IMPROVING THE USABILITY AND UTILIZATION OF CANCER REGISTRY DATA: THE NEED TO IDENTIFY A CORE DATA SET

A Dissertation

Presented to

The Faculty of the Graduate School

At the University of Missouri

In Partial Fulfillment

Of the Requirements for the Degree

Informatics

By

IRIS ZACHARY

Dr. Suzanne Austin Boren, Dissertation Supervisor

December 2012

The undersigned, appointed by the dean of the Graduate School,

have examined the Dissertation entitled

IMPROVING THE USABILITY AND UTILIZATION OF CANCER REGISTRY DATA:

THE NEED TO IDENTIFY A CORE DATA SET

Presented by iris Zachary

A candidate for the degree of

Informatics

And hereby certify that, in their opinion, it is worthy of acceptance.

Suzanne Austin Boren, Adviser

Lanis Hicks, Health Management and Informatics

Eduardo Simoes, Health Management and Informatics

Jeannette Jackson-Thompson, Health Management and Informatics

---

Wade Davis, Health Management and Informatics

Sanda Erdelez, School of Information Science & Learning Technologies

To Isaac, Christopher, Abraham, Kristina, my mom and Rudy

ACKNOWLEDGMENTS

I would like to thank all the people that I met on my way to a doctorate in health informatics who encouraged, helped, shaped, and motivated me to go on with this project. I particularly want to express my gratitude to my adviser, Dr. Suzanne Boren. Throughout my journey, she provided encouragement, sound advice, excellent teaching, and many good ideas. Dr. Boren's diligence, knowledge, and support inspired me and made my research come alive. I also want to thank my committee members, Dr. Lanis Hicks, Dr. Eduardo Simoes, Dr. Jeannette Jackson-Thompson, Dr. Wade Davis, and Dr. Sanda Erdelez, for their valuable and constructive criticism during my PhD training. With support from my advisor and other members of my committee, my thoughts have been transformed into products that can be used and published.

I am grateful to all the professors, colleagues, collaborators, and friends who have helped me with my work and research. I want to thank present and past members of the Missouri Cancer Registry and Research Center (MCR-ARC) for providing a place where new ideas are born and supported and for sharing their knowledge and experience with me. I especially want to thank Saba Yemane, MCR-ARC Database Manager, for teaching me everything I know about the MCR-ARC database and who has been my biggest support through all these years. I want to give special thanks to Dr. Jeannette Jackson-Thompson, MCR-ARC Director, for letting me gain experience, grow, and learn to become who I am and get where I am today.

Each of my committee members has provided me with invaluable gifts for my future that I will always carry: Dr. Hicks with her belief in me; Dr. Simoes with the vision of always dreaming bigger; Dr. Jackson-Thompson for shaping me and making it all possible; Dr. Boren who showed me that everything is achievable through precise knowledge and diligence; Dr. Erdelez for completing the picture and providing a vision of creativity; and Dr. Davis who put things into perspective for me and challenged me early on. I am also grateful to Dr. Chi-Ren Shyu, Director of the MU Informatics Institute, for his genuine advice, enthusiasm, expectations and guidance through my years at MU Informatics Institute.

Dean George Justice has given me invaluable insight and understanding for academics through Preparing Future Faculty (PFF) which helped me to get ready for research, teaching, and service. During this time I met Dr. Vicky Sauter from the University of Missouri-St Louis; both she and Dr. Tony Kazic gave me insight and courage to look ahead.

I am also grateful to my mother, Kristina; her support and outlook of never giving up hope and caring will always inspire me. I have saved the most important acknowledgements for last. I owe everything that I am today to my husband Rudy and my family. This work is dedicated to my children, Isaac, Christopher, and Abraham, who are growing up while I have been in school. I hope you will have a vision and dream for your life. Dream big!

Finally, I hope I can contribute in a small way to further dreams and encourage others in their lives like you have all inspired me.

TABLE OF CONTENTS

Acknowledgments ii

List of Tables vii

List of Figures viii

Abstract ix

Chapter 1: Introduction 1

Core Data Set 7

Coding and Classification of Cancer 9

Informatics Tool and Data Access 10

The Problem and the Goal for the Core Dataset 11

Use and Impact of the Project 13

Health Informatics Tools to Support Disease Registries 14

Data Request Tool 16

Glossary 17

References 20

Chapter 2: Article Systematic Review 23

Development and Use of Information Technology in the Field of Public Health and Cancer Registration: A Systematic Review 23

Abstract 23

Introduction 24

Methods 26

Search Strategy *27*

Study Selection and Data Extraction *27*

Results 29

Discussion 38

Conclusion 41

References 42

Chapter 3: Article Tool 46

Research in Cancer Registries: Utilization of Informatics Tools for Registry Data Access and Management 46

Abstract 46

Introduction 47

Purpose and Importance of the Study 49

Data Security 50

Data Requests 51

Methods 53

Study Approach 53

System Design and Development 54

User Interface Development 57

Definitions of Terms 58

Results 58

Discussion and Conclusion 59

References 61

Chapter4: Article Survey 63

Effective Information Management in Cancer Registries: Evaluating and Addressing the Needs for Cancer Research and Data Collection 63

Abstract 63

Introduction 65

Background 65

Project Objective *67*

Methods 69

Survey Development *69*

Survey Content *69*

Survey Sample *70*

Survey Administration *70*

Limitations *71*

Results 71

Cancer Registry Survey *72*

Cancer Research Survey *78*

Discussion 81

Conclusion 83

References 84

Acknowledgements *86*

Bibliography 87

Vita 95

Appendix 96

Research Cover Letter *97*

Timeliness, Completeness, Accuracy *98*

Timelines for Coding, Standards and Staging Starting in 1976 ..... *99*

Cancer Registry Questionnaire *100*

Cancer Research Questionnaire *126*

LIST OF TABLES

Table 1: Systematic Review 30

Table 2: Publications on Cancer Data Incidence and Mortality..... 48

Table 3: Type of Data Requests 52

Table 4: Cancer Registry Region and Location..... 73

Table 5: Cancer Registry Data Request..... 75

Table 6: Cancer Registry Data and Data Availability..... 77

Table 7: Cancer Research Region and Location 79

Table 8: Cancer Research Availability of Variables and Missing Data 80

Table 9: Cancer Research Data 81

LIST OF FIGURES

Figure 1: Research Data Request Flowchart 56

Figure 2: US Census Regions 73

Figure 3: NCI Cancer Centers 78

ABSTRACT

Cancer registries in the US and Canada have a long history of data standards and data collection that have developed from a minimal dataset to the standard dataset that is used now. Central Cancer Registries (CCRs) are good resources for cancer data, but are often underutilized. CCRs are recognized for high quality data standards by the Centers for Disease Control and Prevention (CDC) National Program of Cancer Registries (NPCR) or the National Cancer Institute (NCI) Surveillance, Epidemiology, and End Results (SEER) Program and receive certification from the North American Association of Central Registries (NAACCR). Each year, there are many changes to the data that are collected in the cancer registry field. Standards, requirements, and medical knowledge change frequently. The changes in the data collection process cause interference and decrease in quality of data fields, but also delays in the timely collection of cancer registry data. The objective of this study is to identify what essentially needs to be collected and what can be collected optionally in a cancer registry. The goal is a robust dataset that can be used for other disease registries, cancer data surveillance, public health, and research. CCRs and Cancer Centers (CR) were surveyed to identify and describe the data items that are collected and needed to achieve a dataset that can serve cancer surveillance and research. The surveys were analyzed to identify overlaps of common and special interests, as well as barriers. The results showed that cancer registries have data available, but need to look at the timely release of a core dataset for use in cancer surveillance and research. The surveys also evaluated the barriers to data use from cancer registries and barriers for data use of collected datasets to identify the initial data request process. Data in the cancer registry are in a format that can easily be adopted by public health, surveillance, and research. The requesting process needs to be accessible, understandable, and streamlined to enable successful use of the data.

CHAPTER 1: INTRODUCTION

The areas of health informatics, cancer, and cancer registry have been developing rapidly over the past few years. Development and research have specifically occurred in the areas of treatment, research, data collection, and disease reporting.[1] Bernstam, Smith, & Johnson (2010)[2]state that definitions of informatics are based on information, data, and knowledge that in the context of a domain, become different areas of informatics. For example, health informatics is the science of information applied in the clinical and biomedical fields. Health informatics encompasses clinical, biomedical, and practice data that, in the field of cancer, include demographics, diagnostic findings, tumor information, treatment, and follow-up information. The collection of these data is an interdisciplinary process, including the fields of biology, pathology, treatment, imaging, clinical trials, epidemiology, and data collection.[3] Therefore, with this variability of data, a common standardized core data set for collection of cancer data is needed that can be shared among disciplines and utilized by researchers.

Cancer registries have many partners that provide the information needed to get a complete case. The source providers follow a variety of standards and need to work together to make collected cancer data useful, standardized, and organized to support clinical practice, public health, and research. The development of a standardized core dataset can enable researchers to use data easily from the cancer registry and exchange data with less.

Interoperability among data is very important and almost inevitable with the use of cancer data because of the many sources that are needed to collect a cancer case.[4]

Data collection is changing every year, with new data items added continuously, thus making it harder to get a complete dataset published in a timely manner and have relevant valid data available for research. Dimick (2010)[5] points out that updates to cancer registry standards and changes in rules and codes for reporting and collecting of data are taking place every couple of years, but the year 2010 required numerous changes in the cancer registry field, including major updates to coding rules.

Through development of a core dataset, cancer data collection can be more systematic, and scientific cancer data can be used more effectively by a variety of researchers and institutions in a common standardized format. The need and gap is to connect the research community and public health to the data and to enable CCRs to fulfill data requests easily, but still maintain data security and confidentiality. In order to achieve these goals, capturing the data requests and the information about the researcher is very important. The identification of a core dataset is critical in an era where the electronic health record (EHR) is required.

The Nursing Minimum Data Set (NMDS) provides a formal structure for electronic data sets to support nursing care in all settings, with 16 elements that include nursing care elements, client elements, and service provider elements.[6] In order to have a useful data set, and to make cancer registry data accessible, the emphasis needs to be on a core dataset that can be easily maintained and retrieved. It is possible to make the core dataset available and conduct special studies for additional data items that are of interest, and complete data in a timely manner.

Central cancer registries (CCRs) collect a vast amount of information that is based on knowledge that is changing continuously when new medical discoveries are made and new coding and disease treatments are used based on new knowledge.[5] The data set collected by CCRs is growing every year. Bray and Parkin (2009)[7] emphasize, in their articles, that population-based cancer registries have become more than a source that collects, stores, and provides information on the incidence of cancer; they have evolved to cancer information and research data centers.[1, 7] In Missouri alone, the data set has been growing, from a data set that included 25 required data elements in 1985 to a data set that requires more than 200 data elements in 2010. This large increase in data fields and data items is very difficult to manage.

Changes on the dataset include new content in four different reference books that are used for cancer reporting. These major changes were announced by the American Joint Committee on Cancer (AJCC), and used to code primary site, histology, and stage. Collaborative staging, which is a combined stage field, resulted in almost 150 additional fields for registrars to capture, learn and report; it also included revisions and changes to software.[5] Implementation of new hematopoietic and lymphoid neoplasm case reporting and coding rules and implementation of the AJCC Rapid Reporting Requirement that is intended to shorten the abstracting time and additional changes were introduced by the Commission on Cancer (COC).[5] The record length of the previous dataset was 6694 bytes. It is now expanded to a 22,824 byte record after the changes from 2010 were applied. Additionally, the date fields are changing from MM/DD/YYYY to YYYY/MM/DD, to make cancer registry data compatible with national standards.

Text fields in the new dataset were expanded to 1000 characters each, where they previously ranged from 250 to 500 characters. This large amount of data is a challenge to manage, not only for small facilities with limited resources, but also for larger facilities and CCRs. Health information technology is critical in achieving the goals of health care reform in the next few years.[9] The new file size does not allow for easy manipulation of the data, including generating reports and data cleanup and corrections. The size of the MCR database has nearly doubled, which results in increased costs for storage, maintenance, and data backup.

CCRs, mandated in every state, are funded partially by the state. In most states, CCRs are primarily funded through the Centers for Disease Control and Prevention (CDC) National Program of Cancer Registries (NPCR), established by the U.S. Congress in October 1992 (Public Law 102-515, titled the Cancer Registries Amendment Act) and/or by the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results (SEER Program), established in 1971.[10] Although reporting requirements have increased, funding for NPCR registries has not. Funding will not necessarily increase in the future in an environment of shrinking public health resources and unstable funding.[11]

CCRs can ensure funding by engaging in research projects and special studies, if the data collected are complete, accurate, and timely to market their services and make information available in a timely manner to be useful for ongoing cancer surveillance, cancer control, and the possibility of collaboration of multidisciplinary groups.[12]

The three elements most important in the cancer registry are timeliness, completeness, and accuracy. There is a need for assessment of what is collected and what needs to be collected. In the current dataset, many data items are meeting the 95% standard for completion, but others, particularly detailed treatment information, are not. The data items that do not meet the standard for completion are not very useful for research, public health, or surveillance, because only data that are timely, accurate, and complete can be used effectively for any type of research or public health and surveillance activities.

The two programs (NCI/SEER and CDC/NPCR) that collect national cancer incidence data and the organization that certifies CCRs (NAACCR) are standard setters for all registries. Registries follow SEER, or NPCR or both SEER and NPCR requirements as well as NAACCR guidelines. The SEER Program requires that 95% of cases need to be reported within 22 months of the end of the diagnosis year.[13, 14] Until recently, NPCR cases were reported 24+ months after the end of a diagnosis year; since 2009, NPCR has required submission of data at or before 23 months. In addition, NPCR-funded registries are required to report 90% of cases within 12 months of the end of a diagnosis year.[13]

NAACCR, the organization that certifies CCRs, requires submission of data from CCRs within 23 months of the close of the diagnosis year.[13] The Cancer Common Ontologic Representation Environment (caCore group), which consists of the NCI, the Cancer Biomedical Informatics Grid (caBIG), and the NCI Community Cancer Centers Program, has developed a common infrastructure for cancer informatics.

One part of this infrastructure that has been developed by the National Cancer Institute (NCI) is the cancer Biomedical Informatics Grid (caBIG) Information Network (https://cabig.nci.nih.gov/) in order to facilitate data sharing and data exchange.[15] This group is also working on the clinical oncology requirements for the EHR that include guidelines for clinical data elements and interoperability. The group was able to identify almost 200 data elements that support treatment summaries, and is currently working on a core dataset for physician offices. Covitz, et al 2003)[16], which is a short and limited dataset. The caBig makes open access tools available to all users of the network.

Foley (2011)[17] recently published an article that lists some of the problems with the cancer research network. These problems include high cost, with not many cancer centers adopting the data management tools or using the infrastructure, and hard to use software that lacks user support. Other concerns are that the tools are open access tools, which make them anti-competitive and over-marketed. Also, the process of integrating the caGrid within existing databases or legacy systems is not an easy process. Data are collected in many different ways and different programs follow various standards which makes data exchange and data aggregation difficult.[15] The group has developed common data elements that are used to collect and report data to support the sharing of cancer research information through data standardization, research access, and tools.

Core Data Set

The three main indicators for quality of a cancer registry dataset are accuracy, completeness, and timeliness to make the dataset useful and valuable for public health, surveillance, and research.[10] With the increased number of data elements, there is a concern regarding data quality with some data items. The need for data quality does not change, despite the fact that it gets increasingly difficult to maintain quality of all data elements for completeness and accuracy in a timely manner.

Accuracy is necessary to minimize under- and over-estimation of cancer incidence, and to get a quality dataset that is reliable and consistent. Timeliness of the data is needed for the creation of a national up-to-date dataset for research and evaluation. Completeness of the dataset is important to be able to calculate true cancer incidence rates. Accuracy includes, for example, the accuracy of coding; if a cancer is coded to the wrong site, this cancer will be over counted in one site and under counted in another.

Hutchison (et al 2004)[10] describes timeliness as the measure for the data collection processes that includes the reporting on the time span that is necessary for accuracy and usefulness of the data. Completeness is the comprehensiveness of the data and the dataset that are collected; for example certain values of a field and data elements that are not coded as unknowns, or blank values that are not blank or coded as unknown.[10]

Timeliness, completeness, and accuracy are essential for cancer data collection. These qualities of cancer data are needed to achieve a high quality dataset for surveillance data and to establish needs and gaps of public health programs and effective intervention to respond to new developments and treatments in the area of cancer.

The sources for data collection in the cancer registry field are hospitals, pathology laboratories, nursing homes, freestanding cancer centers, and clinical laboratories. Cancer data that are collected are 85% from hospitals, 10% from non-hospital sources, and 5% from other sources, such as DCO (Death Clearance Only) cases that are identified through death certificates and have not been reported by any other source. Hospitals (ACOS approved) are required to have an established cancer registry. Hospitals with less than 200 cases are not required to have a certified tumor registrar (CTR), but must have a designated person to take care of the data collection. Non-Hospital sources are encouraged to report electronically through Public Health Information Network (PHIN). A core dataset would enable facilities with limited resources to report more easily to the central registry. Different facilities have different resources available to them; the goal would be a dataset that all facilities and parties are able to fulfill the requirements of the core data set, even with limited resources to ensure a uniform core dataset. A core data set can serve the registry and the researcher with meeting both interests and needs for quality information.

Coding and Classification of Cancer

Accurate coding of cancer data that are collected is essential, since the use of the data depends on retrieval of the cancer cases using codes. Therefore, the accuracy and completeness of the coded cases are extremely important. Cancer Registries report their data to NPCR and SEER, and are reviewed and rated based on quality, with the main quality indicators of accuracy, completeness, and timeliness. Only registries that meet the quality standards, accuracy, completeness, and timeliness, are included in the national dataset.

