By Amal A. Alzu’bi, PhD; Valerie J. M. Watzlaf, MPH, PhD, RHIA, FAHIMA; and Patty Sheridan, MBA, RHIA, FAHIMA

Abstract

The purpose of electronic health record (EHR) abstraction includes collection of data related to administrative coding functions, quality improvement, clinical registry functions and clinical research. This article examines the different abstraction methods, such as manual abstraction, simple query, and natural language processing (NLP). It also discusses the advantages and disadvantages of each of those methods. The process used for successful EHR abstraction is also discussed and includes the scope and resources needed (time, budget, type of healthcare professionals RHIA, RHIT, etc.). The relationship between EHRs and the clinical registry is also examined with a focus on validity of the data extracted. Future research in this area to examine abstraction methods across hospitals who do data abstraction are being finalized for a future publication.

Keywords: Electronic health record, EHR, abstraction, natural language processing, NLP, query, quality improvement, patient safety.

Introduction

The widespread adoption of electronic health record (EHR) systems makes it possible to retrieve patient records digitally and to extract useful clinical data. Therefore, several secondary applications have become accessible such as quality management, health management and translational research. All of these secondary applications aim to improve patient care. The overall quality of healthcare and patient treatment depends heavily on the quality of data. Therefore, having inaccurate, incomplete, and inconsistent data and documentation can result in errors and adverse events that may affect patient safety, limit health information exchange (HIE), and hinder clinical research.

In the remainder of this article, we discuss the methods of data abstraction. Advantage and disadvantages of data abstraction, the key factors for successful abstraction within EHRs and registries will also be discussed. We also discuss the importance of having a health information management professional at the forefront of clinical data abstraction methods.

Background

The Centers for Medicare & Medicaid Services (CMS) developed the Meaningful Use program, now called Promoting Interoperability, to help healthcare professionals and hospitals improve quality, efficiency, and safety of patient health through the use of certified EHRs. Thus, it is increasingly important to get meaningful and efficient methods for data collection, sharing and reporting. Furthermore, it is important to have efficient methods for abstracting the needed clinical data from
EHR systems and other clinical documents. Three methods that can be used for data abstraction include, manual abstraction, search engines, and abstraction using natural language processing (NLP). Several measures can be used to evaluate the quality of each one of these methods, including, completeness, correctness, concordance, currency, and plausibility.9

**Abstraction Methods**

**Manual Record Abstraction.** Manual abstraction is the process of collecting important information from a medical record and transcribing it into discrete fields within the new EHR. Structured data parts (coded data) in the EHR such as, medications, diagnoses, and an active medication allergy list, helps to abstract the needed information from EHR systems.10 Manual abstraction can be performed by health information management professionals, nurses, physicians, or other individuals who have training in data abstraction. The review of the entire medical record allows one to collect more specific clinical details, especially for the information that is not readily coded using the existing coding systems.11

Manual abstraction helps to integrate discrete patient data into the EHR and make them readily available for healthcare providers. It also allows for triggering some decision support alerts that are related to information integrated into the EHR.12

Although the manual abstraction method is convenient and easy to understand, it has several limitations:

1. Some outcome measures, such as those related to cancer recurrence, are usually documented in unstructured notes and reports. Thus, it will be hard to extract all the required information since there are limited structured parts.13,14

2. Manual abstraction can be time consuming and expensive.

3. Manual abstraction methods threaten the privacy of patient information.15

4. Manual abstraction may increase errors. For compliance with CMS regulations, clinical data abstractors need to review patient records to identify the ones that meet the guidelines.16 Reviewing this large volume of data increases the risk of making errors. The situation will be worse when dealing with narrative information that lacks standardization. Therefore, data abstractors may get unreliable results.

**Search Engines (Simple query).** Several studies have shown that physicians prefer to enter their comments in some unstructured free-text entries even if there are options for structured coding elements in the system.17,18 Additionally, unstructured free text entries are always required for some
complex tasks such as, clinical trial recruitment. Extracting some useful clinical data from such unstructured free-text entries is a complex task that face several challenges including; the physician tendency to use some acronyms, abbreviations, negation, and hedge phrases. Lack of standard grammar and punctuation may lead to ambiguity and misunderstanding. The difficulty in automatically processing some context-sensitive meanings and temporal relationships is another challenge. There is a significant need for searching full text medical records. Having some simple SQL queries and search engine tools can help in conducting this full-text search.

