

Evolution of the Cancer Registrar in the Era of Informatics

Kelly W. Merriman, MPH, PhD, CTR¹; Ronda G. Broome, MS, MSHMI, CTR²; Giordana De Las Pozas, MD¹; Lisa D. Landvogt, BA, CTR³; Ying Qi, MS, CTR⁴; and Judith Keating, CTR⁵

The cancer registrar reports accurate, complete, and timely abstracted cancer data to various healthcare agencies. The data are used for understanding the incidence of cancer, evaluating the effectiveness of public health efforts in the prevention of new cases and improving patient care outcomes and survival. There are increasing demands placed on registrars for additional data points with real-time submission to reporting agencies. To that end, registrars are increasing the use of informatics to meet the demand. The purpose of this article is the role of the registrar in the collection and reporting of critical cancer data and how registrars are currently using informatics to enhance their work. This article describes how informatics can be leveraged in the future and how registrars play a vital role in meeting the increasing demands placed on them to provide timely, meaningful, and accurate data for the cancer community.

JCO Clin Cancer Inform 5:272-278. © 2021 by American Society of Clinical Oncology

BACKGROUND: HISTORY OF THE CANCER REGISTRY PROFESSION

A cancer registry is an information system that allows for the collection, management, and analysis of information on patients with cancer.¹ The first modern hospital-based cancer registry was established at Yale-New Haven Hospital in Connecticut in 1926, and the first state central registry was established in Connecticut in 1935 (Table 1).²

In 1956, the American College of Surgeons (ACoS) required a cancer registry for approved cancer programs in the United States. In 1971, the National Cancer Act budgeted monies to the National Cancer Institute (NCI) for research, detection, and treatment of cancer, and in 1973, the SEER Program of NCI established the first national cancer registry. Finally, in October 1992, Congress established a National Program of Cancer Registries (Public Law 102-515, The Cancer Registries Amendment Act). There are three types of registries: hospital, population-based central registry, and special purpose registries.³ Most registries include reports of all malignant cancers, except for nonmelanoma skin cancers. Reported data can come from various medical facilities, including hospitals, physicians' offices, and pathology laboratories. The data are collected for multiple reasons, including understanding the incidence of cancer, researching patterns of cancer treatment, and improving patient care outcomes and survival.⁴

Abstracting cancer data is a highly skilled profession that requires comprehension of cancer information,

and appropriately coding those data and processing them into an international format. Possessing this knowledge allows cancer registries to provide data users with the necessary facts for that facility, time period, and specific cancer. The specifications of the coding process are critical for the quality of the output based on the increased knowledge and continuing education of the cancer registrar. Often, registrars serve as data subject matter experts and are responsible for bridging the gap between clinicians and statisticians. The registrar collects the data and ensures the accuracy, completeness, and quality of the data. They comply with state and federal laws for reportable cases and maintain the database that houses the cancer program's cancer cases. The ability to follow and report the patient's cancer journey from diagnosis through treatment and survival information in a cancer registry system is one of the many core strengths of the cancer registry.

The cancer registry profession has been valued since the early 20th century, gaining further recognition and prominence through the efforts of ACoS to include it as a required component of the accredited cancer program. In addition, the National Cancer Registrars Association (NCRA) has helped to further promote the registry profession through NCRA's Council of Certification by creating the certified tumor registrar (CTR) credential in the 1980s.² In the 21st century, the CTR profession has increased complexity because of the challenges of locating patient information from multiple sources and abstracting the required information

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on January 19, 2021 and published at ascopubs.org/journal/cci on March 19, 2021; DOI <https://doi.org/10.1200/CCI.20.00123>

CONTEXT

Key Objective

To describe how informatics can be leveraged by cancer registrars to capture essential cancer data.

Knowledge Generated

Informatics is currently being used by cancer registrars for case finding and follow-up. In the future, entering data into structured fields in the Electronic Health Record and other initiatives will allow for more timely and accurate data reported to the state and National Cancer Database.

Relevance

The data that are abstracted by the cancer registrar have significant clinical relevance for multiple reasons. First, the data collected provide a census of cancer cases that can be used to define cancer incidence and mortality (eg, American Cancer Society Facts and Figures Reports). Second, the data are used to update staging versions over time. Finally, the data are used to improve patient care outcomes and survival.

into the growing number of defined data fields. Cancer registrars must abide by the national standards of reporting accurate, complete, and timely cancer data to various healthcare agencies.⁵ The purpose of this article will be to review how registrars leverage informatics to meet the ever-increasing demand of additional variables and report this information in real time to numerous external reporting agencies. In addition, the evolving role of the registrar will be explored.

