May 2019

The Assessment of Technology Adoption Interventions and Outcome Achievement Related to the Use of a Clinical Research Data Warehouse

Katie A. McCarthy
University of Wisconsin-Milwaukee

THE ASSESSMENT OF TECHNOLOGY ADOPTION INTERVENTIONS AND OUTCOME ACHIEVEMENT RELATED TO THE USE OF A CLINICAL RESEARCH DATA WAREHOUSE

by

Katie A. McCarthy

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Health Sciences

at

The University of Wisconsin-Milwaukee

May 2019

ABSTRACT

THE ASSESSMENT OF TECHNOLOGY ADOPTION INTERVENTIONS AND OUTCOME ACHIEVEMENT RELATED TO THE USE OF A CLINICAL RESEARCH DATA WAREHOUSE

by

Katie A. McCarthy

The University of Wisconsin-Milwaukee, 2019
Under the Supervision of Dr. Jake Luo

Introduction: While funding for research has declined since 2004, the need for rapid, innovative, and lifesaving clinical and translational research has never been greater due to the rise in chronic health conditions, which have resulted in lower life expectancy and higher rates of mortality and adverse outcomes. Finding effective diagnostic and treatment methods to address the complex challenges in individual and population health will require a team science approach, creating the need for multidisciplinary collaboration among practitioners and researchers.

To address this need, the National Institutes of Health (NIH) created the Clinical and Translational Science Awards (CTSA) program. The CTSA program distributes funds to a national network of medical research institutions, known as "hubs," that work together to improve the translational research process. With this funding, each hub is required to achieve specific goals to support clinical and translational research teams by providing a variety of services, including cutting edge use of informatics technologies. As a result, the majority of CTSA recipients have implemented and maintain data warehouses, which combine disparate data types from a range of clinical and administrative sources, include data from multiple institutions,

and support a variety of workflows. These data warehouses provide comprehensive sets of data that extend beyond the contents of a single EHR system and provide more valuable information for translational research.

Although significant research has been conducted related to this technology, gaps exist regarding research team adoption of data warehouses. As a result, more information is needed to understand how data warehouses are adopted and what outcomes are achieved when using them. Specifically, this study focuses on three gaps: research team awareness of data warehouses, the outcomes of data warehouse training for research teams, and how to measure objectively outcomes achieved after training.

By assessing and measuring data warehouse use, this study aims to provide a greater understanding of data warehouse adoption and the outcomes achieved. With this understanding, the most effective and efficient development, implementation, and maintenance strategies can be used to increase the return on investment for these resource-intensive technologies. In addition, technologies can be better designed to ensure they are meeting the needs of clinical and translational science in the 21st century and beyond.

Methods: During the study period, presentations were held to raise awareness of data warehouse technology. In addition, training sessions were provided that focused on the use of data warehouses for research projects. To assess the impact of the presentations and training sessions, pre- and post-assessments gauged knowledge and likelihood to use the technology. As objective measurements, the number of data warehouse access and training requests were obtained, and audit trails were reviewed to assess trainee activities within the data warehouse.

Finally, trainees completed a 30-day post-training assessment to provide information about barriers and benefits of the technology.

Results: Key study findings suggest that the awareness presentations and training were successful in increasing research team knowledge of data warehouses and likelihood to use this technology, but did not result in a subsequent increase in access or training requests within the study period. In addition, 24% of trainees completed the associated data warehouse activities to achieve their intended outcomes within 30 days of training. The time needed for adopting the technology, the ease of use of data warehouses, the types of support available, and the data available within the data warehouse may all be factors influencing this completion rate.

Conclusion: The key finding of this study is that data warehouse awareness presentations and training sessions are insufficient to result in research team adoption of the technology within a three-month study period. Several important implications can be drawn from this finding. First, the timeline for technology adoption requires further investigation, although it is likely longer than 90 days. Future assessments of technology adoption should include an individual's timeline for pursuing the use of that technology. Second, this study provided a definition for outcome achievement, which was completion of activities within a data warehouse needed to achieve an intended research outcome as identified by the research team. While this definition is a good baseline, it needs to be refined with input from research teams. Finally, this study confirmed previous findings related to technology adoption, which indicated that time, ease of use, support, and data availability are important factors. Additional work is needed to identify the significance and correlation of these factors with technology adoption.

This study provides important findings regarding attainment of technology knowledge and its links to actual technology use, the correlation between self-reported likelihood to use and actual technology use, the timeline of technology adoption, and foundational protocols for objective measurement of technology use and adoption. Future research should focus on refining the objective measurement of "outcome achievement," understanding the timing of technology adoption, and measuring the significance and correlation of factors influencing data warehouse use.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Definition
A	Attitude
AMIA	American Medical Informatics Association
ARRA	American Recovery and Reinvestment Act
BCW	BloodCenter of Wisconsin
BI	Behavioral intention
BMI	Biomedical informatics
BRAIN	Brain Research through Advancing Innovative Neurotechnologies
BTRIS	Biomedical Translational Research Information System
CBA	Cost-benefit analysis
CBC	Complete blood count
CDC	Centers for Disease Control and Prevention
CHW	Children's Hospital of Wisconsin
CITI	Collaborative Institutional Training Initiative
CRDW	Clinical Research Data Warehouse
CTSA	Clinical and Translational Science Award
CTSI	Clinical and Translational Science Institute
D&M IS Success Model	DeLone and McLean Information Systems Success Model
DOI	Diffusion of Innovation
EFPIA	European Federation of Pharmaceutical Industries and Associations
EHR	Electronic health record
EHR4CR	Electronic Health Records for Clinical Research
FDA	Food and Drug Administration
GPRD	General Practice Research Database
HIT	Health information technology
HITECH	Health Information Technology for Economic and Clinical Health
i2b2	Informatics for Integrating Biology & the Bedside
i~HD	European Institute for Innovation through Health Data
IC	Information commons
ICU	Intensive care unit
IRB	Institutional Review Board
IS	Information system or information systems
IT	Information technology

Abbreviation	Definition
MCW	Medical College of Wisconsin
MHS	Marine Hospital Service
MU	Marquette University
MSOE	Milwaukee School of Engineering
NAS	National Academy of Sciences
NCATS	National Center for Advancing Translational Sciences
NCI	National Cancer Institute
NHGRI	National Human Genome Research Institute
NIH	National Institutes of Health
NLM	National Library of Medicine
NLP	Natural language processing
OHDSI	Observational Health Data Sciences and Informatics
PDA	Personal digital assistant
PEOU	Perceived ease of use
PHS	Public Health Service
PMI	Precision Medicine Initiative
PU	Perceived usefulness
REDCap	Research Electronic Data Capture
SEM	Structured equation modeling
SOA	Service Oriented Architecture
TAM	Technology Acceptance Model
TRA	Theory of Reasoned Action
UWM	University of Wisconsin-Milwaukee
UK	United Kingdom
US	United States
VA	Clement Zablocki VA Medical Center

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to Dr. Jake Luo, who served as my advisor and committee chair. Dr. Luo stepped in at a difficult point in my journey and provided essential advice and support. Without his guidance and timely assistance, this dissertation would not have been possible.

I would like to thank my University of Wisconsin-Milwaukee (UWM) committee members, Dr. Ron Cisler, Dr. Rohit Kate, and Dr. Timothy Patrick. Throughout this process, you challenged me to consider different perspectives and methods, while also providing much needed recommendations and support. I will be forever grateful for the time and effort you have expended in this process.

I am thankful for the Biomedical Informatics (BMI) department within the Clinical and Translational Science Institute (CTSI) of Southeast Wisconsin, especially Bradley (Brad) Taylor, Kristen Osinski, and Mark Oium. I greatly appreciate Brad's willingness to sponsor this project, and all the time he spent providing direction and removing roadblocks. Kristen has been an amazing research partner, always available to answer questions, discuss challenges and potential solutions, and laugh and commiserate when most needed. Finally, Mark's REDCap knowledge and skills have been crucial to capture and maintain the data for this project, and I will always be thankful for his patience and timely support.

I am grateful to my Milwaukee School of Engineering (MSOE) colleagues, Dr. Michael Payne, Beth Slayman, Melinda (Mindy) Davies, and Dawn Blessing. I will never be able to express how much your support and encouragement has meant to me, both personally and professionally.

From the hours spent listening, commiserating, and providing advice to reviewing my drafts and providing crucial critiques, each of you played a critical role in this process and I am blessed to have each of you in my life.

I am thankful for the expertise of my care team, especially Dr. Angela English and Jennifer Newkirk, which has been essential in ensuring my physical and mental well-being throughout this process.

I am blessed to have amazing friends, especially Jill Skaletski and Dawn Drinkwine, who provided the necessary compassion, support, laughter, and "sanity checks" throughout this process. Jill, you are an amazing, compassionate individual, and I know that I can always count on you for encouragement and perspective. Dawn, while we have rarely lived in the same city or state, we always find a way to connect. Our conversations and laughter were critical to maintaining my sanity. Without each of you, I could not have completed this journey.

Most importantly, I am blessed by a loving God. "I have told you this so that you might have peace in me. In the world you will have trouble, but take courage, I have conquered the world." John 16:33 (NABRE)

Finally, I dedicate this project to my parents, Jean and Patrick McCarthy, and brother, Kevin McCarthy. Without you, I would not be who I am today. You have provided constant love, support, and laughter throughout this process, and I am continually blessed by your presence in my life.