The coding system that is used for the coding of cancer is ICD-O3 (Coding of primary site and histology) and ICD-10 CM codes used to code cause of death. The codes for cancer coding are continuously updated and revised, depending on changes in the field, knowledge discoveries, new developments, and treatment modalities. Cancer Data, Health Information and Technology, and Disease Collection. The Missouri Cancer Registry reviews about 30,000 records per year. This includes nearly 27,000 Missouri cases that are ascertained, reviewed, and coded. About 90 percent of the cases originate from hospitals and 10 percent are reported from non-hospitals facilities. The time until cases are ready to be released and published for research and public health and surveillance is two years from the date of diagnosis. Additionally, it takes time to perform any research or study, resulting in analyzed data that are relatively old when it is released and published. Faster access to the collected information is essential, especially in the field of oncology where new treatments and new modalities are discovered frequently and need to be reflected in the data collection process.[18]

Quality data require access to timely information that is accurate and complete. In order to record data precisely, and reliably collect, document, store, and retrieve data fields, data elements must be comparable and consistent. Further, the data items and fields that do not meet completeness and accuracy standards in the cancer registry database are not useful for research studies, public health, and surveillance and do not add to the quality of the dataset. Therefore, a shortened, clinical cancer core dataset would be very useful for immediate research, public health, and surveillance, with the goal of a complete and accurate minimum dataset.

Informatics Tool and Data Access

The better the data that are collected and the faster the data are available for research, public health, and surveillance, the better the research and outcome can be offered based on the cancer data. Research data tracking involves following data from the point the research data request is made to the final state when the data are released to the investigator. In order to support the process of data requests, a tool was developed to support the process of data requests, the tracking of data requests, and the analysis of the type of data requests that are submitted. The research data request process is a manual paper based process that is not standardized for central cancer registries. At the most the researcher has the option to print or download a request from a website. Many CCRs list forms, data item lists, data user agreements, and IRB forms on their website but the actual process of requesting data is still a paper based process.

The process of requesting data needs more attention, if the request includes patient data. There is also a difference if the request includes patient identifiers or if the data request uses de-indentified data.

The Problem and the Goal for the Core Dataset

Cancer data need to be collected and counted, but only complete, accurate, and timely data make a significant difference in the cancer field. The data request process for central cancer registry data is currently cumbersome and not very efficient. The basic questions are: What data elements are needed for public health surveillance, and what data elements are needed for research? By constantly adding data elements, are we getting too specific versus complete? Data elements, for example, for treatment are only of value if the fields are complete, accurate and timely. The data that are collected in the registry field are most beneficial for data analysis and research, public health, and surveillance when the data are accurate, timely, and complete. Fields and/or data elements that are inaccurate or incomplete are not beneficial to cancer surveillance or research. Therefore, in order to get a dataset with the goal of accuracy, completeness, and timeliness, the current dataset needs to be evaluated and data elements and fields identified that are necessary and can be completed accurately. Efforts to standardize data collection procedures (e.g., standard data elements with standard codes in a standardized layout) intensified with the establishment of NAACCR in 1990 and the establishment of NPCR in 1992.[10]

Starting in 1994, NPCR required all CCRs funded by CDC to follow the established reporting guidelines for incidence cases that include registry operations, training, publication of an annual report, and case-finding and re-abstracting audits at reporting facilities.[13]

In order to get the most accurate and complete data, the development of a core dataset that can be used for research, public health, and surveillance and for central cancer registry is beneficial and can easily be expanded to meet the need of special studies or specific study goals. Core datasets are not only developed for cancer registries; Lenti (2008)[19] describes the development of a minimum core data set for stroke registries.

It is more efficient to work with a core minimal dataset and request more data if needed than to request more information regularly that is possibly not complete or unknown. The improved dataset can aid not only cancer registries and researchers, but can also be utilized in cancer surveillance and decision making for public health. The proposed core dataset can enable states and registries that may lack the funding and resources to be able to meet standards and be included in national datasets, leading to greater coverage of the areas and regions where data are collected. The development of the core dataset can not only be used in cancer registration, but also be utilized in other disease registries and expanded to public health and epidemiology.

The development of a core dataset that includes common data elements that have been identified through the survey administered in this study allow for the development of a distinct readable phrase or sentence associated with a data element within a data dictionary.[20]

The NAACCR data standards and data dictionary include the required data elements that are collected by central cancer registries. The core dataset can utilize the NAACCR data dictionary and the Cancer Biomedical Informatics Grid (caBIG), along with the results, to determine a suggested core dataset for registries. Fangxia (2006)[21] states that caGrid is a good choice and has many possibilities for cancer centers to be able to apply the necessary adjustments, but mentions that it will not work in all settings. Ash et al (2008)[4] describes the development of common data elements enforced through a controlled vocabulary, ontology, and semantic modeling methodology. The common data elements are different data types, such as demographic data elements, clinical history data, pathology data, treatment data, recurrence, vital status, and epidemiologic data.

The common data elements result in an enhanced dataset for researchers and the ability to make the system interoperable among several facility types. The University of Virginia maintains a clinical data repository with web access through their website to their clinical data repository that includes key data elements.[22]

Use and Impact of the Project

The aim of the core dataset is to make the data that are collected readily usable and accessible without compromising the quality or confidentiality of the data. Findings from the survey aid in identifying the data items and elements that are necessary for research, surveillance, and data collection.

Currently cancer registries are part of many different areas, like epidemiological and clinical research, cancer prevention and control, screening programs and evaluation and monitoring of cancer programs, follow up of cancer patients, and studies relating to advancements in treatment of cancer.[7] The core dataset can be used, in addition to an extended dataset, but can be made available in a more timely fashion for research and surveillance. In addition to the core data set, a data request tool was developed and implemented to further enhance and improve the process of connecting the CCR data with researchers and research facilities that are in need of data.

Health Informatics Tools to Support Disease Registries

Shortliffe (1984)[23] states that medical information science is the science of using system analytic tools to develop procedures and algorithms for decision making and analysis of medical knowledge, including management and process control of medical knowledge. Today, with the field expanding even more than a few years ago, informatics tools are the solution to many modern day requirements and challenges. Informatics tools are enabling standardization and management of otherwise cumbersome processes.

The enhanced development of informatics tools is emphasized by the HITECH Act in support of the conversion from paper to electronic processes in all clinical settings. The data request tracking tool supports the capture of the data requests electronically, and makes the process of requesting data from registries very clear and transparent.

Data requests are a critical part of any disease registry, and are often a time-consuming and labor-intensive process. We have identified this area as needing and benefitting most from a standardized, streamlined, and electronic process.

With the designation of the Missouri Cancer Registry as a Research Center, growing interest is expected. Informatics tools need to aid the exchange of information and knowledge to support intraoperability and interoperability within and between various organizations.[24] It is critical for registries to become not only information providers, but actual providers in support of data and research needs. In addition to providing data that are needed for operations at the registry and for research, a CCR can become more of a data center that enables the sharing of quality data. Many areas in informatics have developed web-based data request tools that can aid the registry in tracking the data requests.

The Research Data Request tool tracks the data request type, date, data type (de-indentified), signed consent forms, information about the investigator, any revisions to the original data request and release of the data, date, and facility/investigator. The data are entered in the database by category (e.g., cancer type/site, time period, geographic area, project / request name) to allow for queries on those features.

The data tracking tool facilitates meeting HIPPA requirements. HIPPA requires Protected Health Information (PHI) to be maintained for six years from the date of its creation or the date when it was in effect, whichever is later, with the common rule to keep data and records for a minimum of three years after completion of the research project.

Throughout the process, the tool can generate status updates for the investigator and involved CCR staff to view where the data request is in the process. Further, the information is collected in a standardized format. The process of filling data requests includes a data release/ use agreement that includes a standard phrase that researchers/data requestors must include in any data use or publication. The use of that common phrase allows the registry to find how and when the data are used.

Data Request Tool

The data request tool was developed to support the data request process at the CCR, to capture and track data requests, and to identify data requestors. This tool improves the efficiency and timeliness of data requests processed at the CCR. The way data requests are received varies from one CCR to another. Requests may be received verbally, as a written request, or in an electronic form.

The process at MCR-ARC was improved in 2006 to incorporate a tracking system and the design of a data request form based on SEER*Stat to clarify data requests. MCR-ARC no longer accepts verbal requests. The University of Utah Health Sciences Center (UUHSC) has implemented and assessed an open source tool for research data tracking.[24]

The tool is used to identify cohorts of patients and possible subjects for research with expectations of filling the request without having to modify the initial data request.[24] The major difference in a data request is the request for patient or demographic data. Additional review is required to ensure the correct handling of this type of data, which would require pre-

processing of the data in order to fill the data request.[24] The main difference described is the request requires patient data in addition to counts. The most frequently requested data elements that Deshmukh (2009)[24] described were demographics, diagnosis, procedures, medications, and laboratory tests. The data requests that could not be filled by the tool without significant modifications were requests for one or more institution-specific criteria and had one or more temporal criteria such as ranges of dates, or more complex requests such as a specific week of each month or quarter that need to be calculated and run sequentially.[24]

Glossary

ARRA - American Recovery & Reinvestment Act of 2009

caBIG -The cancer Biomedical Informatics Grid

CCR - Central Cancer Registry

caCore - Cancer Common Ontologic Representation Environment

CoC - Commision on Cancer

CDC - Centers for Disease Control and Prevention: Federal agency of the Department of Health and Human Services.

CTR - Certified Tumor Registrar

Data Repository - A logical and sometimes physical partitioning of data where multiple databases that apply to specific applications or sets of applications reside.

E-Health - transfer of health resources and health care by electronic means encompassing three areas delivery of health ie. Tele- medicine, using IT to improve health services, the use of e-business and e-commerce
http://www.who.int/trade/glossary/story021/en/index.html

EHR - Electronic Health Record

EMR - Electronic Medical Record

HIE - Health Information Exchange

HIO - Health Information Organization

HITECH - Health Information Technology for Economic and Clinical Health Act

HIPAA - Health Information Portability & Accountability Act

ICGC - International Cancer Genome Consortium

NAACCR - North American Association of Central Cancer Registries

NCRA - National Cancer Registrars Association

NCI - National Cancer Institute

NCDB - National Cancer Database

NIH - National Institutes of Health

NPCR National Program of Cancer Registries

MCR - Missouri Cancer Registry

MO-HITECH - Missouri Office of Health Information Technology

PHIN - Public Health Information Network

PII - Personally Identifying Information

PHI - Protected Health Information

SEER – Surveillance, Epidemiology, and End Results

Type of Data Request –

Non-sensitive: Non-confidential data elements in aggregate form, e.g. case counts by county, race, or sex. PHI(patient, physician or facility) removed.

Sensitive: Record-level data without names or identifying information; zip codes and county of residence may be included. Cells that show 5 or fewer in a category are suppressed.

Confidential: Data include individual personal identifiers such as name, social security number, or street address which directly link an individual with a

diagnosis. For cancers that are relatively rare, cancer site/type may also be categorized as “confidential.” The individual, record-level data with personal identifiers may be used for the purposes of record linkage, but not direct patient contact.

Confidential (patient contact): Data include individual personal identifiers such as name, social security number, or street address which directly link an individual with a diagnosis. If a research proposal includes contact with patients, justification for contact and IRB approved consent/release of contact information form is required. Cancer patient must provide written permission to MCR before contact information is released to researcher.

References

[1] Parkin DM. The evolution of the population-based cancer registry. *Nat Rev Cancer*, 2006; 6(8), 603-612.

[2] Bernstam EV, Smith JW, Johnson TR. What is biomedical informatics? *J Biomed Inform*. 2010 Feb;43(1):104-10. Epub 2009 Aug 13.

[3] Hersh W. A stimulus to define informatics and health information technology. *BMC Medical Informatics and Decision Making* 2009, 9:24.

[4] Ash JS, Anderson NR, Tarczy-Hornoch P. People and Organizational Issues in Research Systems Implementation. *JAMIA*, Volume 15, Issue 3, May-June 2008, Pages 283-289.

[5] Dimick C. Sweeping Revisions for Cancer Registries. *Journal of AHIMA* 2010. (Accessed 02/11/2011) http://journal.ahima.org/2010/07/01/sweeping-revisions-for-cancer-registries/

[6] NASN National Association of School Nurses. (2004). Position statement: Nursing Classification Systems: North American Nursing Diagnosis Association (NANDA), Nursing Interventions Classification (NIC), and Nursing Outcomes Classification (NOC). Available at http://www.nasn.org/ (Accessed 02/16/2010) http://www.nasn.org/PolicyAdvocacy/PositionPapersandReports/NASNPositionStatementsFullView.

[7] Bray F, Parkin DM. Evaluation of data quality in the cancer registry: Principles and methods. Part I: Comparability, validity and timeliness. *Eur J Cancer* 2009;45(5):747-55.

[8] Bray F, Moller B. Predicting the future burden of cancer. *Nat Rev Cancer* 2006;6:63-74.

[9] Custodio R, Gard A, Graham G. Health Information Technology: Addressing Health Disparity by Improving Quality, Increasing Access, and Developing Workforce. *Journal of Health Care for the Poor and Underserved*. Volume 20, Number 2, May 2009

[10] Hutchison CL, Menck H, Burch M, Gottschalk R. Cancer Registry Management Principles and Practice, NCRA Inc.,2004; 389-440.

[11] Hartman AM, Thun MJ, Ballard-Barbash R. Linking TobaccoControl Policies and Practices to Early Cancer Endpoints: Surveillance as an Agent for Change. *Cancer Epidemiol Biomarkers Prev* 2008; 17:2215-2219.

[12] Biancioni F, Brunori V, Valigi P, Stracci F, LaRosa F. Cancer registry and information technology: a new management system for integrating cancer registry and oncology departments. 2010 IEEE Workshop on Health Care Management, WHCM 2010.

[13] Hofferkamp J (Ed). Standards for Cancer Registries Volume III: Standards for Completeness, Quality, Analysis, Management, Security and Confidentiality of Data. Springfield (IL): North American Association of Central Cancer Registries, August 2008. Chapter 6: Security and Confidentiality.

[14] Havener L, Thornton M, editors, Standards for Cancer Registries Volume II: Data Standards and Data Dictionary, Thirteenth Edition, Version 11.3. Springfield, IL: North American Association of Central Cancer Registries, April 2008.

[15] Saltz J, Oster S, Hastings S, Langella S, Kurc T, Sanchez W, Kher M, Manisundaram A, Shanbhag K, Covitz P. caGrid: design and implementation of the core architecture of the cancer biomedical informatics grid. *Bioinformatics*, Vol.22 no.15, 2006, pages 1910-1916.

[16] Covitz PA, Hartel F, Schaefer C, De Coronado S, Fragosos G, Sahni H, Gustason S, Buetow KH. caCore: a common infrastructure for cancer informatics. *Bioinformatics*. 2003 Dec 12;19(18):2404-12.

[17] Foley J. Report Blasts Problem-Plagued Cancer Research Grid. InformationWeek. April 08, 2011 12:45pm. www.informationweek.com/news/government/enterprise-architecture/229401221 April 29, 2011.

[18] Thobourn KK, German RR, Lewis M, Nichols PJ, Ahmed F, Jackson-Thompson J. Case completeness and data accuracy in the Centers for Disease Control and Prevention's National Program of Cancer Registries. *Cancer*. 2007 Apr 15;109(8): 1607-16.

[19] Lenti G, Agosti M, Massucci M, Zampolini M, Paolucci S, Franeschini M. Developing a minimum data set for stroke patients assessment: the Protocollo di Minima per l'Ictus (PMIC) as a starting point towards an Italian stroke registry. *European Journal of physical &rehabilitation medicine*. 2008 Sep;44(3):263-9.

[20] Mohanty SK, Mistry AT, Amin W, Parwani AK, Schmandt L, Winters SB, Miliken E, Kim P, Whelan NB, Farhat G, Melamed J, Taiolo E, Dhir R, Pass HI, Becich MJ. The development and deployment of Common Data Elements for tissue banks for translational research in cancer – An emerging standard based approach for the Mesothelioma Virtual Tissue Bank. *BMC Cancer* 2008, 8:91.

[21]Fangxia L, Smalley KJ, LondonJW. The Challenges of Integrating caGRID in the Cancer Center Environment. 2006 Kimmel Cancer Center, Thomas Jefferson University (TJU), Philadelphia, PA.

[22] University of Virginia. UVa Clinical Data Repository. Retrieved March 2011.

(http://www.healthsystem.virginia.edu/internet/Copy%20of%20phs/informatics/cdr.cfmenter)

[23] Shortliffe EH, Sondik EJ. The public health informatics infrastructure: anticipating its role in cancer. *Cancer Causes Control* 2006;17(7): 861-9.

[24] Deshmukh VG, Meystre SM, Mitchell JA. Evaluating the informatics for integrating biology and the bedside system for clinical research. *BMC Medical Research Methodology* 2009, 9:70.

CHAPTER 2: ARTICLE SYSTEMATIC REVIEW

DEVELOPMENT AND USE OF INFORMATION TECHNOLOGY IN THE FIELD OF PUBLIC HEALTH AND CANCER REGISTRATION: A SYSTEMATIC REVIEW

Abstract

Cancer registries play an important role in public health surveillance and research, which is enhanced and supported through the use of information technology. The intent of this review is to show the development and use of information technology in the cancer registration field, and the role and impact it has on public health. The authors searched MEDLINE, COMPENDEX/ GEOBASE and IEEE Xplore from 1990 to 2012. Articles were identified if they included the use and discussion of information technology in the cancer registry and public health fields. The authors identified five categories that emerged through the review to be addressed and met in order to be successful in the electronic health information era: 1. Standards, 2. Interoperability/ Collaboration, 3. Access, 4. Outcome measures, and 5. Cost effectiveness. The reviewed articles showed that cancer registries are among the leading users in various information technologies and can be a measure to demonstrate e-health use by various partners and participants.