CURRENT WORK OF THE REGISTRAR AND INFORMATICS

By law, hospital registrars report cancer cases to their state cancer registry within 6 months of their facility's date of first contact with a patient. In addition, for hospitals that are voluntarily accredited by Commission on Cancer (CoC), the shift to concurrent abstracting is changing the timeliness of reporting. In the past, registrars were required to submit to the National Cancer Database (NCDB) on an annual basis. Registrars are now focusing on abstracting in phases (concurrent abstracting) after certain time periods of the patient's cancer care. Finally, the 2018 data standards for collection of cancer registry data saw an unprecedented

number of changes in reporting data variables being required for submission. New data fields were created in response to the American Joint Committee on Cancer (AJCC) 8th edition staging manual and included 137 site-specific data items (SEER SSDI). These additional data items further enhanced the value of registry data, yet also increased the complexity and time required to abstract new cancer cases. Thus, it is critical to leverage informatics to help with the registrars' processes, beginning with follow-up data from the Electronic Health Record (EHR) and the institutional data warehouse (IDW).

Information collected during follow-up includes patient vital status through the lifetime of the patient. Cancer registrars perform, at minimum, annual follow-up for each patient with cancer as this provides accurate and updated survival information. Because of the recent advances in the electronic exchange of information, automated follow-up takes full advantage of electronic resources and reduces the time and labor expended in data collection. The data include using structured fields for completed appointments for physical visits, the use of the billable service date as the last contact date, and the latest encounter date in the EHR.

Another challenge for registrars is case finding, which involves identifying all the cancer cases that are reportable to the state of all patients seen at the facility. Reportable cancer cases are those cases that are required to be submitted to state registries and the NCDB. As the demand for more real time reporting increases, a registrar cannot possibly review all their cases manually in the larger cancer facilities. Thus, it is imperative to leverage additional tools to extract information for case finding and if possible, auto-populate the database from identified variables in the EHR. This ranges from querying the IDW for identifying cases using billing data and the use of Natural Language Processing software.

Natural language processing (NLP) allows users to convert free text into machine readable, structured data, which are a large part of the data in EHRs currently today. The main

TABLE 1. Timeline of Cancer Registry or Profession

1926—First hospital-based cancer registry at Yale in New Haven, Connecticut
1935—First state central registry was established in Connecticut
1956—American College of Surgeons required cancer registry for approved cancer programs
1973—First national cancer registry established by National Cancer Institute within the SEER Program
1983—National Tumor Registrars Association established certified tumor registrar credential
1992—Congress established the Cancer Registries Amendment Act
1993—National Tumor Registrars Association changed to National Cancer Registrars Association
2017—Health Information Technologists and Medical Registrars standard occupational code established

FIG 1. Screenshot of commercialized tumor registry system pathology data in CAP format.

data leveraged using NLP are College of Pathology templates (CAP) from the EHR to derive the International Classification of Diseases for Oncology (ICD-O) codes for site and histology as well as diagnosis date and AJCC stage. [Figure 1](#) demonstrates how registrars can autopopulate fields from a CAP protocol into a commercialized Tumor Registry system after verifying the information, in this example, an astrocytoma of the frontal lobe. This example demonstrates how a registrar's job may become more of quality assurance than manually entering data into a database. Additionally, some site-specific factors can be determined such as Estrogen and Progesterone Receptors status for breast cancer.

