural language processing and how to pull from specific sections of documents within the electronic record. The I2B2 findings and software solutions could be adapted to many different use cases and customizing to fit the needs of the electronic record.

4.5 "Grey" Literature Search

4.5.1 Overview

A review of informal (i.e., not peer-reviewed such as trade publications, news articles, or op-eds) "grey" resources was performed to ensure a broad scope and review process that accounted for industry-led efforts outside of research-funded institutions and academia. Outside of peer-reviewed journals and publications, our research team looked to grey resources to provide information and literature produced

by various stakeholders including academia, government, business and industry. Google searches using the project relevant keywords produced hundreds of thousands of hits ranging from topic-specific scholarly articles, government and public health agency (Centers for Medicare & Medicaid Services [CMS], Centers for Disease Control and Prevention [CDC]) Web pages, payer, health IT vendor marketing materials, and relevant health IT news and blog posts.

To provide a glimpse into the information available with a simple Google search using the key word, "EHR data" produced 16.4 million results, while "EHR Surveillance" produced 335,000 hits. From looking into grey-literature resources, it was apparent that academics, health system executives, Federal entities, public health organizations, and health IT industry vendors are all highly invested in leveraging data from electronic health records.

Since the enactment of the HIT for Economic and Clinical Health (HITECH) Act of 2009 and the availability of Federal incentives for the adoption and use of health IT, there has been a spur in innovations, solutions, and services to leverage data to improve efficiency, efficacy and outcomes. As more data are captured, the value of this data is starting to be realized as the landscape of health care moves towards keeping patients healthy, improving quality of care and value-based payment models.

The shift from fee-for-service to payment models based on quality and value of care is also spurring the need to define baselines, capture trends, and benchmarking to relay progress and outcomes. The CMS EHR Incentive Program incentivized hospitals to not only accelerate EHR adoption, but to use the technology to improve health care delivery and outcomes through the flow and exchange of health information to minimize gaps in care and highlight needs.[33] A culture of data use for quality measurement, improvement, and reporting is rapidly becoming commonplace, with the expectation that EHR data will be mined, analyzed, and used to inform care delivery and improve patient safety.

4.5.2 Quality Measure Reporting

To qualify for upcoming Meaningful Use Stage 3 incentives, eligible professionals are required to use certified EHRs with the capability of reporting adverse events to Federal agencies and public health organizations including the FDA and the CDC.[34] Clinical quality measures (CQMs) are tools that measure and track quality of healthcare services provided by eligible professionals, eligible hospitals and Critical Access Hospitals have expanded beyond manual abstraction to electronic capture, calculation, and transmission using certified EHR technology. Providers are required to submit electronic CQM (eCQM) data from EHRs to receive incentive payments for Meaningful Use.[35] While the incentives for extracting quality data out of the volume of data available in EHRs is immense, so is the need to automate systems to sort for quality data and to reduce the burden of data collection. Incidentally, one of the biggest challenges in realizing the potential of EHR data is in abstracting the information from disparate systems with varying architecture and standards.[36]

4.5.3 Public Health Reporting and Surveillance

Traditionally, public health agencies collect health information to prevent outbreaks, analyze public health trends, and educate the population.[37] Providers reported to public health agencies, using paper or electronically, through registries. The public health incentives in the CMS EHR Incentive Program is driving the development of public health infrastructure and use of health IT to more rapidly report information to agencies. There is the move from paper-reporting and unidirectional electronic reporting to bidirectional data exchange. Meaningful Use is increasing the volume of public health data collected as it is required, as part of compliance, for eligible providers and hospitals to submit immunization information, electronic lab results, syndromic surveillance, and reporting to cancer registries and specialized registries. [37]

To gather useful information, the move towards interoperability and the adherence to nationally recognized standards and certifications criteria attempts to reduce the challenges of the many disparate health IT tools, technologies, and EHR systems. To improve the quality of data and standardize and garner the most use of available data, vendors, industry and academics alike are developing new innovations, tools, and service offerings that can be used for surveillance and Federal public health reporting requirements.

The CDC Center of Excellence in Public Health Informatics at the Harvard Pilgrim Health Care Institute developed the ESP (Electronic Medical Record Support for Public Health), which is an open-source disease surveillance software for extracting data from any modern EHR system for analysis. Each evening, data files are transferred from Atrius Health's Epic EHR to ESP; the data are a predefined subset of data fields from EHRs. These data are then downloaded to the ESP database where detection algorithms are run and cases of interest are identified, including adverse events.[38,39] ESP runs behind the host practice's firewall and only permits external users to run approved retrievals or analyses.[39] ESP automates disease detection and reporting by extracting structured data from EHR systems and organizing them into a standard format, mapping them to disease categories using algorithms, and analyzing and transmitting the data to public health agencies via secure HL7 messages.[39]

Vendors are utilizing analytics to leverage the most of EHR data collection and support quality reporting and surveillance. Realizing the potential of EHR data in surveillance is evident as cloud-based vendors have the capability to compile summary statistics in real time from the available EHR clinical documentation from physicians real-time.

4.5.4 NLP Technology

Due to the diversity in EHR systems and continued evolution of standards, stakeholders are working to overcome the challenges of abstracting key data elements from the EHR. While content and data standards for structured and coded data for problems, medications, allergies, lab results, immunizations, and procedures are commonplace within certified EHRs, much of the needed data for medical coding, research, and patient safety surveillance such as QSRS adverse event detection is contained within clinical notes. Getting the data needed from EHRs can be complicated as EHRs not only provide multiple areas to enter data, but also allow for free text to be entered. Academics and vendors alike are moving towards utilizing natural language processing (NLP) and data mining tools to convert free text into computable formats. More and more, EHR vendors are incorporating analytics platforms and reporting data warehouses into their EHR solution offerings.[40]

The global health care NLP market has been forecasted grow from $1.10 billion in 2015 to $2.67 billion in 2020.[41] The NLP market is driven by the tremendous increase in unstructured clinical data accounting for 80 percent of data in EHRs.[42] NLP technology leverages a patient's EHR information to convert free-text notes into structured data, parses clinical notes to extract diagnosis and billing codes, and augments clinical decision support by analyzing references in response to clinical queries.[43] Using algorithms to extract data from free-text and auto-coding medical record data, NLP has the potential to help improve patient care by analyzing, normalizing, and aggregating patient data that may be documented in various locations within a chart, and may use different terminologies, formats, colloquialisms, or shorthand expressions. In providing more accurate and uniform information in a centralized location, NLP can also leverage EHR data to facilitate quality and safety monitoring.[44] Bio-surveillance studies at the Mayo Clinic showed that NLP, when used to evaluate an entire encounter note, provided superior data to when data was extracted from the chief complaint field alone.[43]

Various practices are being applied and tested to extract quality data from an EHR. For researchers participating in the Academy Health's Health IT for Actionable Knowledge project, extracting quality data is a priority. In an effort to cultivate "research-grade" data, participating institutes have an established IT infrastructure that allows research departments to set up and manage separate data warehouses that allow researchers to clean data for research and surveillance purposes.[45] Entities, including the University of Michigan Health System and the Parkland Center for Clinical Innovation, among others, are also using NLP technology to gather data from multiple sources including laboratory, EMR, diagnostics, ePrescribing, and claims systems to reduce false positives based upon data collected only within a single system.[36]

Whereas early use of NLP was spearheaded by academic research organizations discussed within the Peer-Reviewed Literature Review section of this report, a growing marketplace of vendors accessible to community hospitals and standalone hospitals (vs. large integrated delivery networks or academic medical centers) has been emerging to bring capabilities in computer-assisted coding and facilitation of clinical documentation improvement beyond the research community. Leading vendors of NLP technology are

working to innovate and improve their technologies to facilitate meaningful analytics to identify patterns, determine risk and quality, and transform data into valuable information.

4.6 Discussion and Recommendations

The environmental scan yielded a rich overview of the current computational methods for extracting information to identify adverse events from electronic health records as well as other data sources. Both the searches of the peer-reviewed and grey literature sources point to extensive use of algorithmic approaches to process information found within the EHR and a field that is advancing with increased investments for research and increased investments by private sector companies to develop commercial products.

4.6.1 Limitations With Current Studies

The peer-reviewed literature pointed to limitations with current studies that are applicable to the automation of QSRS. First, few studies applied developed computational algorithms across multiple provider sites with different EHRs. Consequently, little information can be gleaned from the peer-reviewed literature on the ability to have a set of algorithms broadly applicable to different providers that are using different EHRs. In addition, each of the studies reviewed by the research team reported accuracy of their algorithmic approach using different measures and techniques that prevented direct comparison.

4.6.2 Feasibility of Automating QSRS

Based on the environmental scan and an assessment of the QSRS questions, it is clear that there is variation in the types of questions included in QSRS, and this variation will impact the ability to successfully apply computational methods for partial or full automation. The QSRS questions that seek to identify a single concept are well-suited for partial automation using computational methods. Similarly, questions that seek to identify a single value are also suited for partial automation. However, many of the questions in QSRS are more complex than single concept identification and numeric extraction. Many questions require the identification of multiple concepts, identification of contingencies, identifying temporal sequences, reasoning with subjective concepts and open-ended questions. The more complex questions present challenges to many of the computational methods currently available. Consequently, if current algorithmic approaches are applied the results may not meet the desired accuracy standards. There are, however, several different researchers working to develop solutions to many of the current challenges as evidenced by the NIH-sponsored I2B2 informatics center.

4.6.3 Applying Commercial Solutions

The availability of EHR-vendor led and independent health IT solutions designed to analyze both structured and unstructured EHR data is highly applicable to the notion of automating QSRS. Configurable and advanced rules engines that populate EHR-vendor data warehouses should be evaluated to determine ability to automate QSRS algorithms. For analysis of unstructured text, NLP solutions that have historically been used to assist coding efforts that provide a side-by-side view of the EHR source text from which a coding determination was made, may be configured to support QSRS abstraction from EHR data.

4.6.4 Implications for Feasibility Study

The findings from the environmental scan were used to determine the appropriate vendors and stakeholders to interview and analyze in subsequent tasks within this study, as well as to develop the topic areas to focus discussion questions on. Questions included—

1. Are there any (partially) shared data frameworks across vendors?

2. How much variation is there within vendor products?
   – Different implementation and customization
   – Differences in local use of product

3. How can a complete record be transmitted electronically?

4. How have third-party NLP tools been applied?
   – How much variability is there across different EHR vendor solutions?
   – How much variability is there across different implementations of the same vendor EHR?
   – Coding abstraction applicability to QSRS

5. How does local practice variation across hospitals affect the availability of both structured and unstructured data and the ability to isolate and identify QSRS data sources?
   – Are multiple approaches required for each QSRS adverse event?
   – Is there alignment with Federal and payer-led initiatives in a consistent set of standardized data that can reduce abstraction time and enable processing of higher chart volumes?

4.6.5 Additional Reference Information

As a companion to this section, originally submitted in November 2015 as part of "Task 2 – Environmental Scan Report," the following documents and supporting information were provided to further inform and support the QSRS Automation Feasibility Study (AFS):

- **Appendixes:** Journals researched, commercially available NLP vendor solutions, citations, and references in this report are provided in Section 6: Appendixes.
- **Environmental Scan PowerPoint Presentation:** Presented to members of the AHRQ Center for Quality Improvement and Patient Safety (CQuIPS) on November 23, 2015, this PowerPoint presentation summarizes the QSRS AFS Environmental Scan Report, and provides additional background on machine learning algorithmic approaches and commonly used language systems. An updated, final version of this presentation was delivered on November 25, 2015.
- **Coded Articles Analysis:** The coding scheme applied to articles reviewed and tagged as relevant is provided as an accompanying set of reference materials for this report.
- **Articles Reviewed:** PDF copies of articles retrieved, reviewed, and tagged as relevant are provided as an accompanying set of reference materials for this report.
- **Draft Research Plan:** Presented to members of AHRQ CQuIPS on October 5, 2015, delivered on October 8, 2015, and revised on November 10, 2015, detailed the methods and approaches planned in conducting this Environmental Scan Report.

5 Stakeholder Findings (Task 3)

As previously mentioned, the third and fourth tasks of work were completed in parallel following the completion of tasks 1 and 2. Initially these tasks of work were to be conducted sequentially; however, as the project team conducted tasks 1 and 2 and developed plans for tasks 3 and 4, it was determined that the two tasks were complementary and should not be completed separately.

The third task of work involved identifying stakeholder groups and conducting interviews with representatives from each group. The fourth and final task of work involved an extremely detailed analysis of the QSRS system in the context of two identified electronic health record (EHR) developers and one natural language processing (NLP)/analytics vendor.

5.1 Stakeholder Selection, Outreach, and Discussions

The primary objective of task 3 was to gain additional information on existing relevant capabilities as researched in the Environmental Scan, such as the ability to identity adverse events in health care settings in general, and the types of adverse events specified in Quality and Safety Review System (QSRS) in particular using electronic systems. Although the project team was limited to speaking with no more than nine stakeholders for this task, publicly available sources of information were researched and secondary research (such as attendance at industry conferences and meetings related to the topics of EHR data for patient safety reporting and surveillance) was conducted as needed.

The work completed in tasks 1 and 2 guided the strategy for identifying stakeholder groups and selecting the appropriate representatives from each group to reach out to, as well as the topic areas to address in the interviews.

Based on the project team's deep experience implementing and optimizing EHRs with hospitals and health systems, there was a keen awareness of the inherent variability that exists within product readiness and strategies in EHR and health information technology solutions that support different market segments.

Three stakeholder groups were identified as integral to this feasibility study:

- **Health systems/hospital providers:** This stakeholder group – end users of EHR vendors and analytics vendor products considered above – was evaluated to develop an understanding of the real-world feasibility of implementing vendor-defined – and vendor marketed – capabilities. Provider organizations also provided insight into their own initiatives for adverse event (AE) detection and surveillance, lending information about level of effort and effectiveness of EHR vendor and analytics solutions.

- **EHR vendors:** To gain an understanding of EHR standardization, configurability, clinical decision support, and reporting architecture, and alignment through base and expanded solution offerings that align with QSRS modules and algorithms.
- **NLP/analytics vendors:** Given the continuing maturation of use of NLP for analytics of health data and advanced analytics and offerings from "big data" vendors, a more detailed review of vendor offerings to address QSRS algorithm data sources within text and narrative fields within EHRs was conducted.

Speaking with representatives from these three stakeholder groups provided an inclusive overview and understanding of both the capabilities of, and needs from, vendors to support partial or full data abstraction from EHRs. For each stakeholder group identified, a short list of possible organizations was created and reviewed. To determine which three organizations would ultimately be chosen to reach out to and hold discussions with, criteria included the following:

- Market share representation
- Diversity among stakeholder group
- Likelihood of engaging stakeholder and obtaining participation

5.1.1 Providers/Health Systems

It was essential to speak with providers in the health care setting. This provided the opportunity to identify and understand current capabilities in use for adverse event reporting and tracking. It also allows for the comparison between available vendor capabilities and what is ultimately implemented and used at the end-user level.

The three provider/health systems selected to represent this market segment were—

- Provider A
- Provider B
- Provider C

These three health systems are extremely diverse in their size, locations, patient populations, and EHRs, providing as wide-ranging and inclusive a representation of the provider community as can be obtained by speaking with only three organizations.

Provider A

Provider A is a large not-for-profit health care system in the South, and includes nearly 50 hospitals, nearly 1,000 patient care sites, more than 6,000 active physicians, over 40,000 employees, and a health care insurance provider.

As a recently merged health system, Provider A is representative of many other health systems across the country that have multiple EHRs and are in the process of integrating those systems. Several EHR systems are in use at the health system, including one predominant EHR within the inpatient areas and another EHR within Emergency Departments. The remaining hospitals are using another EHR, which the outpatient facilities are transitioning to as well. Some of Provider A's specialty practices use other EHRs.

Provider B

Provider B is a not-for-profit health care organization in the eastern United States serving approximately 1 million residents within its region. Provider B includes 6 hospitals, including academic, community, and critical access hospitals, over 200 practice sites, 700 physicians, and over 12,000 employees. Provider B has been using its EHR since 2008 and is fully live on all modules across the health system.

Provider C

Provider C is one of the largest integrated health care systems in the United States consisting of medical centers, outpatient clinics, community living centers, and other support centers.

Provider C uses its own EHR to provide an integrated inpatient and outpatient electronic health record for patients. It offers enhanced capabilities and flexibility to adapt to health care and technology innovations, and continually improve health care. Provider C possesses an extensive health care informatics team that estimates over 2 petabytes of data for analysis and analytics.

5.1.2 EHR Vendors

Discussions with EHR vendors were intended to shed light on capabilities the vendors are currently providing in terms of adverse event reporting in general, as well as in particular to what QSRS specifically tracks. Conversations with the EHR vendors also provided insight into what customers (i.e., providers/health systems) are looking for/requesting in terms of AE tracking and reporting. In addition, the EHR vendors provide insight on the possibility of abstracting data from their EHRs as well as the associated barriers and challenges to doing so.

To be representative of the EHR market while only speaking with three vendors, we prioritized vendors representing the largest or expanding market share serving enterprise hospitals and the midmarket setting. In particular, the vendors (EHR A, EHR B, EHR C) were selected for their expanding installation bases.