In order to solve the problem with abstraction and obtain useful information from the unstructured clinical notes, researchers at the University of Michigan have developed the Electronic Medical Record Search Engine (EMERSE). This is a full-text search engine that is mainly designed to extract useful information from the narrative clinical notes in the EHR systems. CLSearch is another tool for searching free-text reports within EHR systems.

**Abstraction Using NLP**

NLP can be defined as computation algorithms for analyzing machine readable unstructured text. NLP can conclude the meaning behind the words. Automated data abstraction using NLP has the potential to convert the unstructured text-free notes into structured and codified format. Thus, NLP is an efficient alternative for manual abstraction. NLP-based systems can reduce the time and efforts of manual abstraction in large-scale population-based studies.

NLP has the potential to extract all the needed information and perform some complex multivariable queries. It tags every word and puts it into a discrete format that can be used for reporting. Additionally, NLP can recognize related words and phrases. For example, high blood pressure and hypertensive can be considered as fitting in the overall description of the term hypertension. Experts agree that meaningful use/promoting interoperability may be the largest driving factor behind NLP adoption. This is due to the ability of the NLP system to search through a large volume of documents and extract information that are related to meaningful use data elements, such as, a problem list, procedures, medications, allergies, vital signs, social history, and quality measure information.

NLP has been successfully used to abstract useful information in several applications including; emergency medicine physician visit notes, pathology reports, identifying individuals based on cancer screening, abstracting findings from imaging, conducting pharmacogenomics research,
extracting cancer stage information from narrative EHR data\textsuperscript{38}, and identification of breast and prostate malignancies described in pathology reports\textsuperscript{39}.

**Advantages and Disadvantages of Abstraction**

Advantages of the abstraction process include\textsuperscript{40}:

1. Ensure correct placement of data into their intended field in the EHR.
2. Speed up the go-live process for physicians since the abstraction can help to provide easy and rapid access to patient data.
3. Save electronic storage space since abstracting only the needed information requires less storage space than whole clinical documents.
4. Abstraction is a source of supplemental information that supports claims information, which in turn provides more specific evidence for clinical care\textsuperscript{41}. For example, for some measures, claims information is incomplete. So, information from the abstraction process can be used to supplement evidence of the service provided, to verify the population that is being measured.
5. Abstraction supports key processes such as coding and reimbursement, quality improvement, billing audits, and clinical research\textsuperscript{42}.

Also, abstraction can have some disadvantages since it may take extra time and resources in order to enter all the patient information into the EHR. **Table 1** provides a summary of the advantages and disadvantages for each type of clinical data abstraction method.

**Discussion**

**Successful Abstraction Process**: One example of successful data abstraction was provided by Care Communications, Inc. which was a leader in providing data abstraction services and is now a part of Ciox Health. Based on their experience in data abstraction it is important to satisfy some key factors including\textsuperscript{43}: Increasing the medical records procurement rate, enhancing data integrity using an inter-rater approach and working with specialists in the field of health information management who are familiar with HIPAA, ongoing status reporting, and personalized project management.

Several decisions should be taken in order to guarantee a successful abstraction process, such as\textsuperscript{12}:

1. Determine the scope of the abstraction, which means deciding what data should be abstracted and when and whether there are some special abstraction needs for sub-specialists.
2. Determine the time required to do the abstraction.
3. Determine the budget for the abstraction process based on the scope of the abstraction.
4. Determine who will do the real abstraction of data and how the abstractors will be trained.

**Scope of Abstraction:** Examples of abstracted data include: demographics, scheduled appointments, active orders, allergies, medications, immunizations, chronic conditions, problem lists, hospital discharge summaries, special studies (echocardiograms, pulmonary function tests, etc.), and patient history (medical, surgical, social and family). The chart abstraction process may also include the identification of key paper clinical documents that need to be included in the new EHR by scanning those records into the electronic chart prior to bringing the new EHR live.