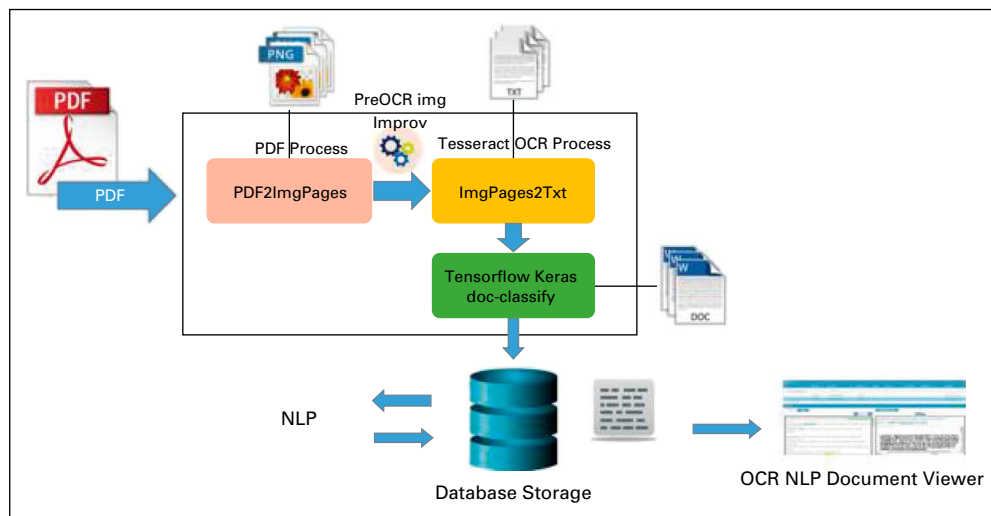
One can also apply NLP to scanned documents using optical character recognition (OCR), which allows registrars to increase efficiency by using informatics tools. For example, a major challenge in the cancer registry at larger cancer hospitals such as MD Anderson Cancer Center is that more than half of the new patients managed at the facility have received previous treatment, relapsed (non-analytical), and present with a large volume of outside scanned records. MD Anderson's information technology team was able to develop an OCR NLP pipeline for text retrieval and cancer information extraction and summarization. First, the resolution of the image of a scanned document is enhanced and text is retrieved from the image using OCR. Then, the pipeline uses clinical language annotation, modeling, and processing (CLAMP) as the primary NLP component for named entity recognition against OCR-ed text ([Fig 2](#)). It is based on a machine learning model (conditional random forest), specifically trained on clinical

data. The development of CLAMP Cancer Modules involved building cancer-specific NLP components and extending CLAMP interface that allows end users to customize different components into an NLP pipeline for their specific needs.⁶ Some additional rules are built post-NLP to extract event dates related to named entities and to suppress noises from nonpatient data such as genetic test reference data or patient's family cancer history. Finally, the extracted cancer diagnosis, diagnosis date, and treatment information are organized into oncology specific data folders and presented to the end user in a web application. This process has matured, and there now exists a search engine that is EHR embeddable. The application also has the capacity to ingest very large volumes of either PDF or electronically transferred documents, returning a detailed list of cancer diagnoses, stage, genomic information, and treatment information with dates for rapid review. The platform tags the location within the source documents for rapid validation of the information in a side-by-side window format. This tool can improve the overall efficiency of data abstraction and processing for registrars. For example, MD Anderson's registrars' review of outside scanned documents on non-analytic cases was reduced from approximately 45 to 15 minutes per cancer case. Registrars have been actively involved in this work, helping to improve the NLP accuracy and the time it takes to review a cancer case with numerous scanned documents.

THE FUTURE OF THE REGISTRAR AND INFORMATICS

Some data elements in the EHR, such as demographics and tumor characteristics, are also needed in the registry

FIG 2. OCR/NLP diagram. NLP, natural language processing; OCR, optical character recognition.



system. To avoid duplication of effort, there is a need for interoperability between the two systems. Interoperability is the ability of computer systems or software to exchange and make use of information.⁷ Interoperability allows for delivery of the right information at the right time to the right place. To improve efficiency and accuracy of data collection in the registry using interoperability, there is a need to leverage the information in the EHR and directly link this information into the registry software. The current method of manual extraction of patient information from EHRs and manually entering this information into the registry software is not sustainable.

One data standard effort that has been developed is Health Level Seven (HL7) to empower global health data interoperability within EHRs.⁸ Just like in the cancer registry with usage of coding manuals, HL7 provides a protocol of specifications in addition to a framework of standards for the exchange, integration, sharing, and retrieval of electronic health information, which is what we need.⁹ Electronic data sharing between healthcare organizations through interoperable processes would enable coordinated care, decrease costs, and improve patient outcomes.⁷ If EHRs were able to interact and information was shared with the registry software, then the registrar's role would become more of quality assurance of the data. Hopefully, interoperability will also aid the registrar in obtaining treatment information from outside facilities, since patients often receive fragmented treatment at multiple facilities. In addition, obtaining detailed outcome information such as recurrence will improve with an increase in interoperability between numerous EHRs.

