

CLINICAL DATA ABSTRACTION: A RESEARCH STUDY

Posted on March 15, 2021 by Matthew

Category: [Spring 2021](#)

By Valerie J. M. Watzlaf, PhD, MPH, RHIA, FAHIMA; Patty T. Sheridan, MBA, RHIA, FAHIMA; Amal A. Alzu'bi, PhD; and Laura Chau, BS in HIM

Abstract

This is the second part in a two-part research study on clinical data abstraction.¹ Clinical data abstraction is the process of capturing key administrative and clinical data elements from a medical record. Very little is known about how the abstraction function is organized and managed today. A research study to gather data on how the clinical data abstraction function is managed in healthcare organizations across the country was performed. Results show that the majority of the healthcare organizations surveyed have a decentralized system, still perform the abstraction in-house as part of the coding workflow, and use manual abstraction followed by natural language processing (NLP) and simple query. The qualifications and training of abstractors varied across abstraction functions, however coders followed by nurses and health information management (HIM) professionals were the three top performers in abstraction. While, in general, abstraction is decentralized in most enterprises, two enterprise-wide abstraction models emerged from our study. In Model 1, the HIM department is responsible for coding, as well as all of the abstraction functions except the cancer registry and trauma registry abstraction. In Model 2, the quality department is responsible for all of the abstraction functions except the cancer registry, trauma registry, and coding function.

Keywords: Abstraction, electronic health record, clinical, descriptive research study, natural language processing, query, models

Introduction

Clinical data abstraction is the process of identifying and capturing key administrative and clinical data elements. The purpose of abstraction includes the collection of data related to administrative coding functions, quality improvement, patient registry functions and clinical research. A previous review of the literature which includes the abstraction methods, advantages and disadvantages of those methods, the abstraction process, and the connection between EHR abstraction and patient registries have been summarized in a previous manuscript.¹ From this review, we found very little on how the enterprise-wide clinical data abstraction function is managed. Therefore, we have conducted a descriptive study that includes multiple methods of data collection such as qualitative interviews and quantitative surveys of healthcare professionals on how the abstraction function is managed in their organizations today. The results of this study will be shared as well as best practice models that can be used in the organization and management of the abstraction function.

Specific Aims

The goals or specific aims of this research study include the following:

1. Examine how the abstraction function is organized and managed today.

2. Examine the differences in quality and efficiency between manual and automated abstraction.
3. Provide best practices on how healthcare systems are organizing the abstraction function.

Methodology

A descriptive research design was conducted and included multiple methods of data collection to include qualitative interviews with healthcare professionals responsible for the abstraction function. This was followed up with a quantitative-based questionnaire to a larger sample of healthcare professionals.

Participants for the interviews were gleaned from the AHIMA Engage Community, AHIMA Component State Associations (CSAs), coding and quality communities, LinkedIn, and emails to colleagues of the authors who were managers of clinical data abstraction. Twenty-one responses were received, and eight healthcare organizations were interviewed in depth. [Appendix A](#) provides the questions asked during the interview.

All interviews were conducted by the primary researchers and were recorded and transcribed. Content analysis was performed on the transcripts by searching for patterns and themes and then summarizing the findings by discussing them with the research team. The quantitative section of the study included building a survey that collected responses from 50 Ciox Health clients on how they manage the abstraction function. The questions for the survey are listed in [Appendix B](#). IRB approval through the University of Pittsburgh at the exempt level was obtained and the IRB number is PRO15110055. The survey was administered through Qualtrics and initial analysis was done through Qualtrics and more specific graphs and tables were conducted by the researchers using Excel.

Results

Qualitative Interview Results:

Demographic Information of Interview Participants

Participants of the interviews were from large healthcare organizations with complex medical systems and employed a range of abstractors from five to 15 full-time employees (FTEs). The range of years of experience of the leaders of the clinical abstraction function ranged from 10 to 36 years and were primarily in senior level positions.

Management of Clinical Data Abstraction

Management of the clinical data abstraction varied from centralized and decentralized abstraction that were separate from the medical record coding and some that were not separated from coding. The breakdown is displayed below in [Figure 1](#) and shows that 38 percent (3/8) implemented a centralized abstraction function but the leader of that function did not oversee coding. Twenty five percent (2/8) did not have the abstraction function separated from coding but did oversee the coding function and 38 percent (3/8) did separate abstraction from coding and did oversee the

coding function.

Advantages and disadvantages for separating abstraction from the coding function were discussed during the interview and included increased productivity and data quality, process standardization, leveraging electronic data sources, a focus on skill sets and not on a multidisciplinary person that can do it all. Disadvantages include that abstractors find things that the coder may miss (and vice-versa) so that a second check is not there, multiple people are handling the record multiple times, leveraging structured fields in the EHR to reduce manual abstraction means potentially eliminating abstractor FTEs, which can be seen as a disadvantage since the abstractor could lose their job.