5.1.3 NLP/Analytics Vendors

The third market segment identified as essential to this feasibility assessment is that of NLP/analytics vendors. Discussions with these groups were intended to shed light on what capabilities the vendors are currently providing in terms of data abstraction and how the available technologies could be applied to QSRS. Conversations with the NLP/analytics vendors also provided insight into what customers (i.e., providers/health systems) are looking for/requesting in terms of data abstraction and analytics.

To be representative of the NLP/analytics market while only speaking with three vendors, vendors that provide NLP offerings ranging from computer-assisted-coding to predictive analytics to Patient Safety Organization support and surveillance were selected (Vendor A, Vendor B, Vendor C). These three distinct focus areas for NLP were identified as each vendor brings a unique approach and service offering – and therefore associated algorithms and capabilities – that could be leveraged to apply NLP to EHR data in support of automation of QSRS.

5.1.4 Process, Outreach, and Discussions

While the individual stakeholder organizations were being identified, the project team worked with the AHRQ team to develop the appropriate outreach and communication strategy. To ensure that the project was appropriately represented to the stakeholders, a one-page project summary was created. This document was reviewed and approved by the AHRQ project team and signed by the AHRQ and project team leads. This document provided a standardized overview of the project in terms of scope and objectives, in particular those which were to be addressed during conversations with stakeholders. In addition to the project one-pager, a standardized outreach email was created to ensure that accurate information was consistently communicated to stakeholders.

Outreach communications were managed by the project's project manager to ensure that communications were consistent, timely, responsive and otherwise appropriately managed. Conversations with stakeholders were set up as Web-based conference calls to allow for screen-sharing as needed, be it by the project team or the stakeholder being interviewed.

Prior to the conversations being held, a formal Discussion Guide was developed. This Discussion Guide went through several iterations with AHRQ review and input. The Discussion Guide consisted of topics and questions that provided consistency across the interviews and allowed for analysis and synthesis of collected findings.

Semi-structured discussions were held with the representatives from each group, using the Discussion Guide. Calls were approximately 1 hour in length and were attended by a minimum of one project team member, but usually at least two individuals. Notes were taken in real time during the conversations. If additional information was identified as being needed after the calls, email communications were used.

5.2 Results Overview

The conversations with the nine selected stakeholders from the three stakeholder groups provided great insight into the current state of capabilities related to abstracting data from EHRs in general and as it relates specifically to QSRS and the AEs addressed.

5.2.1 Providers/Health Systems

The health systems we spoke with reported similar experiences and challenges with AE reporting and tracking. In general, they are doing very little related to AEs within their EHRs.

AE tracking is documented in systems separate from the EHR(s) in a manual process that requires a voluntary action by the end-user. Providers A and B use two distinct systems.

None of the health systems reported using AE tools provided by the EHR vendors. One health system expressed interest in using EHR vendor tools, but indicated that deviations taken in the organization's EHR system's build make it difficult to take advantage of such tools. One health system reported not being aware of any AE tools available by the vendor(s).

For Provider B, the process for implementing content and design changes (in general and specifically related to AEs) involves first looking at what EHR A provides and determining if (1) they can actually use it (due to their customized build) and (2) if it's content that they find useful. For example, they are currently building out Nursing Dashboards, the first being for catheter-associated urinary tract infections (CAUTI) and the second being central line-associated blood stream infections (CLABSI). They looked to see what EHR A provides, but found it not to be useful (because only provides CAUTI rate) nor actionable for their needs. Since they cannot use what is provided by EHR A, this requires that they use internal resources (which are limited) to build new content. They would like to build additional dashboards for pressure ulcers, falls, and others, but this requires a full-time dashboard resource, which they do not have the staffing or financial priority to acquire or train for the use of at this time.

For Provider A, in addition to documenting AEs within an adverse event tracking application, the organization uses a manual process for AE tracking within the EHRs, using a classic trigger tool method. This involves retrospectively randomly sampling 1–3 percent of charts and reviewing each for no more than 20 minutes to flag AEs. The events are scored in terms of preventability and compared against what is voluntarily reported in their adverse event tracking system. The biggest barriers of this process are cost and credibility with only 30 charts a month when the organization has thousands of patient visits within a month. Provider A believes having an automated process, provided by the EHR or another technology, for this would be highly beneficial.

For Provider B, when AEs are entered into their tracking system, the health analytics team receives this information to follow up on. The team reviews the occurrence, completes a "deep dive" into clinical information within EHR B to perform a root-cause analysis and determine how to fix/prevent the issue in the future. This is a reactive and retroactive process rather than proactive—which they would prefer. Provider A also reported that they are working with an adverse event tracking system vendor for more real-time monitoring, using structured data.

In terms of using NLP technologies, neither health system is currently doing so. Provider A looked into using NLP tools to flag triggers but did not pursue. For Provider B, they are just starting a project with Virginia Tech's School of Medical and their research pathology that utilizes NLP to analyze pathology reports.

In terms of the specific Adverse Events tracked by QSRS, in looking at the EHRs and the data that is included, Falls is the most commonly/likely to be built, documented and tracked in a way that could possibly be abstracted (without NLP or other analytics). For Provider B, the Nursing Falls Assessment uses SmartText which contains data elements that could be pulled. That said, as Progress Notes are still often used for documentation – even when not required – relevant data is not always in the available, structured locations, and therefore would not be captured by non-NLP processes.

In another example from Provider B related to QSRS specifically, while medication administration is documented in a structured method in the eMAR, related AE information is not documented in that location, but rather is done in a Progress Note in a non-structured format. This means that it is not possible to pull medication AE information from a standard or custom report (or other non-NLP process).

In speaking with both health systems, while it is helpful to be able to pull reliable, accurate retrospective reports, ideally it is preferred to have real-time reports with information and data that can be acted on while the patients are still in the hospital. Both health systems reported overall data challenges. These include—

- Data captured in multiple locations
- Data not structured (especially for physicians/providers)
- Lack of trust in accuracy of data, reports
- Workflow, documentation variability
- EHR customization prevents usage of standard reports, dashboards, and clinical decision support rules and alerts
- Hard to get consensus on clinical decision support rules, alerts

Other challenges that both organizations reported included those related to lack of resources to—

- Build reports, dashboards
- Educate end-users

When provided with an overview of QSRS and its intended use both at the CDAC and in pilot for use within health systems, health systems expressed concern regarding implementing QSRS for surveillance without significant automation. Health system representatives cited the very limited profit margins—often under 1 percent—within hospitals and health systems and challenges in obtaining resources to support configuring needed reports and dashboard within their EHRs. Coding resources are particularly scarce, and health systems did not foresee the availability or value in diverting limited resources to review EHR chart data for surveillance without external funding to support the ability to perform a study with a meaningful number of charts.

As noted above, health systems that have made significant investments in EHR technology are now directing their limited resources to optimization of that technology and associated workflows. In reviewing research, working with quality improvement efforts, and working with their EHR vendors, they find a significant number of best practices and lessons learned in implementing improved documentation, more actionable alerts, and provider training in support of prevention of AEs. Health systems are prioritizing resources and funding to implement near-real-time alerts and reporting and implement predictive analytics to proactively identify potential adverse events, manage severity, or prevent avoidable readmissions.

Health systems also expressed concern over additional potential Federal reporting requirements and fatigue from measurement overload. Participants inquired whether our project was a predecessor to future Federal reporting requirements and were informed that the project goals are focused on automation feasibility and not part of upcoming requirements. In reviewing select QSRS modules with the health systems and seeking feedback for feasibility of automation, we found some of the evident challenges and barriers to automating QSRS using data from their EHRs include—

- Lack of structured data elements in general/overall
- Lack of required fields that correspond to QSRS questions
- Multiple locations to document same information (which is not linked, built the same way, and often charted by multiple provider types)
- Significant documentation completed in Progress Notes (narrative, unstructured text)
- Variability in end-user documentation and related workflows
- Inability to take advantage of vendor standard build, content, and reporting architecture due to health-system specific customization or legacy content

- Data contained within nursing flowsheets or nursing assessments which have health-system specific value sets rather than a vendor-provided or standards-based value set which may be difficult to normalize or standardize across hospitals and/or vendors

While some of the above challenges could be addressed by changes to the EHR in terms of design and build (e.g., more structured data fields, more required fields, etc.), as long as there are still multiple locations to document, many of which are free-text Progress Notes, it is not likely that QSRS could be automated without the use of NLP or other data analytics technologies.

Clinical informatics representatives from provider organizations further noted that the numerous options and efforts supporting prevention of certain adverse events addressed by QSRS (such as pressure ulcers, CLABSI, and CAUTI in particular) have already established quality improvement programs and industry or vendor best practices for workflows and documentation practices using their EHR. Vendor reporting modules provide the ability—when implemented—to provide quick and near-real time information for decisionmaking at the point of care. Implementing abstraction-based or post-discharge surveillance monitoring systems for these adverse events did not seem to yield value. A review of QSRS modules to identify adverse events with the least amount of maturity in terms of available EHR reporting and decision support rules that have a significant quality or cost impact may be beneficial to prioritize for either chart abstraction or automation.

Provider C abstracts approximately 3 million charts a year using manual chart abstraction. Informatics leaders interviewed expressed the same desire regarding automation as noted as the goals of this study. Provider C has implemented NLP for analysis of two different data types: ejection fraction and microbiology. The team reported microbiology data to be very difficult to analyze and would not recommend it as an area to begin NLP due the text-based nature and complexity of the data. Provider C's team expressed that automation and NLP would be less complex across a single site where documentation may be similar, but as each Provider C site has its own nuances and system implementation customizations, automation requires a significant amount of computing power and resources.

Provider C cited major challenges for NLP for adverse events to be that adverse event language resides in multiple locations (e.g., physician notes, nursing notes, respiratory therapy notes, laboratory reports, etc.), where each location has its own defined (or undefined) vocabularies. The Provider C team indicated the technology for abstraction automation is improving, but estimated 10–15 years for it to reach full automation.

Quality Reporting and Chart Abstraction

As the focus of this study was the feasibility of automation of QSRS through use of EHR data, the study did not analyze in depth the data contained in other hospital systems and data sources that are not automatically (i.e., electronically) populated with the EHR as the data source. Examples of systems not analyzed for this report included, but are not limited to: voluntary adverse event/safety reporting systems (e.g., RL systems, Quantros) core measures/quality abstraction systems (e.g., Xerox Midas+, 3M SoftMed, Dolbey), and data aggregators for analytics and measurement (e.g., UHC/Vizient, Quintiles).

Most health systems utilize clinical nurse abstractors in support of The Joint Commission and CMS Core Measures, and hospital specialty accreditation, advance certification levels (e.g., stroke, trauma, other disease-specific certification). However, these resources are typically housed within Performance Improvement (PI), Quality Improvement (QI), or Risk departments. Providers interviewed indicated that although EHRs may offer useful data for proactive and day-to-day management of patient care, government and various certification/accreditation programs dictate the use of manual chart abstraction. All providers interviewed expressed an interest and desire to reduce the reliance on manual chart abstraction and welcome partial automation approaches to reduce the manual processes and need to staff high-cost/specialized resources. (Generally, hospitals utilize clinical teams such as nurses, and not coders, for chart review and abstraction for these purposes.)

Another key finding was the distinct segmentation of QI/PI departments and personnel from clinical informatics and EHR build/optimization teams. With the emergence of Chief Nursing Informatics Officers(CNIOs), there has been an increase in coordination between EHR teams and QI/PI. CNIOs and physician IT/informatics leaders reported rarely or never seeing reports derived from chart abstraction from QI/PI teams. Health system safety leaders indicated their data sources to be solely voluntary reporting systems, and not EHR-generated reports.

Some hospital quality teams perform chart abstraction for charts flagged for abstraction by their data providers. For example, some solution partners analyze claims/billing data from member hospitals, and identify charts for abstraction for measurement programs. This type of partial automation—triggered by claims/billing—could be an approach for consideration by AHRQ for QSRS partial automation. As certain diagnoses (e.g., pressure ulcers, CAUTIs) may be coded by billers post discharge, the availability of coded billing/diagnoses data that is interfaced and populated post coding within the EHR, can be used as a trigger to identify charts for abstraction and may serve as another approach in lieu of "entry questions" for QSRS modules.

5.2.2 EHR Vendors

The EHR vendors we spoke with provided us with information that was in line with what the health systems reported as well as the information the project team had gathered during the other project tasks. The EHR vendors are currently providing little if no AE tracking or reporting capabilities out of the box. However, vendors are offering reporting tools and best practice workflows that leverage vendor-specific EHR capabilities to support adverse event reporting. Given the amount of QSRS source data contained in nursing documentation, which is more structured than physician documentation, it is usually only coded and/or configured for specific purposes (e.g., reporting to a registry, for quality reporting).

The vendors interviewed all indicated that their customers typically use an adverse event reporting system for voluntary reporting of adverse events and noted that their customers preferred certain event reporting details to be contained within third-party systems that are confidential and not discoverable in legal proceedings. Vendors preferred to focus their adverse event development efforts on real-time reporting, rules, alerts, and other predictive analytics to support prevention and reduction of harm.

Vendors reported considerable burden in developing reports, report mapping tools, analytics, or best practices in support of required reporting to support customer participation in payer incentive programs and particularly value-based payment. Upon reviewing select QSRS algorithms, vendors noted that the feasibility of implementing automated functionality to support QSRS questions could range from "easy to implement but might have a lower level of accuracy" to "this information is text based, but the customer could create a structured documentation field to capture this information" to "very difficult, because the information is contained within narrative text." Examples of these are provided below:

- **HEALTHCARE-ASSOCIATED INFECTIONS (HAI)–CAUTI [GENERICEQ25] – Did the patient have a urinary tract catheter inserted during the stay?**
  - *Easy to implement with a lower level of accuracy:* Determining whether an order placed is relatively simple, using structured order types
    - May not be accurate: An order may have been placed, but the catheter may not have been inserted
  - *Alternative – more difficult, but more accurate*: Catheter insertion cannot be confirmed without checking nursing notes (typical current state)
    - Structured: Hospital may have implemented a flowsheet or nursing documentation form that is structured that captures this information (custom or hospital-specific implementation)
    - Structured: Hospital may have a flowsheet value that can inform if a catheter was inserted (custom or hospital-specific implementation)

- ▪ Unstructured text: Hospital may not have this within a flowsheet, but implemented this within nursing notes. (custom or hospital-specific implementation)
- **HAI – SSI [SSIEQ1] – Which if any did the patient experience?**
  Answer Options: Skin and subcutaneous tissue infection without deeper involvement; Infection of deep soft tissue; Infection without mention of superficiality; Infection within the organ space; None; Can't Tell
  - o Difficult: There is no defined dataset for collecting information for surgical site infections. Confirmation of infections would most likely be found in free text format in progress notes or nursing assessments.

Adverse Event tracking and reporting is challenging from the vendor-side for a number of reasons including:

- Data may appear to be structured, but is not stored as "reportable" – structure is more for ease of documentation or to facilitate data capture, but is not available for reporting
- Multiple locations for documentation – too much variability to support mapping
- Free text documentation
- Build can be customized, requiring hospitals to build their own reports or manually determine mapping for reporting modules

Nursing flowsheet documentation—although structured—is not often coded, especially nursing flowsheet content/documentation implemented in the 1990s or early 2000s. Most dropdown/picklist values are user-/customer-defined, and there is not an industrywide standards data/value set for flowsheet values. Adoption of electronic physician documentation (e.g., progress notes) is still limited. Progressive organizations with highly advanced capabilities for computerized order entry, results reporting, and nursing documentation may still maintain physician progress notes on paper or transcribed via dictation.

5.2.3 NLP/Analytics Vendors

The use of commercial NLP solutions is gaining in use and adoption. In most cases, NLP capabilities are not marketed as such, and are more embedded capabilities within predictive and advanced analytics solutions. The push for payment from volume to value, and increased focused on readmissions risk and reduction has accelerated the development and availability of analytics capabilities that are now accessible to most provider organizations, not just academic medical centers and large integrated delivery systems.

Analytics vendors ranged from large "big data" organizations to computer-assisted coding solutions expanding their solutions to support point-of-documentation clinical documentation improvement (CDI) initiatives to startups focused on a targeted set of outcomes.

All of the vendors interviewed focused on a defined subset of outcomes; however, the approaches used by the vendors to apply NLP or other analytics techniques to measurement could be applied to QSRS algorithms. Computer-assisted-coding solutions apply NLP to highlight and analyze relevant text based upon machine learning and algorithm development that "trains" the software. Predictive analytics companies are looking to datasets beyond the EHR (e.g., census, credit reporting bureaus) and data from hundreds of hospitals (using different EHR vendor systems) to refine their algorithms for predicting adverse events, readmissions, and other key outcomes and events.

Although the academic community is continuing its research using open source or "big data" vendor partnerships, the availability (and affordability) of advanced analytics and NLP to hospitals is increasing. The vendors reviewed—as well as other commercial vendors in this space—should be considered for automation pilots and feasibility testing for QSRS. These vendors, as depicted in the Vendor B example below, apply automation engines using NLP and analytics to EHR and other data sources to partially automate coding, CDI, and/or quality abstraction workflows.

6 Automation Feasibility Analysis (Task 4)

6.1 Approach Detail

6.1.1 QSRS Algorithm/Question-by-Question Analysis

Introduction/Summary

The fourth and final task of work was completed in parallel with the third task of work and involved two main activities. The first was an extremely in-depth and detailed analysis of each of the 205 Quality and Safety Review System (QSRS) questions, module by module. This analysis was in two parts, the first from a clinical/informatics perspective and the second from a research/science perspective.