Introduction

Public health encompasses disease prevention, health promotion and surveillance for a defined population. Public health, according to the United States Public Health Service (USPHS)[1] is the field of health science that is concerned with safeguarding and improving the physical, mental, and social well-being of the community as a whole. A big part of informatics in healthcare focuses on innovative development and use of technology to process information with application at every stage of healthcare, from basic research to delivery of care.[2] Health informatics consists of clinical and biomedical data and clinical practice; in the field of cancer, this includes public health surveillance and research. Data elements collected generally include demographics, diagnostic findings, tumor information, treatment, and follow-up data, but may also include items such as co-morbidities, personal or family history of cancer, and risk factors etc.[3] Public health informatics is described as the application of information science and technology to public health practice and research.[4] The collection of data about cancer is an interdisciplinary process including the fields of biology, pathology, treatment, imaging, clinical trials, and epidemiology.[5]

The public health field has a long history but is only recently emerging and developing within the informatics arena.[6] Kenneth Thorpe presented at the 2011 public health informatics conference kickoff: What can we do with technology: health information technology infrastructure that facilitates the flow of information, allowing us to look at health across communities that enables immediate feedback and quality assessment necessary to succeed in

having information infrastructure not only for both clinical care and preventive care, therefore permitting transformation of the healthcare system.[7]

Cancer registries collect incidence and prevalence data of cancer in a specific geographic area following the CDC's National Program of Cancer Registries (NPCR), the North American Association of Central Cancer Registries (NAACCR), NCI's Surveillance, Epidemiology & End Results (SEER) Program and the American College of Surgeons (CoC) standards. [8] Reporting is mandated in all states from reporting such sources as hospitals, both inpatient and outpatient, nursing homes, surgery centers, cancer treatment centers, pathology laboratories, physician offices, other state cancer registries, and death certificates. Cancer registries provide mortality data, data and linkages to cancer screening and intervention programs, cluster investigations if needed, and data for informatics and clinical research.[9] The role of cancer registration is expanding in the public health field, with a long history and over ten years of quality data that meet timeliness, accuracy, and quality standards for research. Not only is it providing surveillance for health promotion and disease prevention, but also support for research and health programs.[3] Cancer registration, with its long history of data collection, takes the lead in the field for providing accurate, timely, and reliable data in a standardized data layout.[10]

The field of public health needs to make use of all of its resources, and take a multidisciplinary approach to address future challenges. Healthcare costs are rising and medical errors and adverse events are estimated in the thousands, of which half of them are potentially preventable.[11] One of the strategies to reverse these alarming trends is to bring the health care information system up-to-date and transform the healthcare information exchange.[11, 12]

This new emphasis on health informatics and health information exchange offers a unique opportunity for public health to play an even larger role in prevention and health promotion through effective surveillance. One main problem for public health and also for cancer registries, is to be able to access the growing number of variables from various sources[3]. Only through the use of information technology and the advances in informatics can public health meet the requirements of effective public health practice that require timely, accurate, and authoritative information from many different sources.[13]

This systematic review analyzes and describes the role of information technology in the field of public health with emphasis on cancer registration, for the last two decades (1990 to 2012). The goal of the review is to demonstrate 1) the importance of information technology in the field of public health with focus on cancer registration and 2) to identify categories that are necessary for information technology to ensure success in healthcare delivery and service for public health, surveillance, and research. This review is part of a project that includes data set development, data use in cancer registration, and improvement in providing the necessary data access to customers.

Methods

Data sources: Searches in three electronic databases, MEDLINE (1990-2012), Compendex (1990-2006)/ Geobase (1990-2012), and IEEE Xplore were performed (1990-2012). All publication types were included. The searches were limited to English language.

Search Strategy

The search strategy included Medical Subject Headings (MeSH) and keywords in combinations, including public health (MeSH), informatics (MeSH), registries (MeSH), cancer (keyword), disease (keyword), and information technology (keyword). Inclusion and Exclusion Criteria: The inclusion criteria were any article that discussed the development and use of information technology within the field of public health and cancer registry and its outcomes. Articles were included if they were published from 1990 to the present and if they had significant information to contribute to the specific development and use of information technology and informatics in the registry field. Each article was reviewed for information and data in the key areas of use of information technology, informatics, cancer registry, and public health. Articles were excluded if they focused on clinical applications or clinical studies only. Articles that described clinical care or clinical trials including the cancer registry were excluded.

Study Selection and Data Extraction

The reviewer evaluated the titles and abstracts of the identified articles and screened the articles based on the inclusion and exclusion criteria described. Data abstraction was performed by one investigator independently (IZ) using a structured abstraction process, and the abstractions were independently reviewed by another investigator (SAB). The abstraction form is available by authors upon request. Any discrepancies between the two investigators were resolved through discussion and consensus. The articles were included if they met the eligibility criteria. Information was abstracted from the articles regarding the information technology used, the methods, and the conclusions.

Articles were also classified into categories that emerged while abstracting information from the articles: 1) standards, 2) interoperability/collaboration, 3) access, 4) outcome measures, and 5) cost effectiveness. The standards category includes data standards, data dictionaries, standard vocabulary, and metadata that are used in the cancer registry field. The interoperability/collaboration category incorporates any article that included the ability of two or more systems or components to exchange information and to use the exchanged information.[14] Articles were also selected for this category when they described collaborations of systems or groups to improve effectiveness, efficiency, and facility of electronic exchange. Articles that discussed improved or enabled access to electronic health information and data were included in the access category. Articles were selected and included if they discussed enhancement of outcomes or cost effectiveness (CE). For example, outcome measures in cancer registration are survival and quality of life.[15] The articles were included in the category for cost effectiveness if they demonstrated cost savings or a decrease in program costs by success in the program outcomes. The articles were screened and assigned a number depending on the criteria that were discussed or used. Furthermore we selected all keywords and mesh subject headings assigned to the articles and identified 63 different keywords from a diverse range of topics.

Results

This comprehensive literature search identified 461 articles. The titles and abstracts were read and forty-six (46) articles were identified as relevant: Eighteen (18) articles from Medline, eighteen (18) articles from Compendex/ Geobase, and ten (10) articles from IEEE Xplore. The full text of the articles was then read and twenty-one (21) articles met the criteria for inclusion (Table 1).[2-5, 8-24] The studies included use of information technology in the format of software and programming, development of informatics tools, opinion papers, debates, and reviews that advance and support the information technology in public health and cancer registry. The articles were from the United States, the United Kingdom, Australia, Sweden, Germany and Italy.[8, 10, 14, 16, 23, 24] Different types of registries and information technology tools that are or could be used in a registry environment were included. Tools were used in five (5)[10, 11, 12, 15, 23] studies. The tools range from the use of health grid technology, a critical feature of caGrid to create semantic interoperability among data resources, a registration tool that can effectively manage different data sources to decision support systems for a hospital cancer registry.

Table 1: Systematic Review

Authors	Year	Country	Information Technology/ Informatics	Focus of Article	Areas of Need Identified	Methods	Categories*
Yasnoff et al[5]	2000	US	Public health in the information age now and in the future	Public health in the information age	Transformation of public health in the information age	Review	1
Hersh[20]	2002	US	Medical Informatics: Improving health care through information	Discussion of standards	Improve healthcare through information, Develop systems that are easy to use and provide demonstrable benefit		1, 4
Tafazzoli et al[23]	2002	Germany	Integrated decision support in a hospital cancer registry	Informatics within the registry	Efficient use of information for healthcare, Shortcomings in the knowledgebase	Dev. of tool	2, 3, 4
Fenstermacher et al[15]	2005	US	Review of the Informatics Grid (caBIG), one feature is to create semantic interoperability among data resources	Discussion of standards	The need for sharing computational architectures and tools, Informatics Grid (caBIG)	Dev. of tool	1, 2
Green et al[16]	2006	Canada	Identify critical success factors enabling the	Discussion of standards	Information system support as a critical success factor	Case Study	2, 4

Authors	Year	Country	Information Technology/ Informatics	Focus of Article	Areas of Need Identified	Methods	Categories*
			translation of clinical and operational knowledge about effective and efficient chronic care management into primary care practice				
Drolet et al [3]	2007	US	Development of an framework for registries (MDR-OK)	Categorizing of world registries	Cancer registration framework and standards	Review	1
Eckman et al [14]	2007	US	Technologists guide to healthcare interoperability and information infrastructure	Discussion of standards	Interoperability and flexible architecture	Review	2, 4, 5
Kukafka et al [18]	2007	US	The growth and development of public health and its future advancing knowledge management in the public health informational environment	Discussion of standards	Need for public health informatics development	Review	1, 4

Authors	Year	Country	Information Technology/ Informatics	Focus of Article	Areas of Need Identified	Methods	Categories*
Ochs et al [22]	2008	US	Information systems needs for cancer research	Discussion of standards	Information systems needed for cancer research	Review	1, 2, 3
Araujo et al [9]	2009	US	Literature review on public health informatics	Discussion of standards	Interoperability and the diversity of public health	Systematic Review	1, 2
Hersh [2]	2009	US	Discussion on use of terminology in biomedical and health informatics and health information technology	Discussion of standards	Need clear of definitions for informatics and health information technology	Review	1
Parkin [20]	2006	UK	Complete review of population based cancer registries including the history of cancer registration, variables recorded by cancer registries, legal and confidentiality in cancer registration and the role of	Informatics within the registry	Registration of cancer has greatly expanded	Review	1

Authors	Year	Country	Information Technology/ Informatics	Focus of Article	Areas of Need Identified	Methods	Categories*
			population based registries in research.				
Shortliffe et al [22]	2006	US	The public health informatics infrastructure: anticipating its role in cancer	Informatics within the registry	Future for surveillance and information in the culture of cancer care	Review	1
Parkin [7]	2008	UK	The role of cancer registries in cancer control	Informatics within the registry	The role of cancer registries and the development of research as a component of cancer control	Review	1, 4
Taktak et al [24]		UK	Decision support systems in cancer	Informatics within the registry	Use of tools in the registry	Dev. of tool	2, 3
Castro [13]	2009	US	Comparison on use of health informatics in the US and the UK	Discussion of standards	The benefit of health informatics medical data in a timely and efficient manner	Review	2, 3

Authors	Year	Country	Information Technology/ Informatics	Focus of Article	Areas of Need Identified	Methods	Categories*
Bianconi et al [10]	2010	Italy	Web technology Cancer Registry and information technology	Informatics within the registry	Timeliness of data	Dev. of tool	2
Boyd et al [12]	2010	US	Use of Public Health Grid (PHGrid) technology with four main components: Data format, Data storage, Data query, Data visualization	Use of Public Health Grid Technology	Interoperability and timeliness and secure data exchange of information	Dev. of tool	1, 2, 3
Savel et al [21]	2010	US	Public Health Grid (PHGrid) high level architectural framework	Discussion of standards	Data, information and knowledge exchange	Framework	1, 2, 3
Adolfsson et al [18]	2011	Sweden	National Diabetes Register Survey to determine successful reporting from primary health care providers	Informatics within the registry Discussion of standards	Need for outcome measures	Survey	1, 4
Bigus et al [11]	2011	US	Information technology for healthcare transformation to	Outcome measures	Need for outcome measures		2, 4, 5

Authors	Year	Country	Information Technology/ Informatics	Focus of Article	Areas of Need Identified	Methods	Categories*
			measure outcomes				

*Categories: 1=standards, 2=interoperability/collaboration, 3=access, 4=outcome measures, 5=cost effectiveness

Many of the information technology and informatics tools described within the field of public health focus mainly on surveillance, but some are available in cancer registries[15, 23], including the integrated decision support in a hospital cancer registry and the caDSR (cancer Data Standards Repository), a central part of creating semantic interconnections as part of the caGrid architecture. Several articles focused on the need to move forward and include information technology in public health. Seven (7) [7,8, 10, 20, 22-24] articles focused on informatics within the registry.

Registries showed the use of information technology and informatics in the area of common vocabularies and standards, interoperability and information exchange, basic concepts in cancer registration, access, and outcome measures, however some of those potential applications of information technology to public health have yet to be implemented[7]. Information technology is evolving to enable automated reporting from various sources to allow for timely availability of data.[10] This can increase the use of cancer registry data and enhance collaboration and improvements in disease surveillance and prevention.

The eligible articles are listed in Table 1.[2-5, 8-24] Twelve (12) [9-15, 18-22] of the twenty one (21) articles focused on interoperability, information and data exchange, healthcare information infrastructure, and five articles involved both informatics and information technology within the study. New and improved information systems are needed that can meet new challenges related to emerging infections, bio terrorism, and disease reporting that facilitate complete and higher reporting rates.[5,8]

Thirteen (13)[2, 3, 5, 7-9, 12, 15, 18-22] articles included the discussion of standards for the field of informatics and public health. Standards are very specific, and most of the coding systems and standards that are currently used did not consider public health data needs and do not meet the needs of healthcare organizations5. Six articles (6) [12, 13, 19, 21, 23, 24] emphasized and discussed the need for access to data and necessary information. The main problem with expanding the cancer registry scope is the difficulty in accessing an increasing number of variables from a number of sources.[10] Eight (8)[7, 8, 11, 14, 16, 18, 20, 23] articles included the discussion and need for outcome measures, and three (3)[11, 14, 18] articles discussed cost effectiveness and the effective dual use for systems of surveillance developed for specific diseases that can also be used as a surveillance system of adverse effects, and control system for communities.[18] The example of using a system for surveillance as well as for use by clinical providers shows an enabling linkage of clinical care providers and public health.[18] Further, with focus only on individuals at high-risk, the impact on population outcomes may be limited.[18] The field of public health informatics needs to consider the diverse discipline of public health practice to be able to address and focus on outcomes for the population at large.[18] Outcome-based payment models and rewards help in the evaluation of the meaningful use of evidence to improve health outcomes and lower costs.[11] Cost, quality, and productivity are described as the driving force behind the transformation of healthcare to enable improvements for patient care and healthcare delivery.

We identified 63 different keywords from a diverse range of topics. Six main keywords identified were public health, healthcare, informatics (including medical informatics, health informatics, public health informatics), computing, and information technology. The many different keywords indicate the wide and various areas that encompass public health.[2, 9, 18] The analysis of publications showed that information technology and informatics tools are needed and desired in the field of public health; however, they still do not meet the various and broad requirements for all its participants. We specifically included a wide variety of databases from various disciplines to account for the interdisciplinary nature of public health. Cancer registries have a long history of collecting health information from various sources that include paper-based reports, electronic records, HL7 messages, etc., that are getting combined within one cancer registry database. Cancer registries are experienced in bringing health information together in various sources and can provide expertise and experience in the transformation in the field of health informatics to electronic health informatics.

Discussion

Public health needs the use and innovative approach of information technology and informatics to move forward, especially in the era of the EMR, electronic reporting, and health information exchange. One of the problems with the use of information technology in the field of public health and specifically cancer registry is interoperability and information exchange, which needs to be in the same format.

Data integration, data exchange and data access need to work to use compatible data that are available in comparable format for the disciplines to move forward and provide what is needed now and in the future. This is particularly important for the cancer registry where an increasing number of variables from many different sources need to be collected, transmitted and exchanged, and consolidated into one record.[8] Most studies that were selected and met the inclusion criteria are all within the last few years but show only selective and very specific use of information technology. The available information technology has surpassed the actual use or implementation of available information technology. Implementation of available information technology lags behind, even though the trend for medical research that relies on mathematical modeling or data intensive and high speed computing is increasing.[2] In summary, we are dealing with old systems that need to be brought up to standard, with hybrid systems, and with new and up-to-date systems that are not functioning to the best potential within the healthcare system, and therefore, are not able to fulfill the mandate of public health to prevent disease and promote health.[21] Goldin (2006)[28] states that many of these systems and instruments that have been created are excellent research tools, and with funds that are getting smaller, we all need to reuse and use the sources we have available. However, they have evolved separately and, therefore, present an incoherent, fragmented landscape with a great need for integration of existing resources to build an informatics platform for cancer research.[28]

In this systematic review, the authors analyzed eligible articles and content based on use and discussion or recommendations of information technology for the field of public health and cancer registry. We identified five categories that need to be addressed and met in order to be successful in the electronic health information era: Standards, interoperability/collaboration, access, outcome measures, and cost effectiveness. Most articles overlapped with one or two other categories, demonstrating that the categories are interrelated. Interoperability is made possible by the implementation of standards and access that allow for outcome measures that enhance cost effectiveness. The traditional cancer registry is described as retrospective and limited to variables that are determined in health archives that make it more of an information silo, contrary to its role of an interactive information sharing, interoperable, and user-centered provider of cancer data.[10, 12, 25]

We need to look at approaches in different disciplines and where those five categories are already implemented in practice. We need to think large scale -- not national but international -- in order to meet the challenges of the future for public health and cancer registry. One example that demonstrates these categories implemented and met is the international cancer genome network, where ten countries and two European consortia have initiated cancer genome projects under the umbrella of the international cancer genome consortium (ICGC).[29]

Castro (2009)[2], comparing the Unites States and the UK, states that in order to benefit from health informatics, the US needs to develop and support the capability of sharing medical data and knowledge for authorized research in a timely and efficient manner. We suggest that only if the five categories are met at the same time can health information exchange (HIE) be achieved.

Conclusion

In this paper, we have presented development in the field of public health and cancer registry, and emphasized what is important for future development in order to meet the challenges and opportunities that e-health for all participating disciplines presents. Over the next years, there will be large investments and a lot of development, but the challenges can only be met when e-health is addressed on a large scale with all players and participants working together as the ICGC (International Cancer Genome Consortium) demonstrates is possible.[2, 11] The electronic health era does not need more standalone or discipline-specific tools and systems, but rather collaborations and portals that work for most of its participants. We found that if the five identified categories are met, collaborations and requirements for health care information exchange (HIE) and e-health are possible. Therefore, public health informatics cannot only meet and move the current challenges forward, but can also become a leader in the new era of electronic and information based health.

References

[1] United States Public Health Service USPHS Public health definition [Accessed February 7, 2012]. Available at <http://www.usphs.gov/ USPHS definition http://medical-dictionary.thefreedictionary.com/public+health>

[2] Castro D. Meeting national and international goals for improving healthcare: The role of information technology in medical research. *Science and Innovation Policy*, 2009 Atlanta Coneference on Science and Innovation Policy.

[3] Bray F, Parkin DM. Evaluation of data quality in the cancer registry: Principles and methods. Part I: Comparability, validity and timeliness. *Eur J Cancer* 2009;45(5):747-55.