CHAPTER 1: INTRODUCTION

In 2002, the United States (US) National Institutes of Health (NIH) announced development of the NIH Roadmap to accelerate research progress.[1] In creating this roadmap, leaders identified three themes on which to focus national strategies regarding research acceleration: New Pathways to Discovery, Research Teams of the Future, and Reengineering the Clinical Research Enterprise.[1,2] One of the strategies to address the final theme was the launch of the Clinical and Translational Science Awards (CTSA) program. Announced in October 2005, the program's goals are to eliminate barriers between basic and clinical research, address the increasing complexities in conducting clinical research, and create academic homes for clinical and translational science.[2]

Administered by the National Center for Advancing Translational Sciences (NCATS), the CTSA program distributes funds to a national network of more than 50 medical research institutions, also known as "hubs," that work together to improve the translational research process.[3,4] In fiscal year 2016, a total of more than $487 million was awarded to 57 CTSA hubs.[4] With this funding, each hub is required to achieve five goals related to translational research. One of these goals is to "advance the use of cutting-edge informatics."[3] In addition, a high priority is the use of information technology (IT) to integrate clinical research and clinical workflows,[5] and the NIH has called for increasing the secondary use of electronic health record (EHR) data for clinical research purposes.[6]

Based on these priorities, many CTSA hubs have implemented and maintain data warehouses, which are "'…subject-oriented, integrated, time-variant, non-volatile collection[s] of data…'

[that address] data management, integration, and access issues."[7] According to a 2010 CTSA survey, 86% of respondents indicated that their organization had one or more data warehouses.[8] These data warehouses combine disparate data types from a range of clinical and administrative sources such as EHRs and registries and may include data from multiple institutions.[8] The result is a comprehensive set of records that extends beyond the data contained within a single EHR and provides more valuable information for translational research.

1.1 Research Problem and Overall Aims

Although significant research is being conducted about and with data warehouses for current and future clinical and translational research purposes, critical gaps exist in the literature corpus and real-world implementation and practice. First, little is known regarding the effectiveness of interventions in increasing researcher and research team awareness of data warehouses and their use in clinical and translational science. Second, while CTSA hubs often provide data warehouse training for research teams, little is known about the effectiveness of this training and subsequent use of data warehouses. Finally, a critical gap exists in defining objective measures for determining if research teams achieve their intended outcomes when using data warehouses.

To study these knowledge gaps and identify areas that require new solutions, this study developed a presentation intervention to raise research team awareness of data warehouses and assessed and measured the outcomes of this intervention. In addition, the study assessed and measured the outcomes of existing data warehouse training. Finally, the study identified potential factors that influence the achievement of intended outcomes when using a data warehouse.

1.2 Study Conceptual Frameworks

To achieve these aims, the study approach was based on the five innovation-decision stages (knowledge, persuasion, decision, implementation, and confirmation) identified in the Diffusion of Innovation (DOI) theory.[9] In addition, the Technology Acceptance Model (TAM) and DeLone and McLean Information Systems Success Model were used to identify critical content for the awareness presentation intervention and existing training and to focus assessment on key aspects influencing technology adoption and acceptance. Finally, the Bloom's Taxonomy of Educational Objectives was used to correctly structure assessment items for the awareness intervention and training learning objectives. The details of these frameworks are provided in *Chapter 3: Conceptual Frameworks*.

1.3 Scope and Approach

This research study was performed within the Clinical and Translational Science Institute (CTSI) of Southeast Wisconsin in collaboration with the CTSI Biomedical Informatics (BMI) department. The CTSI BMI department provides a Clinical Research Data Warehouse (CRDW) that contains data from a variety of EHRs, clinical information systems, and registries, similar to data warehouses implemented by other CTSA hubs. Subsequently, this study focused on the CRDW and its use as representative of similar data warehouses for clinical and translational research purposes. Data for this study was collected through assessment tools, BMI databases, and CRDW audit trails. The research questions and hypotheses, as well as specific aspects of the study's methodology, design, and procedures are provided in *Chapter 4: Methods*.

1.4 Significance of the Study

While funding for research has declined since 2004, the need for rapid, innovative, and lifesaving clinical and translational research has never been greater.[10] In the US population, 68% of adults age 65 or older have at least two chronic health conditions, and a third of adults are considered obese.[11] In comparison to other developed countries around the world, the US has lower life expectancy, higher rates of mortality for ischemic heart disease, and higher rates of adverse outcomes from diabetes.[11] Finding effective diagnostic and treatment methods to address the complex challenges in individual and population health will require a team science approach, creating the need for multidisciplinary collaboration among practitioners and researchers.[12]

Individuals trained in biomedical informatics are uniquely suited for this research approach. Biomedical informaticists have competencies in the creation and application of models for biomedical data, information, and knowledge.[13] In addition, they are expected to understand concepts and data that span the translational medicine spectrum (e.g., biology, clinical care, epidemiology, and health services). Because integrating clinical and research data from diverse data sources and numerous multidisciplinary teams will be critical for cost-effective and efficient translational research, biomedical informatics approaches will be required "…to manage, organize, and integrate heterogeneous data to inform decisions from bench to bedside to community to policy."[12]

Although technologies have already been implemented to support this requirement, more information is needed to understand how these technologies are adopted and what outcomes are

achieved when using them. By assessing and measuring the CTSI BMI CRDW adoption and use, this study aims to provide greater understanding of data warehouse adoption and the outcomes achieved. With this understanding, the most effective and efficient development, implementation, and maintenance strategies can be used to increase the return on investment for these resource-intensive data warehouse technologies. In addition, technologies can be better designed to ensure they are meeting the needs of clinical and translational science in the 21st century and beyond.

1.5 Summary of Remaining Chapters

This chapter provides an overview of the study aims and a summary of the research approach, scope, and significance. Subsequent chapters provide the following information:

- *Chapter 2: Literature Review* provides the basis for the research aims, as well as the potential contribution of this study to the existing literature corpus.
- *Chapter 3: Conceptual Frameworks* identifies the theoretical models used to design the study components within this research project.
- *Chapter 4: Methods* describes the research questions, design, hypotheses, and associated study components; the CTSI and associated target population; and project elements such as timeline and budget.
- *Chapter 5: Results* provides the study findings, including the acceptance or rejection of each of the study's hypotheses.
- *Chapter 6: Discussion* reflects on the findings based on the study's research questions and outlines future research opportunities.

CHAPTER 2: LITERATURE REVIEW

The focus of this research study is the use of data warehouses for clinical and translational research. This chapter describes several areas within literature that informed this study.

- Section 2.1, *The National Institutes of Health (NIH)*, provides an overview of the NIH, its role in the creation of the CTSA program, and the need for biomedical informatics within this program.
- Section 2.2, *Use of Electronic Health Record Data for Research*, describes the use of EHRs within research, including the challenges that have led to the development and implementation of data warehouses (integrated data repositories).
- Section 2.3, *Summary of Research Themes in Current Literature*, summarizes the literature relating to data warehouses and their uses for research purposes.
- Section 2.4, *Importance of Data Warehouses in Future Research*, outlines key areas of future research requiring the use of data warehouses and associated skills.
- Section 2.5, *Key Gaps Within Existing Literature*, identifies the key gaps that exist within the literature corpus related to data warehouses and their use for clinical and translational research.

2.1 The National Institutes of Health (NIH)

In 1887, the NIH began as a one-room laboratory within the Marine Hospital Service (MHS), the predecessor agency to the US Public Health Service (PHS).[14,15] Since then, the NIH has grown to include 27 institutes and centers, including the National Cancer Institute (NCI), the National Library of Medicine (NLM), the National Human Genome Research Institute (NHGRI), and the NCATS.[16] With funding of more than $32 billion in fiscal year 2016, the NIH has become the

world's largest funder of biomedical research, supporting more than 300,000 scientists and technical personnel working at more than 2,500 universities, hospitals, medical schools, and other research institutions in the US and throughout the world.[17,18] Supported by this funding, research advances have led to increased average life expectancies for Americans, reductions in disability rates, and reductions in all-cause mortality rates, including rates due to cardiac disease, diabetes, stroke, and cancer.[2,17] Overall, NIH-supported research has been a primary source of new discoveries, drugs, devices, and clinical procedures that have contributed to the health and longevity of individuals and populations around the world, resulting in more than eighty Nobel prizes.[10,15] The following subsections describe the development of the NIH roadmap to accelerate research, the subsequent creation of the CTSA program, and the need for biomedical informatics within CTSA hubs.

2.1.1 Development of the NIH Roadmap

Supporting this type of research has not been without challenges, especially in the early 2000s. During that time, the US faced deep federal and trade deficits, increased spending for homeland security, and unexpected economic and financial devastation from natural disasters such as Hurricane Katrina.[2] From a population perspective, the US faced an aging population suffering from predominantly chronic diseases, as well as emerging public health challenges in the form of obesity and diabetes.[2] Within the scientific research community, costs were rising, yet funding was beginning to decrease after years of increases.[1,2] By 2004, NIH funding reached its peak, and then began to decline nearly 2% per year, resulting in a 13% decrease in NIH purchasing power.[10] In response to these challenges, the NIH needed to devise a strategy to improve the return on its investment by reducing costs and refocusing the clinical research enterprise.