In terms of documentation of the analysis, for each QSRS module an assessment document was created that included relevant information for each question as well as areas for documenting the analysis once completed. This resulted in a total of 20 assessment documents being created and completed. To facilitate the analysis process in terms of tracking progress and documenting and summarizing findings, a tracking document was created and used.

Clinical Analysis

Each question in the QSRS survey was evaluated individually on a variety of parameters to get an in-depth understanding of every question in QSRS, best determine the feasibility of automation for each question, and determine the most likely location in the electronic chart for potential automation.

The questions and their associated guidelines in the QSRS Manual were first evaluated by an M.D. and an R.N. provider. Then a series of questions were asked for each QSRS entry. Below are the analyses performed on each QSRS survey question:

- What is the primary or most likely chart location of the answer to this question? (always include reliable locations that could be in structured formats)
- Who is the most likely author of the charting for this location?
- Is the answer in a structured, free text, structured text or numeric value format?

- If the format is not structured or a numeric value, what complexity of natural language processing (NLP) would be required to automate this answer?
- What, if any, are any specific issues or problems associated with this question and potential automation of its answer?
- What, if any, are any questions for EHR/NLP vendors regarding this question and potential automation of its answer?

Answers that met all criteria found in the QSRS manual and were found in structured or numeric value format were deemed capable of achieving automation with simple rules-based algorithms. The rules-based algorithms would entail no more than at most matching names with preprovided lists, evaluating the value of numbers to determine which is the highest or lowest, or measuring time between time/date stamp entries. Natural language processing was considered for an entry if it was thought that an answer could not reliably be found in a structured or numeric value format. Therefore, the free-text format was where the NLP evaluations centered around.

Certified EHR Technology Analysis

Our team analyzed the structured and narrative datasets available via certified health information technology (IT), such as Consolidated Continuity of Care Documents (C-CDA), HL7 interface specifications, and availability of standards-based valuesets and codesets to address QSRS algorithms.

Data Analytics and NLP Analysis

The use of data analytics solutions and toolsets to facilitate automation of unstructured and text-based data contained within EHRs was evaluated to identify feasible approaches for QSRS automation. Health systems looking to leverage the availability of electronic health data within EHRs to proactively identify high-risk patients, adverse events, or other outcomes-related finding may employ a variety of analytics approaches. Some organizations have started to analyze historical data to perform not only surveillance, but develop and design algorithms that can be used to predict future outcomes or identify potential adverse events. This technology may be applied to retrospective EHR chart analysis in support of partial QSRS automation.

- **Predictive analytics** has been described as a combination of *data analytics* with a focus on *prediction.*
- *Data analytics* is a combination of modeling techniques and a data type. Modeling techniques range from regression techniques (linear, logistical, classification trees, etc.) to more complex machine learning techniques (neural networks, support vector machines, etc.). These techniques can be applied to different data types (free-text, categorical/ordinal, continuous, time series, geospatial, etc.).
- *NLP* is generally seen as the application of any of these modeling techniques to free-text.
- *Prediction* is the process of generalizing or modeling "new" data with existing or "known" data. Often the challenge with prediction modeling is not over-fitting nor over generalization of the model. Most any modeling technique and data type can be used for prediction.

The Gartner Data Analytics Maturity Model™ in Figure 7 provides a visual depiction of the types of analytics applied for retrospective vs. prospective information, below.

Figure 7. Gartner Data Analytics Maturity Model ™
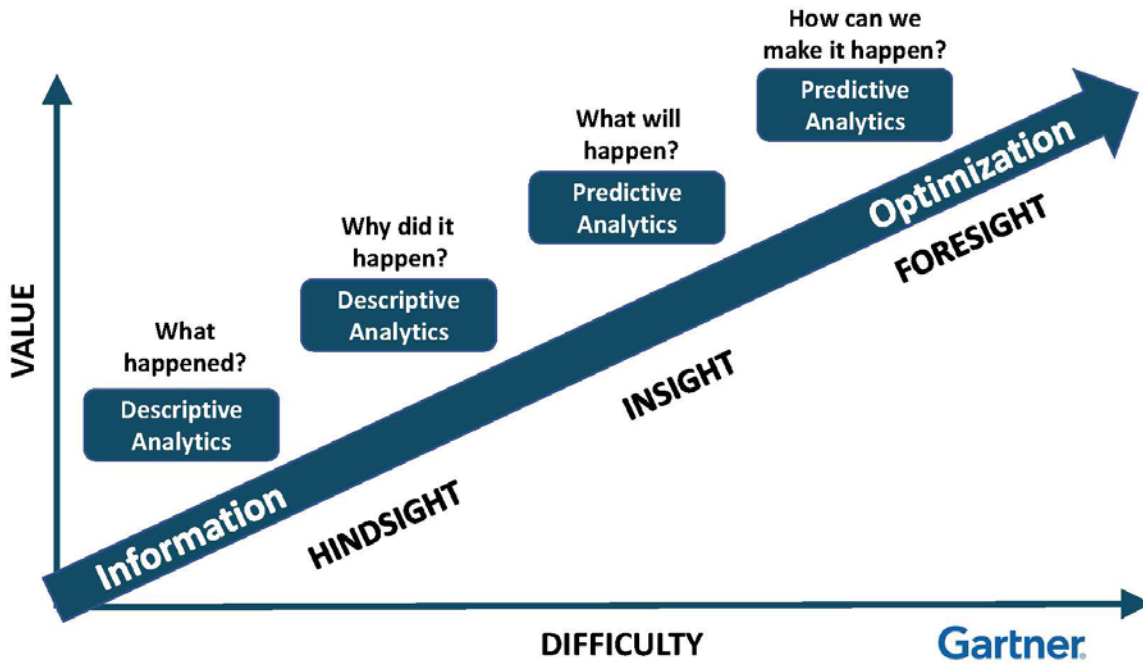
Natural language processing was the analytics approach considered for automation of free text EHR data in support of QSRS automation. QSRS questions that could be answered using information frequently documented within the EHR using structured or coded data were excluded for NLP consideration.

The questions that were classified to require NLP analysis were further categorized by the anticipated complexity of NLP application to identify the necessary information from the patient's chart. The complexity of an algorithm in the Computer Science field is defined by the amount of constraints, dependencies, and time needed to process the data and complete the algorithm. We leveraged the complexity hierarchy, based on our literature search, to categorize each of the QSRS questions that were noted as requiring NLP. Each of these questions was marked as requiring low, medium, or high levels of NLP complexity to extract information to successfully address the question. The mapping of the complexity hierarchy to the low, medium, and high rankings are described in Figure 8.

Figure 8. Mapping of Complexity Hierarchy to Low, Medium, and High Complexity Ratings

Increasing Algorithm Complexity

- Presence of a concept/entity
- Numeric value extraction
- Multiple concept detection

→ **LOW**
Answer requires extracting information from an event, encounter, etc. that occurred

- Specific temporal occurrence of a concept
- Contingency

→ **MEDIUM**
Answer requires extracting information from something happening on the condition of something else

- "Fuzzy" concepts
- Open-ended responses

→ **HIGH**
Answer requires subjective phrasing or narrative response from the user

QSRS Question Assumptions

Each QSRS question has a temporal element given that the question is asking about the current hospital encounter under review and not information about prior admissions or past medical history. However, there are some questions that require a more specific time component, such as asking if a medication was administered within 2 hours following a procedure. To differentiate between the general time component and the specific time component, we assumed that each question required a basic temporal understanding of the current admission and this basic temporal element would not impact NLP complexity. However, when specific temporal elements had to be addressed the complexity would reflect this additional temporal contingency. Thus, NLP complexity was increased only if a specific temporal element was part of the particular QSRS question being ranked.

The research team made one important assumption while coding in that there would be the ability to develop or obtain a comprehensive value-set (such as list of devices, drugs, etc.) to use when conducting the search for questions seeking a particular type or types of items. Additional research may be needed to determine if value-sets could be developed by industry initiatives/standards bodies or whether hospitals/health systems would need to refine/customize mappings to value-sets within their EHRs or configuration of QSRS reporting information.

QSRS NLP Complexity Assignment Process

Each question was coded reviewing the QSRS manual, the team's clinical expertise notations, and the complexity scale to determine the complexity level of each question. The QSRS manual frequently included important information in the Abstraction Notes and Instructions sections of each question that highly influenced the categorization of complexity level and was not solely apparent in the question and answer categories. Each QSRS question could fit into one or more category types, and often multiple types were assigned. To categorize the complexity of the multiple-type questions, we created an algorithm in which adding multiple complexities scaled up to the next level: LOW + LOW = MEDIUM and MEDIUM + any complexity type = HIGH.

6.1.2   Vendor Selection for Analysis

As noted above, the fourth and final task of work involved two main activities. The second involved working closely with two EHR vendors and one NLP/analytics vendor. The general findings from the environmental scan and the stakeholder discussions were combined with the findings from the question-by-question, module-by-module analysis of all 205 QSRS questions, and reviewed with the vendors to accurately determine what capabilities exist in current technologies, both EHRs and NLP/analytics, that support the abstraction of QSRS data from EHRs.

Initially, all three of these in-depth reviews were to occur with three EHR vendors. However, following the information gathered in the previous tasks, and given the need for NLP and advanced analytics to automate QSRS, it was determined that it was more useful to speak with two EHR vendors and one NLP/analytics vendor.

In terms of selection of the specific three vendors, per the contract the first two EHR vendors were to be "selected from those vendors with an estimate of at least 5 percent or greater of the EHR market...and must meet ONC's EHR Certification requirements." To this end, EHR A and EHR B were selected. These two vendors were chosen as they are two of the largest and well-known EHR vendors in the hospital setting, the healthcare space that QSRS is used in.

As noted above, the contract originally stated that the third vendor would also be an EHR vendor, but it was determined that working in greater detail with an NLP/analytics vendor would provide more useful findings and results and ultimately provide a more accurate and inclusive feasibility study. Vendor C was selected in particular as it is an NLP vendor and a health care analytics vendor that provides innovative technologies that are in line with the AEs that QSRS tracks. Our rationale for selecting Vendor C to represent the NLP/analytics vendor includes—

- NLP/Analytics Company with AE Expertise: Vendor C is an innovative health care predictive analytics company that has spent the past years developing and refining its algorithms that support 20 use cases, which include pressure ulcers, sepsis, CLABSI, and CAUTI.
- Predictive Analytics: Vendor C's phenotype model combines clinical data with nonclinical data (e.g., population and census data) to look at 4,000 dimensions to determine what patients are at risk for. The Vendor C platform and the science it is built on results in high accuracy of predictions, which are provided real time.
- EHR Vendor Neutrality: Vendor C uses various data sources including free text notes, coded information, structured and unstructured data, HL7 messaged, ADT feeds, etc., and is integrated into various EHRs including EHR A, EHR B, and other EHRs.

6.2 Results Overview

6.2.1 Clinical Analysis

A majority of questions could not be answered reliably via structured entries in the EHR. Thus, the majority of answers could not be ascertained via automated rules-based algorithms.

The precision and detail of many of the QSRS questions make it difficult to allow for more rules-based automation to occur. Though some QSRS questions may seem straightforward on initial glance, details and criteria for answers, often listed in the manual, make them more complex and thus more difficult to find structured entries as reliable answers for.

Also, though structured entries may exist as potential answers to many questions, they are often not reliable enough to say with any degree of certainty that the answer there would be correct. A good example of this is that often physician orders can be considered a structured entry that could potentially serve as a place to find an answer. However, clinical realities tell us that even though something is entered as a physician order, it is occasionally not completed by nursing or other staff or modified when actually administered. Medications, orders for insertion or removal of central lines and urinary catheters, and intubation/ventilation orders are often modified at the bedside for a variety of reasons (patient refusal or an inability to successfully perform procedure are some of the common reasons why). Thus, physician orders are often not reliable entries when looking for more definitive locales for answers to certain QSRS questions.

Lastly, though many different locations for reliable answers to questions often exist, the structured entries were considered first. And, only if it was deemed that these structured entries were not reliable, were other chart locations evaluated.

Nursing assessments and nursing "flowsheets" proved to be a very good source for finding reliable structured answers to QSRS questions. The fact that they are done at regular intervals and almost all of their entries are either in structured or structured text formats were key to allowing for rules-based automation.

Acceptable accuracy rates for QSRS answers (if one exists) will determine how far-reaching automation can go. As comfortably with some uncertainty increases, more automation can be considered. However, determining the accuracy of automation and comparing it to the current state where the answers are abstracted by humans may not be easy. Pilot studies and head-to-head comparisons are likely warranted.

A consideration of a deeper look into the entry questions and two or three of the most common modules may be warranted as those are the ones that will always/often get used when the QSRS surveys are conducted. This deeper look could entail a closer examination to what exactly it is about these questions that prevents automation and if a consideration of changes could be made to achieve some level of automation.

In performing the clinical review of data sources—with both clinician users and vendor input—our team classified source data format(s) for each QSRS question. More than one data format may be applicable or assigned for a question as the source data may reside in multiple locations (e.g., nursing flowsheet and physician documentation and laboratory result report).

Figure 9 depicts the analysis of questions in aggregate, across 205 questions, and Figure 10 depicts the analysis by module, assigning the following source data formats:

- NV: Numeric Value
- SC: Structured and Coded
- SU: Structured and Uncoded
- ST: Structured Text (Structured field with free text)
- FT: Free Text (Narrative unstructured text documentation)

Figure 9. QSRS Source Data Analysis

Numeric

- YES: 14
- NO: 191

Structured and Coded

- YES: 63
- NO: 142

Structured and Uncoded

- YES: 77
- NO: 128

Structured Free Text

- YES: 64
- NO: 141

Unstructured Free Text

- YES: 132
- NO: 73

Requires NLP for Automation
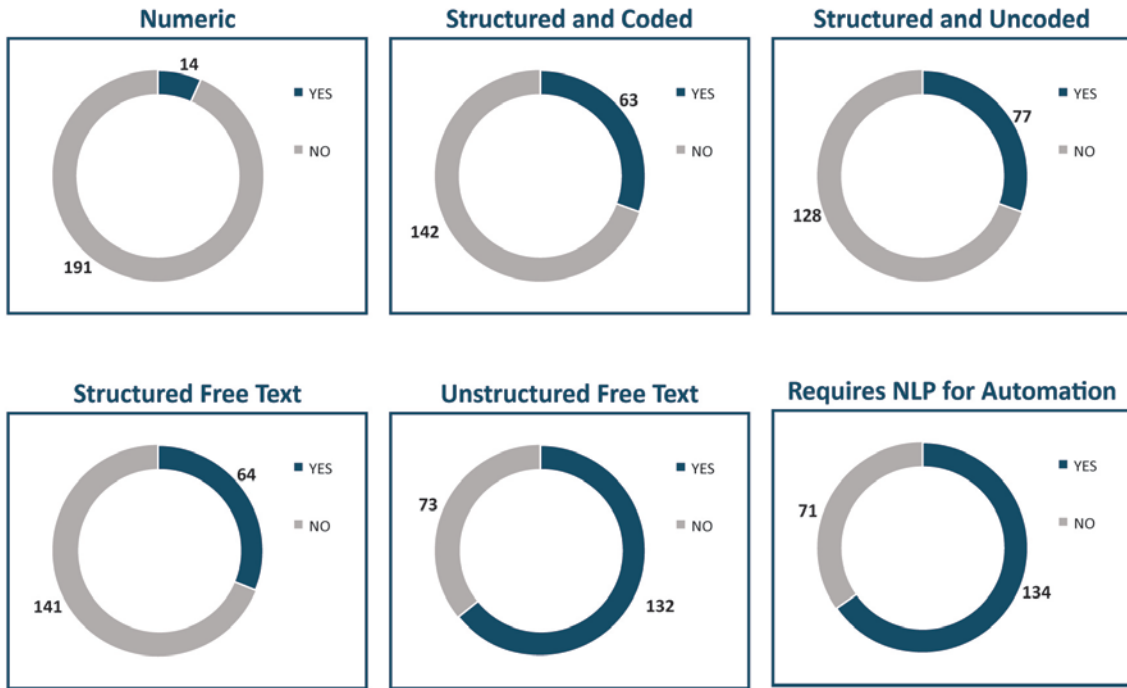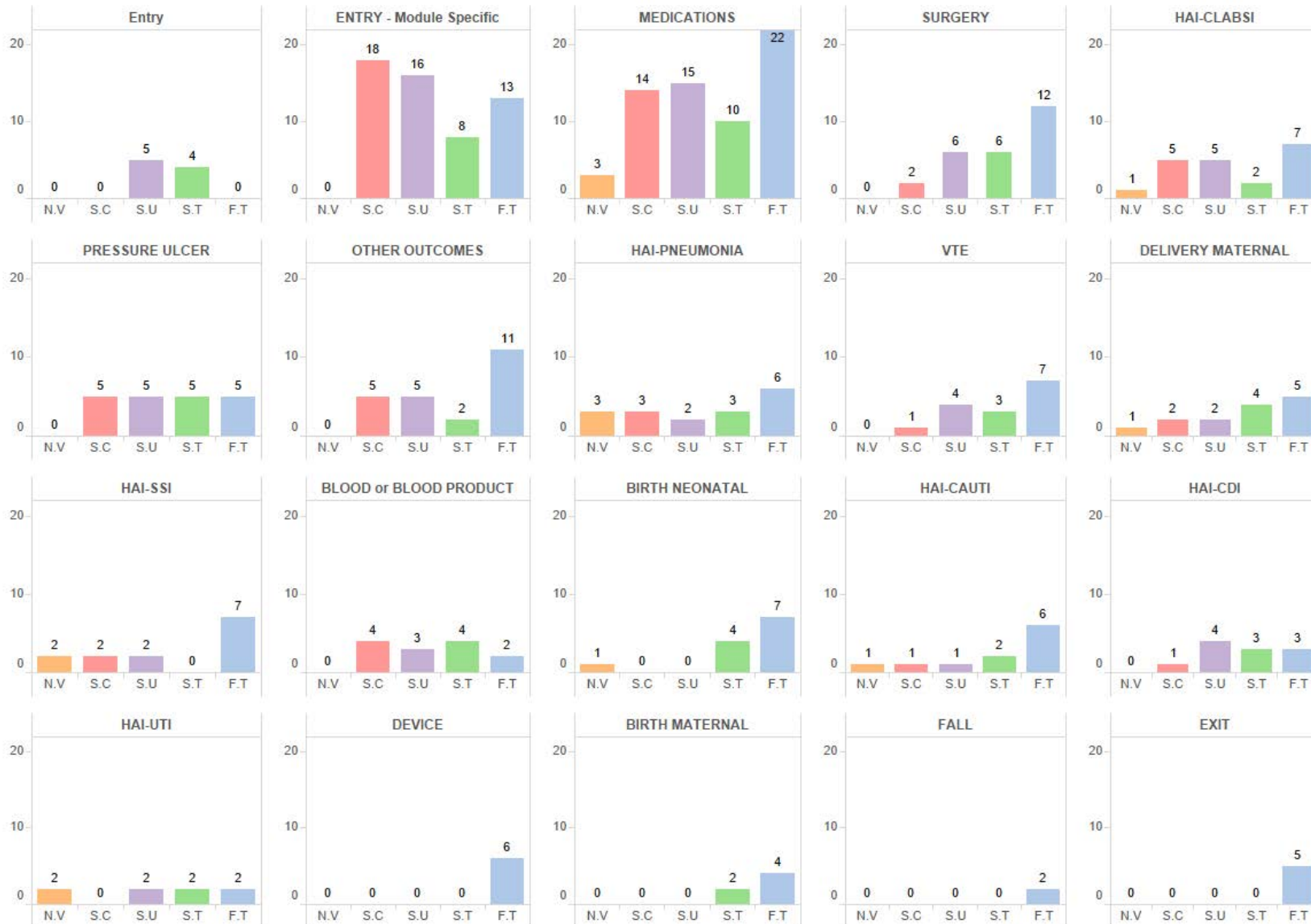
- YES: 134
- NO: 71