[4] Friede A, Blum HI, McDonald M. Public Health Informatics: how information-age technology can strengthen public health. *Annual Review of Public Health* 1691995) 239-252.

[5] Hersh WR. A stimulus to define informatics and health information technology. *BMC Med Inf Decis Mak* 2009; 9:24.

[6] Araujo J, Pepper C, Richards J, Choi M, Xing J, Li W. The profession of public health informatics: still emerging? . *Int J Med Inf* 2009;78(6): 375-85.

[7] Public Health Informatics Conference 2011 Kicks Off: What Can We Do With Technology? Public Health Informatics 2011 Conference August 22, 2011 [Accessed July 7, 2012]. Available at http://blog.rwjf.org/publichealth/2011/08/22/public-health-informatics-conference-2011-kicks-off-what-can-we-do-with-technology/

[8] Biancioni F, Brunori V, Valigi P, Stracci F, LaRosa F. Cancer registry and information technology: a new management system for integrating cancer registry and oncology departments. 2010 IEEE Workshop on Health Care Management, WHCM 2010.

[9] Drolet BC, Johnson KB. Categorizing the world of registries. *J Biomed Infor* 2008;41(6):1009-20.

[10] Shortliffe EH, Sondik EJ. The public health informatics infrastructure: anticipating its role in cancer. *Cancer Causes Control* 2006;17(7): 861-9.

[11] Kohn LT, Corrigan JM, Donaldson MS. Eds. To Error is Human: Building a Safer Health System. Washington, DC: *National Academy Press*; 2000.

[12] Bigus JP, Campbell M, Carmeli B, Cefkin M, Chang H, Chen-Ritzo C-H, et al. Information technology for healthcare transformation. *IBM Journal of Research and Development* 2011;55(5):.6:1-6:14.

[13]Yasnoff WA, O'Carroll PW, Koo D, Linkins RW, Kilbourne EM. Public health informatics: improving and transforming public health in the information age. *Journal of public health management and practice*. JPHMP 2000; 6(6):67-75.

[14] The Institute of Electrical and Electronic Engineers (IEEE) IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries (New York, NY: 1990).

[15] Parkin DM. The role of cancer registries in cancer control. *Int J Clin Oncol*. 2008;13(2):102-11.

[16] Hersh WR. Medical informatics: Improving health care through information. *JAMA*. 2002;288(16):1955-8.

[17] Tafazzoli AG, Altmann U, Burkle T, Holzer S, Dudeck J. Integrated decision support in a hospital cancer registry. *Artif Intell Med* 2002;24:243-55.

[18] Fenstermacher D, Street C, McSherry T, Nayak V, Overby C, Feldman M. The Cancer Biomedical Informatics Grid (caBIGTM), Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference.17-18 Jan. 2006. 743-6.

[19] Green C, Fortin P, Maclure M, Macgregor A, Robinson S. Information system support as a critical success factor for chronic disease management: Necessary but not sufficient. *Int J Med Inf* 2006;75:818-28.

[20] Eckman BA, Bennett CA, Kaufman JH, Tenner JW. Varieties of interoperability in the transformation of the health-care information infrastructure. *IBM Systems Journal*, 6 Apr 2010.

[21] Kukafka R, Yasnoff WA. Public health informatics. *J Biomed Inform* 2007; 40, 365-369.

[22] Ochs MF, Casagrande JT. Information systems for cancer research. *Cancer Invest* 2008;26(10):1060-7.

[23] Parkin DM. The evolution of the population-based cancer registry. *Nat Rev Cancer* 2006;6(8):603-12.

[24] Taktak AFG, Eleuteri SP, Lake SP, Fisher AC. Decision support systems in cancer. Advances in Medical, Signal and Information Processing, 2008. MEDSIP 2008. 4th IET International Conference.

[25] Boyd T, Savel T, Kesarinath G. The use of public health grid technology in the United States Centers for Disease Control and Prevention H1N1 Pandemic Response. 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops.

[26] Savel T, Kesarinath G, Lee B, StinnJ. The use of public health grid technology in the United States Centers for Disease Control and Prevention H1N1 pandemic response. 24th IEEE International Conference on Advanced Information Networking and Application Workshops, WAINA 2010, p974-8.

[27] Adolfsson ET, Rosenblad A. Reporting systems, reporting rates and completeness of data reported from primary healthcare to a Swedish quality register- The National Diabetes Register. *Int J Med Inf* 2011; 80(9):663-8.

[28] Goldin L. Bioinformatics integration for cancer research-goal question-analysis. IEEE International Conference on Information Technology: Research and Education. 2006.

[29] Hudson TJ, Warwick A, Aretz A, Barker AD, Bell C, Bernabe RR, et al. International network of cancer genome projects. *Nature* 2010;464(7291):993-8.

[30] Armstrong BK. The role of the cancer registry in cancer control. *Cancer Causes Control* 1992;3:569-79.

CHAPTER 3: ARTICLE TOOL

RESEARCH IN CANCER REGISTRIES: UTILIZATION OF INFORMATICS TOOLS FOR REGISTRY DATA ACCESS AND MANAGEMENT

Abstract

Cancer registries are information systems that collect data on demographics, tumor characteristics, and treatment. The data collection of cancer diagnosis includes data from pathology, any treatment, clinical trials, epidemiology, survival, and radiology. Development of a cancer registry data request data capture process designed with REDCap (Research Electronic Data Capture) for electronic data requests online. The online access makes the request process more efficient and aids in making the data available in a more timely and efficient manner. Two important gaps in the cancer reporting field are that data that are collected are not adequately used and available in time. Cancer registries have a longstanding history of collecting data, but are not used to their full potential because of the two barriers of availability and access. Both of those need to be addressed to make cancer registry data more useful and marketable for all purposes, including public health, surveillance and research. Cancer registries hold a great amount of information about all cancer sites and are available for research and public health. Registry data need to be protected and available at the same time. In the previous years we have focused on protection and safeguarding of cancer registry data, but not enough on availability and access to this rich data source.

Introduction

A cancer registry is a system for the collection, storage, analysis and interpretation of cancer incidence data by patient.[1] Incidence is defined as the number of new occurrences of cancer cases in a defined population over a specific time period. A population-based cancer registry attempts to collect all newly diagnosed cancer cases within a specific geographic area. Central cancer registries (CCRs), e.g., regional or state registries, receive information from hospital-based registries and from other sources, such as pathology laboratories, freestanding cancer clinics and treatment centers, physician offices, long-term care facilities, and ambulatory surgery centers; state CCRs receive information about residents of their state diagnosed outside the state from other state CCRs.[2, 3] Data from the above sources are edited for quality and consolidated to remove duplicate cases. Data that are at least 90% to 95% complete and with correct evaluation of primaries i.e. incidence are essential to conduct epidemiological studies and research studies, and to evaluate the effectiveness and appropriateness of cancer prevention and control efforts and cancer trends.[4]

CCRs respond to state and local questions and concerns about cancer and provide information to residents, legislators, health professionals, and the public about the burden of cancer to residents of a particular area, or to sub-populations within the area. While cancer registries are currently the only disease registries mandated in every state, registries are increasingly being established to track other medical conditions such as asthma, trauma, diabetes, and congestive heart failure.[5]

Disease registries are an important part of intervention and improvement of care and outcomes in chronic disease, which makes this project of interest to all registries.[6] Registries have a responsibility not just to collect data, but also to make data available for use in public health surveillance and research, within the limits established to protect patient privacy and confidentiality. Data are reported to and made available through standard setters and national programs. MCR-ARC's data are utilized and available through submissions to NPCR/ NAACCR on their websites.

Table 2: Publications on Cancer Data Incidence and Mortality

Table 2 Publications on Cancer Data Incidence and Mortality

Publications	Availability	Web
United States Cancer Statistics		http://apps.nccd.cdc.gov/uscs/
CDC Wonder	online query system	http://wonder.cdc.gov/
Cancer in North American CINA	(Cina online)	http://cancer-rates.info/naaccr/
Cancer in North America	(publication available online)	http://www.naaccr.org/DataandPublications/CINAPubs.aspx
CINA deluxe Analytic File	available to NAACCR member and researchers	By request
Annual Report to the Nation (NAACCR, NCI, NPCR, ACS)		http://www.naaccr.org/DataandPublications/ARN.aspx
IARC – International Agency for Research on Cancer	(numerous publications)	http://www.iarc.fr/en/publications/index.php
Cancer Control P.L.A.N.E.T	online	http://cancercontrolplanet.cancer.gov/
Annual update of MICA on Missouri DHSS services website	website	http://health.mo.gov/data/mica/MICA/

Missouri State Tumor Registrars Association Annual Meeting 2011 (MoSTRA) presentation September 2011

Purpose and Importance of the Study

The purposes of this study are to demonstrate that easy access can make a difference in the use of cancer registry data and the importance of data access. CCRs must not only have high-quality, complete, and timely data; but the data also need to be available and accessible in order to make a difference for public health, surveillance, and research. Therefore, we decided to work on the process for data requests and make an access point available on our website that will allow for an initial electronic data request online. Informatics tools facilitate standardization and management of cumbersome processes.[7, 8]

The University of Utah Health Sciences Center (UUHSC) has implemented and assessed an open source tool for research data tracking.[9] The tool is used to identify cohorts of patients and possible subjects for research with expectations of filling the request without having to modify the initial data request.[9] The major difference in this type of data request is that the request for PHI makes additional review required in order to ensure the correct handling of that type of data, in addition to counts that require preprocessing to fill the data request.[9] The data request tool makes the data request process streamlined and efficient for the CCR and the data more rapidly available and accessible for the requestor. Faster access to the collected information is essential, especially in the field of oncology where new discoveries, new treatments, and new modalities are discovered frequently and need to be reflected in the data collection process.[10]

The objectives of this study are: 1) to assess the cancer registry data request process; 2) to develop recommendations for the data request process; and 3) to design a data request tool with a framework for a web access point for the process of data requests that can be used in data repositories, specifically in registries. It is critical for registries to become not only information providers, but actual providers in support of data and research needs. Specifically, in addition to providing data that are needed for operations at the registry and for research, a CCR can become a data center that enables the sharing of quality data in support of public health, surveillance, and research.

Data Security

Cancer registry databases contain patient data that include protected health information (PHI) that is covered under the American Recovery and Reinvestment Act (ARRA), Health Information Technology for Economic and Clinical Health Act (HITECH), and Health Insurance Portability and Accountability Act (HIPAA) guidelines. Cancer data comprise PHI from patients including data from vulnerable populations, such as children, prisoners, and individuals with illnesses.[11] Every registry should have a chief technology officer who works directly with the registry director who is responsible for data security at the registry.[2] CCRs are required to maintain the same standards of confidentiality that apply to the doctor-patient relationship and for all PHI indefinitely.[2] Registries are required to protect the privacy of the individual patient, the privacy of the reporting source, assure ethical data use, and abide by confidentiality and privacy rules.[2]

The HITECH Act emphasis on functional electronic medical records (EMRs) includes the need to standardize processes and the development of tools that are linked to the EMRs.[12] Cancer registration is not only collecting and managing cancer data, but also involves securing the data and providing data for public health surveillance and research, which makes the registry an important link in demonstrating electronic availability and accessibility.

Data Requests

The process of data collection involves various steps that start with the data request to the registry. Currently the research data request process is a manual paper-based process that is not standardized for CCRs. The Centers for Disease Control and Prevention (CDC) is currently developing a standardized process for access to cancer data at the central cancer registries.[15] Data requests can be received orally (over the phone), as written requests, or in electronic form. Informatics methods, tools, and systems not only can be used to improve outcomes in patient care, but also in processes related to care.[13,14] Many CCRs list forms, data item lists, data user agreements, and IRB forms on their website or on the state health department's website, but the actual process of requesting data is still a paper-based process.

Data requests to the cancer registry include data linkages, requests for de-identified data, requests for aggregate data, requests that include confidential data, and requests that include confidential data with patient contact. These requests contain variables, data ranges, and years.

Some data requests can be filled by resources that are available and are accessible online. For example, over 50% of central cancer registries have data available online in interactive tables on their websites. For more specialized cancer data requests, an individual request to the central cancer registry has to be made. The central cancer registry receives data requests for data linkages and data requests that encompass statistical data. Tracking of research data starts at the point the research data request is made to the final state when the data are released to the investigator.

Table 3: Type of Data Requests

Table 3 Type of Data Requests

	aggregate	Deidentified (non-confidential)	Confidential	Confidential with patient contact	Data Linkage
Non sensitive	e.g. Case counts by county	Yes, no counts under #6			Yes
Sensitive	Record-level data (includes zipcodes etc…)		Data including PHI such as name, SSN, date of birth, etc.	Data include PHI and patient contact	Yes

Each state in the U.S. has its own process for requesting data from the central cancer registry.[15] In order to be able to provide data, data requests to the central registry need to be completed. The CDC Cancer Registry Data Access for Research Project involves the collecting of information for all states on human subjects review (IRB) application content and processes, required consent processes, data access processes, data linkage processes, and release of identifiers for patient contact by researchers.[15]

Nineteen states assess fees for data requests from the central cancer registry, and have varying time frames in which to fill data requests that range from under two months to 2-6 months. The Data Access Research Project Group divided the states in three categories of less complex, middle complex and more complex process to request data; about half (25) of the states do have a more complex data request process in place.[15]

Methods

Study Approach

For this study a database and data access entry was designed to capture the data requests and requestor. The data dictionary was created and imported into REDCap (Research Electronic Data Capture), which was used to create the database. In order to make the data request process more efficient and streamlined, we assessed the data request process at the Missouri Cancer Registry and Research Center (MCR-ARC) and identified the steps and fields that are necessary to make the data request easy and simple to use. The data request process is described in the figure below. The process starts with the requestor initiating a request through the web portal of the tool. The submitted request triggers an initial review from MCR staff to determine what level of review is required. Requests involving PHI or data at the zip code level require more detailed inquiry than requests for aggregate data.

The overall goal was to make this a generic process that can be used in different settings, and speed up the initial time for data requests response and make the process electronic. The following fields are included in the data access tool demographics about the requestor and descriptive data about the data request.

System Design and Development

We identified a tool and developed a data request process that can aid and assist the research data request and enhance the use of the data request process for the registry with a database that captures the requests. The application is initially developed to support and make the data request process more efficient for researchers and to support the MCR-ARC research group to manage data requests for aggregate, deidentified non confidential, confidential, confidential with patient contact, and data linkages. We chose to use the browser-based, metadata-driven Electronic Data Capture (EDC) software REDCap (Research Electronic Data Capture) to design our metadata tracking data base.[16] REDCap uses a SQL database via a secure web interface.[16] REDCap was initially designed to address common issues in academic biomedical research, therefore, data requestors and especially researchers are familiar with the REDCap interface.[16] The program uses questionnaires for collecting the requestors information and request information, which is linked to the data collection database. This allows for easy expansion or adjustment of the request tool. If considered necessary, the database can be expanded for additional data items.

Redcap was chosen for this project because: 1. the availability to users, no charge for academic use of REDCap, 2. the requestor has access over the web, web-based, 3. accessible via application programming interface (API), 4. shared library access[16]; Redcap can handle surveys as well as electronic data collections, and 5. Support through the REDCap consortium and various working groups[16]. The data request form is for the data requestor and is available on the MCR-ARC website. The functional requirements for our data model for the database were the usability, the ease to replicate, the designs ability to represent various data types, and the data should be organized to allow for easy data retrieval. The unique identifiers for the database include study_id and requestor_id, and are linked to the data request. The database includes fields for: 1. name of requestor, 2. description of data request, 3. time interval of the data request, 4. type of data requested, 5. date of request, 6. organization, 7. new request. The data request is entered into the MCR-ARC Data Request form; the study ID and requestor ID are unique values, the date of the initial request gives the request a time stamp that allows for tracking and analysis of time needed to fill request and make data available to the requestor.