Beginning in 2002, the NIH began working on the "NIH Roadmap" with the goal of defining a set of limited priorities that could lead to the acceleration of research across the institute missions.[1] As a national clinical research enterprise, several key challenges needed to be addressed in this roadmap. First, research study participants would need to increase in number and diversity (such as by gender, age, and ethnicity) to support the speed of discovery. Encouraging public participation would require addressing concerns related to conflicts of interest, safety, and privacy.[19] Second, information systems would need to be developed and implemented to support effective and efficient data capture, maintenance, and retrieval. Addressing this challenge would require creation of data and exchange standards, as well as educational programs to produce a qualified workforce in biomedical informatics.[19] Third, accelerating the rate of research requires an adequately trained workforce of many kinds of investigators such as clinicians, basic scientists, computer programmers, and engineers who can work together on the complex issues facing health and healthcare. Funding sources and programs would need to be created to support the work of young researchers, as well as incentivize multidisciplinary projects.[19] Finally, overall funding (government, industry, and foundations) would need to increase to support the identified initiatives.[19]

After consulting more than 300 of the nation's biomedical leaders from academia, government organizations, and industry regarding these challenges, three themes emerged and became the foci for the NIH Roadmap: New Pathways to Discovery, Research Teams of the Future, and Reengineering the Clinical Research Enterprise.[1,2] The New Pathways to Discovery theme focuses on initiatives to address the technologies and approaches necessary to understand complex biological systems and address contemporary research issues such as building blocks and pathways, molecular imaging, bioinformatics, computational biology, and structural

biology.[1] The Research Teams of the Future theme focuses on the creation of new team science organizational models that provide mechanisms for high-risk strategies, interdisciplinary research, and public and private research partnerships.[1] Finally, the Reengineering the Clinical Research Enterprise theme focuses on creating integrated networks of academic centers and communities to support clinical trials and research, as well as creating the training programs and informatics infrastructure necessary to make these networks effective, efficient, and trustworthy.[1]

2.1.2 Creation of the CTSA Program

Once identified, these themes were translated into specific NIH awards and initiatives. In fiscal year 2005, for example, the roadmap initiatives resulted in 345 individual awards at 133 institutions in 33 states.[2] One of the initiatives related to the Reengineering the Clinical Research Enterprise theme was the development of the CTSA program. Announced in October 2005, the program's goals are to eliminate barriers between basic and clinical research, address the increasing complexities in conducting clinical research, and create academic homes for clinical and translational science.[2] The program focuses on translational research, which has two approaches. The first, known as "bench-to-bedside" research, is the translation of laboratory discoveries to clinical practice with the end goal of new treatments that can be used clinically or commercially.[8,20] The second, known as health services research, is the translation of research and treatments into practice, making sure that these interventions reach the patients and populations for whom they are intended as well as using clinical practice to inform new studies. This type of translational research looks to improve the quality of care by increasing access, coordinating systems of care, providing information and decision support for clinicians, and strengthening the patient-provider relationship.[8,20,21] With new knowledge about disease having

a 15- to 25-year gestation from basic discovery to clinical application and most research being adopted by only 50% of clinicians within 17 years, the need to accelerate both forms of translational research is critical to addressing current and future challenges in health and healthcare.[10,22]

Administered by NCATS, the CTSA program distributes funds to a national network of more than 50 medical research institutions, also known as "hubs," that work together to improve the translational research process.[3,4] For example, in fiscal year 2016, a total of more than $487 million was awarded to 57 institutions through the CTSA program.[4] With this funding, each hub is required to achieve the following CTSA program goals:

- "Train and cultivate the translational science workforce.
- Engage patients and communities in every phase of the translational process.
- Promote the integration of special and underserved populations in translational research across the human lifespan.
- Innovate processes to increase the quality and efficiency of translational research, particularly multisite trials.
- Advance the use of cutting-edge informatics."[3]

2.1.3 Need for Biomedical Informatics within CTSA Hubs

Since the inception of the CTSA program, the NIH has recognized "… the critical need to apply biomedical informatics theories and methods to enable the collection, exchange, management, analysis and dissemination of multidimensional datasets and knowledge collections."[23] As defined by the American Medical Informatics Association (AMIA), "biomedical informatics

(BMI) is the interdisciplinary field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving, and decision making, driven by efforts to improve human health."[13] Although biomedical informaticists will typically specialize in one or more application domains such as biology (bioinformatics), clinical care (clinical informatics), research processes (research informatics), or public health (public health informatics), the BMI core competencies have broad applicability inside and outside of biomedicine.[24] In addition to fundamental scientific skills such as problem definition, data analysis, solution generation and implementation, collaboration, and discussion and dissemination of ideas, such core competencies include familiarity with biological, biomedical, and health concepts and problems; creation and application of models for biomedical data, information, and knowledge; knowledge of data structures, algorithms, programming, mathematics, and statistics; and understanding and application of the fundamentals of social, organizational, cognitive, and decision sciences.[13]

Proficiency in these core competencies makes biomedical informaticists uniquely suited for translational research. Because integrating clinical and research data is critical for cost-effective and efficient translational research, BMI approaches are required "…to manage, organize, and integrate heterogeneous data to inform decisions from bench to bedside to community to policy."[12] Additionally, achievement of translational innovations requires a team science approach, creating the need for multidisciplinary teams of practitioners and researchers to conduct studies.[12] Because biomedical informaticists are expected to understand concepts that span the translational medicine spectrum (e.g., biology, clinical care, epidemiology, and health services), they have the ability to interact and communicate effectively with a variety of team

members, as well as serve as translators for the entire team.[12] Subsequently, BMI expertise and resources are critical for CTSA hubs to achieve program goals.[23]

2.2 Use of Electronic Health Record Data for Research

For most CTSA hubs, BMI programs have focused on informatics training, database design/hosting, data warehouses, data sharing infrastructure, and complex data analyses.[23] Among these activities, a high priority is the use of IT to integrate clinical research and clinical workflows.[5] In addition, the NIH has called for increasing the secondary use of EHR data for clinical research purposes.[6] Based on these priorities, BMI programs began to focus on the use of existing data sources, specifically EHRs, for clinical and translational research purposes. The following subsections describe the impact of government incentives on the rate of EHR adoption, the challenges in using EHR systems for research purposes, and the creation of data warehouses to address these challenges.

2.2.1 Impact of Government Incentives on EHR Adoption

Although EHRs contain a wide array of clinical data that have tremendous potential for comparative effectiveness and outcomes research, direct use of this data has proven problematic.[25] One initial challenge was the low adoption rate of health information technology (HIT) and EHRs. In 2008 (just two years after the first CTSA hubs received funding), less than 10% of non-federal acute care hospitals had adopted basic EHRs that allowed entry and retrieval of patient demographics; problem lists; medication lists; medication orders; discharge summaries; and lab, radiology, and diagnostic test results.[26] While researchers estimated that adopting EHRs could result in saving more than $77 billion annually, the adoption of HIT was

modest compared to other industries and showed little sign of increasing significantly.[27] Recognizing the need to reengineer the way healthcare data are collected, stored, and used, the US government passed the Health Information Technology for Economic and Clinical Health (HITECH) Act in 2009.[28] Part of the American Recovery and Reinvestment Act (ARRA) of 2009, the HITECH Act set aside $29 billion over ten years to support the adoption of EHRs.[28] A key component of this legislation was incentive payments for eligible professionals and hospitals that demonstrated meaningful use of these EHRs.[29] By 2015, these incentives resulted in more than $20.9 billion in payments to professionals and hospitals.[30] In addition, 96% of non-federal acute care hospitals possessed EHR technology, with almost 84% of these hospitals having implemented basic EHR functionality.[26]

2.2.2 Challenges in Using EHR Systems for Research Purposes

While the HITECH Act has increased the adoption of EHRs, EHR systems and underlying databases are not structured for research purposes, which poses inherent challenges when attempting to use EHR systems for capturing, storing, and retrieving research data. While some organizations have been successful in configuring EHRs for both clinical and research purposes,[31] most have struggled to overcome basic conflicts between the information and workflow needs of clinicians and researchers.[25]

One fundamental issue is that EHR systems are optimized to quickly store and retrieve information based on a single patient, while research requires querying data over multiple patients.[32] Within some EHR systems, only limited functions may be available to construct multi-patient queries and/or these queries may require significant time to complete.

Additionally, EHR system contracts may create problematic organizational boundaries. Generally, the purchase, implementation, and maintenance of an EHR system is governed by a contract between the EHR vendor and the purchaser. Contractual terms may limit the number of licenses that can be used to access the EHR and/or the organizations that may enter data within the system. The result is an EHR system that includes information for a single healthcare organization, while research may require data from multiple institutions.[25]

A third issue is that EHR systems are designed to capture unstructured, narrative data that involve no or a minimally specified data model. While EHR systems may include some data rules and validation (such as not allowing entry of birthdates with a year 1800 or limiting entries to a defined list of options), a significant amount of EHR data cannot be validated by the system, nor do most systems provide a significant number of features for data validation and analysis. In contrast, research databases often involve well-defined and structured data in discrete fields to allow the system-supported validations and analyses required to answer specific research questions.[25]

Finally, most EHR systems capture data such that clinicians are not required to clean or transform data prior to its use in clinical contexts. As a result, most EHR systems lack the necessary functions and features to perform the type of data cleaning and/or transformation required for research purposes. In addition, even if these features were provided within an EHR system, performing data cleaning and/or transformation within the primary clinical database could prove harmful to patient care if done incorrectly.[32]

In an attempt to address these challenges, EHR system vendors have developed functions and features to allow organizations to share their EHR systems with partners (such as independent physician groups) and to extract EHR data into databases specifically for reporting purposes. Unfortunately, these advancements do not address the lack of an underlying data model in the originating system, nor the need for competing health systems to participate in the same research project. As a result, CTSA hubs have investigated alternative approaches for using EHR data within translational research.

2.2.3 Creation of Data Warehouses for Research

In response to these challenges, many CTSA hubs have implemented and maintain data warehouses (also known as integrated data repositories), which are "'…subject-oriented, integrated, time-variant, non-volatile collection[s] of data…' [that address] data management, integration, and access issues."[7] For example, according to a 2010 CTSA survey, 86% of respondents indicated that their organizations had one or more data warehouses.[8] These data warehouses combine disparate data types from a range of clinical and administrative sources such as EHRs and registries and may include data from multiple institutions.[8] Additionally, instead of being built to only support one workflow (such as clinical care), data warehouses typically support a range of heterogeneous users including researchers, clinicians, and administrators.[8] From a data quality perspective, data warehouses can store metadata from each of the originating systems that explain the context in which the data were captured and their meaning, which increases the likelihood of correctly translating the data. Because these data warehouses also contain data from multiple sources, researchers can access more complete information for an individual patient (such as having access to records from multiple hospitals

for a single patient) and/or link clinical information with financial, utilization, and quality data to verify consistency of records (such as ensuring that gynecological exams were only conducted on biologically female patients).[33] The result is a more comprehensive set of data that extends beyond the contents of a single EHR system and provides more valuable information for translational research.

2.3 Summary of Research Themes in Current Literature

The creation of data warehouses for clinical and translational research has resulted in studies related to their development and implementation. In addition, clinical researchers have begun to publish studies based on the use of EHR and other data from these data warehouses. Finally, another category of literature relates to improving data warehouses for research purposes. The following subsections provide examples of these studies and summarize the key findings.