Qualifications and Training

The qualifications and training of the clinical data abstractors varies across abstraction functions. Registries tend to hire credentialed and educated professionals in the field, e.g. cancer registry (CTR), trauma registry (CSTR), cardiac and vascular registry (RN, LPN, RHIA) etc. Quality management departments hire educated and/or experienced abstractors with credentials, such as the RHIA, RHIT, RN, LPN. All abstractors should have an expertise in computer skills, be detail oriented, and have medical record knowledge (data sources, medical terminology, anatomy and physiology, pharmacology, pathophysiology).

Data Elements Abstracted:

The data elements that are most commonly abstracted that are separate from the coding function include:

1. CMS quality reporting measures
2. National Quality Forum (NQF) quality measures
3. The Joint Commission quality measures
4. Patient registry functions (trauma, stroke, cancer, thoracic surgery, general surgery, cardiac etc.)
5. Clinical research studies

The common data elements that are abstracted as part of the coding process include:

1. Any providers involved in the patient's care
2. Date, time of procedure, surgical suite, time of anesthesia, results
3. CDI recommendations
4. Date of POA indicators
5. Admission/Discharge Dates
6. Discharge Disposition

Clinical Data Abstraction Methods Used

The clinical data abstraction methods used by the eight interviewees included five used manual abstraction, two used simple query, and one used NLP. One of the comments by the participants stated, "We tried to do some NLP selection, optical character recognition, discrete data point download, but we found that the medical record is so varied in responses and we have very specific data definitions that the error rate was higher than what we were willing to tolerate, and we have a much better success rate with the visual validation and abstraction."

Data Validation Methods

There can be multiple methods of data validation in the abstraction process. The most common methods used were inter-rater reliability and most of the organizations do this concurrently and as well as retrospectively. The gold standard for the accuracy rate is 95 percent and was made part of the abstractor's job description and is a quality metric for performance evaluations.

Quantitative Survey Results

A variety of healthcare professionals from a number of different healthcare facilities gave valuable information on how the abstraction function is managed in their organization. The majority of the respondents (58 percent) hold the position of HIM director. Seventy percent are from large comprehensive healthcare systems ranging from 100 to 500+ beds. The majority (58 percent) of participants reported that they employed a range of abstractors from zero to nine FTEs with some reporting as high as 30 or more (8 percent). Also, 41 percent of those that performed the abstraction function were coders, followed by nurses (27 percent) and then HIM professionals (8 percent). ([Table 1](#)).

Other findings ([Figure 2](#)) show that manual abstraction (58 percent) is the primary abstraction method while NLP was 18 percent; simple query 12 percent; and another 12 percent said that they used their EHR systems/encoder to run reports.

Results also showed that the way the abstraction function is organized across an organization is fairly split evenly between centralized and decentralized (48 percent answered that abstraction is decentralized across the organization and takes place in different departments and 44 percent said that abstraction is primarily centralized with some decentralization). The other 8 percent had varied answers on the topic ([Figure 3](#)).

Seventy percent of the respondents said that abstraction is performed in-house and 78 percent said that it is performed as a part of the coding workflow process ([Figure 4](#) and [Figure 5](#)).

Additionally, retrospective validation using a convenience sample is the most popular validation tool to ensure the abstraction data quality (50 percent).

The results showed that most of the healthcare organizations have fragmented abstraction

functions. Based on the results, it seems that there is inconsistency among healthcare organizations on how best to manage abstraction. Furthermore, while centralized abstraction services are prevalent, more (48 percent) of the health systems surveyed have decentralized abstraction functions.

Discussion

Results from this descriptive study that incorporated both qualitative and quantitative data about clinical data abstraction, found that most of the healthcare organizations interviewed and surveyed have a decentralized system but some said they were moving toward a centralized system. The majority of healthcare organizations still perform the abstraction as part of the coding workflow, and 70 percent of those surveyed do it in-house. The majority of those healthcare organizations use manual abstraction followed by NLP and simple query.

The qualifications and training of clinical data abstractors varies across abstraction functions as registries and quality reporting tend to hire credentialed and educated professionals, whereas abstraction for coding related data need less education and skills. Most common data elements collected include required quality reporting measures, patient registry functions, clinical research studies and data collected as part of the coding process such as POA indicators, discharge disposition etc.

Two enterprise-wide abstraction models emerged from our study. In Model 1 ([Figure 6](#)), the HIM department is responsible for coding, as well as all of the abstraction functions except the cancer registry abstraction which is normally housed under the oncology department. In Model 2 ([Figure 7](#)), the quality department is responsible for all of the abstraction functions except the cancer registry abstraction and is not responsible for the coding function. Model 1 is centralized under HIM and still includes coding, administrative data elements abstracted, quality measures, special study data abstraction, and registry data abstraction. Model 2 is centralized under the quality department and includes everything in the first model except coding.