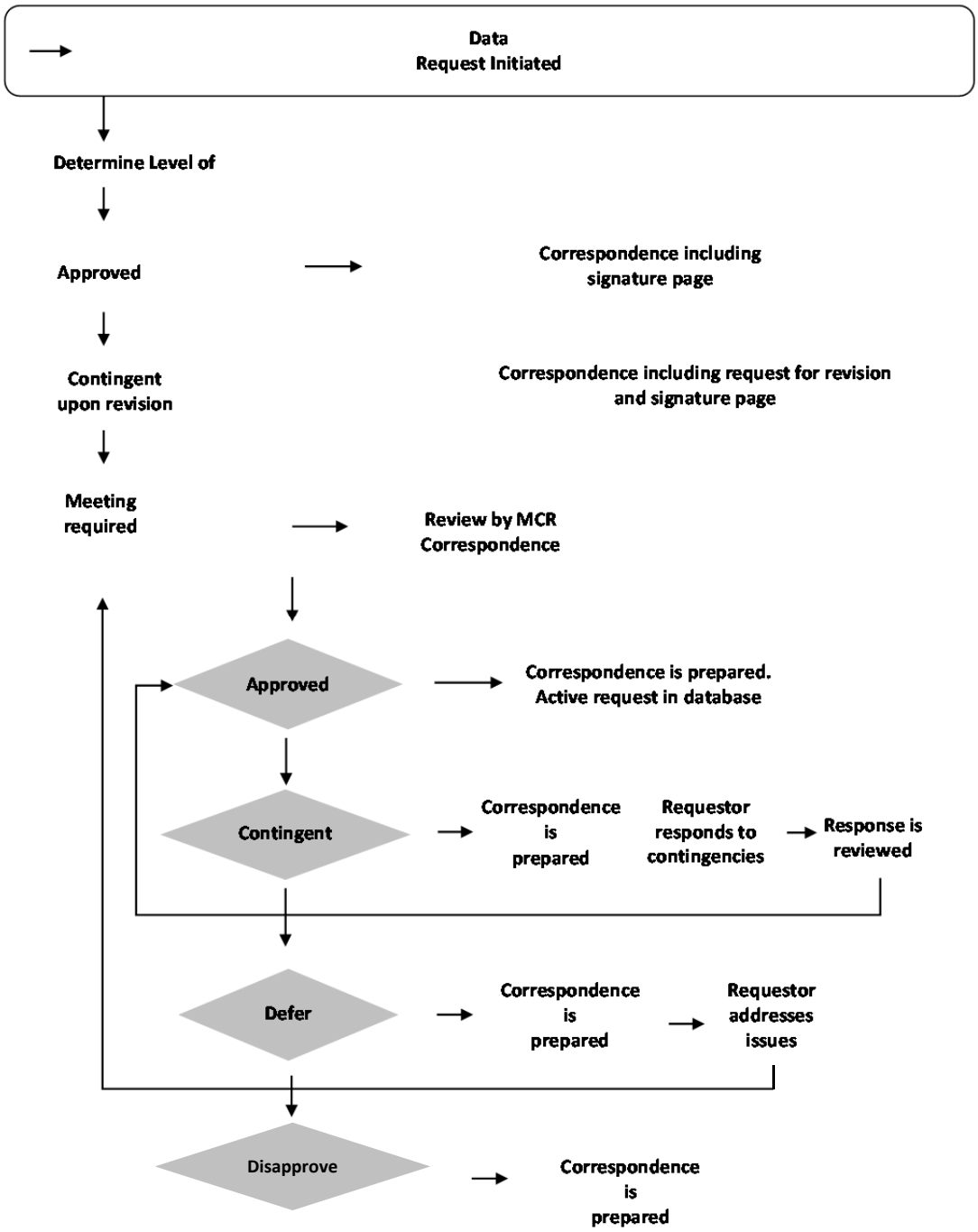

Figure 1: Research Data Request Flowchart

Research Data Request Flowchart Source: [adapted by MCR-ARC from

www.slu.edu/Documents/provost/**irb**/**IRB_Protocol_Flow_Chart**.ppt]

The data dictionary was developed based on our current research request database and includes demographics from the requestor and detailed descriptions and data about the data request including field names, data types and data range etc. for each dimension in their case report form. The data tables are linked to the working web-based EDC forms and application environment. The data request tool is the model of the web application and is testing with our research unit. We reviewed tested our data access internally to be able to validate and improve the process.

User Interface Development

For the first round of testing we used the questions from our paper based forms. These forms were developed based on a review of the database and on experience with direct data requests from the MCR-ARC research group. It includes explanations of different types of requests and offers an option to make the initial request online based on the type of data request. The user interface is currently in the testing phase. From the results of testing the form internally for usability, we decided to include additional fields that allow the requestor to upload their protocol if applicable and/ or available and their IRB protocol if available and applicable to the request. We decided that only three fields are necessarily required fields, Name, Email and Purpose of Request, to be able to get back to the requestor if needed. More usability testing is planned after internal testing based on usability factors.[17]

Definitions of Terms

Reportable Cases MO: All inpatient cancer cases diagnosed and/or treated for cancer in a facility after August 28, 1984 (192.650 RSMo) must be abstracted and reported to the Missouri Cancer Registry, all other sources after 1999

Data Linkage: Data gathering from different sources

Data Request: Request for data for any of the available variables

Informatics: Procedures, protocols, algorithms

Data Access: Approach data, entry to data, make use of data

REDCap Research Electronic Data Capture: Electronic data capture software

Results

The development of an electronic database and web-based application to make the data requests process online allows for easy tracking of data requests, and the type of data that are requested; this allows identification of the need for data in certain areas, like site, and histology or treatment. The central cancer registries can identify who requests their data, monitor new and ongoing requests, and plan for resources if needed. Further, the electronic process ensures that data are kept confidential through an improved and enlightened process; it gives the registry the opportunity to make sure that the data source is acknowledged by requiring a standard form (i.e.: data provided by the Missouri Cancer Registry and Research Center MCR-ARC).

The data request tool makes the process easier through online access to the request by the central cancer registry, with the ability to collect necessary forms and signatures. It allows the user to make the request online, and it provides overview and assistance for the request process. It notifies the registry that a request has been made and serves to streamline the process so data requests can be answered quickly, like some of the requests for linkages or de-identified data. Based on the review of the MCR-ARC research team, the data request entry form was revised to include additional information and give requestors the opportunity to describe the study protocol.

Discussion and Conclusion

The problem of making cancer data accessible but secure at the same time is a process that involves numerous steps and partners in the process. We found that making the process electronically available and streamlining a process helps and encourages the use of the process enormously for interested users. Therefore, it is important and beneficial to have the initial process available electronically, even if the process still requires parts that need to be done and submitted through email or fax, like IRB forms and approvals. Developing and designing this process and tool has become very beneficial for MCR-ARC to determine what data are requested and what data requests are more prevalent to allocate time and resources in the essential area. Tracking of the data requests includes the requestor and the actual data request, which also allows for some administrative use of the tracking data.

We found that making a manual process available online enhanced use and usability for the user and encourage use of the data. The user interface has an intrinsic setup and is easy to use; therefore, it promotes use when tested for usability. The research request tool has a variety of potential users, which include the State of Missouri, but also at the national level, as all registries are receiving data requests. The data request tool would be more useful if certain steps could be tied to the tool. For example, to manage all research forms related to a study, certain approval steps could be linked. Or, for example, if there is another study partner, they could become a user of the request database, and have access to the central location of the data request. MCR-ARC is planning to further develop this tool and make it more user-friendly through some usability studies and electronically accessible by mobile application.

References

[1] Wingo PA. A national framework for cancer surveillance in the United States. *Cancer Causes Control* 2005:16: 151-70.

[2] Hofferkamp J (Ed). Standards for Cancer Registries Volume III: Standards for Completeness, Quality, Analysis, Management, Security and Confidentiality of Data. Springfield (IL): North American Association of Central Cancer Registries, August 2008. Chapter 6: Security and Confidentiality.

[3] Parkin DM. The role of cancer registries in cancer control. *Int J Clin Oncol*. 2008;13(2):102-11.

[4] Havener L, Thornton M, editors. Standards for Cancer Registries Volume II: Data Standards and Data Dictionary, Thirteenth Edition, Version 11.3. Springfield, IL: North American Association of Central Cancer Registries, April 2008.

[5] Casalino L, Gillies RR, Shortell SM, Bodenheimer T, Robinson JC, Rundall T, Oswald N, Schauffler H, Wang MC. External incentives, information technology, and organized processes to improve health care quality for patients with chronic diseases. *JAMA* 2003; 289:434-41.

[6] Metzger J. Using Computerized Disease Registries in Chronic Disease Care. Oakland, CA: California Healthcare Foundation, 2004.

[7] Yasnoff WA, Overhage MJ, Humphreys BL, LaVenture M. A National Agenda for Public Health Informatics. *J Am Med Inform Assoc*. 2001 Nov-Dec; 8(6): 535–545.

[8] Hersh W. A stimulus to define informatics and health information technology. *BMC Medical Informatics and Decision Making* 2009, 9:24.

[9] Deshmukh VG, Meystre SM, Mitchell JA. Evaluating the informatics for integrating biology and the bedside system for clinical research. *BMC Medical Research Methodology* 2009, 9:70.

[10] Thobourn KK, German RR, Lewis M, Nichols PJ, Ahmed F, Jackson-Thompson J. Case completeness and data accuracy in the Centers for Disease Control and Prevention's National Program of Cancer Registries. *Cancer*. 2007 Apr 15;109(8): 1607-16.

[11] Protection of Human Subjects - Vulnerable Populations, Institutional Review Boards, Informed Consent - Research, Participants, IRB, and Federal - StateUniversity.com http://education.stateuniversity.com/pages/2073/Human-Subjects-Protection.html#ixzz2Bn8fU9ie

[12] Health Information Technology: Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology, 2014 Edition; Revisions to the Permanent Certification Program for Health Information Technology.

[13] Balas EA, Austin SM, Mitchell JA, Ewigman BG, Bopp KD, Brown GD. The clinical value of computerized information services. A review of 98 randomized clinical trials. *Arch Fam Med*. 1996;5:271–8.

[14] Hersh WR. Medical informatics: improving health care through information. *JAMA*. 2002;288:1955–8.

[15] Eheman C. CDC Cancer Registry Data Access for Research Project. National Center for Chronic Disease Prevention and Health Promotion, Division of Cancer Prevention and Control
http://www.cdc.gov/cancer/npcr/pdf/Cancer_Registry_Research_Data_Access.pdf

[16] Harris PA, Taylor R, Thileke R, Payne J, Gonzalez N, Conde J. Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009 Apr; 42(2):377-81.

[17] Yen PY, Bakken S. Review of health information technology usability study methodologies. *J Am Med Inform Assoc* 2012 19:413-422.

CHAPTER4: ARTICLE SURVEY

EFFECTIVE INFORMATION MANAGEMENT IN CANCER REGISTRIES: EVALUATING AND ADDRESSING THE NEEDS FOR CANCER RESEARCH AND DATA COLLECTION

Abstract

Cancer registry data collection at a minimum involves collecting data on demographics, tumor characteristics, and treatment. A common, identified and standardized set of data elements is needed to share and make data available quickly and efficiently to all users. This study identifies data elements that are collected and needed to support researchers and public health and surveillance, and to develop a suggested core data set. Surveys were developed for central cancer registries (CCRs) and for researchers and research programs to identify data needs and barriers. Based on the focus of the research, the evaluation of the research registries and databases, and systematic review of the literature, the questions were developed in the following topics: 1.Research, 2. Data collection, 3. Database/ repository, 4. Use of data, 5. Additional data items, 6. Data requests, 7. New data fields, and 8. Cancer registry data set. The review of the surveys showed that cancer registry data are used for public health surveillance (100%) and research (96%). Data are available online in interactive tables by over 50% of CCRs and 87% of CCRs have more than 10 years of data available. CCRs report that treatment data are not complete.

Cancer researchers indicate that they are interested in treatment variables from CCRs. Over 70% of CCRs agree that there are too many required data elements. Cancer registries have data available for use, but need to review what data are needed and used and build collaborations and partnerships to connect common interests in the data.

Introduction

Cancer registries are information systems designed for the collection, management, storage, and analysis of data on individuals diagnosed with cancer.[1] Cancer registries collect data elements that include demographic, diagnosis, tumor, treatment, and outcome information.[2] There are three types of cancer registries: facility-based registries collect information about patients at that facility; specialty registries collect information on one type of cancer (e.g., brain tumors, pediatric cancers); and central cancer registries (CCRs) collect information about cancer patients in a particular geographic area (e.g., a region, state or territory).[3, 4, 5] Cancer registries have a long history of collection and maintenance of data. Standard setters for cancer registries include the American Joint Committee on Cancer (AJCC), the North American Association of Central Cancer Registries (NAACCR), the American Cancer Society (ACS), the Centers for Disease Control and Prevention's National Program of Cancer Registries (CDC-NPCR), the NCI's Surveillance Epidemiology and End Results Program (NCI-SEER), the National Cancer Registrars Association (NCRA) and the World Health Organization (WHO)[6].

Background

The first hospital cancer registry was established at Yale in New Haven, Connecticut in 1926. The first CCRs were established in 1935 in Connecticut and 1946 in California. Public Law 92-218, the National Cancer Act of 1971, directed NCI to "collect, analyze, and disseminate all data useful in the prevention, diagnosis, and treatment of

cancer".[6] This led to establishment of the Surveillance, Epidemiology and End Results (SEER) program in 1973; although covering only 10% of the US population, it was the first national cancer registry in the US. Congress established the National Program of Cancer Registries (Public Law 102-515) in 1992 to be administered by the CDC;[7] the purpose was to establish a CCR in states without one and enhance registries in states that had an existing registry. Together, the two programs cover the entire nation; high-quality cancer incidence data are now available for over 96% (check %) of the US population.

Cancer data need to be collected and counted, but only complete, accurate, and timely data make a significant difference in the cancer field. Parkin (2008) emphasizes that the role of cancer registries has expanded in the last two decades to include not only the collection of cancer diagnosis and treatment data, but also planning and evaluation of cancer control activities, the involvement in patient care and survival data. With the expansion of cancer registration, it is to become a big part of a global activity in the fight against cancer.[8, 9, 10]

Cancer registration is a rapidly changing field that needs to stay in pace with the changing medical field, expansion of the medical knowledge base, and the coding of disease and treatment.[8, 11, 12, 13] The evaluation, surveillance, and prevention of cancer rely on the statistics that are obtained from cancer registry data. The data set that is collected is meaningful when it meets certain criteria, like timeliness, completeness, and quality of data.[14, 15]

This project evaluates the dataset that is collected for completeness and quality, and determines whether problems with certain data items or the need for certain data elements can be identified so a core dataset can be developed, maintained, and stored. The data in the CCR are needed and important, but are only used when it is securely accessible and the need for data items and elements is met. In 2011, the surveys for the CCRs and the research groups were developed and distributed. The surveys addressed data collection, research, questions about missing and incomplete data elements, research requests, quality measures, data elements, and data quality for CCRs and cancer researchers. The cancer registry and cancer research surveys asked questions about needs and additional barriers that exist in the use of cancer registry data. All US CCRs were included in this survey.

Project Objective

The study was developed to be able to identify data elements that are collected and data elements that are needed to support researchers and public health and surveillance to aid in the development of a suggested core data set. The standardization of cancer registry data is an important part of cancer registration, and has developed from approximately 25 required data elements to more than 200 required data elements within the last two decades. The basic questions are: What data elements are needed for public health surveillance, and what data elements are needed for research?

By constantly adding data elements, are we getting too specific versus complete in the data collection? Data elements, for example, for treatment are only of value if the fields are complete. The data that are collected in the registry field are most beneficial for analysis and research, public health, and surveillance when the data are accurate, timely, and complete.[16] The second reason was to identify the overlap and gaps in data that are collected and data that are needed. For example, data collected on comorbidities can be used to impact treatment and impact decisions on prognosis for patients.[17] The surveys were distributed to all CCRs in the US and all NCI designated comprehensive cancer centers and NCI cancer centers. NCI recognizes two types of cancer centers, comprehensive cancer centers and cancer centers, based on the type of grant received. NCI-designated cancer centers meet certain criteria for programs in multidisciplinary cancer research, and dedicate development, funds and resources to research. Based on the surveys the project evaluates the dataset that is collected for completeness and quality, and determines whether problems with certain data items can be identified to be able to develop and maintain recommendations for a core dataset.

Methods

Survey Development

A systematic review of the literature was conducted. For CCRs, we looked at required data elements, both long-term and newly-introduced.

Two surveys were developed by the investigators, with input from experts in the field of cancer registry, survey research, health information technology, and cancer research. Based on the focus of the research, the evaluation of the research registries and databases, and systematic review of the literature, the survey was developed. The study consists of two surveys, one for the CCRs and one for cancer researchers. The surveys answer the questions of what are the essential data elements necessary for the cancer research community and what are the barriers and needs for data for the cancer registry and the cancer researcher. The surveys' content was reviewed internally and the surveys were approved by the Health Science Institutional Review Board at the University of Missouri.

Survey Content

The investigators developed two surveys, one for CCRs and one for researchers, with content specific to the cancer registry based on what data are collected, and a survey with content specific to the researcher on what data are wanted. The instrument for CCRs contained eight topics: 1. Research, 2. Data collection, 3. Database/ repository, 4. Use of data, 5. Additional data items, 6. Data requests, 7. New data fields, and 8. Cancer registry data set.

The survey for researchers contained topics 1 through 6. The pilot survey was sent to 5 internal and 10 external participants, 5 internal and 6 external participants responded and gave suggestions for clarifications of some of the questions in the surveys. The surveys include 41 questions for cancer registries and 32 questions for cancer researcher surveys.

Survey Sample

There are two participant groups for the surveys. One group included every central cancer registry in the United States (51). We sent the CCR survey to all 50 states CCRs and the District of Columbia. All states were included. The other group included researchers at NCI-designated Comprehensive Cancer Centers (41) and National Cancer Institute Cancer Centers (26). We identified researchers at 41 NCI-designated comprehensive cancer centers and 24 cancer centers throughout the U.S and based on referral to the American Medical Informatics Association (AMIA) Clinical Research Informatics working group and AMIA Public Health Informatics working group. The participants were informed of the research study. There was no randomization because of the small sample size.

Survey Administration

The two surveys were administered between October 2011 and March 2012. Both groups received the surveys by email. Non-responders received two reminders each after four weeks.

For the group of the cancer research surveys the survey was also sent to two scientific working groups to open it up to a broader audience, these were sent in February 2012. The survey was closed in May 2012. Data were collected and checked for errors, missing data, and editing and entry errors. Descriptive statistics were applied to the data. We had 43 cancer registries respond to the cancer registry survey and 28 respond to the cancer research survey.

Of the responders, 35 (88%) were NPCR registry responders, 4 (10%) were SEER registry responders and 1 (2%) was both SEER and NPCR funded registry responders. The response rates to questions from NCI designated research centers were more limited. One drawback identified involved finding the appropriate contact person listed, and may have contributed to the lower response rate from cancer researchers.

Limitations

The study had some limitations. One of them was the low response rate for the cancer research group which resulted in opening up the survey to the AMIA clinical research informatics and public health working group.

Results

This study summarizes the survey results and identifies the overlap of data elements/items between cancer registries and researchers to connect the common interest in the collected data. Forty-three out of 51 registries were represented. Survey respondents are divided into two groups, a total of 51 registries and regional registries were included in the survey.

Findings from the survey did clarify the barriers for the use of cancer registry data and clarify data elements that are necessary for research, surveillance, and data collection. Based on the review of the surveys, it became clear that a gap exists between the data elements that are collected and the data elements that are used and needed for surveillance and research. The process of requesting data from a cancer registry is currently cumbersome and not very efficient. There are different data requests that require different handling and processing.