2.3.1 Data Warehouse Development and Implementation

Literature focusing on data warehouse development and implementation provides insight into the benefits and challenges of implementing these platforms. Common themes include the creation of a common data model and/or mapping that can be used to transform data from local systems into the data warehouse; technical architectures that can be scaled to integrate data from a variety of sources; and the need to protect patient privacy while reducing regulatory burden. Below are several examples of data warehouse projects described in the literature that represent these themes.

The Observational Health Data Sciences and Informatics (OHDSI) program is "an international collaborative whose goal is to create and apply open-source data analytic solutions to a large network of health databases to improve human health and well-being."[34] Coordinated through a center housed at Columbia University, the OHDSI team of academics, industry scientists, health care providers, and regulators have developed a common data model that can be used to transform databases into a common format that can be centralized into a single data warehouse for research purposes.[34,35] Once transformed into this structure, researchers can apply the open source analytic tools for data exploration (ATLAS), data quality assessment (ACHILLES), feasibility assessment (CALYPSO), and drug exposure visualization (DRUG EXPOSURE EXPLORER) that have been developed by OHDSI.[34,35] Additionally, researchers can join the OHDSI research network and collaborate on research projects.[35] Continued development of this platform will support observational studies that inform clinical practice using the unprecedented amount of patient data currently available.[34]

European researchers have developed the Electronic Health Records for Clinical Research (EHR4CR) project, which is "…an innovative platform capable of transforming conventional clinical research processes by enhancing protocol feasibility assessment, patient identification for recruitment, and clinical data exchange."[36] Funded by the European Commission and the European Federation of Pharmaceutical Industries and Associations (EFPIA), the EHR4CR project involves 34 academic and private partners working together with the goal of developing adaptable, reusable, and scalable solutions for reusing EHR data for research purposes.[36,37] The platform is based on a Service Oriented Architecture (SOA) that allows service providers and consumers to dynamically connect to the platform through clearly defined interfaces.[37] Data from clinical data warehouses or EHRs at local sites are interfaced and mapped to the EHR4CR

database and then provided through different services to the end user.[37] These services are focused on particular use cases such as querying of eligibility criteria, recruiting patients, and mapping terminology.[37] Security is maintained through implementation of standards such as SOAP (web service interactions), WS-Security (secure messaging), WS-Trust (relationship brokering), and SAML (end user authentication).[37] As a result of this work, the European Institute for Innovation through Health Data (i~HD) has been formed to guide and catalyze "…the best, most efficient and trustworthy uses of health data and interoperability for optimizing health and knowledge discovery."[38]

The Oncoshare project was founded in 2008 as a collaboration among Stanford University, the Palo Alto Medical Foundation, and the Cancer Prevention Institute of California with the goal of developing a shared database for translational research and outcomes analysis specifically for women treated for breast cancer.[39,40] This shared data warehouse integrates data from registries, EHRs, genomic sequencing laboratories, and patients, providing the necessary breadth and depth to identify care pathways that provide the best outcomes for patients.[39,40] To create this integration, records that meet the inclusion criteria are transformed into a standard data model and then aggregated into a shared, anonymized data warehouse.[39] From this data warehouse, researchers can create data marts that are specific to their particular research questions.[39] These data marts contain a subset of data from the data warehouse and are focused on a specific research question or subject area.[41] The Oncoshare project has resulted in more than 20 publications and presentations, and ongoing research continues with funding from the NCI and the NIH CTSA program, among others.[40]

In 2008, the NIH Clinical Center established a data warehouse known as the Biomedical Translational Research Information System (BTRIS), which contains data on over 500,000 research subjects.[42] While researchers can extract both de-identified and identified data from the BTRIS, practical and administrative challenges exist related to patient privacy regulations.[42] Because one approach to reducing this regulation is to use data from deceased patients, researchers from the NIH Clinical Center used BTRIS to identify deceased patients and extract them into a separate data warehouse.[42] While continued work is needed to clarify the regulatory guidance for de-identifying decedent records and to determine the effect of bias that may exist (such as having a proportionately larger number of older, sicker patients than the BTRIS), a data warehouse consisting of deceased patient records could prove an important tool to reduce regulatory burden when accessing data for research purposes.[42]

2.3.2 Use of Data from EHRs and Data Warehouses

Literature also exists that describes the results of using data from EHRs and data warehouses for clinical and translational research purposes. One category of literature addresses the ability to more effectively and efficiently perform research tasks such as cohort identification and feasibility assessment using data warehouses when compared to traditional methods. A second category includes literature that attempts to replicate clinical research through use of a data warehouse. While these studies have identified efficiency gains and some ability to replicate results, caution is advised when considering a complete replacement of traditional methods with data warehouse supported methods. Below are several examples from the literature of these types of research studies.

Based on EHR data, Mayo Clinic researchers compared two methods for prospective recruitment of patients with heart failure – one using natural language processing (NLP) and one using predictive modeling – to traditional methods that rely on coding information.[43] Reliance on coding information is problematic for some diseases, such as heart failure, which cannot be diagnosed with a single test and may have a variety of symptoms due to its syndromic nature.[43] In addition, coding is a manual process that introduces delays in identifying patients for recruitment. To find a more effective and efficient process for prospective recruitment, researchers developed algorithms for these two approaches and applied them to EHR clinical notes. They found that the NLP algorithm identified all the heart failure patients that could be identified by the traditional coding approach, but also identified additional cases that the traditional approach had not identified, which could be helpful as a screening mechanism for observational studies.[43] The predictive modeling algorithm had a positive predictive value of 82%, which indicates a high likelihood of identifying patients that truly have heart failure, making this method a good screening mechanism for clinical trial inclusion and exclusion criteria.[43] A key benefit of using these screening mechanisms is that they rely on clinical notes, which are typically available more quickly than traditional coding; thus, researchers can identify patients more quickly for participation in research studies.[43]

Using the EHR4CR platform, researchers performed a cost-benefit analysis (CBA) to assess the value of the EHR4CR compared to current research processes for hypothetical Phase II and Phase III oncology trials.[36] Three scenarios were analyzed: protocol feasibility assessment, patient identification for recruitment, and clinical study execution.[36] In performing this analysis, researchers found that using EHR4CR would translate into faster time to market (20% reduction in average cycle time) and reduced resources and costs (50% reduction in person time and

costs).[36] Using the EHR4CR platform, these benefits were achieved through reducing the efforts required for refinement of inclusion and exclusion criteria, accelerating patient recruitment, and optimizing clinical trial execution.[36]

Researchers in the United Kingdom (UK) used a comprehensive longitudinal electronic clinical database, the UK General Practice Research Database (GPRD), to replicate a set of six previously completed randomized controlled trials related to cardiovascular drug treatments.[44] The GPRD is a data warehouse that contains over eight million patient records from a representative sample of 5.7% of the UK population from 1990-2000.[44] While some limitations were identified, such as missing or limited lab results, vitals, and history, the results of this work suggested that "…observational studies using databases might produce valid results concerning the efficacy of cardiovascular drug treatments."[44]

2.3.3 Improvement of Data Warehouses for Research Purposes

Although literature contains examples of the benefits of using data warehouses for clinical and translational studies, considerable challenges exist before data warehouses are optimized for research purposes. Broadly, these challenges involve data quality, data models and standards, data structure, and ethical and legal considerations. Below is a summary of these challenges and the work being done to improve data warehouses.

Data quality is one of the most significant challenges in the use of data warehouses for research purposes, in part because the concept of data quality is multi-dimensional and lacks a commonly accepted definition.[45–48] Within the literature, the concepts of accuracy and completeness are most often used in data quality descriptions.[45–47,49–55] Accuracy, which can be defined as the

extent to which data conform to the truth, often relies on correct and careful documentation, which can be challenging in a busy patient care setting.[45,55] The result can be wrong diagnoses on a problem list, inclusion of documentation for two patients within the same record, and implausible documentation of services (such as gynecological services for biologically male patients).[46,55] The definition and measurement of data quality concepts, though, can be context dependent, meaning that data considered high quality for one use may be considered poor quality for another use.[48] For example, completeness in a clinical setting may be defined as documenting all of the observations from a patient encounter, while completeness when reusing this same data for research purposes may be defined as containing all the data elements necessary for the research study.[6,49,55–58] Without a standard definition of data quality, efforts to improve data warehouses will produce marginal results, if any. In addition, perceptions of poor data quality could prove to be a significant barrier in the adoption of data warehouses in clinical and translational research. To address this challenge, current research focuses on identifying frameworks and ontologies for data quality[45–47,52] and assessing data quality using these tools.[56,57,59]

Additionally, the lack of a single, accepted data model for clinical and translational research has hampered interoperability across institutions and countries.[60] As a simple example, a laboratory test such a comprehensive blood count (CBC) can be coded as a CBC in one system and code 1568 in another system. When attempting to integrate laboratory tests into a single data warehouse, they must to be mapped to a single, common concept. Without this mapping, a researcher would need to know to use CBC and 1568 as codes to extract CBC results, which is an unrealistic expectation and likely to cause data to be missing from the data extraction. To address this challenge, various international certification and standards bodies continue to pursue

the goal of creating the necessary models and standards. Once identified, significant work will be required to implement these standards in EHRs, registries, and other data sources so that the overhead of transforming and mapping these data into data warehouses can be reduced.[60]

From a data structure perspective, a significant portion of data from clinical sources is maintained within unstructured narratives, such as admission notes, treatment plans, and patient summaries.[60] While abundantly available, this type of data is computationally the most difficult to analyze due to its heterogeneous nature; lack of conformity with grammar and sentence structures; and use of abbreviations, acronyms, and idiosyncratic language.[60] Additionally, information may be duplicated within the data structure, making it difficult to determine if these are new findings or a restatement of existing findings.[59] For example, a clinical note may include a statement of a concerning trend in high blood pressure readings that is already captured within the vitals data fields as discrete data. While the discrete data may be easier to manage computationally, the clinical note may provide the most relevant information for a particular research study. To address this challenge, current research focuses on optimizing NLP algorithms for automated coding of this data into more discrete forms and configuring EHRs such that data can be used for clinical and research purposes.[31,43]

Legally and ethically, organizations that capture, store, and retrieve patient information have a duty to ensure the privacy of this data, a task that is particularly challenging in the age of numerous security breaches to IT systems. To overcome this challenge, de-identification is often used within data warehouses, but the lack of identifiers can reduce the value of the data for research purposes.[60] Additionally, researchers need to ensure appropriate consent was obtained for the data included within the warehouse, which can create a significant administrative burden

and potential bias in the patient population included.[60] Without this consent, public trust in the research community can erode and this, in turn, can hinder public good.[61] To encourage participation in studies, researchers must have a trustworthy process for obtaining consent and maintaining privacy. To address these challenges, researchers are evaluating the use of data from deceased patients as a potential alternative.[42] In addition, the 21st Century Cures Act includes provisions to improve privacy protections for research volunteers.[62]

2.4 Importance of Data Warehouses in Future Research

As the quality of data warehouses continues to improve based on these studies, competency in using data warehouses will be a critical skill for research teams participating in clinical and translational research. Future research in genomics, precision medicine, and big data analytics will require use of large datasets containing diverse data from a variety of sources. The following subsections describe the promise of future research in these areas, the role of data warehouses and large datasets within these research areas, and the need for research teams to demonstrate the necessary skills for taking advantage of these resources.