Figure 10. QSRS Source Data Analysis

6.2.2 NLP Analysis

Of the 205 questions in the QSRS Manual, we determined that 134 (65%) required NLP analysis to extract an answer from the electronic medical record, as indicated in Figure 11. Of the NLP-related QSRS questions, 37 (28%) were categorized as low complexity, 46 questions (35%) were medium complexity, and 51 questions (38%) were high complexity.

Only one module, the Generic Entry Question module, contained questions that were not determined to require NLP analysis for feasible automation. Six of the 19 modules (Device, Fall, HAI-CAUTI, HAU-UTI, VTE, and Exit) required 100 percent NLP analysis. The Medication and Module Specific Entry questions had the highest number of questions requiring NLP analysis, though also the highest number of questions overall. The range of questions determined to require NLP analysis was 2–23, with an average of 7 questions per module. Of the 19 modules having at least 1 question that required NLP analysis, the overall distribution between low-, medium-, and high-complexity questions was roughly 1/3.

Figure 11. QSRS NLP Complexity Profile
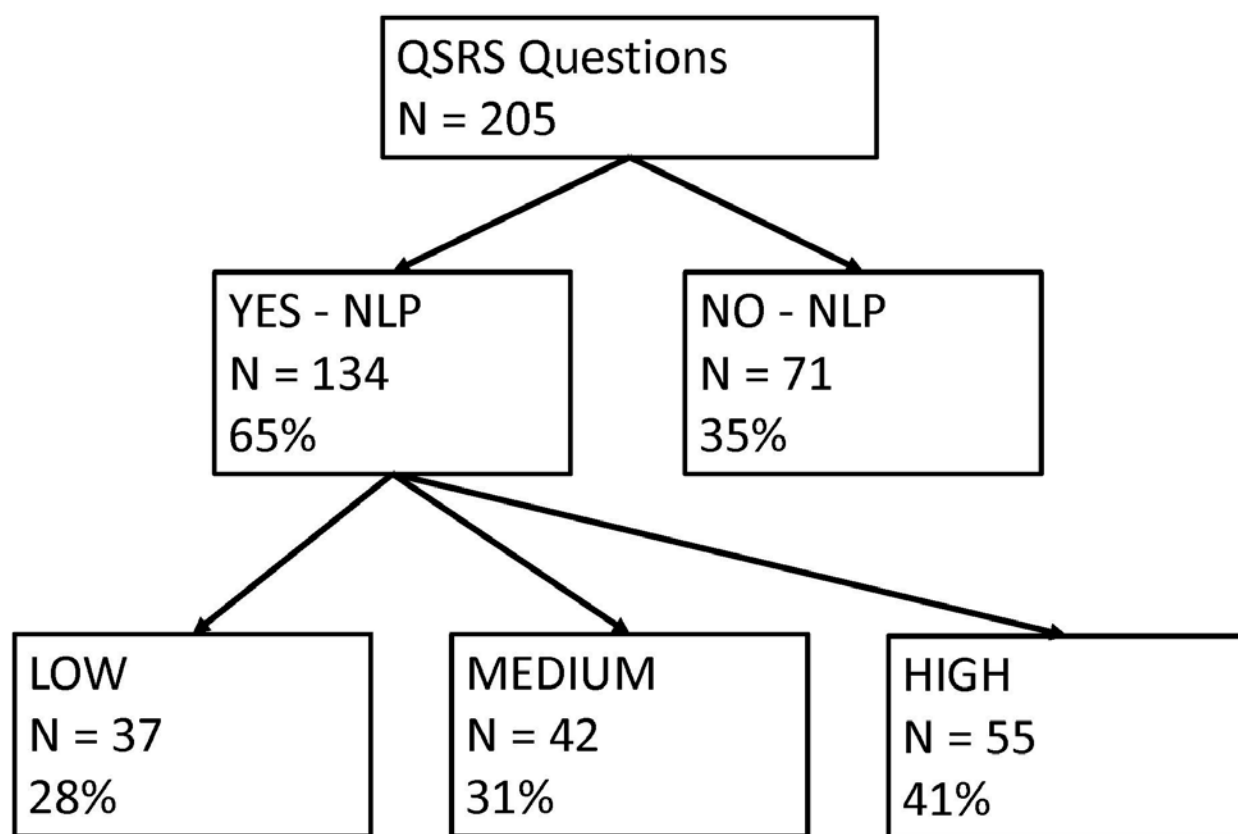
We further subdivided the NLP-related QSRS questions by general entry and specific modules, as shown in Figure 12. (For this analysis, entry and exit questions were each classified as additional "modules.") The Module Specific Entry questions and Medication module had the highest number of low complexity questions (6 and 9), though 12 of the 19 modules that had questions determined to require NLP analysis had at least one low-complexity question. The average number of low-complexity questions was 1.6 per module. The Medication and VTE modules had the highest number of medium complexity questions (14 and 5), while all other modules had four or less medium complexity questions. The average number of medium complexity questions was 2.4 per module. Sixteen of the 19 modules had at least one high complexity question, with Blood or Blood Product, HAI-CDI, and Pressure Ulcer modules being the exceptions. Surgery and HAI-SSI had the highest number of high-complexity questions (8 and 7), and HAI-UTI and HAI-SSI modules contained only high-complexity questions. The average number of high-complexity questions was 2.7 per module.

There are several important takeaways from this analysis. First, the low-complexity NLP approaches for the entry questions is beneficial since there is greater likelihood that NLP can be implemented effectively to address these questions and therefore prevent the need for coders to address these modules. Second, the medications and VTE modules with the highest number of medium complexity NLP questions may be driving the overall average medium complexity percentage. Third, the surgery and HAI-SSI modules with the highest number complex questions may benefit from alternative strategies that are discussed later in this report.

Figure 12 below is a summary of NLP complexities for each question that required NLP. The data is organized by module. This visualization of the 19 existing modules (classifying entry and exit questions as additional modules as noted above) highlights where complexities exist within each module from entry to exit. Each chart is of a single module and displays how many low-, medium-, and high-complexity cases occur. When interpreted from a larger view with all modules the modules with significant amount of a specific complexity can be located. This can assist the coder/reader in seeing which module has the highest relevant complexity and not only how it compares with the other two complexities but also compares with other modules that may or may not be similar. When looking for an intervention strategy regarding any sort of NLP, the strategy can prioritize the specific complexity, low, medium, or high, from this visualization and pull those modules most similar.

Figure 12. QSRS NLP Difficulty
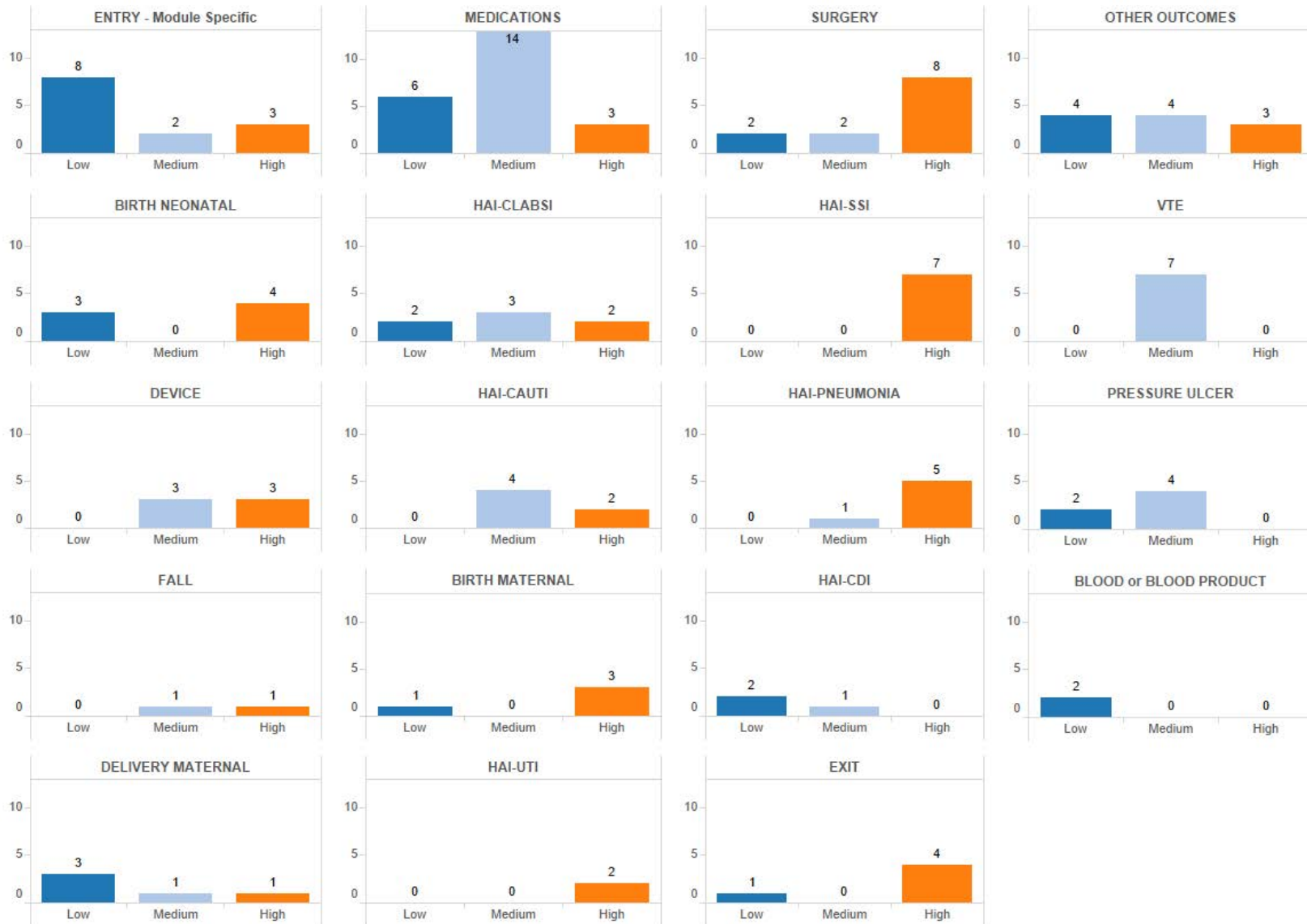
QSRS Data Sources

Knowing which QSRS module or series of questions will require coders to "jump around" the EHR is important, as this might indicate areas where better visualization designs and information displays can help coders reduce the total time on a record. This analysis might also suggest alternative ways of ordering questions to make the search process more linear. The more sources needed for a module, the more investigation and searching may be required of the coder/reader, which could significantly increase the time needed to complete the module. If time is important, then referencing this table could assist in prioritizing which module to complete first.

Which QSRS modules require coders to look for the greatest versus least number of unique data sources? Figure 13 lists the total number of unique data sources listed in the QSRS manual for each module as well as the entry questions for modules. If a data source is listed under several questions in the same module, it will only be counted once. This allows us to understand how many unique sources you will need to check for any specific module.

Figure 13. Unique Data Sources by QSRS Module

Module	Unique Sources
Surgery	37
Other Outcomes	29
Medications	28
Entry Question VTE	28
Delivery Maternal	22
HAI-Pneumonia	22
Entry Question	21
Birth Maternal	21
Blood or Blood Products	18
VTE	15
HAl-CLABSI	13
HAI-SSI	13
Entry Question Blood or Blood Product	12
Entry Question Fall	12
Entry Question Pressure Ulcer	12
Entry Question Device	11
Entry Question UTI	11
HAI-CAUTI	11
Device	10
Pressure Ulcer	10
HAI-CDI	9
Fall	8
Entry Question CDI	8
Entry Question CAUTI	7
Entry Question CLABSI	7
HAI-UTI	6
Entry Question Pneumonia	6
Entry Question SSI	6
Entry Question Surgery	6
Exit Questions	6
Entry Question Birth Maternal	4

Which data sources contain the answers to the most QSRS questions? The answer to this question will help us better understand which data sources provide the most information about adverse events and where coders potentially might spend most of their time. This will help us identify high value areas for automation and the application of other automated techniques. For each data source, the table below counts how many times (questions) it was listed as a data sources. For example, "Nursing" was listed as a data source for 169 questions. The results are depicted in Figure 14 below.

Figure 14. Data Sources for QSRS Questions

Data Source	Count of Questions
Nursing	169
Consult	163
Progress	140
Discharge	128
History	110
Physician	99
ER Note	99
Lab	54
Anesthesia	50
Flow	44
Admin	37
MAR	33
Rad	31
Operative	28
LD	19
Code	17
Procedure	15
Dialysis	14
Aph	13
PT	13
Transfer	13
Skin	10
Wound	9
Transfusion	8
Intervention	7
Respiratory	7
Amb	6
Diabetic	5
Record	4
Autopsy	3
C-section	3
Circulator	2
Medication	2
PACU	2
Toxicology	2
Angiography	1
Vitals-Pulse Ox	1
Device	1
Face	1
Diagnostics	1
NICU	1
Nuclear	1

Data Source	Count of Questions
Order	1
Risk	1
DVT	1
Ventilation	1
VTE	1

Figure 15 below is provided as a mapping of the data source names as they appear in the QSRS manual and the abbreviated and collapsed short name to make visualization and aggregation easier and more meaningful. The analysis by "Shortened Source Names" was provided above in Figure 12.