The different types of data requests are non-confidential, confidential, and confidential with patient contact. The data elements needed for public health surveillance and the data elements needed for research are overlapping, but are distinct in their needs. Questions that include not sure or don't know are included in the response count.[18, 19] For this study we calculated the average based on the number of people who responded to each question. We calculated the response rate for each question by number of people who answered the question divided by the total number of people who completed the question.[19]

Cancer Registry Survey

The overall response for CCRs was 43 out of 51 surveys (84.3%). The respondents were evenly distributed over the United States: 11 (27%) from the Northeast, 12 (29%) from the Midwest, 10 (24%) from the South and 8 (20%) for the West. The states included in the regions are listed in the graphic.

Thirty-five (88%) respondents are funded by NPCR, 4 (10%) are funded by SEER and 1 registry is funded by both NPCR and SEER. Twenty-three (66%) registries are located at a State Health Department and 12 (34%) are located at a University.

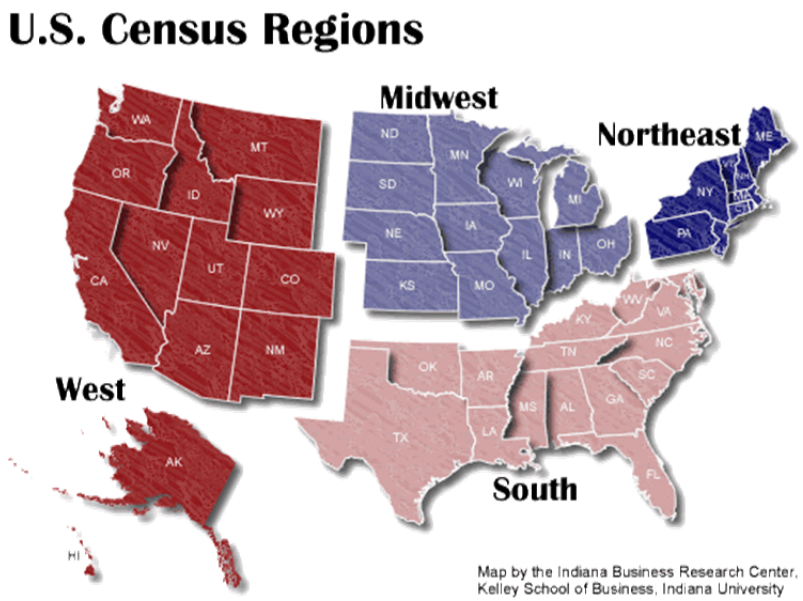

Figure 2: US Census Regions

(http://www.stats.indiana.edu/maptools/maps/boundary/census_regions_main.gif)

Table 4: Cancer Registry Region and Location

Table 4 Cancer Registry Region and Type

Survey Item	N	Response
In which geographic region is your registry located?	41	
Northeast		11 (27)
Midwest		12 (29)
South		10 (24)
West		8 (20)
Where is your registry located?	35	
State Health Department		23 (66)
University		12 (34)
Not sure		0

Registries have good data on demographic and stage, size and histology that are included as part of silver or gold certification, but are lacking data on treatment variables, and as some respondents pointed out, passing silver and gold certification does not necessarily guarantee that the data that are needed for surveillance are available for all types and sites of cancer.

For example some respondents pointed out that it depends on site and histology for how accurate and complete data are even for gold and silver certified registries; capturing of all cases can be challenging for sites that are treated outside of hospital settings, such as melanoma of the skin, prostate, and blood disorders and cancers. In one question, the registries were asked about the top five state specific items collected in addition to the NAACCR/SEER required fields. Twelve (100%) respondents collect tobacco, tobacco history, tobacco years, last name and first name, address at diagnosis, blood quantum, and 1 of the 12 responded not collecting any additional state items. When asked how many years of data the CCRs have available, 20 (87%) respondents said that they had more than 10 years of data available, 2 (9%) had between 6 and 10 years of data available, and only 1 respondent's registry has between 0 to 5 years of data available. Fifteen (68%) respondents reported that they receive updated information on vital status and tumor status for each case. One respondent indicated that they receive vital status only through linkage with state vital records.

Table 5 shows the use of cancer registry data; the respondents were able to choose multiple answers. When asked if the cancer registries could fill all data requests they receive, 15 (68%) answered that they could not fill all data requests, 6 (27%) answered that they could fill all data requests, and 1 (5%) was not sure if they could fill all data requests.

Table 5: Cancer Registry Data Request

Table 5 Cancer Registry Responses Data Requests

Survey Item	N	Response		
Can you fill data requests the cancer registry receives?		Yes	No	Not sure
	22	6 (27)	15 (68)	1 (5)
If you cannot fill all data requests, specify the reason(s)?				
	12			
Data elements are not collected		12 (100)		
Data elements are not available		5 (42)		
Data elements are not reliable		7 (58)		
Date elements are not complete		8 (67)		
Date elements have missing /unknown value		4 (33)		

They were asked if they could not fill all data requests what the reason was. Twelve (100%) said that one reason was that data elements are not collected, 5 (42%) said that data elements requested are not available, 7 (58%) said that data elements are not reliable, 8 (67%) answered that data elements requested are not complete, and 4 (33%) responded that data elements have missing and unknown values. Several respondents mentioned to not have enough staff to fill certain data requests or the necessary approval for the data requests is missing.

When asked what they think of the number of data items they are required to collect, 17 (77%) answered too many, 5 (23%) answered about right, none answered too few, and some respondents placed emphasis not on the number of data items but on type and quality of those items. When asked if they were interested in additional data items, 19 (86%) responded no and 3 (14%) responded yes, pointing out that areas such as socioeconomic factors, family history, genomic assays, tumor and bio markers are needed to keep up with the development in the field.

One of the respondents suggested that the collection for cancer registries has shifted, which brings up a very good point: do registries collect what is needed. Should we look at our dataset and evaluate the utility and use/ need of what is collected in the registry, which comes back to the discussion--should we have a core data set that then has additional needed fields based on use for public health, surveillance, and research. Cancer registry respondents indicated that the data are used for public health surveillance (23; 100%), database linkages that include programs like Breast and Cervical Cancer Control Programs (22; 96%), and research (21; 91%), followed by cancer inquiries (20; 87%), special projects (19; 83%), next-of-kin requests (11;48%), and clinical trials (4;17%), and a few named tissue repository of bio-specimens patterns of care studies, FDA monitoring projects, and program planning. Cancer registry data are available online in interactive tables at the county level by 13 (59%) of CCRs.

Researchers focus on treatment and need treatment variables for many studies, where public health and surveillance focus is on the actual cancer diagnosis and follow up, vital status and follow up to determine planning, survival and cancer control.

By constantly adding data elements, one question is: are we getting too specific versus complete? Data elements, for example, for treatment are only of value if the fields are complete. The data that are collected in the registry field are most beneficial for data analysis and research, public health, and surveillance when the data are accurate, timely, and complete. Eighty-seven percent (87%) of cancer registry respondents have more than 10 years of population based data available for use.

Table 6: Cancer Registry Data and Data Availability

Table 6 Cancer Registry Data

Survey Item	N	Responses		
Cancer registry data and availability		Yes	No	Not sure
Do you receive updated information on vital status and tumor status for each case?	22	15 (68)	7 (32)	0
Do you make the cancer registry data available for data requests?	22	22 (100)	0	N/A
Are registry data available online in interactive tables?	22	13 (59)	9 (41)	N/A
Can you fill all data requests the cancer registry receives?	22	6 (27)	15 (68)	1 (5)
Do you consider NAACCR silver certification as research quality data?*	20	14 (70)	6 (30)	N/A
Are you interested in additional data items or elements that are not mentioned. If yes, what additional data items are you interested in?	22	3 (14)	19 (86)	N/A

*Silver certification meets: Case ascertainment has achieved 90% or higher completeness. A death certificate is the only source for identification of fewer than 5% of reported cancer cases. Fewer than 0.2% duplicate case reports are in the file. All data variables used to create incidence statistics by cancer type, sex, race, age, and county are 97% error-free. Less than 3% of the case reports in the file are missing meaningful information on age, sex, county. Less than 5% of the cases in the file are missing meaningful information on race (US only).The file is submitted to NAACCR for evaluation within 23 months of the close of the diagnosis year under review.

(Appendix: Timeliness, Completeness and Accuracy)

Cancer Research Survey

We had 28 overall respondents for the cancer research survey of 66 NCI designated cancer centers, including AMIA Public Health Informatics working group and AMIA Clinical Research Informatics working group. The respondents were fairly even distributed over the United States, 6 (21%) from the Northeast, 9 (32%) from the Midwest, 7 (25%) from the South and 4 (14%) from the West. Eighteen (72%) of the respondents are affiliated with a university or teaching hospital, 5 (20%) with a hospital, 2 (8%) with a physician group and others were state health department, federal government agency, clinical research company or board of health.

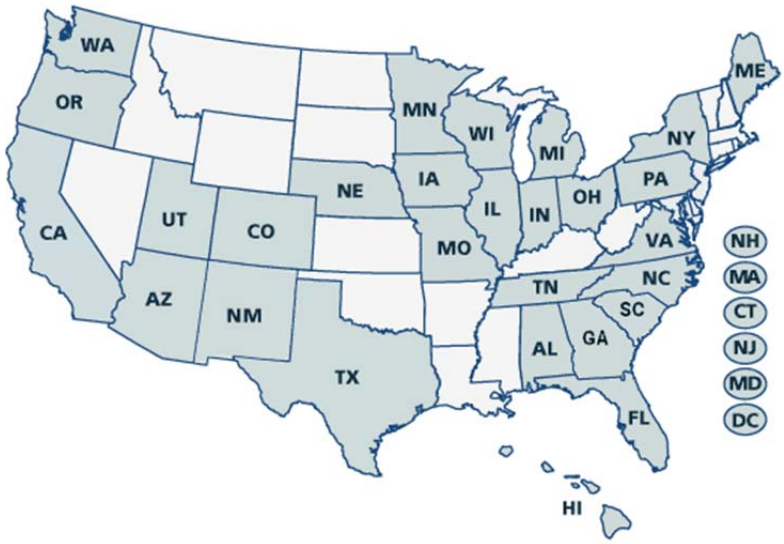

Figure 3: NCI Cancer Centers

(http://cancercenters.cancer.gov/cancer_centers/cancer-centers-list2.html)

Table 7: Cancer Research Region and Location

Table 7 Cancer Research Region and Location

Survey Item	N	Response
In which state is your facility located?	26	
Northeast		9 (32)
Midwest		9 (32)
South		7 (25)
West		4 (14)
Are you affiliated with?	25	
Hospital		5 (20)
University/ Teaching Hospital		18(72)
Physician Group		2 (8)
Not applicable		2 (8)

When asked how many years of data they have available in their database, the majority responded with greater than 5 years. The cancer researchers do collect data on demographic data items, mainly race, ethnicity, and data of birth, and have less than 5% missing or unknown values in those fields, respectively; 50% (8) for date of birth; 46.7% (7) and 33.3% (5) for ethnicity.

Table 8: Cancer Research Availability of Variables and Missing Data

Table 8 Cancer Research Variables

Survey Item	N	Response					
Which of the following demographic data items do you collect?		Yes	No	Not sure			
Name	16	8 (50)	6 (38)	2 (12)			
SSN	16	6 (38)	8 (50)	2 (12)			
Date of Birth	16	12 (75)	3 (19)	1 (6)			
Ethnicity	16	12 (75)	3 (19)	1 (6)			
Race	16	12(75)	3 (19)	1 (6)			
Adress at Diagnosis	16	8 (50)	6 (38)	2 (12)			
Current Address	16	6 (38)	8 (50)	2 (12)			
What percent of the variables have missing or unknown values?							
		<5	10-May	25-Nov	26-50	>50	Not sure
Date of birth	16	8	2	0	0	0	6
SSN	15	4	1	0	0	1	9
Race	15	7	2	0	1	0	5
Name	15	6	1	0	0	0	8
Sex	16	11	0	0	0	0	5
Ethnicity	15	5	2	1	1	0	6
Date of Birth	15	9	1	0	0	0	5
Social security number	15	4	1	0	0	1	9
Address at diagnosis	14	4	1	0	0	0	9
Current address	14	4	0	0	0	0	10
Other (please specify)	6	Other (please specify)	Other (please specify)	Other (please specify)	Other (please specify)	Other (please specify)	6

The cancer research respondents when asked what other data elements or fields they are interested in, listed treatment, staging, outcomes, diagnosis information, number of hospital stays, occupation and industry, comorbidities, family history, biomarkers, diet, exercise, treatment failure, AJCC staging, diagnostic evaluations and cause of death. At least a few of them like the staging, diagnostic evaluations, and diagnostic information cancer registries can provide.

Cancer research respondents have information on tumor and treatment variables, but have missing or unknown values in both categories. The majority does not know how many data elements have missing or unknown values in these two categories. Eight (62%) of the respondents have less than 5% missing or unknown values for site.

Table 9: Cancer Research Data

Table 9 Cancer Research Data

Survey Item	N	Response		
		Yes	No	Not sure
Do you collect information on vital status and tumor status?	14	10 (71)	1 (7)	3 (21)
Are your data complete for all fields in most of your studies?	11	7 (63)	3 (27)	1 (9)
Are the data standardized?	12	11 (92)	1 (8)	0
Are the data deduplicated?	12	8 (67)	2 (17)	2 (17)
Do you know what data elements are available from the state?	12	8 (67)	2 (17)	2 (17)

Discussion

Data collected in the cancer registry field are most useful for research, public health surveillance, evaluation, etc., when the data are accurate, timely, and complete. Basic questions and challenges include difficulties in collecting complete, accurate and timely data elements for all data elements, including all types and histologies of cancer, and to identify the data elements that are needed and to identify common data elements.

Responses point out some data elements that are essential for the cancer research community, as well as barriers and needs for the cancer registry and the cancer researcher. An important issue is the constant addition of new data elements to the required data set by national standard setters.

Are we asking for too many data items with too much specificity at the cost of getting less complete and accurate basic data elements? For example, treatment is only of value if all the fields related to treatment are complete. Over 70 percent of CCRs agree that too many data elements are required, yet data requests often cannot be filled because researchers want data elements that are not collected or not available. More dialogue is needed.

Respondents precisely asked for more than there is available from the cancer registries, but the question is not only more but what. Do cancer registries with new developments in gene discoveries, tumor, and biomarkers have to rethink their collection and dataset? Can cancer registries become new leaders by going hand in hand with the bioinformatics laboratories in forming partnerships that rely on data exchange for necessary research and work in the cancer field? The gap and challenge of meeting the needs of public health, surveillance, and research for the future opens up new opportunities for new collaborations and partnerships for health and bioinformatics.

Conclusion

Most CCRs have complete, timely, and accurate data for all the fields that are required by standard setters, but may lack other variables important to research. A possible solution is that population based registries focus on collecting a core data set that is timely, accurate, and complete that fills the need for public health and surveillance.

Additionally, CCRs collect data fields that are state-specific and dependent on need for clinical research to fill the need for research (i.e., special studies for a specific cancer site for a preset amount of time). CCRs have valuable data that should be available and accessible, not only for public health and surveillance, but also for research. Cancer registries have data readily available for use that are needed and wanted by cancer researchers, and continuous collaborations and new partnerships can be beneficial to the cancer registries and to the cancer researchers.

Based on the written responses, there is a need for a more involved study on need and use of data elements for both groups. Furthermore, the role of CCRs is expanding with advancements in genomics, tumor marker information that is dependent on site, pathways and predictive markers that have prognostic significance and play important roles for cancer surveillance, cancer control, and research. All this information and change needs to be considered when developing a core data set for registries.

References

[1] SEER Training Modules, Brief History of Cancer Registration. U. S. National Institutes of Health, National Cancer Institute. 10/15/2012 http://training.seer.cancer.gov/

[2] Wingo PA. A national framework for cancer surveillance in the United States. *Cancer Causes Control 2005*:16: 151-70.

[3] Menck H, Smart C (eds). *Central Cancer. Registries, Design, Management, and Use*. (Hargood Academic Publishers, Switzerland, 1994).

[4] Hutchison LL, Roffers SD, Fritz AG (eds). *Cancer Registry Management, Principles and Practice* (Kendal Hunt, Iowa, 1997).

[5] Austin DF. Cancer registries: A tool in epidemiology. Rev Cancer Epidemiol, ed. Lilienfeld AM, Elsevier North-Holland, Pub, NY, NY 2:119-140, 1983.

[6] Havener L, Thornton M, editors. Standards for Cancer Registries Volume II: Data Standards and Data Dictionary, Thirteenth Edition, Version 11.3. Springfield, IL: North American Association of Central Cancer Registries, April 2008.

[7] Volkers N. National Statistics Source Reaches 2oth Anniversary. *Journal of the National Cancer Institute*, Vol. 85, No. 23, December 1, 1993.

[8] Parkin DM. The role of cancer registries in cancer control. Int J Clin Oncol. 2008;13(2):102-11.

[9] Wagner G. History of cancer registration. In: Jensen OM, Parkin DM, Maclennan R, Muir CS, Skeet RG, eds. Cancer registration, principles and methods. Lyons: International Agency for Research on Cancer, 1991:3-6. (IARC Scientific Publication No 95.)

[10] Heinze O, Ihls A, Bergh B. Development of an open source provider and organization registry service for regional health networks. HEALTHINF 2010 - 3rd International Conference on Health Informatics, Proceedings, p 535-537, 2010.

[11] Parkin DM. The evolution of the population-based cancer registry. *Nat Rev Cancer*, 2006; 6(8), 603-612.

[12] Bray F, Moller B. Predicting the future burden of cancer. *Nat Rev Cancer* 2006;6:63-74.

[13] Biancioni F, Brunori V, Valigi P, Stracci F, LaRosa F. *Cancer registry and information technology: a new management system for integrating cancer registry and oncology departments*. 2010 IEEE Workshop on Health Care Management, WHCM 2010.

[14] Bray F, Parkin DM. *Evaluation of data quality in the cancer registry: Principles and methods*. *Part I: Comparability, validity and timeliness*. *Eur J Cancer* 2009;45(5):747-55.

[15] Das A. Cancer registry databases: an overview of techniques of statistical analysis and impact on cancer epidemiology. Methods Molecular Biology 2009;471:31-49.