2.4.1 Genomics

From a classical Mendelian perspective, organisms contain genes that are inherited, and these genes define a variety of observable traits depending on gene dominance.[63] The goals of the human genome project are to identify 100,000 different genes in humans and understand their expression and function.[63] Researchers involved in this project and related research "…study genes, their interactions, their mutations, and the relationships they reveal between normal function and disease."[63] To achieve this understanding, research teams need genomic data

(typically, genome sequences), as well as phenotypic (clinical) data for individual patients and populations.[64] A key challenge is the development of a system to house these two distinctly different types of data.

While effectively integrating genomic data into EHRs remains a goal for vendors and health care systems, one of the key existing features of data warehouses is the ability to integrate data from a variety of sources. As a result, a successful data warehouse implementation can be seen as a precondition for integrating clinical and genomics data.[65] To drive the development of more accurate classifications of diseases and to enhance diagnosis and treatment methods, research teams will need "a data network that integrates research data on the molecular makeup of diseases with clinical data on individual patients…."[65] Such a data warehouse will need to include appropriate storage of this data; a framework that enables scientists and researchers to explore the data and generate hypotheses; and information from other databases and data sources for cross-referencing.[65]

Effective use of this data warehouse will require "…better integration of genomics and biomedical informatics into curricula for clinical researchers and providers."[63] Researchers and clinicians will need to participate in the development of standards and models for capturing and storing this information, as well as serve as advocates in debates related to the perplexing ethical, social, and economics issues that will occur when this data is integrated into a single platform.[63] Active and informed participation will require experience with capturing, storing, and retrieving data from a data warehouse to effectively represent the benefits and challenges of using such a technology.

2.4.2 Precision Medicine

In his January 2015 State of the Union Address, President Barack Obama announced the Precision Medicine Initiative (PMI), which is meant to bring together clinical, genomic, environmental, lifestyle, and other related data to provide a comprehensive view of a patient's state over time and customize disease prevention, detection, and treatment based on the individual patient.[66–68] While individual variability (such as blood type) has been considered within some medical treatments, this initiative calls for a broad, evidence-based approach using innovative tools (such as mobile health technologies), large-scale biologic databases (such as the human genome sequence), powerful patient characterization methods (such as proteomics, metabolomics, and genomics), and computational tools for analyzing large sets of data to customize treatment and prevention and improve the overall effectiveness and quality of patient care.[67,68]

In 2016, the 21st Century Cures Act was passed, which provides funding for PMI-related initiatives, among other initiatives.[62,69] Worth $4.8 billion over ten years (starting in fiscal year 2017), this funding will provide the NIH "…with critical tools and resources to advance biomedical research across the spectrum from basic, curiosity-driven studies to advanced clinical trials of promising new therapies."[62,69] Initiatives funded by this Act are meant to reduce administrative burdens that slow the progress of science, enhance data sharing and privacy protections for study participants, improve support for the next generation of biomedical researchers, and ensure the inclusion of diverse populations in clinical research.[62,69] Additionally, support is provided for four highly innovative scientific initiatives: the All of Us Research Program, the Brain Research through Advancing Innovative Neurotechnologies

(BRAIN) Initiative, the Cancer Moonshot[SM], and the Regenerative Medicine Innovation Project.[62]

The All of Us Research Program, formerly known as the PMI Cohort Program, is a longitudinal cohort study meant to involve one million or more Americans who volunteer to provide biologic specimens (e.g., cell populations, proteins, and genome sequences); clinical data from EHRs; and lifestyle and behavioral information.[67–69] This information will be used to create a "..transformative research infrastructure that will enable and simplify research across all diseases."[67,69]

The goal of the BRAIN Initiative is to understand how the brain functions, which will inform efforts to transform the ways in which neurological and mental disorders are diagnosed and treated.[62] The initiative will focus on building technology and knowledge across disciplines to understand how circuits in the brain function in real time and what goes wrong when disease occurs.[69]

The Cancer Moonshot[SM] will accelerate cancer research to improve prevention, detection, and treatment.[70] The goal is "…to double the rate of progress in the fight against cancer, making more therapies available to patients, while also improving [the] ability to detect and prevent cancer."[69]

Finally, the Regenerative Medicine Innovation Project is coordinated through the NIH and the Food and Drug Administration (FDA) and supports clinical research using adult stem cells to further the field of regenerative medicine.[62] Regenerative medicine is a field of science focusing

on the use of stem cells and other technologies to repair or replace damaged cells, tissues, or organs.[71]

All of these initiatives will require the collection, storage, and retrieval of multi-parametric data from a variety of sources.[72] Dedicated efforts will be required to integrate the multidimensional data from EHRs, large-scale genomic-wide data, and information from mobile health technologies into a single data warehouse to support the analytic and bioinformatics needs of clinicians and reseachers.[33] Recently, to achieve this goal, the National Academy of Sciences (NAS) has called "…for the development of an information commons (IC) that amasses medical, molecular, social, environmental, and health outcomes data for large numbers of individual patients."[73] Effective use of such an IC or data warehouse will require changes to the training and education models for clinicians and researchers to include competencies related to quantitative reasoning, ability to access just-in-time information, and practice integrating multiple parameters for a holistic view of the patient and/or disease.[66,68,72]

2.4.3 Big Data and Analytics

Big data refers to complex datasets that challenge traditional data management systems due to the high volume of data, the diversity of the data within the dataset, and the rapid rate at which the data changes.[74–76] Within healthcare, the adoption of EHR systems, digitization of medical images and videos, as well as other related information has resulted in an estimated 150 exabytes (150 billion gigabytes) of data in the US healthcare system by 2011.[77] Should current growth rates continue, big data in healthcare will reach zettabytes (10^{21} gigabytes) or yottabytes (10^{24} gigabytes).

To effectively gain information from this data, scientists apply analytics methods. These methods include "…the use of mathematical and algorithmic processing of data resources, as well as techniques such as text mining and natural language processing, and visual analytics to generate descriptive, predictive, and prescriptive models to analyze and derive insight from data."[75] Application of these methods represents a significant opportunity for health care delivery systems, researchers, and clinicians. Currently, clinical operations are estimated to include $165 billion in waste, which could be reduced by applying big data analytics to determine the most clinically relevant and cost-effective ways to diagnose and treat patients.[77] Additionally, health systems could use analytics to identify the estimated 5% of patients that account for 50% of all US health care spending and determine interventions to reduce these costs while maintaining the quality of care.[76] Research and development is estimated to include $108 billion in waste that could be reduced by using predictive modeling to produce leaner, faster, and more targeted research for drugs and devices, as well as improve design of clinical trials.[77] Application of these methods will also be required to effectively use the data gathered by the initiatives identified within the 21st Century Cures Act.[69,74] Finally, clinicians need efficient and effective ways to apply the latest evidence in their practices, and big data analytics can be used to better predict disease for and match treatments to individual patients.[77] From a public health perspective, these methods are already being applied by the Centers for Disease Control and Prevention (CDC) to inform clinicians about the spread of influenza.[74]

To participate in the use of big data analytics, research teams will need to be versed in informatics, as well as their chosen disciplines.[75,78] Using technologies like data warehouses can provide valuable experiences in applying informatics concepts to their existing research interests.

Additionally, experience leveraging this technology can improve the effectiveness and efficiency with which technologies are adopted within research programs.[75]

2.5 Key Gaps Within Existing Literature

Although significant research is being conducted about and with data warehouses for current and future clinical and translational science purposes, critical gaps exist within the literature corpus.

First, little is known regarding interventions for increasing researcher and research team awareness of data warehouses and their use in clinical and translational science. For example, is a presentation regarding basic data warehouse functions and processes adequate to raise awareness or is a more significant intervention required? Without appropriate interventions, a generation of research teams may be unaware of the data warehouses available to them, resulting in ineffective and inefficient research processes as well as a lack of experience with the necessary informatics methods to face future research challenges.

Second, while CTSA hubs often provide data warehouse training for research teams, little is known about the outcomes of this training. For example, at the end of training, can research team members demonstrate accurate knowledge regarding the data warehouses available to them? Can trainees perform activities within the data warehouses that support their intended study outcomes? Without assessment of these activities, CTSA hubs may be investing significant time and money into training programs that are not achieving their intended learning objectives.

Finally, a critical gap exists in defining objective measures for determining if research teams perform the necessary data warehouse activities to achieve their intended research outcomes. Without this information, the NIH and CTSA hubs are unable to determine if a return on investment is occurring for the significant time and resources involved in implementing and maintaining data warehouses.

As a result, critical needs exist to identify interventions that increase research team awareness of data warehouses, to assess outcomes of data warehouse training, and to define objective measures for determining if research teams perform the necessary data warehouse activities to achieve their intended research outcomes.