Limitations

There were some limitations to our descriptive study as listed below:

1. Even though we received 21 responses for our interviews, we were only able to interview 8 individuals and therefore their responses could be different than what we may have received if we were able to connect with the entire group that responded.
2. The quantitative responses from the survey we developed was limited to just 50 Ciox clients who responded to that survey. This sample may not be a good representation of the entire population of clients who oversee clinical data abstraction and therefore their views may then be different than the entire population.
3. Due to our limited sample size, it was difficult to do more than basic descriptive statistics with the

data received.

Future Research

Future research in this area is needed to focus more on the technologies that may now be used and how they compare to human abstraction in relation to efficiency, accuracy, and cost. There is still limited information in this area, so more research is needed to determine the best methods for abstraction as well as the best organizational and management methods around the abstraction function since it varied across organizations. Also, more research is needed on the best qualifications and training that are needed for abstractors since those performing this function varied across organizations and new technologies could lead to more thorough and extensive training methods. Clinical data abstraction is such a vital function that more research in this area world-wide could determine high quality methods of implementation that can then be used by healthcare organizations across the globe to improve the workflow and the quality of the data collected which in turn will lead to better health outcomes for patients.

Conclusion

There is room for improving the quality of healthcare data abstracted and centralizing the abstraction services. This could possibly be tackled by creating and implementing policies and procedures that can outline how to and who performs the abstraction function. Ensuring that the staff follow the abstraction policy might lead to a more consistent process among healthcare organizations which will result in better healthcare reporting and documentation. **Figure 8** shows our root cause analysis regarding the problem of fragmented abstraction functions. Furthermore, the advances in technology have also improved the clinical data abstraction function. NLP and machine learning systems are able to understand the language of the textual variables within the medical record and produce them so that the abstractor can audit them for inclusion, if appropriate. Over time the accuracy of machine learning systems improves as larger sets of data are reviewed. There have been several studies that have found that the use of NLP and machine learning enhance clinical data abstraction.²⁻⁵ As more healthcare organizations use NLP, the efficiency and quality of clinical data abstraction will increase and the need for health information management professionals in this area at an analyst or auditor level will be needed as well. Education and training in the areas of artificial intelligence and machine learning is important to provide to healthcare and health information professionals so that they understand and use these tools to enhance the clinical data abstraction function within their healthcare organizations.

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. Our partnership focused on conducting research and objective analysis in the field of healthcare data quality and health information management to determine innovative

best practices that when adopted can improve the efficiency and effectiveness of the U.S. healthcare system.

Author Biographies

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, Department of Health Information Management, in Pittsburgh, PA

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

Amal A. Alzu'bi, PhD, (aazoubig@just.edu.jo) is assistant professor, Department of Computer Information Systems, Jordan University of Science and Technology, in Irbid, Jordan.

Laura Chau, BS in HIM (llc31@pitt.edu) is associate software engineer, UHS, in Mechanicsburg, PA.

References:

1. Alzu'bi A, Watzlaf V, Sheridan P. Electronic Health Record (EHR) Abstraction, *Perspectives in Health Information Management*, Submitted May 3, 2020
2. Beck JT, Rammage M, Jackson G, Preininger A, Dankwa-Mullan I, MC, Roebuck, Torres A, Holtzen H, Coverdill S, Williamson MP, Chau Q, Rhee K, and Vinegra M. Artificial Intelligence Tool for Optimizing Eligibility Screening for Clinical Trials in a Large Community Cancer Center JCO. *Clinical Cancer Informatics*; 2020;4(50-59).
3. Tignanelli C, Silverman G, Lindemann E, Trembley A, Gipson J, Beilman G, Lyng J, Finzel R, McEwan R, Knoll B, Pakhomov S, Melton G. Natural language processing of prehospital emergency medical services trauma records allows for automated characterization of treatment appropriateness. *Journal of Trauma and Acute Care Surgery*. May 2020;88(5):607-14. doi: 10.1097/TA.0000000000002598.
4. Zhu R, Tu X, and Huang J. (2020) Using Deep Learning Based Natural Language Processing Techniques for Clinical Decision-Making with EHRs. In: Dash S, Acharya B, Mittal M, Abraham A, Kelemen A. (eds) *Deep Learning Techniques for Biomedical and Health Informatics. Studies in Big Data*, vol 68. Springer, Cham
5. Al-Aiad A. and El-shqeirat T. Text mining in radiology reports (Methodologies and algorithms), and how it effects on workflow and supports decision making in clinical practice (Systematic review). 2020 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 2020, pp. 283-287, doi: 10.1109/ICICS49469.2020.239506.

There are no comments yet.