Six important categories should be recorded when doing abstraction including:

1. Impact on clinical outcomes (length of stay, morbidity, mortality, validated measure of health-related quality of life (HRQOL) or functional status, adverse events).
2. Impact on health care process outcomes (preventative care ordered/completed, clinical study ordered/completed, treatment ordered/prescribed, impact on user knowledge).
3. Impact on workload, efficiency, and organization of health care delivery (number of patients seen/unit time, clinician workload, efficiency).
5. Impact on economic outcomes (cost).
6. Impact on health care providers (HCP) use and implementation (HCP acceptance and satisfaction, implementation of clinical decision support system (CDSS)).

In general, coders abstract Present on Admission (POA), Hospital-Acquired conditions (HAC), some patient safety indicators (PSI) and the Core Measures. Additionally, many facilities require their coders to check the charges for services or enter charges altogether based on the type of record they code. Based on a survey done by Himagine solutions (www.himaginesolutions.com) on their field coders, the coders reported that there are many more elements that are currently being abstracted in an effort to capture data, streamline the process, and assure the accuracy of input.

**Time of Abstraction.** The time required to complete the abstraction depends on the clinical practice (NextGen Healthcare). Generally, patients see their physician three to five times per year. Thus, the abstraction volume will decrease in the first two months. However, the abstraction volume will keep increasing when having new patients, and thus the time will also increase.

**Budget.** Generally, the data abstraction process is labor intensive and requires solid data validation and quality control mechanisms. The budget for abstraction and the needed information varies depending on the scope, size, and the needs of the clinical practice. Thus, the budget can range
Who Will Do the Abstraction? Based on the NextGen Healthcare experience, the abstraction process needs to be delegated to either: nurses or medical abstractors which can include health information management professionals. Some of the NextGen Healthcare clients have used their current health information staff to do the abstraction. One benefit of using the current HIM staff as data abstractors is to reduce the time required to do abstraction since they will be familiar with the practice and the EHR. NextGen Healthcare clients have suggested that there might be a need to hire some temporary abstractors for the first two to three months. Through time, the amount of information that needs to be abstracted will decrease. Additionally, HIM staff, physicians, nurses will become more proficient with the abstraction process, and thus, they will be able to keep up with the abstraction.

Using credentialed HIM professionals (RHIAs and RHITs) and Registered Nurses (RNs) to do the data abstraction will be better than assigning the abstraction task to clinical coders. The reason for this claim is that health information management professionals (RHIAs and RHITs) and RNs are consistently focused on clinical data integrity in their day-to-day tasks. Thus, they will be able to provide the most valuable details about the continuity of patient care. Furthermore, RHIAs, RHITs, and RNs can understand patient data in a broader way and they will be able to extract the critical details since they understand all the different clinical components that shape the picture of the individual’s whole health.

Organizations can use their health information management professionals, internal nurses, and physicians to do the abstraction, or they can outsource the abstraction to other organizations that have some clinical experts who can do the abstraction. Although it seems that doing the abstraction internally can be feasible and cost effective it may reduce the productivity of the internal staff.

The ideal clinical abstraction team can include:

1. Project manager who can monitor all the abstraction project components such as budget and timeframe.
2. Research manager who can monitor the quality of the abstracted data. He/she needs to have a high clinical and technical expertise.
3. Lead abstractor who can monitor the daily details of the abstraction process and supervise the abstractors.
4. Abstractors who will conduct the actual abstraction and they should have experience in clinical data abstraction and familiarity with the EHR.
Ideally, data abstractors need to have the following qualifications:\(^53\)

1. Experience with retrospective data collection from the EHR
2. Clinical and research experience relevant to the study being conducted
3. Advanced educational preparation in a health information and health care profession.

It is important to identify the required resources, budget, and time constraints ahead and before the real abstraction process. Some studies are resource intensive that need high-level planning for all the steps of the abstraction process. For example, the abstraction of charts in the study of screening lung cancer that was performed by Care Communications is very complex and challenging\(^54\) and requires high level project management and clinical experience. This study requires screening thousands of medical records within more than twenty hospitals in the nation.