Figure 15. Data Source Mapping to Shortened Source Names

Shortened Source Names	Data Sources Listed in QSRS
Admin	Admission Note, Admission Note (only if completed after delivery), Admission Notes, Admission Notes (if admitted from OR), Admission nursing notes, Intake and output sheets
Amb	Ambulance Record, Ambulance Records
Anes	Anesthesia Notes, Anesthesia Notes for later procedure, Anesthesia Pre-assessment, Anesthesia Record, Anesthesia record for the operative procedure, Anesthesia Record/Flowsheet, Anesthesia Records, Anesthesia/operative/procedure reports, Anesthesiologist Note, Anesthesiologist Notes, Post-Operative Anesthesia Assessment, Pre-anesthetic assessment performed prior to the procedure., Pre-Operative Anesthesia Assessment
Angio	Angiograms, Magnetic Resonance Venography (MR)
Aph	Apheresis Note, Apheresis Record
Autopsy	Autopsy Report, Post-mortem/autopsy Report
Circulator	Circulator notes
Code	Code Sheet, Coding Sheets
Consult	Consult Note, Consult Notes, Consultation Notes, Consultations, Consults
C-section	Cesarean Section Operative Reports, Cesarean Section Operative Reports
Device	Device (implant) Records
Diabetic	Diabetic Flow Sheets
Diagnostics	Impedance Plethysmography (IPG)
Dialysis	Hemodialysis Note, Hemodialysis records
Discharge	Discharge Summary
DVT	The location of the DVT may be most easily found in the report of the required confirmatory
ER Note	Emergency Department Record, Emergency Room, Emergency Room Notes, ER Notes, ER Record, ER Record/Note, ER Records
Face	Face sheet (only if completed after delivery)
Flow(sheet)	Critical Care Flow Sheets, Flow Sheets, Flow sheets used to record the procedure, Flowsheets, Graphic/flow sheets, I/O Flowsheets, ICU Flowsheets, IV Flowsheets
History	History and Physical, History and Physical (H&P), History and Physical Notes
Intervention	Interventional Note, Interventional Report
Lab	Bacteriology reports, Laboratory Records, Laboratory Reports, Microbiology Lab Reports, Microbiology Reports
LD	Labor and Delivery Note, Labor and Delivery Record, Labor and Delivery Records
MAR	Medication administration record, Medication Administration Record (MAR), Medication Administration Records, Medication Administration Records (MARs)
Medication	Medication Orders
NICU	NICU Records
Nuclear	Nuclear Medicine Reports

Shortened Source Names	Data Sources Listed in QSRS
Nursing	Nurses Note, Nurses Notes, Nursing Admission Assessment, Nursing Admission Assessment/Notes, Nursing Assessments, Nursing Care Plan, Nursing Flowsheets, Nursing Notes, Nursing Notes/Admission Assessment, Nursing/Admission Assessment Notes, OR Nurses Notes, OR nurses Notes for later procedure, OR Nursing Record, OR Nursing Records, Preoperative nursing notes, Vital Signs/Nursing Flow Sheets
Operative	Brief Operative Note, Intraoperative Record, Operating Room Flowsheets, Operative Note, Operative Note for later procedure, Operative Notes, Operative Record, Operative Reports, Operative/procedural Reports
Order	Order Sheets
PACU	PACU Notes, PACU/Recovery Room Record
Physician	ED Physician's note, Physician Consultation Notes, Physician Notes, Physician Order Sheets, Physician Orders, Physician Progress Notes, Physician progress notes regarding pre-mortem tests only, Physician restatement of a Chest-X-ray or image test
Procedure	Procedure Note, Procedure Notes, Procedure Report, Procedure Reports
Progress	Progress Notes, Progress Report
PT	Physical Therapy, Physical Therapy Notes
PulseOx	Blood gas or Oximetry levels
Rad	Chest X rays, Contrast Venography, CT Scans, Duplex or Doppler Ultrasonography, Imaging Reports, Radiological/Image Tests, Radiology Notes, Radiology Nursing and Technician Notes, Radiology Reports, Radiology reports including, Radiology Reports/Notes, V/Q Scans, X-ray/Radiology Reports
Record	Entire medical record
Respiratory	Respiratory Therapy Notes
Risk	Risk assessment flow sheet
Skin	Skin Integrity Sheets/Notes
Toxicology	Toxicology Records
Transfer	Transfer Notes, Transfer records
Transfusion	Transfusion records, Transfusion slips/Notes
Ventilation	Ventilation Records
VTE	VTE Protocol or Orders
Wound	Wound Care Notes, Wound Care Team Notes

6.2.3 EHR Vendor Data Availability

Automation is often associated with simplicity. The availability of electronic health information within an EHR can facilitate easy identification of information; however, this is easy only if—

- There is a single place to document this information
- There is a defined field within the EHR for documenting this information (structured data)
- There are a defined set of values for documenting this information (coded data)
- It documented by a single care provider type (e.g., attending physician, consulting physician, nurse, respiratory therapist)
- It is documented once or at defined intervals

The likelihood of all of the above being true or well-defined is little to none. The fact that clinical documentation can—and usually is—documented by multiple provider types combined with the fact that most hospital stays exceed a single "shift," ensure that multiple individuals exist for each provider type.

QSRS questions ask the chart reviewer whether something is present in the chart. Although there may be multiple "preferred" sources (e.g., nursing notes, physician progress notes, nursing flowsheets), the abstraction manual indicates that the reviewer may locate the information to support answering a question in any number of sources (e.g., emergency department notes, History & Physical, Consult Notes).

Whereas using paper records, searching for this information would require a page-by-page review of many pages of documentation, in an electronic world, this information may still reside in any of these same electronic, multiple locations. Excluding a discussion on text-based documentation (this is addressed within the 6.2.2. NLP Analysis section above), searching for a structured entry within an electronic record is presumably much easier than manually reviewing each page of a paper record. The following section provides an example of how automating the finding of a single, coded concept (What problems/diagnoses does the patient have? Was the patient treated for a particular condition?) is more complex than may be assumed when considering availability of EHR data.

Example: Meaningful Use Problem List Reporting

The CMS Meaningful Use (MU) EHR Incentive Program allowed both problems and diagnoses charted using either SNOMED-CT or ICD-9 to meet the goal for MU Stage 1. Problems and diagnoses entries are structured, and coded using terminology standards such as SNOMED-CT, ICD-9, or ICD-10.

- Problems: EHRs may offer a mechanism to classify Problem Lists within the EHR, as there are multiple Problem Lists, based upon the provider classification (e.g., Medical, Interdisciplinary, Nursing, Dietary) or the source of information (e.g. patient-stated problem list). Typically, the Medical Problem list was selected by hospitals for meeting the MU objective and required mapping within MU tools or reporting modules.
- Diagnoses: EHRs offer a multitude of options for charting diagnoses, including and not limited to: admitting diagnosis, discharge diagnosis, working diagnosis, billing diagnosis, referring diagnosis, and reason for visit. Hospitals have the option of selecting which of the diagnoses is to be counted for the measure within MU mapping tools or modules.

Given that EHR A and EHR B have been widely implemented for over two decades, both vendors (along with many vendors in the market today) offer out-of-the-box systems with prebuilt documentation templates or methods to select or import templates from a set of best practices or customer-provided content. An organization may choose to implement these templates as is or modify to accommodate their organizational goals, workflows, or provider preferences. These templates have generally been mapped to reporting solutions or decision support login that can be enabled to access data that has been charted within any of these fields.

Among the challenges encountered by early adopters of health IT that may have legacy modules or content or hospitals that may have implemented custom documentation templates, vendors highlighted that problems recorded as free text or using non-standard vocabularies may require data mapping and conversion utilities to support availability as codified data (e.g., SNOMED-CT, ICD-10). Early adopters of health IT may have legacy modules or content that were implemented using custom documentation templates or predate the availability of certified capabilities and standards. This content may not be available for inclusion in C-CDA documents or for health information exchange without additional mapping or analysis.

Using the MU Problem List example above, there are multiple Problem Lists available within EHRs. Rather than impose a particular workflow upon its customers, vendors have provided flexibility—by making available default reporting options, while enabling customers to map their own selections and workflows to a single reporting element.

If AHRQ provided a QSRS reporting format (discussed further in Section 6.3.1 below), it would enable vendors to develop reports or mapping solutions to facilitate automation of QSRS information. However, vendors indicated both vendors and customers resisting the creation of structured data and documentation to facilitate payer or quality reporting.

Measurement Burden and Need for Measures Alignment

An analysis of almost 50 State and regional measure sets found over 500 unique measures were in use, with only one-fifth used in more than one program.[46,47] In the second quarter of 2014, 33 different CMS programs used over 850 unique measures, with only one-third used in more than 2 different CMS programs.[46,48]

For ease of comparison across vendors participating in the QSRS analysis, we selected three common adverse events addressed by QSRS for review and comparison: Falls, Pressure Ulcers, and Catheter-Acquired Urinary Tract Infections (CAUTI). Additional QSRS modules were reviewed on an ad-hoc basis. All vendor offered "out-of-the-box" support for these three adverse events. This did not mean that reporting capabilities were available to customers without implementation effort; however, all three vendors provided guidance on how to best implement reporting and analytics for these adverse events.

Our findings concurred with statements in the literature noting significant effort and burden on provider organizations to implement effective EHR-reporting. "While recent EHR enhancements have begun to support real-time measurement, these systems currently fall woefully short in meeting the needs of providers; and many systems are unable to generate simple, reliable, and actionable reports.[49] Many measures continue to require meticulous reviews of medical records by trained professionals who otherwise would be directing their expertise to providing and improving patient care."[50]

Both EHR A and EHR B provided guidance and support modules for customers for a set of core measures for quality reporting that included some of the QSRS modules, including the three we reviewed. The vendors provide customers with process flow, documentation tools, and reporting guidance. All vendors encouraged the creation of national algorithms to enable the creation of tools at a vendor level that do not need to be recreated at the local level.

Opportunities for automation cited by EHR vendors included development of structured documentation templates where appropriate. The following examples are provided for CAUTI and Pressure Ulcers to highlight EHR approaches to structured documentation collection to facilitate reporting and decision support for the noted adverse events:

CAUTI Examples

- Central Lines or Tubes Present on Admission (POA) Form that collects using structured data:
    - Lines or Tubes POA
    - Urinary Catheter POA
    - Urinary Catheter Type (with structured responses)
- Guidance to include/develop an insertion and discontinuation date/time field
- Development of alerts to trigger indwelling catheter review daily
- Mapping of multiple documentation field to indicate that a catheter has been removed (order discontinue vs. nursing documentation of indication of catheter removal)
- Development of a "Catheter Justification" field requirement within the provider order

Pressure Ulcer Examples

- Reports for Hospital Acquired Pressure Ulcers by Hospital Service Area
- Workflow Guidance: Flowsheets, Best Practices, Order Sets, Assessments, Intervention
- EHR Implementation and Build Considerations
- Training

The above examples require decision-making and configuration by the health system. Health systems monitor their adverse event rates for CAUTI, Pressure Ulcers, and Falls and the documentation practices to determine needed interventions—or training—to drive consistent workflow and documentation practices, modify protocols, and address alert fatigue. Guidance provided acknowledges that the reports may utilize nursing documentation data for hospital support and that documentation reported as part of Quality Reporting may be based upon physician documentation (Pressure Ulcer example where data is sourced from nursing flowsheets or other nursing documentation). Vendors acknowledged that for supreme accuracy or for public reporting or audits, chart review is recommended as EHR-generated reports will only be looking for data available within specific areas of the record.

Both EHR vendors and their hospital customers are focused on using the EHR to not only monitor their adverse events, but implement approaches to address adverse events. Support for surveillance activities that would require manual EHR chart review via coders was not a priority for vendors, citing effort in streamlining EHR documentation practices would best support automated reporting and real-time decision support.

Document and Exchange Standards

An inherent challenge or unknown in the QSRS abstraction process is whether human abstractors tend to review charts and only review the "most likely" sources of data or review "all" of the potential sources of data or only the "first indication" when responding to a QSRS question. It is still unknown whether a review of an entire chart using NLP or using a human abstractor yields the best results for any QSRS module, as a computer can certainly be capable of ingesting the text of an entire patient encounter to search for answers to a question (with a lower level of NLP difficulty) faster than a human abstractor could review every page of a paper or scanned record. Along the same lines, as document and information exchange standards are continuing to evolve and increase in adoption and use, AHRQ may want to consider a comparison of abstraction vs. use of common available electronic data exchange formats to determine if a smaller subset of data is "good enough" to answer a certain number of QSRS questions.

Hospital clinical documentation can duplicative in many areas - a nurse may identify a pressure ulcer in a flowsheet or wound documentation, but it is also documented in the physician note. Discharge summaries may – and should - contain information also charted within progress notes, nursing assessments, or available within orders, lab results, or other ancillary documentation.

As a comparison of Figure 3 (certified EHR technology Common Clinical Data Set) and Figure 14 (common data sources for QSRS questions) indicate, the Common Clinical Data Set (CCDS) and the documents available electronically using certified EHRs does not map to all—or even most—of the data sources. However, the Continuity of Care Document Architecture (C-CDA) standard does allow for additional information outside of data required by certification to be contained and exchanged using this document format. It is unlikely that all data sources in Figure 14. could be contained in a C-CDA document, but it is possible that one source document that "may" be a potential data source could contain need information.

The certified EHR technology requirements for populating the CCDS and CCDA are not specific, nor restrictive, as to data sources – but are not comprehensive. Although diagnoses-based partial automation maybe feasible to identify patients for inclusion, vendors and providers have flexibility in the data source(s) that are selected for Summary of Care Records.

AHRQ may want to evaluate data available through electronic data exchange standards such as C-CDA documents, Health Level Seven (HL7) standard interfaces, standard-based and coded billing data, application programing interfaces (APIs) being made available by EHR vendors for data access by authorized applications, and the emerging Fast Healthcare Interoperability Resources specifications.

A variety of data exchange and document standards could certainly be leveraged for the "technical automation" of how the data gets from a hospital to CMS/AHRQ/Other. The communication of data and transmission protocol from hospitals was not studied/analyzed for this study. As noted in the recommendations section, an evaluation of use of available standards-based electronic data in comparison to manual abstraction may be a useful next step in automation of QSRS.

Vendor Reporting Using EHR Data

As discussed earlier, two of the EHRs studied and an analytics vendor that use a standard EHR dataset request cover a number of QSRS module AE topics from a point-of-care or real-time notification perspective. The data sources are a subset of the potential data sources for not only QSRS but other required Core Measures, specialty certification, and other quality reporting initiatives—and therefore not used for reporting to government agencies, payers, etc. However, these reports are used heavily by the hospitals from an operational perspective in the delivery and management of patient care.

There is a continued segmentation in hospitals of "EHR-based reporting" and "chart abstraction-based reporting"— and hospitals are increasing the amount of resources dedicated to both EHR reporting and analytics and quality improvement/performance improvement abstraction teams. All of the vendors considered for this analysis provide

reporting for: CAUTIs, falls, pressure ulcers. Hospitals and vendors participating in this study all expressed the desire to focus financial and staff resources on EHR-based and real-time reporting that can be actionable and encouraged AHRQ to consider dataset mappings or electronic reporting requirements where possible in lieu of post-encounter surveillance or chart abstraction.

6.2.4 Vendor Analytics Capabilities

EHR vendors—and both EHR A and EHR B—offer enhanced solutions for reporting and advanced analytics. There was acknowledgement amongst EHR vendors that customers may choose third-party analytics platforms completely decoupled from their EHR vendor and that their more progressive customers have already made investments in advance analytics platforms. Both vendors support third-party analytics and reporting via various architectures, interfaces, and APIs. EHR B markets its cloud-based population health platform for identification, scoring, and predictive risk analysis. EHR B's reporting solution extracts data from both EHR B and non-EHR B systems, de-identifies the data, and maps data to common nomenclature to generate adverse drug event and condition outcomes reports for its data contributors to review and benchmark against other Health Facts contributors. EHR A's data store and data warehouse provides the ability for EHR A data within and across an enterprise to be available for analytics. These platforms can be leveraged to support analysis of EHR data for adverse event reporting and support QSRS using structured data and rules-based algorithms.

EHR-agnostic analytics vendors vary in scale and scope from significant investment and implementation resources to more targeted solutions that provide hospitals with a reporting dataset for a defined set of use cases (e.g., Vendor C). Vendor C was reviewed to investigate its approach for standardizing datasets for hospitals to provide, combining data with socioeconomic, census, and other data sources, and then applying its internally-developed patient phenotype model.

Vendor C's modeling was developed by taking an initial set of 20,000 lives that were preclustered into buckets for certain diseases, illnesses, and adverse events (including pressure ulcers, sepsis, CLABSI, CAUTI). The company took 10 years to build 20 use cases, training its model from a combination of EHR data from more than 400 hospitals, historical claims, population census data (graduation records, ranges, access to pharmacies, density of providers), and Equifax/Transunion data to determine indicators for what patients within its buckets were at risk for.

7   Recommendations

7.1   AHRQ Strategies and Considerations

We developed and applied an automation feasibility framework to assess the degree Quality and Safety Review System (QSRS) questions can be automated. Through this work, we found that 58 percent of QSRS questions (118 out of 205) could be reasonably automated with the use of currently available software and applications either without natural language processing (NLP) or "low complexity" NLP. Upon adding "medium complexity" NLP for consideration, an additional 40 questions can be automated, resulting in 77 percent of QSRS questions that are categorized as not requiring NLP or those as low and medium relative NLP complexity. In summary, we believe that 58 percent of QSRS questions are relatively easy to automate and 77% of QSRS questions are feasible for automation using available capabilities in the market today.

For the remaining 23 percent that are require NLP and are classified as "high complexity," in parallel with the approaches and pilots recommended in this section (which best address the 77 percent of questions that are feasible for automation), AHRQ should consider a review of these questions and determine whether the questions can be further specified to align with current clinical documentation practices and available value sets. AHRQ may identify areas for introduction and engagement with standards development organizations and other standards and specifications bodies, as well as health information technology (IT) certification.

This framework and our approach provide a description of the design space that should be considered in building an automated or partial automated QSRS system. Recommendations can be made on which modules would be easier to automate or which data sources would require the most analysis. This discussion can help prioritize automation efforts as well as potential vendor guidance and standardization.

A hybrid rule-based and machine learning approach could be used to help automate several QSRS questions. A hybrid approach can leverage clinical expertise while accounting for variability between systems. In addition, a human-in-the-loop automation approach would be very beneficial to confirm classifications and abstractions particularly for complex and nuanced in questions.