CHAPTER 3: CONCEPTUAL FRAMEWORKS

To address the critical needs identified in the previous chapter, this study includes a presentation intervention to raise researcher awareness of data warehouses, assessments of the awareness intervention and existing data warehouse training, and definition and use of objective measures for data warehouse activities related to research outcome achievement. The design of these study components is based on four conceptual frameworks related to adoption of technology and assessment of learning objectives: Diffusion of Innovation theory, Technology Acceptance Model, DeLone and McLean Information Systems Success Model, and Bloom's Taxonomy of Educational Objectives. Because these frameworks have been applied in a variety of contexts and have proven to be reliable and robust, they provide a solid theoretical foundation for this study. This chapter provides a summary of these frameworks, as well as examples of their applications in healthcare environments and/or with data warehouse implementation and use.

3.1 Technology Adoption Models

One of the key research areas within information systems is understanding how and why individuals adopt new information technologies.[79] Typically, this research focuses in one of three areas: implementation success at an organizational level, individual acceptance, or task-technology fit.[79] While many theories and models have been proposed based on these areas, several have been widely accepted and researched for various industries, contexts, and systems: Diffusion of Innovation theory, Technology Acceptance Model, and DeLone and McLean Information Systems Success Model. This section will describe the theoretical basis for each these models and provide the key findings in studies that apply them.

3.1.1 Diffusion of Innovation (DOI) Theory

Created by Everett M. Rogers in 1962, the DOI theory is the most widely used theoretical basis for studying technology adoption, which is defined as "…a decision to make full use of an innovation as the best course of action available."[9,80] According to Rogers, "diffusion is the process in which an innovation is communicated through certain channels over time among the members of a social system."[9] Based on this definition, DOI has four main elements: the innovation, communication channels, time, and a social system. An innovation is an idea, practice, or object that is perceived to be new by the individual.[9] A communication channel is the means used to share messages from one individual to another such as mass media and interpersonal channels.[9] Time is defined in three ways: time from awareness to adoption or rejection, time to accept an innovation relative to other individuals within the social system, and the rate of innovation adoption.[9] Finally, a social system is the interrelated individuals, groups, or organizations that engage in joint problem solving to accomplish a common goal.[9]

Using this definition as a basis, Rogers identified a five stage innovation-decision process, which is "…the process through which an individual…[passes] from gaining initial knowledge of an innovation, to forming an attitude toward the innovation, to making a decision to adopt or reject, to implementing new ideas, and to confirming this decision."[9] The following table summarizes these stages.[9]

Table 1. Stages within the Innovation-decision Process

Stage	Description
Knowledge	Individual is made aware of an innovation's existence and gains understanding of how it functions
Persuasion	Individual forms an attitude (favorable or unfavorable) towards the innovation

Stage	Description
Decision	Individual engages in activities that lead to a decision (adoption or rejection)
Implementation	Individual puts the innovation into use
Confirmation	Individual seeks reinforcement for his/her decision and may change this decision if exposed to conflicting messages about the innovation

Within the knowledge stage, individuals may either passively receive knowledge (such as being exposed to innovation information during a department meeting) or actively seek knowledge based on an identified need (such as wanting to gain access to health care data and investigating potential data sources).[9] This first exposure to innovation is known as awareness-knowledge.[9] Based on this exposure, an individual may seek out how-to knowledge (information about using an innovation properly) and principles-knowledge (information about how the innovation works).[9] While those looking to influence innovation adoption often focus on awareness, they may be more effective by also focusing on how-to knowledge delivery.[9]

While the knowledge stage is mostly cognitive, the persuasion stage is more about feelings toward the innovation, which lead to the formation of attitude.[9] Attitude is defined as "…a relatively enduring organization of an individual's beliefs about an object that predisposes his or her actions."[9] In forming this attitude, individuals will actively seek out information about the innovation, determine the credibility of the messages received about the innovation, and interpret the information that is received, all with the goal of reducing the uncertainty that is inherent in adopting innovations.[9] Often, individuals will also mentally apply the innovation to their current or future situations as a way to further reduce their uncertainty.[9]

Based on the attitude developed in the persuasion stage, the individual will decide to adopt or reject the innovation. To cope with uncertainty, some individuals may try out the innovation on a small-scale basis, and those who perform this trial will often move toward an adoption decision.[9] Other individuals may look to a peer to use the innovation before adopting the innovation themselves or to demonstrations of the innovation that align with their needs.[9]

If an adoption decision is made, the next stage is implementation, where the individual puts the innovation to use. To this point, the process has been a mental exercise, but implementation involves true behavior change.[9] At this stage, uncertainty exists, particularly in understanding where the innovation can be obtained, how it is used, and how problems can be solved, so support is needed to help reduce this uncertainty.[9]

Finally, as individuals use an innovation, they may decide to continue or discontinue its use based on their experiences and/or messages received regarding the innovation during the confirmation stage.[9] Typically, decisions to discontinue use of an innovation fall into one of two categories: replacement or disenchantment.[9] Replacement discontinuance occurs when an individual rejects an innovation in order to adopt a better innovation (such as upgrading from the current iPhone version to a new iPhone version).[9] Disenchantment discontinuance occurs when an individual rejects an innovation due to dissatisfaction with its performance, which may be caused by lack of compatibility between the innovation and individual need or misuse of the innovation.[9]

The rate of adoption, measured by the number of individuals who use an innovation in a specified period, can vary significantly.[9] Although the communication channel and social system influence this rate, the majority of this variance can be explained by five attributes of an innovation: relative advantage, compatibility, complexity, trialability, and observability.[9] The following table summarizes these attributes.[9]

Table 2. Attributes of Innovation

Attribute	Description
Relative advantage	The degree to which an innovation is perceived as being better than the innovation it supersedes
Compatibility	The degree to which an innovation is consistent with an individual's existing values, past experiences, and needs
Complexity	The degree to which an innovation is perceived as difficult to understand and use
Trialability	The degree to which an innovation may be used on a limited trial basis
Observability	The degree to which the results of the innovation are visible to others

Within healthcare settings, the DOI theory stages and attributes of innovations have been applied in a variety of contexts such as adoption of integrated care pathways (plans of patient care for specific diagnoses or interventions), telemedicine, EHRs, and health care delivery models.[81,82] In the majority of cases, the DOI theory has proven to be a solid theoretical foundation with consistent empirical support that can be used to explain an individual's adoption process, especially in the knowledge, persuasion, and decision stages.[80] In all contexts, the main criticism of this theory is the focus on individual versus organization adoption of innovation.[80]

3.1.2 Technology Acceptance Model (TAM)

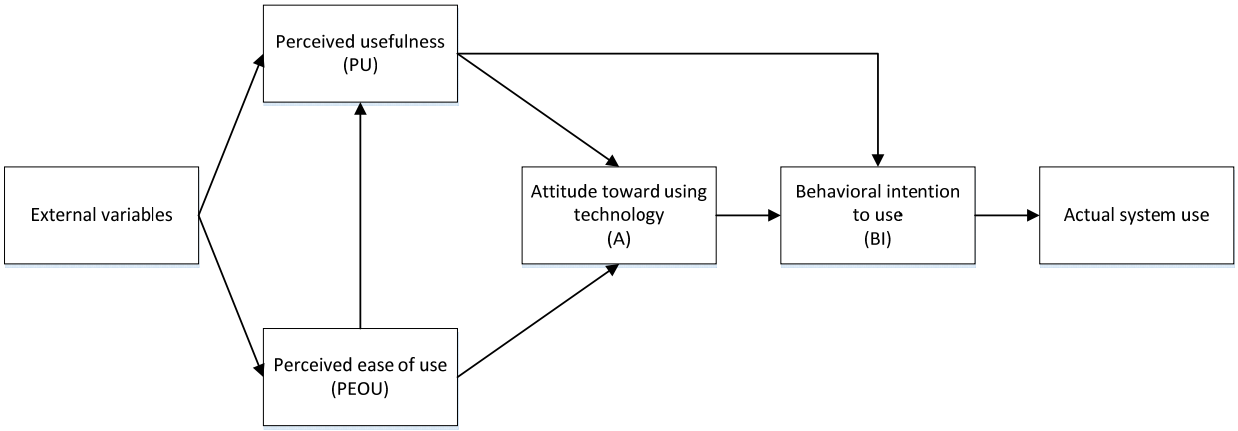
Similar to the concept of adoption, technology acceptance is an individual's psychological state regarding his/her voluntary or intended use of a technology, and studies have found that this intention to use a technology can be used to predict actual use of that technology.[83–85] The TAM is an adaptation of the Theory of Reasoned Action (TRA) specifically for modeling and predicting user acceptance of information systems.[86] The TRA states that an individual's intention to perform a behavior is based on his/her attitude toward the behavior and his/her perception that important individuals to him/her believe the behavior should or should not be performed (known as subjective norm).[83] Although based on the TRA, the TAM does not include the concept of subjective norm and identifies two particular beliefs, perceived usefulness and perceived ease of use, that are relevant to acceptance of technology.[83] The following table summarizes the concepts used within the TAM.

Table 3. Concepts within the Technology Acceptance Model

Concept	Acronym	Definition
External variables	(none)	External factors such as human, social, and system factors[83,87]
Perceived usefulness	PU	An individual's perception that using a particular technology will enhance his or her job performance[79,83,86]
Perceived ease of use	PEOU	An individual's expectation that using the system will be free of effort[79,83,86]
Attitude toward using technology	A	An individual's feelings (positive or negative about performing a behavior[79,83,86]
Behavioral intention	BI	Measure of the strength of an individual's intention to perform a specified behavior[83]

According to the TAM, these beliefs determine an individual's attitude toward the technology. In addition, attitude and perceived usefulness determine behavioral intention. Finally, these beliefs are influenced by external variables such as available training, organizational support, and system features.[83,87] These relationships are shown in the following figure.

Figure 1. Technology Acceptance Model



Adapted from Davis *et al.*, 1989.

The TAM has been well tested in a variety of business and healthcare contexts and has proven to be reliable and robust in predicting technology acceptance.[88,89] The following table provides a summary of several studies and their results within healthcare contexts.