One critical role that most hospital registrars fulfill is the Rapid Quality Reporting System (RQRS) for CoC-accredited facilities. The application of RQRS allows for real-time reporting of quality measures using registry data to actively monitor treatment information.¹⁰ The prospective dashboard that is provided for CoC facilities displays

information regarding treatment and quality measures as documented in the registry. This reporting provides insight into the treatment of the patient with cancer at the facility and allows for real-time alerts if patients are not being treated in a specific time frame.¹⁰ This level of insight allows the cancer program to assess their evidence-based cancer care practice and identify patients who do not receive the anticipated care during the predetermined time frame to allow for intervention. A fundamental change that occurred in late 2020 is the RCRS with NCDB as a response to having timelier data.¹¹ The new IT platform will combine and replace the best of RQRS (at least quarterly) and NCDB (annually) Call for Data and use one portal for all submissions. More importantly, the reporting will be done in phases and along the continuum of the natural course of a patient's cancer. The long-term goal of NCDB with this change in platform is an emphasis on timeliness and agility with an improvement in quality monitoring and quality improvement.¹¹ The greatest challenge in supporting this initiative depends on the healthcare professional entering data in the EHR-supported template structure, so the information can be converted and ingested by the registry software. This requires that the registrar can communicate to technology specialists exactly where the data are located within the EHR.

Furthermore, the registrar is a crucial member of the cancer program. The role has evolved from simply inputting data in the software to creating and generating reports that provide insight into the cancer program. This insight includes, but not limited to, in- and out-migration of the patient with cancer for treatment and services and identifying the top cancer sites served at the facility. Cancer registrars are needed in precision medicine with real-world data abstraction, government programs that use the data, software vendors that are used to collect the data, and on task forces that create and maintain the standards for data collection. Using knowledge, experience, and skills, the registrar is

FIG 3. NCRA strategic management plan.



prepared to collaborate and generate high output and high-quality data to assist in the fight against cancer.

DISCUSSION

As described in this paper, cancer registrars are continuing to leverage informatics in their everyday workflow. Registrars have the knowledge, experience, and skills to accurately abstract the cases. They are leading the way in how informatics can be applied in the cancer registry data collection process. Most cancer registry systems can now integrate with data that reside in the EHR as structured data. The registrar can optimize the structured data and work on prepopulating certain data elements. For example, AJCC staging of the cancer is data element that should be recorded by the clinician before the initiation of treatment or during the cancer journey. If this and other key data elements were stored in EHRs as structured data elements, registrars would be able to pull this into the registry system, which would prevent the manual re-entering of this information.

To generate real-time data, the information should be readily extracted from the EHR, but often it is not. This is primarily due to data not being entered into the structured data fields that are available in these systems. In an editorial response in 2016, Evans et al¹² state that cancer stage should be determined and documented by providers in the EHR, but this often is not done or appears in free text rather than the important structured data fields that are readily available in most EHRs. They commend the work that Warner et al¹³ did with extracting cancer stage information from narrative EMR data using NLP, but they advocate that with such important data, it must be discrete, structured data. Reasonable accuracy is not good enough for such important data items, since staging is such an important indicator for determining treatment and survival outcomes. In addition, registrars are being asked to automate data out of the EHR but struggle with what is the correct stage when there are multiple responses for the same patient's cancer

stage. It takes the registrar longer to determine the correct stage from the numerous responses than to stage the cancer themselves.

Automation from EHRs to cancer registry systems cannot happen unless providers (or their representatives) are willing to enter data into structured data fields. Unfortunately, clinicians are often pressed for time and it is not the best use of their time to enter data discretely in structured fields.^{14,15} In addition, it has been demonstrated that the amount of time that is required for documentation has increased provider frustration with using EHRs. Rosebloom et al¹⁶ describe the tension between the need for structure for the second use of data downstream and flexible documentation for providers. Busy providers value flexibility and efficiency, whereas users downstream often value structure and standardization. The tension between providers and end users differs because the goals are different between the two groups. Providers' goals range from reports of their clinical impressions and actions to justifying the level of service billed and creating a legal record in the case of litigation. However, end users such as registrars require structure and standardization, which limits the providers' ability to describe nuances in a patient's care.¹⁶ One potential solution to help busy physicians document in the EHR is to use registrars to input clinical stage information into the EHR and the physician would just review and sign off on the stage,¹² since registrar staging is considered the gold standard compared with NLP.

Finally, one important initiative led by ASCO and a group of collaborators, including oncologists, researchers, and other experts, in 2018 is mCODE: Minimal Common Oncology Data Elements. The initiative is providing a common data language and a data model for interoperability across numerous EHRs throughout the United States. More importantly, it is developed and maintained by its users in a highly collaborative environment.¹⁷ As mCODE grows, the interoperability capabilities will also work for cancer registry

systems, and registrars will be able to pull the critical data into their systems for reporting purposes.