7.1.1 eMeasures Consideration for QSRS Reporting by Hospitals

To consider full automation, AHRQ may consider replicating the national effort to translate "paper based" clinical quality measures (using human abstractors) to eClinical Quality Measures, or eCQMs (using standards for calculating and reporting from electronic health records [EHRs]). Historically, clinical quality measures have been used for internal quality improvement and benchmarking, as well as for demonstrating compliance with accreditation requirements. Human abstractors reviewed patient charts and counted the patients, or cases, that would populate the numerator and denominator required to calculate the quality measure's percentage, as well as the cases that qualified as exclusions for the measure. Similar to the QSRS algorithm method, abstractors followed measure specifications, developed by measure developers, that defined the clinical conditions for including or excluding a patient from the measure's calculation.

As the costs of manual abstraction increased, industry standards organizations, such as Health Level Seven International (HL7), responded to draft a universal, unambiguous standard to automate the process in EHRs. However, in the absence of a completed and nationally adopted standard, the data for quality measures were abstracted, calculated, and reported across inpatient and outpatient settings with considerable variance. In other words, apples were not compared to apples.

The effort to advance an eCQM calculation and reporting standard accelerated in 2009, when Congress passed the HIT for Economic and Clinical Health (HITECH) Act as part of the broader stimulus package. HITECH authorized the Centers for Medicare & Medicaid Services (CMS) to require standard eCQM calculation and reporting from EHRs certified through the ONC Certification Program to award incentive payments. This milestone propelled two parallel processes, described below, that continue to evolve and undergo refinement.

First, HL7 completed the Health Quality Measure Format (HQMF), used for representing eCQMs in the EHR, and the Quality Data Reporting Architecture (QRDA), used for reporting eCQMs in a standard structure. The ONC Certification Program requires both standards for calculating and reporting eCQMs to CMS. By following these two standards, EHRs calculate and report the same information in the same format, allowing an "apples to apples" comparison.

This comparison is achievable because the HQMF is based on the framework of the Quality Data Model (QDM). As defined by the eClinical Quality Improvement Resource Center, the QDM is an "information model that defines relationships between patients and clinical concepts in a standardized format to enable electronic quality performance measurement." The QDM information framework contains the language necessary to represent a quality measure's individual data concepts in discrete, unambiguous terms using a structure of data types and attributes. For example, if a measure seeks to capture the number of pressure ulcers among patients at discharge, the QDM provides the structure and terminology to define an encounter datatype with attributes of discharge datetime and discharge status, and a pressure ulcer diagnosis datatype with attributes of onset and abatement datetime, anatomical location site, and severity. In this manner, the QDM's structure allows all quality measures to conform to the same language in order to be recognized by the HQMF, and it is updated as quality measures evolve and require new data concepts.

Secondly, measure developers began to translate their measure specifications into "eSpecifications," using a CMS-funded online tool called the Measure Authoring Tool (MAT). The MAT assists measure developers to identify each discrete clinical concept in a quality measure, map the concept to the QDM, and export the measure in the HQMF format. In other words, measure developers use the MAT the design electronic versions of their quality measures and make them consumable by EHRs.

When a health IT vendor constructs a quality measure report for an EHR based on the measure developer's "paper based" specification, the granularity, completeness, and other logic components of the measure may differ from another health IT developer's build of the same measure in a different EHR. Much like manual abstractors, the health IT developers use their own expertise and interpretations of the measure's requirements to represent the measure in their own product. However, when the health IT developers both use the HQMF format of measure, as translated by the measure developers, the calculation will draw on uniformly defined data elements and follow the same logic. When the health IT developers report the

measure using the QDRA standard, the recipient of the report receives the same data in the same structured format.

The HQMF and QDRA ensure standardized quality measure capture and reporting. However, this automation is reliant upon structured data capture in the EHR. The standards are also unable to address variance in clinical workflow that influences information capture in the EHR. **To build measures and measure reports that conformed to the HQMF and QRDA, in many cases, health IT vendors added data fields in the EHR.** Downstream, clinical workflows were modified to ensure the capture of critical measurement data fields. If a portion of a quality measure's data elements are not captured as structured data, they are not reflected in the measure's calculation and the report cannot accurately reflect the care provided to a patient.

To pursue full automation, the QSRS algorithms could be aligned with the industry standards adopted by the ONC Health IT Certification Program. This approach would enable automatic capture of data required by the QSRS algorithms in a format that is machine computable and reportable. Currently, only a small subset of the metadata required by the algorithms is represented in the CEHRT. Converting this metadata to information that can be captured in standard formats may involve supporting efforts to translate nursing flow sheet values and physician progress notes into standards, as the algorithms themselves have undergone substantial clinical vetting. This model takes an approach that leverages EHR data and industry practices to support widespread reporting.

7.1.2 Prioritize Automation of QSRS Entry Questions

In the analysis of both generic and module-specific entry questions, only 30 entry questions, 13 required NLP and 17 did not. Out of the 13 that required NLP, 7 questions have data sources that are likely structured and coded, leaving only 6 questions that require NLP as indicated in Figure 16 and 17 below.

Figure 16. Prioritized Entry Questions

Question	Module	Question ID Number	Numeric Value	Structured, Coded	Structured, Uncoded	Structured Text	Free Text	NLP Analysis Needed
Did the patient have a central venous catheter inserted or accessed during the stay?	ENTRY	GENERICEQ24			X	X		NO
Did the patient have a urinary tract catheter inserted during the stay?	ENTRY	GENERICEQ25			X	X		NO
Did the patient require ventilator support which was not present on admission?	ENTRY	GENERICEQ26			X	X		NO
After the first 24 hours of admission, did the patient require ventilator support?	ENTRY	GENERICEQ27			X	X		NO
Did the patient have an operating room procedure?	ENTRY	GENERICEQ28			X			NO
Did the patient receive a transfusion of blood or a blood product?	ENTRY - BLOOD or BLOOD PRODUCT	BLOODEQ1		X	X	X		NO
Was there a positive CDI Toxin A and/or Toxin B found in the patient's stool sample?	ENTRY - CDI	CDIEQ1		X	X			NO
Was there a toxin-producing CDI organism detected in the patients stool sample?	ENTRY - CDI	CDIEQ2		X	X	X		NO
Was the patient receiving comfort care?	ENTRY - VTE	VTEEQ1		X	X		X	YES
Did the patient have a prior or chronic VTE?	ENTRY - VTE	VTEEQ2		X			X	YES
On admission, did the patient have leg swelling?	ENTRY - VTE	VTEEQ3			X	X	X	YES
On admission, was a VTE prophylaxis ordered??	ENTRY - VTE	VTEEQ4		X	X		X	YES
Does documentation state that prophylaxis was contraindicated?	ENTRY - VTE	VTEEQ6		X	X		X	YES
Was the urinary catheter in place for > 2 days or more?	ENTRY-CAUTI	CAUTIEQ1		X		X		NO
Was the Central line or Umbilical catheter in place for greater than 2 days or more?	ENTRY - CLABSI	CLABSIEQ1		X	X	X		NO
On or within the first 24 hours of admission, was a skin inspection done?	ENTRY - PRESSURE ULCER	PUEQ1		X	X	X		NO
On or within the first 24 hours of admission, was a pressure ulcer risk assessment documented?	ENTRY - PRESSURE ULCER	PUEQ2		X	X	X		NO
Did the patient have any pressure ulcers that were present on admission or acquired during this stay?	ENTRY - PRESSURE ULCER	PUEQ3		X	X	X		NO

Figure 17. Entry Questions Prioritized for NLP Evaluation

Question	Module	Question ID Number	NLP Analysis Needed	NLP Analysis	NLP Complexity
Which, if any, of the following infections did the patient experience?	ENTRY - SSI	SSIEQ1	YES	Multiple concept detection; specific temporal occurrence (after surgery)	High
Was use of a device associated with an adverse outcome(s)?	ENTRY - DEVICE	DEVICE-EQ1	YES	Fuzzy concept	High
Did the patient fall one or more times?	ENTRY - FALL	FALLEQ2	YES	Presence of a concept/entity; Temporal occurrence of a concept (on admission)	Low
Was there a positive urine culture that reported >100,000 cfu/ml with no more than 2 species of microorganisms?	ENTRY - UTI	UTIEQ2	YES	Contingency	Medium
Was there a positive urine culture on Day 1 or Day 2?	ENTRY - UTI	UTIEQ3	YES	Multiple concept detection (N3); specific temporal Occurrence (N4)	High
On admission, did the patient have leg swelling?	ENTRY - VTE	VTEEQ3	YES	Presence of a concept/entity; Temporal occurrence of a concept (on admission)	Low

Focusing on automation of entry question will create a defined and smaller test and evaluation set for AHRQ and will yield the greatest value in that automation can be applied to a large number of EHR charts to process the most amount of records to identify encounters that would "enter" a module and require manual chart review and abstraction.

7.1.3 Identify Lowest Difficulty/Highest Value QSRS Modules

In addition to entry questions, AHRQ may then want to prioritize test and evaluation efforts on the modules that are of either the lowest difficulty (Figure 18) or highest value in terms of adverse events that often go undetected or are underreported in other hospital safety and reporting efforts, cause the most harm, or incur the highest costs, or are most easily prevented. Strategies for further research and evaluation may include:

- Identify QSRS modules that do not have related CMS, value-based care, National Quality Forum-endorsed measures. These modules may be a higher priority due to lack of other data
- In the development of pilots and proof-of-concept projects, prioritize QSRS modules that have the highest incidence AND have limited EHR reporting or have the most negative impact

Figure 18. Lower Levels of NLP Difficulty

## 7.2 Drive Real-World Implementation and Adoption

7.2.1 Leverage Commercial Technologies and Platforms

This study identified many industry efforts and commercially available tools and solutions that offer advanced analytics and/or NLP capabilities to the healthcare and government marketplace. Funding of pilots and efforts that utilize commercially available tools and solutions is encouraged to ensure that the analytics and NLP "learnings" can be applied to future AHRQ/CMS efforts or be available to health systems.

Many research organizations may have advanced in-house NLP expertise and capabilities; however, continued development of proof-of-concepts in research or academic "only" environments may hinder the general availability of solutions for the marketplace.

Figure 19 depicts a Technology Readiness Level approach used by NASA and the Department of Defense (DoD) in considering the evolution of technology from the research lab to the real world. This study identified many real-world solutions and capabilities that, although not available "out of the box" for QSRS automation today, are highly related to the capabilities needed for QSRS automation and should be leveraged. AHRQ is encouraged to fund any future efforts that clearly result in solutions in the higher levels of technology readiness (Levels 6–8), so that the results of pilots and proof-of-concepts funded by AHRQ can achieve the Level 9 goal of proven and successful use in operational environments.
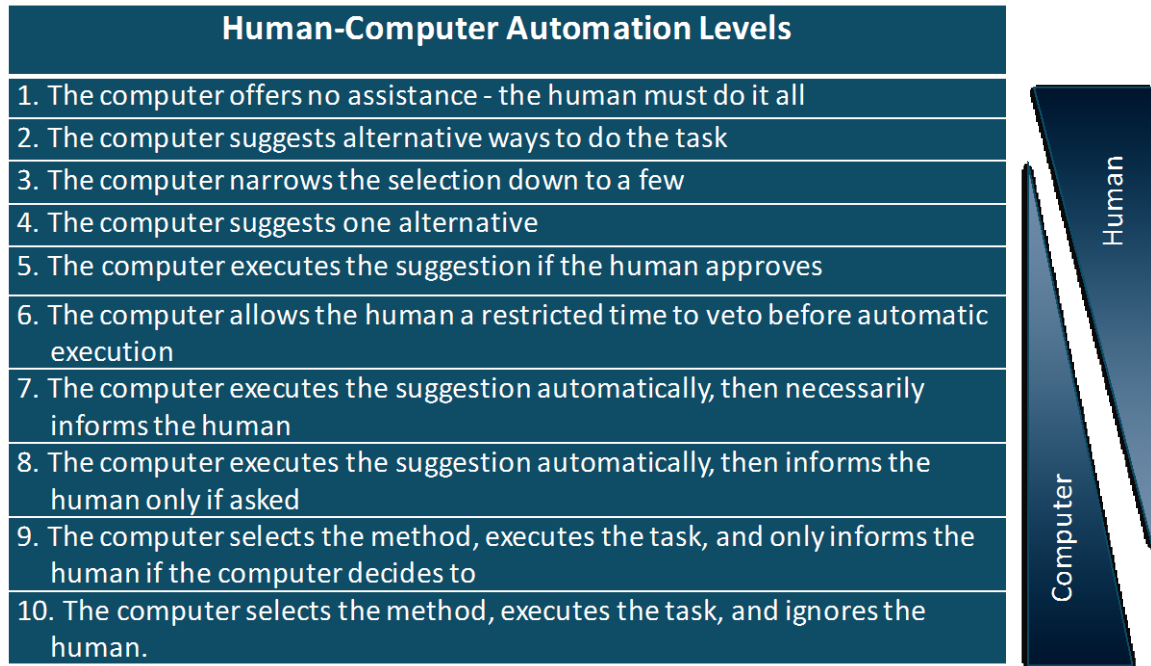
Figure 19. Technology Readiness Level Approach Used by NASA/DoD
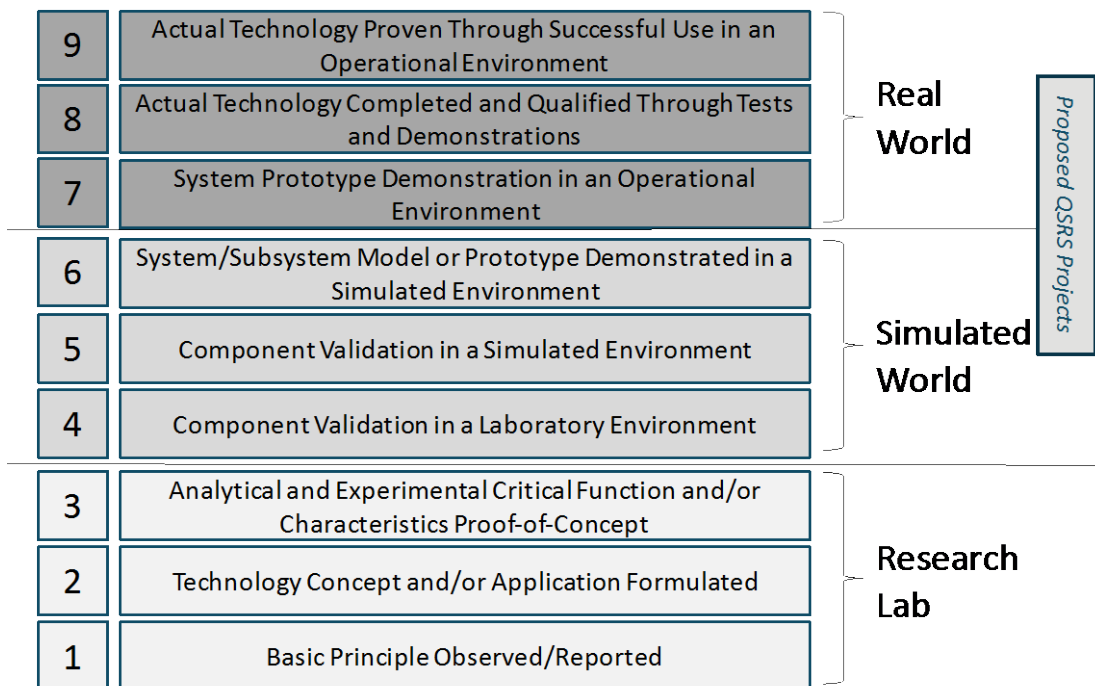
7.2.2 Consider Human-Computer Automation Levels

Humans have been shown to be far superior then computers at certain tasks, such as deciphering complex images. However, computers are much better at processing routine information. There exists a spectrum of human-computer interaction or decisionmaking strategies that could leverage the advantages of both humans and computer, as indicated in Figure 20. On one end of the spectrum the computer offers no assistance and the human must make all the decisions. On the other end, the computer makes all the decision without human involvement. The appropriate level of automation is often context specific and depends on many factors such as the difficulty and criticality of the task.

Figure 20. Human-Computer Automation Levels

Human-Computer Automation Levels
1. The computer offers no assistance - the human must do it all
2. The computer suggests alternative ways to do the task
3. The computer narrows the selection down to a few
4. The computer suggests one alternative
5. The computer executes the suggestion if the human approves
6. The computer allows the human a restricted time to veto before automatic execution
7. The computer executes the suggestion automatically, then necessarily informs the human
8. The computer executes the suggestion automatically, then informs the human only if asked
9. The computer selects the method, executes the task, and only informs the human if the computer decides to
10. The computer selects the method, executes the task, and ignores the human.

Sheridan TB, Parasuraman R. Human-automation interaction. Reviews of Human Factors and Ergonomics. 2005;1(1):89-129.

QSRS automation feasibility is recommended to begin with "human in the loop" models, where there are suggestions from the computer that are validated by the human. Once accuracy levels are verified, AHRQ can consider increasing the level of automation.

As discussed earlier, an area for focus for "full automation" are the entry questions, or leveraging approaches used by UHC/Vizient and Quintiles where billing data from hospitals is used to identify or flag the charts that should be abstracted. In this model, the initial processing of a large number of charts is fully automated, and those that meet specified criteria would then require human interaction.

7.2.3 Sampling

One challenge of model development is find the balance between overfitting and "underfitting" data. This is done by developing models on a sample of data and testing the model on another set of data. Figure 21 below provides an example of a typical machine learning training and testing iteration. Typically, 60–70 percent of the data is selected randomly for model development and the remaining data is used for testing. In addition to random sampling, there are other techniques of selecting and testing data such as k-fold validation[51] and selecting balanced training sets.[52] Choosing between these options depends on several factors such as the complexity of the modeling task and the amount of data available.