Table 4. Results of Using the Technology Acceptance Model in Healthcare Contexts

Population	**Technology**	**Prediction of BI**
Physicians	Telemedicine	42-44% of observed variance in BI[85,90]
Physicians	Internet-based health applications	59% of observed variance in BI[88]
Residents and physicians	Personal digital assistants (PDAs)	57% of observed variance in BI[91]
Physicians, nurses, and medical technicians in Taiwan	Mobile health	70% of observed variance in BI[92]
Medical personnel in Taiwan	Adverse event reporting system	59% of observed variance in BI[93]

Population	Technology	Prediction of BI
Hospital personnel in Greece	Hospital information systems	87% of observed variance in BI[94]
Nurses	Remote intensive care unit monitoring (eICU)	58% of observed variance in BI[95]
Physicians	EHRs	44-56% of observed variance in BI[96,97]

Within these studies, the typical approach is to survey or interview a sample population using a series of items designed to assess the concepts within the model. Each concept is assessed using multiple items that include a mix of positive and negative statements. For example, some ease of use items are "Learning to operate [technology] is easy for me" and "I believe that it is easy to get a [technology] to do what I want it to do."[98,99] Responses to these items are measured using a five- or seven-point Likert-type scale with values ranging from *strongly agree* to *strongly disagree*.[92,99] Likert responses are then converted to numeric equivalents, and data analysis is typically performed using structured equation modeling (SEM) to determine the effect of each of the concepts.

Based on this analysis, the majority of studies using the TAM have found perceived usefulness to be a strong determinant of intention to use.[83,85,95–97,99–102,86–89,91–94] In comparison, studies have shown mixed results for perceived ease of use, with some studies indicating a strong link between PEOU and intention to use[83,86,92–97,101,102] and others indicating PEOU is not a significant determinant.[85,87–89,91,99,100] Generally, perceived ease of use was identified as a secondary determinant to perceived usefulness.[83] As stated by Davis et. al, "users may be willing to tolerate a difficult interface in order to access functionality that is very important, while no amount of ease of use will be able to compensate for a system that doesn't do a useful

task."[83] Interestingly, studies involving physicians were more likely to find that PEOU was not a significant determinant, which researchers postulate is due to a higher level of competence, intellectual and cognitive capacity, adaptability to technology, and operational support for physicians in comparison to other populations studied.[85,89,90,99]

Several themes emerged from these studies. One common theme was the need for users to recognize the technology as compatible with their work tasks and job performance in order to perceive the technology as useful.[84,92,103] Training was also a recurrent theme, indicating that training interventions focusing on the benefits of the technology as well as including content regarding real-life scenarios influenced perceived usefulness and attitude.[87,91,92,94,96] Also identified was the need for appropriate ongoing support for the technology once training was complete.[87,92,102] Another theme was the need to demonstrate the effectiveness of a technology, particularly in comparison to the status quo, to improve attitude and perceived usefulness of a technology.[84,91] Finally, although intention has been shown to be a good predictor of actual use of a technology, researchers identified the need to include objective measures of actual use whenever possible.[86,104]

3.1.3 DeLone and McLean Information Systems Success Model

One of the first models used to measure successful implementation of information systems was created by DeLone and McLean in 1992. Known as the DeLone and McLean Information Systems (D&M IS) Success Model, the model is an attempt to define a dependent variable for information system (IS) research, as the evaluation of practices, policies, and procedures requires an established dependent variable.[105] The original model identified six interrelated dimensions

as categories of IS success.[105,106] In 2003, ten years after the initial publication of this model, DeLone and McLean updated the model based on the research studies conducted during that time period. At that time, two extensions of the model were identified: the addition of service quality and the introduction of net benefits to address all the impacts (including individual and organization) of a system.[106] The following table summarizes these dimensions and typical measurements used.

Table 5. D&M IS Success Model Dimensions

Dimension	Model Year	Description	Example Measurements
System quality[105,106]	1992	Measure of the system characteristics[105,107]	Ease of use, functionality, reliability, flexibility, data quality, portability, verifiability, and integration[106–108]
Information quality[105,106]	1992	Measure of the characteristics of the information output provided by the system[105,107]	Accuracy, timeliness, completeness, relevance, understandability, and consistency[106,107]
Use[105,106]	1992	Measure of the use of the information by decision makers and intended users in a voluntary context[105,107]	Frequency of use, self-reported use, actual use, time of use, appropriateness of use, number of accesses, and usage pattern[106–108]
User satisfaction[105,106]	1992	Measure of the satisfaction experienced by decision makers and intended users in a non-voluntary context[105,107]	Satisfaction ratings for specific systems and scales used to assess attitudes and satisfaction[105,107]
Individual impact[105,106]	1992	Measure of the effect of the information on the behavior of the recipient[105]	Job performance, decision-making performance, and quality of work environment[106,108]
Organizational impact[105,106]	1992	Measure of the effect of the information on the performance of an organization[105]	Costs and benefits, contribution to company profits, and return on investment[105,108]

Dimension	Model Year	Description	Example Measurements
Service quality[106]	2003	Measure of the service provided by IS organizations to end users[106]	Currency, reliability, responsiveness, assurance, and empathy[106,108]
Net benefits[106] (replaces individual and organization impact)	2003	Measure of all impacts (e.g., individual, workgroup, organizational, and industry) and benefits of a system[106,107]	Improved decision making, improved productivity, increased sales, cost reductions, creation of jobs, and economic development[107]

Since its publication, this model has been widely accepted as a framework for IS success measurements.[107,108] Further research has focused on identifying the independent variables that may "cause" this success. Research by Petter, DeLone, and McLean published in 2013 identifies several variables that serve as strong determinants of overall IS success: enjoyment, trust, user expectations, extrinsic motivation, and IT infrastructure.[107] Several other variables were identified as strong determinants for specific IS success dimensions. The following table summarizes these findings.[107]

Table 6. Determinants of IS Success Dimensions

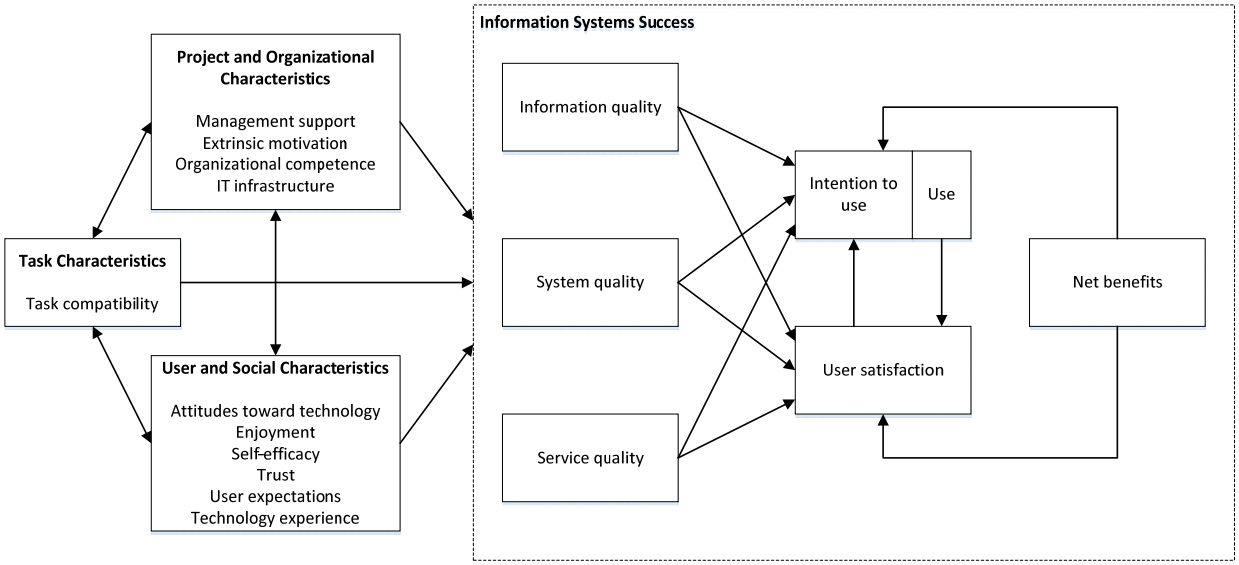
Determinant	Overall IS Success	System Quality	Use	User Satisfaction	Individual Impact
Task compatibility				Strong	
Attitudes toward technology				Strong	
Enjoyment	Strong				
Self-efficacy		Strong			
Trust	Strong				
User expectations	Strong			Strong	
Technology experience		Strong			
Management support					Strong

Determinant	Overall IS Success	System Quality	Use	User Satisfaction	Individual Impact
Extrinsic motivation	Strong		Strong		
Organizational competence			Strong		
IT infrastructure	Strong		Strong		

Adapted from Petter *et al.*, 2013.

These determinants are the interrelated, independent variables for the dependent variable of IS success. The following figure summarizes these relationships.

Figure 2. Determinants and Dimensions of IS Success



Adapted from Petter *et al.*, 2013, and DeLone and McLean, 2003.