**EHRs and Registries**

A registry can be defined as an organized system that uses observational study methods to collect uniform clinical data to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes. Registries are focused on populations and are designed to fulfill specific purposes defined before the data are collected and analyzed.\(^55\) On the other hand, EHRs are focused on the collection and use of individual patient health-related information. Although, it seems that both registries and EHRs overlap in functionality, their roles are different. According to the Institute of Medicine (IOM), (which is now called the National Academy of Medicine), an EHR has four core functionalities: health information and data, results management, order entry and support, and decision support.\(^56\) There are several obstacles to achieve the meaningful communication between systems such as, EHRs and registries. These obstacles are related to confidentiality, security, privacy and data access.\(^55\)

Currently, there is an increasing demand for physicians to participate in the registries in order to manage safety, evaluate effectiveness, and measure and improve the quality of patient care. Therefore, it is becoming increasingly important that EHRs should serve as an interface for several registries with different purposes at the same time. EHRs can enable health care information to be available and accessible to registries. Additionally, EHRs can provide some relevant information from the registry to the physicians such as, information about natural history of disease, safety\(^57\), effectiveness, and quality.\(^58\) Figure 1 demonstrates the relationship between the EHR and the registry.

Navaneethan et al.\(^59\) have described the development of a registry for patients with chronic kidney disease (CKD) that is derived from EHR data. The benefits of this kind of patient registry can range...
from allowing better aggregation of patient data for practice assessment or quality improvement, to facilitating clinical research. The study shows that the quality of data in this registry is comparable to that of the data from a much more labor-intensive and expensive process of human abstraction. This registry can be used for quality improvement, clinical research, and other important tasks.

Conclusion

Abstraction Validity. Medical record abstraction is a primary mode of data collection in secondary data use. Abstraction is associated with high error rates.\(^6\) It is important to validate the abstracted data and ensure that the data are abstracted correctly and consistently. There are several benefits of the validation process\(^4\) including:

1. Enhance the clarity of specification through the identification of specification ambiguities that are related to the abstraction process.
2. Help to ensure abstractor consistency through the ongoing monitoring.
3. Reveal quality of care opportunities.
4. Provide some information for future internal quality improvement. Strategies for improving the validity of data abstracted from medical records include\(^6\) training abstractors, masking abstractors to study hypotheses, assessing interrater reliability and agreement, and the re-abstraction of records.\(^6\)

There are three components of validation\(^4\) including:

1. Validating the currently used tools from different vendors and updating the existing tools.
2. Validating the abstraction process and this can be done during the data collection by taking a convenience sample of records and ensure that all measures are abstracted consistently by different vendors to uncover any specific ambiguities.
3. Validating at the end of data collection in order to ensure the integrity and accuracy of the abstracted data.

Future Research

We have conducted a study to examine abstraction methods across hospitals using interviews of managers of abstraction within their healthcare organizations as well as a survey of clients of a large consulting company (Ciox Health) who do data abstraction. Those results are being finalized and will be provided in Part II of a future publication.

Acknowledgements:
This research study was supported by a unique research entity between CIOX Health and the University of Pittsburgh, Department of Health Information Management, School of Health and Rehabilitation Sciences.

AUTHOR BIOGRAPHIES

Amal A. Alzu’bi, PhD, (aazoubi9@just.edu.jo) is Assistant Professor- Department of Computer Information Systems, Jordan University of Science and Technology in Irbid, Jordan.

Valerie J. M. Watzlaf, MPH, PhD, RHIA, FAHIMA, (valgeo@pitt.edu) is Associate Professor and Vice Chair of Education, Department of Health Information Management, University of Pittsburgh, School of Health and Rehabilitation Sciences.

Patty Sheridan, MBA, RHIA, FAHIMA, (pattytsheridan@gmail.com) is President, Sheridan Leadership Consulting (formerly Senior Vice President, HIM Services, at Ciox Health).

References


10. Floyd, J.S., et al., Use of administrative data to estimate the incidence of statin-related


31. Eramo, l., Natural Language Processing. For The Record, 2011. 23(8).
40. Hundt, A., Legacy Data Management: Strategic Considerations for Successful Clinical Data Abstraction.
42. Medical records abstracting- Harness the power in your clinical data. 2015 ; Available from: http://solutions.3m.com/wps/portal/3M/en_US/Health-Information-Systems/HIS/Products-and-Services/Abstracting/.


49. NextGen. ; Available from: NextGen Healthcare


52. Medical Record Abstraction for Real World Data..Simplified Keeping your sanity when integrating medical record data into large scale research projects. 2014 ; Available from: http://www.slideshare.net/CareCommunications/medical-recordabstractionforrealworlddata.


60. Nahm, M., Data Accuracy in Medical Record Abstraction. UT SBMI Dissertations, 2010.


There are no comments yet.