Figure 21. Human-Computer Automation Levels
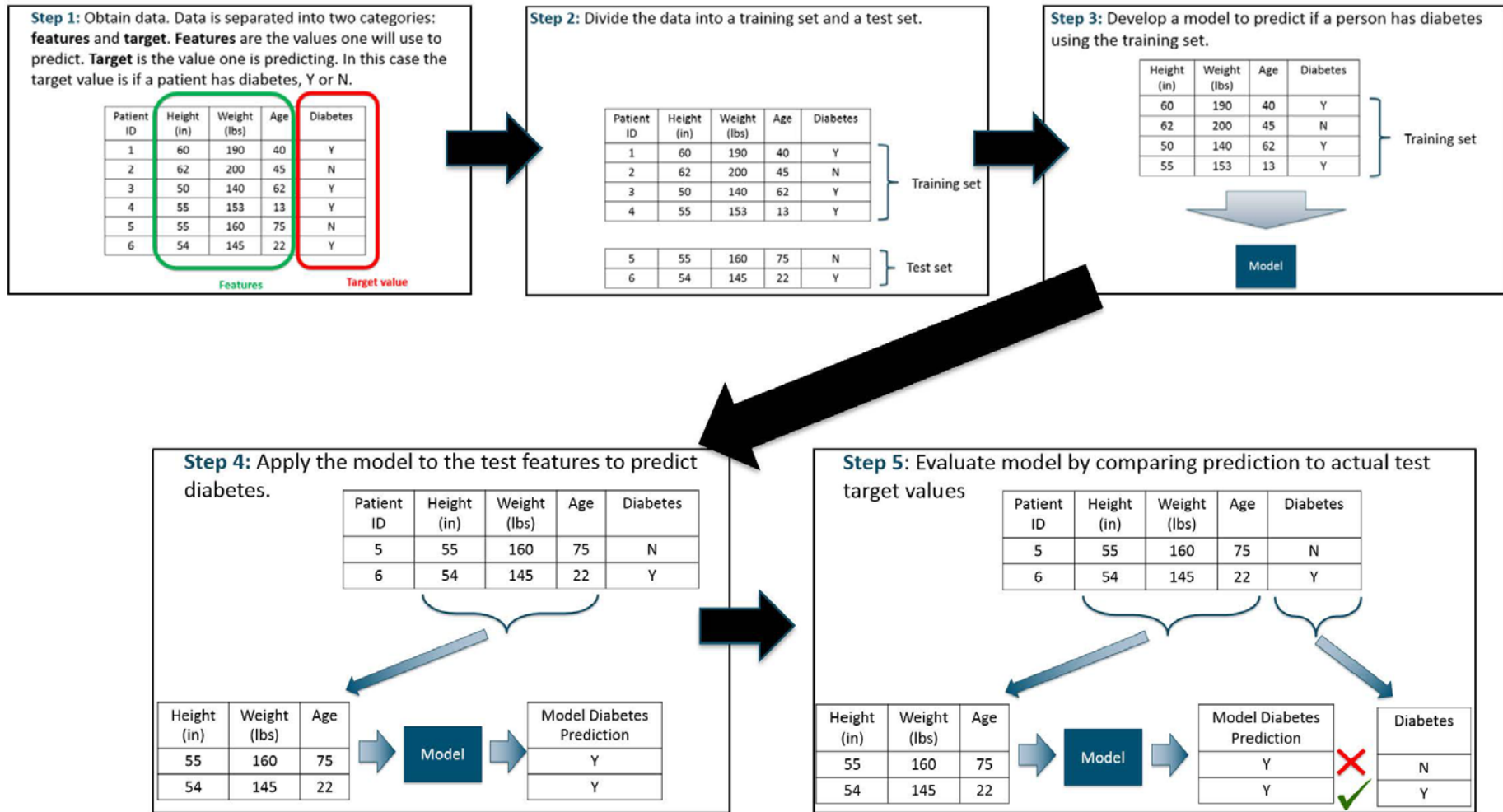
7.3 Pilots and Projects

7.3.1 Proof-of-Concept Project: Test Rule-Based and NLP Approaches To Populate QSRS

Purpose: *Identify effective interface designs to support QSRS automation*. Determine whether potential automated approaches will work by piloting on a sample of data from different EHRs (Figure 22 shows the proposed process).

Dataset: Select 1,000 charts from 3 different vendors and have development teams develop algorithms to populate QSRS.

Metrics:

- Accuracy, Precision, F1-score, Recall/Precision at K
- These outcome metrics measure performance of the model(s)

Outcome:

- The outcome of this pilot are prediction model(s) and their associated performance

Recommendations:

- Apply approach to entry questions
    - Advantages: Gatekeeper questions
    - Disadvantages: Broad scope
- Apply approach to "Blood or Blood Product" and "HAI-CDI"
    - Advantage: Lowest number of Low and Medium NLP questions, specific set of questions, more narrow search space
    - Disadvantage: Opportunity for increase false positives

Estimated Timeframe:

- The complete pilot can take an estimated 0.5 to 1.25 years
- (~2-4 months) Acquiring data (getting organizational agreements signed, security protocols, institutional review board and other approvals, etc.) – variability due to organizational policies and protocols
- (~1-4 months) Annotating data (coder time, establishing interrater reliability (IRR), etc..) – variability due to number of codes and records and difficulties with IRR
- (~2-4 months) Develop and apply model – variability due to data and model characteristics
- (~1-3 months) Validate outputs – variability due to number of codes and records

Figure 22. Proposed Process for Identifying Effective Interface Designs To Support QSRS Automation
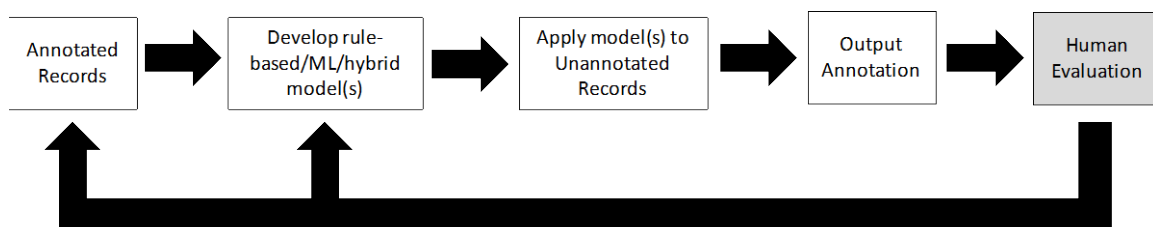
Given the complexities of fully automating the extraction of patient information from the EHR to populate QSRS, other mechanisms to support partial automation and abstractor interaction with QSRS should be examined. The automation of QSRS will likely rely on an algorithmic approach and the algorithms will have an associated probability of detecting the appropriate information from the patient record and will have an associated margin of error. This information can be used in several ways to facilitate the abstractor's task

of examining the patient record to answer QSRS questions. It will be critical to determine the most effective methods for presenting this information to abstractors. We recommend the following:

    A. An in-depth environmental scan of the human-computer interaction literature and current automated systems to determine effective abstractor interaction methods (e.g., highlighting of relevant clinical text with likely information necessary for QSRS questions, providing computer generated responses with associated probability values etc.).
    B. Prototype development of high value methods discovered in the environmental scan
    C. User testing with abstractors to determine usability and effectiveness of identified methods.

7.3.2 Pilot Project: Test Partial Automation Interface-Level Guidance

Purpose: There are different levels of automating QSRS, ranging from automatic identification of the correct information to human-in-the-loop processes like highlighting where the appropriate information *might be found* (Figure 23).

Action: Test different interface level representations on end-users to determine which are most effective. Consider selecting from a set of established vendors or organizations that partner with vendors such as the vendors considered in this report. Select one partner or run multiple parallel pilots to discover how predictive analytics can be applied.

Dataset: 1,000 charts (already annotated charts preferred though not required)

Metrics (Dependent Variable):

- Time to complete, accuracy, subjective response, number of times coder revisits the same data source, number of transitions between data sources, mouse clicks, workload metrics (NASA TLX)
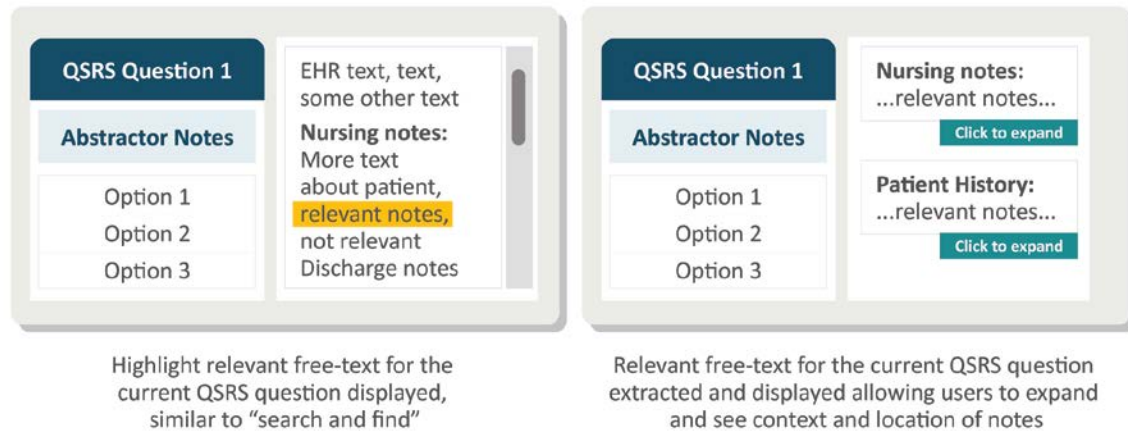- Additional workload metrics could be eye tracking, heart rate monitoring, etc.

Estimated Timeframe:

- The complete pilot can take an estimated 0.5 to 1 years
- (~1-3 months) Acquiring data (getting organizational agreements signed, security protocols, institutional review board and other approvals, etc.) – variability due to organizational policies and protocols
- (~2-5 months) Develop interfaces – variability due to the number and complexity of the interfaces
- (~1-2 months) Run experiment – variability due to recruitment of coders
- (~1-2 months) Data analysis – variability due amount of metrics collected

Competition Approach

- Establish a competition with a common dataset from which different development teams employ their respective approaches to determine which team can perform with the highest accuracy
- Resources/Teams Needed:
  - Lead to manage and run competition
  - Competition participants
  - Lead for evaluation of competition and results

Figure 23. NLP-Driven Human Guidance

Highlight relevant free-text for the current QSRS question displayed, similar to "search and find"

Relevant free-text for the current QSRS question extracted and displayed allowing users to expand and see context and location of notes

To develop a deeper understanding of the feasibility of automating QSRS development teams should actively develop algorithms with actual patient records to discover associated challenges. We recommend having several development teams working to create algorithms given the diverse expertise and various approaches that could be employed. One possibility is to establish a competition with a common dataset from which different development teams employ their respective approaches to determine which team can perform with the highest accuracy. To create a process whereby different development teams create possible algorithmic solutions, we recommend—

A. Creating a diverse set of data from which to work from. This dataset should include at least 1,500 records stemming from different EHRs so that the data reflect the differences in which EHR developers represent information and the different ways providers use EHRs. These data should be manually abstracted so that there is "ground truth" as to which QSRS events are present for each record in this dataset. One thousand five hundred records is the minimum number of records recommended. More data are always better for algorithm development.

B. 1,000 of these records should be available to the development teams along with QSRS and the associated manuals.

C. The development teams should be permitted to employee any approach they deem appropriate to successfully answer the QSRS queries.

D. The remaining 500 records should be used for evaluation. The algorithms developed by each team, based on the 1,000 provided records, should be applied to the 500 records that were not part of the algorithm development. Accuracy should be assessed by comparing the human annotation of those 500 records to the algorithm outputs.

7.3.3   Comparison Study: Automation with One Vendor Versus Manual Abstraction

Consider the approach described above in 7.3.2 across multiple vendors for a single vendor to automate as much of QSRS possible, prioritizing the modules indicating in this report. AHRQ currently does not have much visibility into the human abstraction pilot; funding a study where manual abstraction is performed alongside NLP/Automation approaches—whether human-assisted (highlighting information for human validation), fully automated (using claims data or other EHR data to identify charts for abstraction or analysis).

This approach could yield useful information for AHRQ including—

- Should AHRQ/CMS identify datasets or EHR documentation practices to facilitate information capture where automation feasibility or accuracy is low?
- Can automation yield more accurate results for certain modules in comparison to abstraction or vice versa?

- Should AHRQ/CMS implement automation centrally or leave automation strategies and implementation to individual hospitals and health systems?
- What types of automation are ready for "prime time" and which modules are the best candidates? (Level of automation/NLP)

7.3.4 Comparison Study: C-CDA and HL7 ADT Interfaces Versus Manual Abstraction

As EHR documentation is duplicative in nature, rather than using the approach of "all the possible locations data may reside," conduct a study to determine the efficacy of available electronic data using standards formats such as the C-CDA (Consolidated Continuity of Care Documents), HL7 ADT (Admit-Discharge-Transfer), and billing/claims transactions to determine if the needed information can support QSRS functions and needs. Conduct a parallel portion of the study where manual abstractors are reviewing the same charts that are processed electronically.

- Perform pilot to reverse engineer QSRS questions using only data available within EHR A and EHR B C-CDAs from either Discharge Summary/Summary of Care or Data Portability
- Perform parallel pilot with either CDAC abstractors or other coders who review the same charts using current manual review/abstraction process
- Consider a competition approach where the electronic EHR source data is provided to multiple vendors/participants

7.3.5 Additional Research/Pre-Work

As this study focused on QSRS automation using EHRs as the data source, there are a number of other hospital data sources and systems to consider review to assess potential duplication of QSRS and/or opportunities to align the QSRS information collection process with other reporting programs. AHRQ many want to consider additional research in the following areas:

A. **Mini-Automation Feasibility Assessment That Considers Non-EHR Data Sources:** Scaled down environmental scan to ensure tools used by Performance Improvement, Quality Improvement, Risk Management teams in hospitals are considered
   - Voluntary Reporting Systems for Safety
   - Chart Abstraction Vendors for Quality
   - Core Measure Reporting Systems
   - Claims/billing data-assisted identification of charts for abstraction
   - Core measures, accreditation review and comparison vs. QSRS (Where may QSRS be duplicative?)

B. **Risk Management/Insurance Premium Opportunities for Use of QSRS Surveillance Data**
   - Conduct discussions with multiple health systems and AHRQ

C. **Review of Hospital Resources and Tools for Predictive Versus Retrospective Chart Analysis**
   - Consider how QSRS can be applied or extended for more predictive analysis vs. retrospective review

7.4 Conclusions

The fundamental goals and variance in reporting based upon defined datasets and mapping of EHR data elements versus review of all documentation (performed by anyone, anywhere) is core to the automation feasibility discussion.

As QSRS is used for surveillance purposes, and is retrospective, the value of QSRS for EHR-stakeholders (e.g., provider organization leads for EHR optimization and quality, vendors) was cited to be limited, as EHR-focused resources are prioritized for predictive and real-time information analysis. With the goals of QSRS as stated and currently planned for surveillance, partial automation is feasible in the areas discussed within this section.

Provider organizations cited tremendous burden for quality reporting and chart abstraction and welcomed the notion of automated reporting, yet also indicated the significant burden of reporting and mapping EHR data for multiple measure reporting programs.

AHRQ should consider the areas for further study identified within this report, using the strategies provided to focus and prioritize efforts that may yield the most value in the near-term. NLP and text-based automation is available and improving, but still evolving and not yet optimal to address highest complexity, open-ended QSRS questions.

8 Appendixes

8.1 Appendix A. Environmental Scan Keywords and Journals

Patient safety surveillance, patient safety reporting, EHR surveillance, patient safety event identification, secondary uses of EHR data, natural language processing AND safety, machine learning algorithms, automated adverse event reporting, trigger tools, algorithm automated identification, clinical research informatics, phenotype algorithm, automated electronic search, algorithm AND safety, EHR active surveillance, EHR and chart extraction (and abstraction), adverse report (reporting, and reports), incident report (reporting, and reports)

LIST OF JOURNALS	
INFORMATICS JOURNALS	Advances in Bioinformatics
	Applied Clinical Informatics
	Bioinformatics
	Bioinformatics Oxford England
	BMC Bioinformatics
	BMC Medical Informatics and Decision Making
	Computer Methods and Programs in Biomedicine
	Computers in Biology and Medicine
	Health Information Science and Systems
	Healthcare Informatics Research
	International Journal of Health Care Quality Assurance
	International Journal of Medical Informatics
	JMIR (Journal of Medical Internet Research) Medical Informatics
	Journal of Biomedical Informatics
	Journal of Clinical Informatics
	Journal of Healthcare Risk Management of the American Society for Healthcare Risk Management
	Journal of Medical Systems
	Journal of the American Medical Informatics Association (JAMIA)
	Methods of Information in Medicine
	Online Journal of Public Health Informatics
	Studies in Health Technology and Informatics
SAFETY AND QUALITY JOURNALS	BMJ Quality and Safety
	Journal of Patient Safety
	Journal of Safety Research
	Joint Commission Journal on Quality and Patient Safety

8.2 Appendix B. NLP Vendor Summary Findings

The natural language processing (NLP) market within health care is growing with vast opportunities for advancement for vendors, providers, and consumers. The current health-care–specific NLP market provides an array of products and services that range from supporting enhanced evidence-based decision making to health analytics and quality metrics. NLP tools and services are designed and marketed to advance accuracy, reduce documentation burdens, enhance patient safety, unlock insights from unstructured data, and improve overall quality of care.