[16] Fleurant M, Kell R, Love J, Jenter C, Volk LA, Zhang F, Bates DW, Simon SR. Massachusetts e-Health Project increased physicians' ability to use registries, and signals progress toward better care. Health Aff 2011;30(7):1256-64.

[17] Piccirillo JF, Tierney RM, Costas I, Grove L. *Prognostic importance of comorbidity in a hospital-based cancer registry*. JAMA. 2004 May 26;291(20):2441-7.

[18] Salant P, Dillman DA. (1994) *How to conduct your own survey*. New York: Wiley.

[19] Sorra J, Franklin M, Streagle S. *Survey User's Guide: Hospital Survey on Patient Safety Culture*. AHRQ Publication No. 08-0060, September 2008. Agency for Healthcare Research and Quality, Rockville, MD.
http://www.ahrq.gov/qual/nhsurvey08/nhguide.htm

Acknowledgements

Missouri Cancer Registry and Research Center is supported, in part, by a cooperative agreement between the Centers for Disease Control and Prevention (CDC) and the Missouri Department of Health and Senior Services (DHSS) (#U58/DP003924-01) and a Surveillance Contract between DHSS and the University of Missouri.

BIBLIOGRAPHY

Adolfsson ET, Rosenblad A. Reporting systems, reporting rates and completeness of data reported from primary healthcare to a Swedish quality register- The National Diabetes Register. *Int J Med Inf* 2011;80(9):663-8.

Araujo J, Pepper C, Richards J, Choi M, Xing J, Li W. The profession of public health informatics: still emerging? . *Int J Med Inf* 2009;78(6): 375-85.

Armstrong BK. The role of the cancer registry in cancer control. *Cancer Causes Control* 1992;3:569-79.

Ash JS, Anderson NR, Tarczy-Hornoch P. People and Organizational Issues in Research Systems Implementation. *JAMIA*, May-June 2008; 15(3):283-289.

Austin DF. Cancer registries: A tool in epidemiology. *Rev Cancer Epidemiol*, ed. Lilienfeld AM, Elsevier North-Holland, Pub, NY, NY 2:119-140, 1983.

Balas EA, Austin SM, Mitchell JA, Ewigman BG, Bopp KD, Brown GD. The clinical value of computerized information services. A review of 98 randomized clinical trials. *Arch Fam Med*. 1996;5:271–8.

Bernstam EV, Smith JW, Johnson TR. What is biomedical informatics? *J Biomed Inform*. 2010 Feb;43(1):104-10. Epub 2009 Aug 13.

Biancioni F, Brunori V, Valigi P, Stracci F, LaRosa F. Cancer registry and information technology: a new management system for integrating cancer registry and oncology departments. 2010 IEEE Workshop on Health Care Management, WHCM 2010.

Bigus JP, Campbell M, Carmeli B, Cefkin M, Chang H, Chen-Ritzo C-H, et al. Information technology for healthcare transformation. *IBM Journal of Research and Development* 2011;55(5):.6:1-6:14.

Boyd T, Savel T, Kesarinath G. The use of public health grid technology in the United States Centers for Disease Control and Prevention H1N1 Pandemic Response. 2010 IEEE

24th International Conference on Advanced Information Networking and Applications Workshops.

Bray F, Parkin DM. Evaluation of data quality in the cancer registry: Principles and methods. Part I: Comparability, validity and timeliness. *Eur J Cancer* 2009;45(5):747-55.

Bray F, Moller B. Predicting the future burden of cancer. *Nat Rev Cancer* 2006;6:63-74.

Casalino L, Gillies RR, Shortell SM, et al. External incentives, information technology, and organized processes to improve health care quality for patients with chronic diseases. *JAMA* 2003; 289:434-41

Castro D. Meeting national and international goals for improving healthcare: The role of information technology in medical research. *Science and Innovation Policy*, 2009 Atlanta Conference on Science and Innovation Policy.

Covitz PA, Hartel F, Schaefer C, De Coronado S, Fragosos G, Sahni H, Gustason S, Buetow KH. caCore: a common infrastructure for cancer informatics. *Bioinformatics*. 2003 Dec 12;19(18):2404-12.

Custodio R, Gard A, Graham G. Health Information Technology: Addressing Health Disparity by Improving Quality, Increasing Access, and Developing Workforce. *Journal of Health Care for the Poor and Underserved*. May 2009;20(2).

Das A. Cancer registry databases: An overview of techniques of statistical analysis and impact on cancer epidemiology. *Methods Molecular Biology* 2009;471:31-49.

Deshmukh VG, Meystre SM, Mitchel JA. Evaluating the informatics for integrating biology and the bedside system for clinical research. *BMC Medical Research Methodology* 2009, 9:70.

Dimick C. Sweeping Revisions for Cancer Registries. *Journal of AHIMA* 2010. (Accessed 02/11/2011) http://journal.ahima.org/2010/07/01/sweeping-revisions-for-cancer-registries/

Drolet BC, Johnson KB. Categorizing the world of registries. *J Biomed Infor* 2008;41(6):1009-20.

Eckman BA, Bennett CA, Kaufman JH, Tenner JW. Varieties of interoperability in the transformation of the health-care information infrastructure. *IBM Systems Journal*, 6 Apr 2010.

Eheman C. CDC Cancer Registry Data Access for Research Project. National Center for Chronic Disease Prevention and Health Promotion, Division of Cancer Prevention and Control.
http://www.cdc.gov/cancer/npcr/pdf/Cancer_Registry_Research_Data_Access.pdf

Fangxia L, Smalley KJ, LondonJW. The Challenges of Integrating caGRID in the Cancer Center Environment. 2006 Kimmel Cancer Center, Thomas Jefferson University (TJU), Philadelphia, PA.

Fenstermacher D, Street C, McSherry T, Nayak V, Overby C, Feldman M. The Cancer Biomedical Informatics Grid (caBIGTM), Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference.17-18 Jan. 2006. 743-6.

Fleurant M, Kell R, Love J, Jenter C, Volk LA, Zhang F, Bates DW, Simon SR. Massachusetts e-Health Project increased physicians' ability to use registries, and signals progress toward better care. Health Aff 2011;30(7):1256-64.

Foley J. Report Blasts Problem-Plagued Cancer Research Grid. InformationWeek. April 08, 2011 12:45pm. www.informationweek.com/news/government/enterprise-architecture/229401221 April 29, 2011.

Friede A, Blum HI, McDonald, M. Public Health Informatics: how information-age technology can strengthen public health. *Annual Review of Public Health* 1691995) 239-252.

Goldin L. Bioinformatics integration for cancer research-goal question-analysis. IEEE International Conference on Information Technology: Research and Education. 2006.

Green C, Fortin P, Maclure M, Macgregor A, Robinson S. Information system support as a critical success factor for chronic disease management: Necessary but not sufficient. *Int J Med Inf* 2006;75:818-28.

Harris PA, Taylor R, Thileke R, Payne J, Gonzalez N, Conde J. Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009 Apr; 42(2):377-81.

Hartman AM, Thun MJ, Ballard-Barbash R. Linking TobaccoControl Policies and Practices to Early Cancer Endpoints: Surveillance as an Agent for Change. *Cancer Epidemiol Biomarkers Prev* 2008; 17:2215-2219.

Havener L, Thornton M, editors, Standards for Cancer Registries Volume II: Data Standards and Data Dictionary, Thirteenth Edition, Version 11.3. Springfield, IL: North American Association of Central Cancer Registries, April 2008.

Health Information Technology: Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology, 2014 Edition; Revisions to the Permanent Certification Program for Health Information Technology.

Heinze O, Ihls A, Bergh B. Development of an open source provider and organization registry service for regional health networks. HEALTHINF 2010 - 3rd International Conference on Health Informatics, Proceedings, p 535-537, 2010.

Hersh W. A stimulus to define informatics and health information technology. *BMC Medical Informatics and Decision Making* 2009, 9:24.

Hersh WR. Medical informatics: improving health care through information. *JAMA*. 2002;288:1955–8.

Hofferkamp J (Ed). Standards for Cancer Registries Volume III: Standards for Completeness, Quality, Analysis, Management, Security and Confidentiality of Data. Springfield (IL): North American Association of Central Cancer Registries, August 2008. Chapter 6: Security and Confidentiality.

Hudson TJ, Warwick A, Aretz A, Barker AD, Bell C, Bernabe RR, et al. International network of cancer genome projects. Nature 2010;464(7291):993-8.

Hutchison LL, Roffers SD, Fritz AG. (eds). *Cancer Registry Management, Principles and Practice* (Kendal Hunt, Iowa, 1997).

Hutchison CL, Menck H, Burch M, Gottschalk R.. Cancer Registry Management Principles and Practice, NCRA Inc.,2004; 389-440.

Kohn LT, Corrigan JM, Donaldson MS. Eds. To Error is Human: Building a Safer Health System. Washington, DC: *National Academy Press*; 2000.

Kukafka R, Yasnoff WA. Public health informatics. *J Biomed Inform* 2007; 40, 365-369.

Lenti G, Agosti M, Massucci M, Zampolini M, Paolucci S, Franeschini M. Developing a minimum data set for stroke patients assessment: the Protocollo di Minima per l'lctus (PMIC) as a starting point towards an Italian stroke registry. *European Journal of physical &rehabilitation medicine*. 2008 Sep;44(3):263-9.

Menck H, Smart C. (eds). Central Cancer Registries, Design, Management, and Use. (Hargood Academic Publishers, Switzerland, 1994).

Metzger J. Using Computerized Disease Registries in Chronic Disease Care. Oakland, CA: California Healthcare Foundation, 2004.

Mohanty SK, Mistry AT, Amin W, Parwani AK, Schmandt L, Winters SB, Miliken E, Kim P, Whelan NB, Farhat G, Melamed J, Taiolo E, Dhir R, Pass HI, Becich MJ. The development and deployment of Common Data Elements for tissue banks for translational research in cancer – An emerging standard based approach for the Mesothelioma Virtual Tissue Bank. *BMC Cancer* 2008, 8:91.

NASN National Association of School Nurses. (2004). Position statement: Nursing Classification Systems: North American Nursing Diagnosis Association (NANDA), Nursing Interventions Classification (NIC), and Nursing Outcomes Classification (NOC). Available at http://www.nasn.org/ [Accessed 02/16/2010]

http://www.nasn.org/PolicyAdvocacy/PositionPapersandReports/NASNPositionStatementsFullView.

National Cancer Institute Cancer Centers. [Accessed 11/15/2012] http://cancercenters.cancer.gov/cancer_centers/cancer-centers-list2.html

National Cancer Institute SEER Training Modules Brief History of Cancer Registration, http://training.seer.cancer.gov/registration/registry/].

Ochs MF, Casagrande JT. Information systems for cancer research. *Cancer Invest* 2008;26(10):1060-7.

Parkin DM. The evolution of the population-based cancer registry. *Nat Rev Cancer* 2006;6(8):603-12.

Parkin DM. The role of cancer registries in cancer control. *Int J Clin Oncol*. 2008;13(2):102-11.

Piccirillo JF, Tierney RM, Costas I, Grove L. Prognostic importance of comorbidity in a hospital-based cancer registry. *JAMA*. 2004 May 26;291(20):2441-7.

Protection of Human Subjects - Vulnerable Populations, Institutional Review Boards, Informed Consent - Research, Participants, IRB, and Federal - StateUniversity.com http://education.stateuniversity.com/pages/2073/Human-Subjects-Protection.html#ixzz2Bn8fU9ie

Public Health Informatics Conference 2011 Kicks Off: What Can We Do With Technology? Public Health Informatics 2011 Conference August 22, 2011 [Accessed July 7, 2012]. Available at http://blog.rwjf.org/publichealth/2011/08/22/public-health-informatics-conference-2011-kicks-off-what-can-we-do-with-technology/

Savel T, Kesarinath G, Lee B, StinnJ. The use of public health grid technology in the United States Centers for Disease Control and Prevention H1N1 pandemic response. 24th IEEE International Conference on Advanced Information Networking and Application Workshops, WAINA 2010, p974-8.

Salant P, Dillman DA. (1994). How to conduct your own survey. New York: Wiley.

Saltz J, Oster S, Hastings S, Langella S, Kurc T, Sanchez W, Kher M, Manisundaram A, Shanbhag K, Covitz P. caGrid: design and implementation of the core architecture of the cancer biomedical informatics grid. *Bioinformatics*, 2006;22(15), 1910-1916.

Sorra J, Franklin M, Streagle S. Survey User's Guide: Hospital Survey on Patient Safety Culture. AHRQ Publication No. 08-0060, September 2008. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/qual/nhsurvey08/nhguide.htm

Shortliffe EH, Sondik EJ. The public health informatics infrastructure: anticipating its role in cancer. *Cancer Causes Control* 2006;17(7): 861-9.

Tafazzoli AG, Altmann U, Burkle T, Holzer S, Dudeck J. Integrated decision support in a hospital cancer registry. *Artif Intell Med* 2002;24:243-55.

Taktak AFG, Eleuteri SP, Lake SP, Fisher AC. Decision support systems in cancer. Advances in Medical, Signal and Information Processing, 2008. MEDSIP 2008. 4th IET International Conference.

The Institute of Electrical and Electronic Engineers (IEEE) IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries (New York, NY: 1990).

Thobourn KK, German RR, Lewis M. Nichols PJ, Ahmed F, Jackson-Thompson J. Case completeness and data accuracy in the Centers for Disease Control and Prevention's National Program of Cancer Registries. *Cancer*. 2007 Apr 15;109(8): 1607-16.

United States Public Health Service USPHS Public health definition [Accessed February 7, 2012]. Available at http://www.usphs.gov/ USPHS definition http://medical-dictionary.thefreedictionary.com/public+health

U.S. Census Regions. Indiana Business Research Center. Keiley School of Business Indiana University. Accessed 11/15/2012 (http://www.stats.indiana.edu/maptools/maps/boundary/census_regions_main.gif)

Volkers N. National Statistics Source Reaches 20th Anniversary. *J Natl Cancer Inst*, 1993;85(23): 1897-1899.

Wagner G. History of cancer registration. In: Jensen OM, Parkin DM, Maclennan R, Muir CS, Skeet RG, eds. Cancer registration, principles and methods. Lyons: International Agency for Research on Cancer, 1991:3-6. (IARC Scientific Publication No 95.)

Wingo PA. A national framework for cancer surveillance in the United States. *Cancer Causes Control* 2005:16: 151-70.

Yen PY, Bakken S. Review of health information technology usability study methodologies. *J Am Med Inform Assoc* 2012 19:413-422

Yasnoff WA, O'Carroll PW, Koo D, Linkins RW, & Kilbourne EM. Public health informatics: improving and transforming public health in the information age. Journal of public health management and practice. *JPHMP* 2000; 6(6):67-75.

Yasnoff WA, Overhage MJ, Humphreys BL, LaVenture M. A National Agenda for Public Health Informatics. *J Am Med Inform Assoc*. 2001 Nov-Dec; 8(6): 535–545.

VITA

Iris Zachary received her bachelor degree in information science in Germany 1996 and a master's degree in health informatics from the University of Missouri in 2006. She is currently working on a PhD degree in Health Informatics at the University of Missouri. She is the assistant data base manager for the Missouri Cancer Registry at the University of Missouri. Her research interests include public health, disease registries, health informatics, cancer registry, data collection, cancer registry data, and core data set.

APPENDIX

Research Cover Letter

UNIVERSITY OF MISSOURI-COLUMBIA
HUMAN SUBJECTS INSTITUTIONAL REVIEW BOARD

Cover Letter for surveys to provide informed consent.

Dear Respondent,

I am inviting you to participate in a research project on the data items to be included in a cancer registry database used for research as well as public health surveillance. In addition, this project explores the fields needed for a Core Data Set. Along with this letter is a short questionnaire that asks questions about the data items included in your database and data items in which you are interested. I am asking you to look over the questionnaire and, if you choose to participate, complete it online or send it back to me by mail. It should take about 20 minutes to complete.

The results of this project will be used for my dissertation project on the usability and utilization of cancer registry data. Through your participation, I hope to understand what researchers are interested in and what fields are essential in the collection of data about cancer. I plan to share my results by publishing them in a scientific journal and propose a minimum data set that can be utilized for research.

I do not know of any risks to you if you decide to participate in this survey. Your responses will not be identified with you personally, and all answers will be kept confidential. I will not share any information that identifies you with anyone outside my research group, which consists of me and my committee [Sue Boren *and my cxxxx*]. You should not put your name on your questionnaire. If you do not feel comfortable taking the survey online, you can also print the email and mail it to the following address.

The survey should take you about 20 minutes to complete. Your participation is voluntary and there is no penalty if you do not participate. Please let me know if you would like a summary of my findings. To receive a summary, please send your e-mail address and your name to me at zacharyi@missouri.edu.