In conclusion, a change in workflow can create resistance; however, registrars must be willing to embrace new processes and technology. Informatics must become a part of registry abstraction, and registrars play a vital role in that process. For example, they will be needed from the initial build as the subject matter experts in the rules and standards for each data field to the curation of the imported data (Fig 3). The profession will require additional education for

the registrar to oversee and manage the data that come from a structured EHR. Registrars must take their initial role from decades ago and become part of the solution to the many obstacles. They must realize their potential and the potential of taking the profession to the next level in all possible instances. New drugs, therapies, and surgical techniques are constantly maturing and evolving. Cancer registrars must do the same and drive the changes that need to happen for cancer registries to continue to thrive as they have for nearly a century.

AFFILIATIONS

¹MD Anderson Cancer Center, Houston, TX

²Syapse, Inc, San Francisco, CA

³Henry Ford Cancer Institute, Detroit, MI

⁴Westat, Inc, Rockville, MD

⁵Hartford HealthCare Cancer Institute, Hartford, CT

CORRESPONDING AUTHOR

Kelly W. Merriman, MPH, PhD, CTR, MD Anderson Cancer Center, 1155 Pressler St, Houston, TX 77030; e-mail: kmerriman@mdanderson.org.

AUTHOR CONTRIBUTIONS

Conception and design: All authors

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless

otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians ([Open Payments](http://OpenPayments)).

No potential conflicts of interest were reported.

ACKNOWLEDGMENT

The authors would like to gratefully acknowledge the trailblazing work that the Institutional Analytics and Informatics NLP team at MD Anderson (Jeff Jin, Trey Kell, and Maria Soto) have been doing; and the support, agility, and flexibility that Oncology Data Management Systems (ONCO) has provided cancer registries. Finally, without the tremendous support of the National Cancer Registrars Association (NCRA) certified registrars would not exist nor function nearly as well.

REFERENCES

1. National Cancer Institute: What is a cancer registry. https://seer.cancer.gov/registries/cancer_registry/index.html
2. National Cancer Registrars Association: History. <https://www.ncra-usa.org/About/History>
3. National Cancer Institute: Types of Registries. <https://training.seer.cancer.gov/registration/types/>
4. Parkin DM: The role of cancer registries in cancer control. *Int J Clin Oncol* 13:102-111, 2008
5. White MC, Babcock F, Hayes NS, et al: The history and use of cancer registry data by public health cancer control programs in the United States. *Cancer* 123:4969-4976, 2017
6. Soysal E, Warner JL, Wang J, et al: Developing customizable cancer information extraction modules for pathology reports using CLAMP. *Stud Health Technol Inform* 264:1041-1045, 2019
7. Lareau D: Council post: How interoperability can improve patients' healthcare process. *Forbes*, 2020. <https://www.forbes.com/sites/forbestechcouncil/2020/05/11/how-interoperability-can-improve-patients-healthcare-process/#5b1dd75c358>
8. HL7® Governance and Operations Manual. 2020. https://www.hl7.org/documentcenter/public/membership/HL7_Governance_and_Operations_Manual.pdf
9. About Health Level Seven International. 2020. <https://www.hl7.org/about/index.cfm?ref=footer>
10. Edge SB, McKellar D, Stewart AK: Using the American College of Surgeons cancer registry to drive quality. *J Oncol Pract* 9:149-151, 2013
11. McCabe R: National Cancer Database: The past, present, and future of the cancer registry and its efforts to improve the quality of cancer care. *Semin Radiat Oncol* 29:323-325, 2019
12. Evans TL, Gabriel PE, Shulman LN: Cancer staging in electronic health records: Strategies to improve documentation of these critical data. *J Oncol Pract* 12:137-139, 2016
13. Warner JL, Levy MA, Neuss MN, et al: ReCAP: Feasibility and accuracy of extracting cancer stage information from narrative electronic health record data. *J Oncol Pract*, 12(2), 157-158, 2016
14. Bush RA, Kuelbs C, Ryu J, et al: Structured data entry in the electronic medical record: Perspectives of pediatric specialty physicians and surgeons. *J Med Syst* 41:75, 2017
15. Gilbert JA: Physician data entry: Providing options is essential. *Health Data Manag* 6:90-92, 1998
16. Rosenbloom ST, Denny JC, Xu H, et al: Data from clinical notes: A perspective on the tension between structure and flexible documentation. *J Am Med Inform Assoc* 18:181-186, 2011

17. Osterman T, Terry M, Miller R: Improving cancer data interoperability: The promise of the minimal Common Oncology Data Elements (mCODE) initiative. *J Clin Oncol Clin Cancer Inform* 4:993-1001, 2020