#	Vendor Groupings	Industry	Product/Services	Application of NLP Technology

1	**Speech Recognition/ NLP Vendors**	Technology-Enabled Healthcare Solutions	• Transcription
			• CDI
			• Imaging
			• Practices
			• Coding
			• Dictation
			• Speech Recognition Adoption Services
				• CDI, CAC, speech recognition, transcription
				• Speech Understanding
				  o Combines NLP with speech recognition technology for enhanced evidence-based decision making
				• Clinical documentation improvement (CDI) and information integration
				  o Access information from unstructured report narratives
				  o Aggregates data from clinical systems such as EHR
				  o Quality measure reporting, patient outcomes
				  o Supports CDI by tracking, analyzing, reporting on issues over time and by physician
				• Overall aim to deliver actionable intelligence from report narratives and structure data sources to improve gains in productivity, quality scores and decrease errors

2		Computer Software Technology	• Clinical Language Understanding Technology • Clinical Narrative Capture • Diagnostic Imaging • CDI, Quality & Coding	• CDI, CAC, quality management • Leverages advancements in NLP, medical Artificial Intelligence (AI) and speech recognition technology to create tailored solutions for healthcare informatics • Medical record coding, quality management and CDI services • On-demand speech and clinical language understanding on a worldwide basis

#	Vendor Groupings	Industry	Product/Services	Application of NLP Technology
3	**Predictive Analytics and Risk Management/ NLP Vendors**	IT and Services	• Risk Management Solutions for MA, Commercial/ ACA, Medicare ACO Medicaid • NLP and analytics technology	• Enables automated extraction of usable data from unstructured clinical text to SNOMED • Uses NLP to assist in risk adjustment for ACOs, Medicare Advantage, etc.

4		Cognitive Computing Company	• Risk Management Solutions – HCC profiler
			• Patient Modelling
				• Automated extraction, analytics, coded data, risk adjustment
				• Patented technology platform includes:
				  o Comprised of the latest natural language processing (NLP) and machine learning technologies, the analytics engine uses their Patient Model to profile healthcare, determine risk and quality, assess care decisions, and quantify performance.
				  o Data extraction tools, a scalable analytics pipeline, semantic concept extraction, computer learning infrastructure, flexible micro-service APIs, and optimized expert review workflows

		Predictive Analytics	• ICD-10 Planning and Impact Analysis • Predictive Analytics	• Standard dataset request from hospitals • Apply NLP/model derived from 400 hospitals to perform cohort classification, risk stratification, and apply clinical rules • Real-time alerts delivered at point-of-care to prevent AEs or reduce readmission risk
5	**Querying and Text Analytics/ NLP Vendors**	Computer Software Technology	• Text mining platform	• Intuitive, interactive querying based on natural language processing (NLP): quickly interpreting the meaning of unstructured text sources and returning high value, relevant results • NLP based text mining for high-value knowledge discovery and decision support (CDSS) • Produces software and solutions to help the pharma-biotech and healthcare industries speed up the drug-discovery cycle and improve patient outcomes

#	Vendor Groupings	Industry	Product/Services	Application of NLP Technology
6	**Querying and Text Analytics/ NLP Vendors (continued)**	Health Services and Innovation	• NLP Technology	• CDI, CAC, record review • Uses NLP for record review to pinpoint cases for physician queries and improve clinical documentation output
7		Healthcare Software Developers	• Scribe, CAC	• CDI, CAC, transcription • Uses NLP to code inpatient and outpatient charts • Coder automation, productivity • Query functionality • Interactive edits
8	**Knowledge Solutions and Health Analytics/NLP Vendors**	Mechanical, Engineering	• Natural Language Processing (NLP) Platform • UIMA (Unstructured Information Management Architecture) – open source standardized and NLP solution	• CDI, CAC, quality metrics, health analytics • Coders use knowledge/expertise to transform data into valuable information • Algorithms and models allow machines to start with existing sources of knowledge, analyze new data, improve capabilities • NLP allows for more effective 'human to human' communication • CAC enabling NLP allows useful codes to be proposed and coders to verify and edit • Leverages patient information in clinical documentation to improve communication between caregivers, reduce the cost of working with clinical documentation, and automate coding and documentation process • Hybrid model of NLP of rules-based and statistical models
9		IT and Services	• NLP and Machine Learning Technology Platform	• Automated extraction, analytics, CDSS, open source • Extract key data from documents • Reveal insights, patterns, relationships across data • Analysis of unstructured data • NLP is used to understand grammar and context • Searches documents to find answers • Collects evidence and uses scoring algorithm to rate quality of this evidence • Ranks possible answers based on score of supporting evidence • Advanced image analytics

#	Vendor Groupings	Industry	Product/Services	Application of NLP Technology
10	**Knowledge Solutions and Health Analytics/NLP Vendors (continued)**	IT and Services	• Health Analytics	• Algorithm to process text, health analytics, open source • Efficient algorithms to process texts and make information accessible to computer applications • Uses NLP and machine learning to speed up diagnosis of pneumonia in ICU patients • Use EHR of patients whose pneumonia diagnosis is established by clinical consensus and employ NLP tools to identify critical clinical info – want to test if pneumonia can be diagnosed based on automatic review of digital medical records
11		IT and Services	• Knowledge Solutions • EHR Vendor • Hospitals & Health Systems • Continuum of Care • Medical Devices • Member Engagement • Physician Practice • Pharmacies • Research • Population Health • Workplace Health	• CDI, CAC • Automated collection of patient information • Identifies and interprets complex concepts, phrases and medical terminology from digital free-text clinical documentation • Works with notes from voice recognition systems, transcription, or created from directly entered narrative documentation • Understand grammar, syntax, synonymy and phraseology • Extracts appropriate SNOMED-CT codes, ICD-9, CPT, E&M codes for automating the coding, billing process

8.3 Appendix C. References

1. The Office of the National Coordinator for Health Information Technology. Adoption of Electronic Health Records Systems among U.S. Non-Federal Acute-Care Hospitals: 2008-2014. ONC Data Brief No. 23. April 2015. https://www.healthit.gov/sites/default/files/data-brief/2014HospitalAdoptionDataBrief.pdf.

2. Avillach P, Coloma PM, Gini R, et al. Harmonization process for the identification of medical events in eight European healthcare databases: the experience from the EU-ADR project. *J Am Med Informatics Assoc*. 2013;20(1):184-92. doi:10.1136/amiajnl-2012-000933.

3. Classen DC, Jaser R, Budnitz DS. Adverse Drug Events Among Hospitalized Medicare Patients: Epidemiology and National Estimates from a New Approach to Surveillance. *Jt Comm J Qual Patient Saf*. 2010;36(1).

4. Kahn MG, Ranade D. The impact of electronic medical records data sources on an adverse drug event quality measure. *J Am Med Inform Assoc*. 2010;17(2):185-91. doi:10.1136/jamia.2009.002451.

5. Wang X, Hripcsak G, Markatou M, Friedman C. Active Computerized Pharmacovigilance Using Natural Language Processing, Statistics, and Electronic Health Records: A Feasibility Study. *J Am Med Informatics Assoc*. 2009;16(3):328-37. doi:10.1197/jamia.M3028.

6. Li Q, Melton K, Lingren T, et al. Phenotyping for patient safety: algorithm development for electronic health record based automated adverse event and medical error detection in neonatal intensive care. *J Am Med Inform Assoc*. 2014;21(5):776-84. doi:10.1136/amiajnl-2013-001914.

7. Carrell DS, Cronkite D, Palmer RE, et al. Using Natural Language Processing to Identify Problem Usage of Prescription Opioids. *Int J Med Inform*. 2015;84(12):1057-64. doi:10.1016/j.ijmedinf.2015.09.002.

8. Singh R, McLean-Plunckett E a, Kee R, et al. Experience with a trigger tool for identifying adverse drug events among older adults in ambulatory primary care. *Qual Saf Health Care*. 2009;18(3):199-204. doi:10.1136/qshc.2007.024406.

9. Mull HJ, Nebeker JR, Shimada SL, Kaafarani HM a, Rivard PE, Rosen AK. Consensus building for development of outpatient adverse drug event triggers. *J Patient Saf*. 2011;7(2):66-71. doi:10.1097/PTS.0b013e31820c98ba.

10. Liu M, McPeek Hinz ER, Matheny ME, et al. Comparative analysis of pharmacovigilance methods in the detection of adverse drug reactions using electronic medical records. *J Am Med Informatics Assoc*. 2013;20(3):420-6. doi:10.1136/amiajnl-2012-001119.

11. Iyer S V, Harpaz R, LePendu P, Bauer-Mehren A, Shah NH. Mining clinical text for signals of adverse drug-drug interactions. *J Am Med Informatics Assoc*. 2014;21(2):353-62. doi:10.1136/amiajnl-2013-001612.

12. Li Y, Salmasian H, Vilar S, Chase H, Friedman C, Wei Y. A method for controlling complex confounding effects in the detection of adverse drug reactions using electronic health records. *J Am Med Informatics Assoc*. 2014;21(2):308-14. doi:10.1136/amiajnl-2013-001718.

13. Kilbridge PM, Noirot L a, Reichley RM, et al. Computerized surveillance for adverse drug events in a pediatric hospital. *J Am Med Inform Assoc*. 2009;16(5):607-12. doi:10.1197/jamia.M3167.

14. Ficheur G, Chazard E, Beuscart J-B, Merlin B, Luyckx M, Beuscart R. Adverse drug events with hyperkalaemia during inpatient stays: evaluation of an automated method for retrospective detection in hospital databases. *BMC Med Inform Decis Mak*. 2014;14(1):83. doi:10.1186/1472-6947-14-83.

15. Boytcheva S. Shallow medication extraction from hospital patient records. *Stud Heal Technol Inf*. 2011;166:119-28. doi:10.3233/978-1-60750-740-6-119.

16. Tseytlin E, Mitchell K, Legowski E, Corrigan J, Chavan G, Jacobson RS. NOBLE - Flexible concept recognition for large-scale biomedical natural language processing. *BMC Bioinformatics*. 2016;17(1):32. doi:10.1186/s12859-015-0871-y.

17. Hota B, Lin M, Doherty J a, et al. Formulation of a model for automating infection surveillance: algorithmic detection of central-line associated bloodstream infection. *J Am Med Inform Assoc*. 2010;17(1):42-8. doi:10.1197/jamia.M3196.

18. Donnan PT, Dorward DWT, Mutch B, Morris AD. Development and validation of a model for predicting emergency admissions over the next year (PEONY): a UK historical cohort study. *Arch Intern Med*. 2008;168(13):1416-22. doi:10.1001/archinte.168.13.1416.

19. Freeman R, Moore LSP, García Álvarez L, Charlett A, Holmes A. Advances in electronic surveillance for healthcare-associated infections in the 21st Century: a systematic review. *J Hosp Infect*. 2013;84(2):106-19. doi:10.1016/j.jhin.2012.11.031.

20. Tseng YJ, Wu JH, Lin HC, et al. Rule-based healthcare-associated bloodstream infection classification and surveillance system. *Stud Heal Technol Informatics*. 2013;186(1):145-9. doi:10.3233/978-1-61499-240-0-145.

21. Kaafarani HM a, Rosen AK, Nebeker JR, et al. Development of trigger tools for surveillance of adverse events in ambulatory surgery. *Qual Saf Health Care*. 2010;19(5):425-9. doi:10.1136/qshc.2008.031591.

22. Rosen A, Mull H, Kaafarani H. Applying trigger tools to detect adverse events associated with outpatient surgery. *J Patient Saf.*. 2011;7(1):45-59. doi:10.1097/PTS.0b013e31820d164b.

23. Choi J, Kim H. Enhancement of decision rules to increase generalizability and performance of the rule-based system assessing risk for pressure ulcer. *Appl Clin Inform*. 2013;4(2):251-66. doi:10.4338/ACI-2012-12-RA-0056.

24. Gifford J, Delisle S. Application of an automated surveillance system to evaluate outpatient care of pneumonia. *Online J Public Health Inform*. 2015;7(1):e76.

25. Niland JC, Stiller T, Neat J, Londrc A, Johnson D, Pannoni S. Improving patient safety via automated laboratory-based adverse event grading. *J Am Med Informatics Assoc*. 2012;19(1):111-5. doi:10.1136/amiajnl-2011-000513.

26. Rochefort CM, Verma a. D, Eguale T, Lee TC, Buckeridge DL. A novel method of adverse event detection can accurately identify venous thromboembolisms (VTEs) from narrative electronic health record data. *J Am Med Informatics Assoc*. 2014:155-65. doi:10.1136/amiajnl-2014-002768.

27. Bodenreider O. The unified medical language system (UMLS):integrating biomedical terminology. *Nucleic Acids Reserach*. 2004.

28. Liu S. RxNorm: perscpription for electronic drug information exchange. *IT Prof*. 2005:17-23.

29. Manning CD. *Foundations of Statistical Natural Language Processing*. MIT Press Cambridge, MA, USA; 1999.

30. Uzuner O. Evaluating the state of the art in coreference resolution for electronic medical records. *J Am Med Informatics Assocation*. 2012:786-91.

31. Sun W, Rumshisky A, Ozlem U. Evaluating temporal relations in clinical text: 2012 i2b2 challenge. *J Am Med Informatics Assocation*. 2013:806-13.

32. Hahn U, Inderjeet M. The challenges of automatic summarization. *Computer (Long Beach Calif)*. 2000;33(11):29-36.

33. Centers for Medicare & Medicaid Services. Promoting Interoperability. May 31, 2018. https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentiveProgram s/index.html?redirect=/ehrincentiveprogra ms.

34. Anthony L. In an EHR Era,adverse event reporting takes on a new look. Pharmaceutical Commerce. Dec. 27, 2012. http://pharmaceuticalcommerce.com/legal _regulatory?articleid=26717

35. Centers for Medicare & Medicaid Services. Clinical Quality Measures Basics. May 8, 2018. https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentiveProgram s/ClinicalQualityMeasures.html.

36. Roop ES. Addition by extraction. For The Record. 2013 Oct;25(14):18. http://www.fortherecordmag.com/archives /1013p18.shtml.

37. Wu L. Issue Brief: Health IT for Public Health Reporting and Information Systems. The Office of the National Coordinator for Health Information Technology. April 29, 2014. https://www.healthit.gov/sites/default/files /phissuebrief04-24-14.pdf.

38. EHR Data FAQs. NIH Collaboratory. https://www.nihcollaboratory.org/Pages/e hr-data-faqs.aspx.

39. Electronic Medical Record Support for Public Health. http://esphealth.org.

40. Mining EHR data for quality improvement. Medical Economics. March 13, 2015. http://medicaleconomics.modernmedicine. com/medical-economics/news/mining-ehr-data-quality-improvement?page=full.

41. Monegain B. Natural language processing in high demand. Healthcare IT News. Aug. 14, 2015. http://www.healthcareitnews.com/news/n atural-language-processing-demand.

42 Bonney S. Practical applications of NLP in healthcare: enhancing the value of electronic data. Porter Research. July 30, 2015. http://porterresearch.com/practical-applications-of-nlp-in-healthcare-enhancing-the-value-of-electronic-data/.

43. Roop ES. A powerful combination. For the Record. 2012 Oct. 8;24(18):18. http://www.fortherecordmag.com/archives /100812p18.shtml.

44. Impact Advisors, LLC. Emerging Technologies in Healthcare. May 2014. http://www.scottsdaleinstitute.org/docs/sp onsors/ImpactAdvisors/2014-May.Emerging-Technologies-in-HC.pdf.

45. Finding Value in Volume: An Exploration of Data Access and Quality Challenges. AcademyHealth. March 8, 2012. https://www.academyhealth.org/sites/defa ult/files/publications/files/publications/HIT 4AKAccessandQual.pdf.

46. Measure Applications Partnership. Cross-Cutting Challenges Facing Measurement: MAP 2015 Guidance. March 2015. http://www.qualityforum.org/Projects/i-m/MAP/Measure_Applications_Partnership /Cross-Cutting_Challenges_Facing_Measurement_ Final_Report.aspx.

47. Buying Value Measure Selection Tool. http://www.buyingvalue.org.

48. CMS Measures Inventory. Centers for Medicare & Medicaid Services. March 27, 2018. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/CMS-Measures-Inventory.html

49. Part II: Measure Madness. Measures That Matter.Healthcare Association of New York State (HANYS). Accessed June 28, 2018. http://www.hanys.org/quality/clinical_op erational_oversight/measures_that_matt er/docs/mtm_measure_madness.pdf.

50. Chassin MR, Loeb JM, Schmaltz SP, et al. Accountability measures -- using measurement to promote quality improvement. *New England Journal of Medicine.* 2010 Aug 12;363:683-8. http://www.nejm.org/doi/full/10.1056/NEJMsb1002320.

51. Cross-validation (statistics). Wikipedia. June 21, 2018. https://en.wikipedia.org/wiki/Cross-validation_(statistics).

52. Wei Q, Dunbrack RL, Jr. The role of balanced training and testing data sets for binary classifiers in binary informatics. PLoS One. 2013;8(7):e67863. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3706434.