If you have any questions or concerns about completing the questionnaire or about being in this study, you may contact me at zacharyi@missouri.edu. The University of Missouri Health Sciences Institutional Review Board (IRB) has approved this study. If you have any concerns about your rights as a participant in this study, you may contact the Human Research Protection Office via email (https://irb.missouri.edu/) or by telephone (573-884-8596).
Sincerely,

Iris Zachary

Timeliness, Completeness, Accuracy

MO NPCR-CSS 2011 Data Submission Specifications Diagnosis Year 1995-2009

Criteria	NPCR Advanced National Data Quality Standard(12 month)	NPCR National Data Quality Standard (24 month)	United States Cancer Statistics Publication Criteria	U.S. County Public-Use File Criteria	Measurement Error
Percentage Completeness of Case Ascertainment	>=90%	>=95%	>=90%	>=90%	-1.0%
Age (Missing/ Unknown)	NA	<=2%	<=3%	<=3%	-0.4%
Sex (Missing/ Unknown)	NA	<=2%	<=3%	<=3%	-0.4%
Race (Missing/ Unknown)	NA	<=3%	<=5%	<=5%	-0.4%
County (Missing/ Unknown)	NA	<=2%	NA	<=3%	-0.4%
Percentage Death Certificate Only (DCO)	NA	<=3%	<=5%	<=5%	-0.4%
Unresolved Duplicates (per 1000)	NA	<=1	NA	NA	-0.4
Percentage Passing Coordinated Core Edits	>=97%	>=99%	>=97%	>=97%	NA

Timelines for Coding, Standards and Staging Starting in 1976

1976-1991	ICDO1976	1986-1990	Cancer Program Manual 1986
1977-2000	Summary Staging1977	1988-1994	Data Acquisition Manual
1983-1988	TNM 1983(breast only) 1988	1988-1997	2-digit surgery codes
1988-1991	SEER Extent of disease 1988	1988-1991	SEER Program Code Manual 1st ed
1989-1992	TNM 3rd ed 1989	1991-1995	Cancer Program Manual 1991
1992-2000	ICDO 1992 2nd ed	1992-1997	SEER Program Manual 2nd ed
1992-1997	Seer extent 2nd ed 1992	1993	SEER Selfinstructional Manual Book 8 3rd ed
1993-1997	TNM 4th 1993	1994-1995	Data Acquisition Manual revised
1998	Seer extent 3rd 1998	1996-2002	Registry Operations Manual ROADS
1998-2002	TNM 5th 1998	1998-2002	New surgery Codes
2001	ICDO 3rd ed 2001	2003	Cancer Program Standards revised volume 1
2001	Summary staging 2000(2001)	2003	Data Standards (FORDS)
2003	TNM 6th ed 2003	2003	SEER Program Manual (Treatment Codes only)
2004	Collaborative Staging 2004	2010	Record length increased to 22500 from 6500
2010	Collaborative Staging V 2 2010		

Introduction

The primary purpose of this project is to identify: 1) data items that need to be included in a cancer registry database so that it can be utilized for research as well as for public health surveillance; and 2) data items needed for a core data set. A secondary purpose is to learn more about the characteristics of registries that collect data for different purposes. Through your participation, I hope to understand what researchers are interested in and what fields are essential in the collection of cancer incidence data for public health surveillance, research, clinical trials, etc.

Participation is voluntary and there is no penalty if you do not participate. Regardless of whether you choose to participate, please let me know if you would like a summary of my findings.

Survey Instructions:

1. It would be beneficial if two people in the registry that hold different positions would be able to take the survey, i.e. Registry director and database administrator etc.

2. If you do not complete the survey at once, do not select submit. Otherwise you will not be able to finish the survey.

Registry

1. In which geographic region is your registry located?

- ◯ Northeast
- ◯ Midwest
- ◯ South
- ◯ West

2. What agency or agencies provide(s) funding to your registry? Check all that apply.

- ◯ NPCR
- ◯ SEER
- ◯ Both NPCR and SEER

Other (please specify)

3. What percent of your registry's annual funding for core surveillance activities is provided by NPCR and /or SEER?

- ◯ < 10%
- ◯ 10%-25%
- ◯ 26%-50%
- ◯ 51-75%
- ◯ >75%
- ◯ Don't know
- ◯ Refused

4. Not counting in-kind support, what percent of your registry's annual budget is state funding for core activities?

- ◯ < 10%
- ◯ 10%-25%
- ◯ 26%-50%
- ◯ >50%
- ◯ Don't know
- ◯ Refused

5. Where is your registry located?

- ◯ State Health Department
- ◯ University

Other (please specify)

6. How many funded full-time equivalent (FTE) positions does your registry have?

- ◯ < 5
- ◯ 5 - 10
- ◯ 11 - 15
- ◯ 16-20
- ◯ 21- 25
- ◯ 26-30
- ◯ >30
- ◯ Don't know

7. How many funded positions are currently vacant?

Insert number of vacant positions including job title.

8. Please estimate the percentage of tumor variables that have missing or unknown values.

	< 5%	5%-10%	>10%	Don't know
Stage	☐	☐	☐	☐
Size	☐	☐	☐	☐
Histology	☐	☐	☐	☐

Other (please specify)

9. How many of your staff are certified tumor registrars (CTRs), epidemiologists, research analysts?

Research Analyst	
CTR	
Epidemiologist	
IT specialist	
Other (Please specify)	

10. What position do you hold in the registry?

- ◯ Director
- ◯ Manager
- ◯ Quality Assurance

Other (please specify)

Data Collection

Thinking about the file that your registry submits to NAACCR for certification, please answer the following questions.

11. Please check the level of all demographic variables that have missing or unknown values.

	< 5%	5%-10%	>10%	Don't know
Race	☐	☐	☐	☐
Ethnicity	☐	☐	☐	☐
Sex	☐	☐	☐	☐
Age	☐	☐	☐	☐
Marital Status	☐	☐	☐	☐
SSN	☐	☐	☐	☐
Date of Birth	☐	☐	☐	☐

Other (please specify)

12. Please estimate the percentage of all treatment variables that have missing or unknown values.

	< 5%	5%-10%	11%-25%	26%-50%	>50%	Don't know
Chemo	☐	☐	☐	☐	☐	☐
Radiation	☐	☐	☐	☐	☐	☐
Hormones	☐	☐	☐	☐	☐	☐
BRM	☐	☐	☐	☐	☐	☐

Other (please specify)

13. Other than the required data elements, how many standardized NAACCR demographic data elements does your registry collect?

- ◯ none
- ◯ 1-4
- ◯ 5-9
- ◯ 10-15
- ◯ >15
- ◯ Don't know /not sure

14. Other than the required data elements, how many standardized NAACCR tumor data elements does your registry collect?

- ◯ none
- ◯ 1-4
- ◯ 5-9
- ◯ 10-15
- ◯ >15
- ◯ Don't know /not sure

15. Other than the required data elements, how many standardized NAACCR treatment data elements does your registry collect?

- ◯ 1-4
- ◯ 5-9
- ◯ 10-15
- ◯ >15
- ◯ Don't know
- ◯ None

16. How many state specific data elements does your registry collect?

- ◯ 1-4
- ◯ 5-9
- ◯ 10-15
- ◯ >15
- ◯ Don't know
- ◯ None

17. List the most important top five state specific items that you collect in addition to the NAACCR/ SEER required fields.

Item 1 (Specify)	
Item 2 (Specify)	
Item 3 (Specify)	
Item 4 (Specify)	
Item 5 (Specify)	

18. How complete are your data in the top five additional/ state specific fields that you listed in question 17?

	< 5%	5%-10%	11%-25%	26%-50%	>50%	Don't know
Item 1	☐	☐	☐	☐	☐	☐
Item 2	☐	☐	☐	☐	☐	☐
Item 3	☐	☐	☐	☐	☐	☐
Item 4	☐	☐	☐	☐	☐	☐
Item 5	☐	☐	☐	☐	☐	☐

19. Do you conduct quality control on the additional/ state specific fields?

- ◯ All
- ◯ Some
- ◯ None
- ◯ Don't know

20. In what year was your registry established?

Please specify.

21. How many years of population-based data do you have available?

- ◯ 0-5
- ◯ 6-10
- ◯ > 10

Other (please specify)

22. How many years of data meet NAACCR or NPCR or SEER publication standards

	0-5	6-10	> 10	Don't know
NAACCR	☐	☐	☐	☐
NPCR	☐	☐	☐	☐
SEER	☐	☐	☐	☐

Other (please specify)

23. For how long has your registry been certified, i.e. Silver or Gold certification by NAACCR or SEER?

- ◯ 1 year
- ◯ 2-3 years
- ◯ 4-5 years
- ◯ > 5 years

24. Specify the diagnosis years for data that are available, e.g., 1995 to 2008, 1995-2000, 2000-2005, 2006-2008

25. What data ranges are available for data requests?

Check all that apply.

- [ ] 1 year
- [ ] 3 years
- [ ] 5 years

Other (please specify)

26. Please indicate the range of years for which registry data are suitable for research, e.g. 2001-2007, 1996-2000

Database related

27. Do you receive updated information on vital status and tumor status for each case?

- ◯ Yes
- ◯ No
- ◯ Don't know

Other (please specify)

28. How many records are in your registry's database?

29. How many records do you process on an annual basis?

- ◯ < 5000
- ◯ 5000-9999
- ◯ 10000-24999
- ◯ 25000-49999
- ◯ 50000-74999
- ◯ >75000

30. How many incidence cases for your state do you collect on an annual basis?

- ◯ < 5000
- ◯ 5000-9999
- ◯ 10000-24999
- ◯ 25000-49999
- ◯ 50000-74999
- ◯ >75000

Cancer Registry Data Request

31. Population-based cancer incidence data can be used for a variety of purposes. Please check all of the purposes for which data from your registry are currently being used. Check all that apply

- [ ] Public health surveillance
- [ ] Research
- [ ] Database linkage (e.g., Breast & Cervical Cancer Control Program)
- [ ] Cancer inquiries (reports of excess cancers)
- [ ] Special projects
- [ ] Clinical trials
- [ ] Next of Kin requests

Other (please specify)

32. Do you make the cancer registry data available for data requests?

- ◯ Yes
- ◯ No

Other (please specify)

33. Are registry data available online in interactive tables?

- ◯ Yes
- ◯ No

Other (please specify)

34. At what demographic / geographic level are registry data available to the researcher for data requests? Check all that apply

- [ ] The data are available at the state level
- [ ] The data are available at the county level
- [ ] The data are available at the zipcode level

Other (please specify)

35. How many data requests do you receive annually/monthly, e.g. 2 per month, 40 per year?

Annually specify	
Monthly specify	
Not Sure	

36. Can you fill all data requests the cancer registry receives?

- ◯ Yes
- ◯ No
- ◯ Not sure

Other (please specify)

37. In addition to the standard edits, which of the following quality measures does your registry carry out?
Check all that apply.

- ☐ <100% visual review
- ☐ 100% visual review
- ☐ Random review
- ☐ Reabstracting audits
- ☐ Casefinding audits

Other (please specify)

Quality Measures

38. If you cannot fill all data requests, please specify the reason(s). Check all that apply

- [ ] Data elements are not collected
- [ ] Data elements are not available
- [ ] Data elements are not reliable
- [ ] Data elements are not complete
- [ ] Data elements have missing/ unknown values

Other (please specify)

39. Do you consider NAACCR silver certification as research quality data.
Silver certification meets:
Case ascertainment has achieved 90% or higher completeness
A death certificate is the only source for identification of fewer than 5% of reported cancer cases
Fewer than 0.2% duplicate case reports are in the file
All data variables used to create incidence statistics by cancer type, sex, race, age, and county are 97% error-free
Less than 3% of the case reports in the file are missing meaningful information on age, sex, county.
Less than 5% of the cases in the file are missing meaningful information on race (US only).
The file is submitted to NAACCR for evaluation within 23 months of the close of the diagnosis year under review

- ◯ Yes
- ◯ No

Other (please specify)

Data Elements and Data Quality

40. What do you think of the number of data items that you are required to collect?

- [ ] too many
- [ ] about right
- [ ] too few

Other (please specify)

41. Are you interested in additional data items or elements that are not mentioned. If yes, what additional data items are you interested in? Please specify.

- ◯ Yes
- ◯ No

Other (please specify)

End of survey

Thank you for completing the survey

Introduction

The primary purpose of this project is to identify: 1) data items that need to be included in a cancer registry database so that it can be utilized for research as well as for public health surveillance; and 2) data items needed for a core data set. A secondary purpose is to learn more about the characteristics of registries that collect data for different purposes. Through your participation, I hope to understand what researchers are interested in and what fields are essential in the collection of cancer incidence data for public health surveillance, research, clinical trials, etc.

Survey Instructions:

1. It would be beneficial if two people in the registry that hold different positions would be able to take the survey, i.e. Registry director and database administrator etc.

2. If you do not complete the survey at once, do not select submit. Otherwise you will not be able to finish the survey.

Institution/ Facility

1. In which state is your facility located?

Please specify

2. Are you affiliated with?

- [ ] Hospital
- [ ] University/ Teaching Hospital
- [ ] Physician Group
- [ ] Not applicable

Other (please specify)

3. How many full-time equivalent (FTE) positions are in your department? Please specify the number.

- ◯ <5
- ◯ 5-10
- ◯ 11-15
- ◯ 16-20
- ◯ 21-25
- ◯ 26-30
- ◯ >30
- ◯ Don't know

4. How many FTE equivalent positions are currently working on cancer research?

- ◯ <5
- ◯ 5-10
- ◯ 11-15
- ◯ 16-20
- ◯ 21-25
- ◯ >25
- ◯ Don't know

5. Who is responsible for the data collection process?

Please specify job title and credentials.

6. What position do you hold at your facility?

- [ ] Clinical Trial Coordinator
- [ ] Research Analyst
- [ ] Physician
- [ ] Nurse

Other (please specify)

Research Area/ Data Collection

7. In what area of study / specialty do you collect data?

- [ ] Oncology
- [ ] Radiology
- [ ] Hematology
- [ ] Clinical Trial

Other (please specify)

8. How many years of data do you have in your research database?

- [ ] < 1 year
- [ ] 1-2 years
- [ ] 3-4 years
- [ ] 5 years
- [ ] >5 years

9. Which of the following demographic data items do you collect? Check all that apply.

	Yes	No	Not sure
Name	◯	◯	◯
Social Security Number	◯	◯	◯
Date of Birth	◯	◯	◯
Ethnicity	◯	◯	◯
Race	◯	◯	◯
Address at Diagnosis	◯	◯	◯
Current Address	◯	◯	◯

10. What percent of the variables have missing or unknown values?

	<5%	5%-10%	11%-25%	26%-50%	>50%	Don't know
Date of birth	☐	☐	☐	☐	☐	☐
SSN	☐	☐	☐	☐	☐	☐
Race	☐	☐	☐	☐	☐	☐
Name	☐	☐	☐	☐	☐	☐
Sex	☐	☐	☐	☐	☐	☐
Ethnicinty	☐	☐	☐	☐	☐	☐
Date of Birth	☐	☐	☐	☐	☐	☐
Social security number	☐	☐	☐	☐	☐	☐
Address at diagnosis	☐	☐	☐	☐	☐	☐
Current address	☐	☐	☐	☐	☐	☐

Other (please specify)

11. Do you follow the minimum Office of Management and Budget (OMB) standard for collecting race, i.e.,
American Indian or Native American,
Asian
Black or African American,
Native Hawaiian or Pacific Islander,
White
http://www.ofm.wa.gov/pop/race/racerecommendations.pdf
Please give an example of how you collect race.

- ◯ One race only
- ◯ More than one race
- ◯ Race following Census data collection format
- ◯ Don't know

Other (please specify) and give example

12. Do you collect data on Hispanic ethnicity?

- ◯ Yes
- ◯ No
- ◯ Don't know

Other (please specify)

13. Do you place the Spanish, Hispanic, Latino question before the race question?

- ◯ Yes
- ◯ No
- ◯ Don't know
- ◯ N/A

14. Other than listed, in what other fields or data elements are you interested? Please specify, e.g., types of treatment, AJCC staging etc.

15. What percent of the tumor variables have missing or unknown values?

	<5%	5%-10%	11%-25%	26%-50%	>50%	Don't know
Site	☐	☐	☐	☐	☐	☐
Stage	☐	☐	☐	☐	☐	☐
Histology	☐	☐	☐	☐	☐	☐

Other (please specify)

16. What percent of treatment variables have missing or unknown values?

	<5%	5%-10%	11%-25%	26%-50%	>50%	Don't know
Chemotherapy	☐	☐	☐	☐	☐	☐
Radiation	☐	☐	☐	☐	☐	☐
Hormones	☐	☐	☐	☐	☐	☐
Bio Response Modifier	☐	☐	☐	☐	☐	☐

Other (please specify)

17. Do you collect information on vital status and tumor status?

- ◯ Yes
- ◯ No
- ◯ Not sure

18. Do you use a data layout, manual, data dictionary or code book? Please specify.

- ☐ Standard layout: NAACCR
- ☐ Standard layout: Other
- ☐ No data layout
- ☐ Layout developed by your institution
- ☐ Non-standardized layout
- ☐ None

Other (please specify)

19. Do you follow any data standards for the data elements you are collecting? Please check all that apply.

- [ ] NAACCR
- [ ] SNOMED
- [ ] ICD9
- [ ] ICD10
- [ ] ICD-O3
- [ ] Other
- [ ] None

Other (please specify)

20. On average, how many data elements /data items do you collect in a research project/s? Specify number.

Number	
Other (Please specify)	
Don't know	

21. Are your data complete for all fields in most of your studies?

- ◯ Yes
- ◯ No
- ◯ Don't know

Please specify for example demograhics not complete, tumor variables complete etc.

22. If 'No' which fields are incomplete. Please list fields.

23. Have you ever submitted a data request to a cancer registry? Check all that apply.

- [ ] Yes, hospital
- [ ] Yes, state cancer registry
- [ ] No
- [ ] Not sure

Other (please explain)

24. Have you ever used cancer registry data for research or projects? Check all that apply.

- [ ] Yes, hospital
- [ ] Yes, state cancer registry
- [ ] No
- [ ] Not sure
- [ ] Other please explain

Other (please specify)

Database related

25. How many cases does your database contain?

Please specify the number of cases.

26. Are the data standardized?

- ◯ Yes
- ◯ No
- ◯ Don't know

27. Are the data deduplicated?

- ◯ Yes
- ◯ No
- ◯ Don't know

Data Use and Data Requests

28. How are your data used? Please check all of the purposes for which data from your database are currently being used.

- [ ] Clinical studies
- [ ] Research
- [ ] Special projects
- [ ] Clinical trials
- [ ] Submit data to state central registry

Other (please specify)

29. Rank the quality of your data in terms of its value for research.

- ◯ High quality
- ◯ Moderately high quality
- ◯ Adequate quality
- ◯ Not adequate quality
- ◯ Don't know

Additional Data Items

30. What data items are you interested in from a hospital cancer registry? Please specify.

31. What data items are you interested in from a state cancer registry? Please specify.

32. Do you know what data elements are available from the state cancer registry data in your state?

- ◯ Yes
- ◯ No
- ◯ Not sure

End of Survey

Thank you for completing the survey