The D&M IS Success Model has been applied to numerous IT systems, including the implementation and use of data warehouses.[109] Applying this model to data warehouse projects is particularly important, as studies have reported project failure rates from 41-90%.[110–112]

Typically, studies applying the D&M IS Success Model survey or interview a sample group with the goal of identifying factors that determine successful system implementation and use.[109] Each potential factor is measured using multiple items with responses recorded using a seven-point

Likert-type scale with values ranging from *strongly agree* to *strongly disagree*.[109,113] For example, items measuring data quality include "Users have more accurate data now from the data warehouse than they had from source systems" and "I can get data that is current enough to meet my work needs."[109,113] Likert values are then converted to numeric values and analyzed using a variety of statistical techniques, such as partial least squares, regression, and hypothesis testing.[7,109,110,113]

Based on this analysis, several factors have emerged as critical to the successful implementation and use of data warehouses. First, the quality of the data warehouse (system quality) and the data (information quality) have been found to be key factors associated with net benefits of using the system.[108–110] Also influencing net benefits is the degree to which the features and functions provided by the data warehouse align with business needs and user tasks.[110,114] Management support, training, and appropriate resources to implement and support the data warehouse have been found to improve user satisfaction.[7,108,109,114–118]

3.2 Bloom's Taxonomy of Educational Objectives

Because training was identified as an influencing factor within the TAM and D&M IS Success Model, an understanding of educational objectives also provides a framework for this study. Bloom's Taxonomy of Educational Objectives (known as Bloom's Taxonomy) is a framework for classifying what students are meant to learn as a result of instruction.[119] The goal of this taxonomy is to help educators in two ways: to clarify their intended outcomes and to inform the design of appropriate instruction and assessment methods.[120] Originally envisioned as a method for facilitating the exchange of test items among faculty at various higher education institutions,

this taxonomy has become widely known, resulting in its translation into 22 languages and use in all levels of education.[119]

The original taxonomy, published in 1956, provided six distinct levels of learning: knowledge, comprehension, application, analysis, synthesis, and evaluation.[119] These levels are organized in a cumulative hierarchy, where the ordering is simple to complex and from concrete to abstract, so that each simpler level must be mastered before the next more complex level.[119] In 2001, a revision of the original taxonomy was published, which provided changes to the naming of each level to reflect more modern language.[119–121] The following table describes each revised level.[119,121]

Table 7. Cognitive Levels within the Revised Bloom's Taxonomy

Level	Description
Remember	Able to retrieve relevant information from long-term memory
Understand	Determine the meaning of instructional messages (oral, written, and graphic)
Apply	Carry out or use a procedure in a given situation
Analyze	Break material into parts and detect how the parts relate to one another and the overall concept
Evaluate	Make judgments based on criteria and standards
Create	Put elements together to create a novel, coherent whole or an original product

This revision also identified four types of knowledge that could be assessed using the cognitive levels.[119,120] Three of these types were included in the original taxonomy (factual, conceptual, and procedural) and a fourth (metacognitive) was added in the revision.[119] The following table describes each of these types and typical knowledge covered by each type.

Table 8. Knowledge Types within the Revised Bloom's Taxonomy

Type	Description	Typical Knowledge
Factual	Basic elements that a student must know to be acquainted with and solve problems within a specific discipline[119,120]	• Terminology • Specific details and elements[119]
Conceptual	Interrelationships among the basic elements that allow them to function together[119,120]	• Classifications and categories • Principles and generalizations • Theories, models, and structures[119]
Procedural	• How to do something • The methods of inquiry • The criteria for using skills, algorithms, techniques, and methods[119,120]	• Subject-specific skills and algorithms • Subject-specific techniques and methods • Criteria for determining when to use appropriate procedures[119]
Metacognitive	• Knowledge of cognition, in general • Awareness and knowledge of one's own cognition[119,120]	• Strategic knowledge • Cognitive tasks, including appropriate contextual and conditional knowledge • Self-knowledge[119]

The combination of knowledge type and cognition level provides a two-dimensional representation of learning objectives. For example, a learning objective that states, "identify the four types of knowledge represented in Bloom's taxonomy" would be *factual* knowledge at the *remember* cognitive level. In contrast, a learning objective that states, "Design program-level learning outcomes based on Bloom's taxonomy" would be *conceptual* knowledge at the *create* cognitive level.

Based on these dimensions, assessments are created to determine if the learning objectives have been met. For the first example objective above, a simple multiple-choice test item may be sufficient to confirm that this objective has been met. In the second example, a summative report outlining the research performed to define and determine the outcomes, as well as the outcomes

themselves, may be an appropriate assessment. These assessments should be linked to the objectives as well as the content of the instruction provided.[120]

3.3 Application of Conceptual Frameworks in this Study

This study included a presentation intervention to raise awareness of data warehouses among research teams, assessments of the awareness intervention and existing data warehouse training, and definition and use of objective measures for data warehouse activities related to research outcome achievement. This section summarizes the use of the conceptual frameworks in designing each of these components.

The study approach mirrored the five innovation-decision stages identified in the DOI theory. In the knowledge stage, Rogers recommends providing awareness-knowledge and how-to knowledge to influence innovation and adoption. As a result, this study included an intervention to provide awareness-knowledge and use of an existing training process to provide how-to knowledge.

In the persuasion stage, individuals actively seek out information regarding the innovation and mentally apply an innovation to their current or future states. As a result, this study included a process for individuals to request access and training for a data warehouse. In addition, the request involved in this process required the individual to identify potential applications of the data warehouse in his/her research project.

In the decision stage, individuals who try out an innovation are more likely to adopt that innovation. As a result, part of the training provided within this study included hands-on experience using a data warehouse.

In the implementation stage, individuals need support to put an innovation to use. As a result, this study included distribution of support information (e.g., how to obtain access to the data warehouse and who to contact with questions) during the awareness intervention and training.

In the confirmation stage, individuals decide to continue or discontinue using the innovation. Deciding to discontinue can be due to a replacement innovation or disenchantment with the innovation. As a result, the awareness intervention and training included information about the benefits of using a data warehouse instead of paper chart review or EHR chart review for common research tasks to influence replacement of the existing methods. In addition, the awareness intervention and training provided realistic expectations about how the data warehouse can be used for research purposes. Appropriate expectations were critical to avoid disenchantment with data warehouses.

Based on the TAM and D&M IS Success Model, the awareness intervention and training included the following:

- Realistic expectations for system and information quality.
- Information about the usefulness of a data warehouse for common research tasks (task compatibility).
- Description of the benefits that can be achieved using a data warehouse (net benefits).
- Information for obtaining data warehouse support.

Using Bloom's Taxonomy, the learning objectives for the awareness intervention and training were written. This study included assessment of these learning objectives based on the knowledge type and cognitive level.

This study also included a follow-up assessment 30 days after completion of training to assess factors identified as significant by the TAM and D&M IS Success Model, including ease of use for data warehouse tools, availability of support, and benefits and barriers to using a data warehouse.

Finally, the TAM identifies the need to objectively measure the adoption and use of technology. As a result, this study included measures of the number of requests for data warehouse access, the number of requests for data warehouse training, and the number of individuals completing the associated data warehouse activities to achieve their intended research outcomes.

The following table provides a summary of the conceptual frameworks and their application within this study.

Table 9. Summary of Conceptual Frameworks

Framework	Application to Study
Diffusion of Innovation (DOI) theory	• Provides the stages of innovation adoption, which were used as the basis for the study approach. • Identifies the activities that indicate movement between the stages of technology adoption, which were used assess the progression of individuals from one stage to the next within the study. • Describes reasons for discontinuing the use of an innovation, which were used to inform potential barriers included in the post-training assessment within this study.

Framework	Application to Study
Technology Acceptance Model (TAM) and DeLone & McLean IS Success Model	• Describes key content influencing technology adoption highlighted within the awareness presentation intervention and the existing training. • Identifies key aspects influencing technology adoption/acceptance, which provided the focus for the assessments within this study. • Identifies that assessment of factors influencing technology adoption should be combined with objective measurement of system use, which provided the reason for including the objective measurement component of this study.
Bloom's Taxonomy of Educational Objectives	• Describes a framework for creation of learning objectives, which was used to create objectives for the awareness presentation intervention and existing training. • Provides levels of assessment based on objectives, which was used to determine the appropriate assessment for each learning objective identified for the awareness presentation intervention and existing training.

CHAPTER 4: METHODS

Within this chapter, the research design and approach for this study are described. Details are provided regarding the organization and target population involved; the research hypotheses, procedures, and assessment tools for each of the study components; project information describing the researcher's role and the project's timeline and budget; and the limitations of this research study.

4.1 Research Design

This research study applied a mixed methods approach to evaluate a presentation intervention to raise awareness of data warehouses, to assess existing data warehouse training, and to define and use objective measures for data warehouse activities related to research outcome achievement. While the majority of the data collected within this study were quantitative in nature, some qualitative information was gathered to better understand the reasoning underlying the quantitative responses.

4.2 Organizational Background

This research study was conducted with the Clinical and Translational Science Institute (CTSI) of Southeast Wisconsin. Founded in 2010, the CTSI is a consortium of eight regional organizations including the BloodCenter of Wisconsin (BCW), Children's Hospital of Wisconsin (CHW), Clement Zablocki VA Medical Center (VA), Froedtert Hospital, Marquette University (MU), Medical College of Wisconsin (MCW), Milwaukee School of Engineering (MSOE), and University of Wisconsin-Milwaukee (UWM).[122] Awarded a $20 million CTSA in 2015, the CTSI has a mission of advancing the health of the community through research and discovery.[122]

Currently, the research portfolio includes more than 185 studies, including 47 collaborative research studies.[122] Within the CTSI, the BMI department provides consultative services, solutions for data management (such as REDCap, Confluence, and OnCore), and access to clinical data through the image de-identification service and the Clinical Research Data Warehouse (CRDW).[123] The CRDW and its use was the focus of this research study.

4.2.1 Overview of the CRDW

The CRDW is a data warehouse that contains data from a variety of sources including the Froedtert & Medical College of Wisconsin EHR (Epic), CHW EHR (Epic), MCW physician billing system (GE/IDX), Froedtert Hospital legacy systems (Affinity, Intellidose, and SIS), and the MCW tissue bank.[124] This data includes, but is not limited to, patient demographics, diagnoses, clinical encounters, lab results, medications, procedures, specimen information, enrollment in clinical trials, and information from clinical registries. Currently, the CRDW includes more than 1.9 million patient records, 524 million patient encounters, and 2.9 billion clinical facts.[124]

To access this data, the BMI department provides three self-service tools for querying the CRDW and extracting data:

- Informatics for Integrating Biology & the Bedside (i2b2) can be used to query the CRDW for developing hypotheses, determining feasibility, obtaining cohort counts for grant submissions, and defining cohorts for data analysis. All data queried through this tool is de-identified and is organized in a concept tree structure.

- TriNetX is an additional query tool that is sponsored by the pharmaceutical industry.[124] Similar to i2b2, TriNetX can be used for developing hypotheses, determining feasibility, obtaining cohort counts for grant submissions, and defining cohorts for data analysis. All data queried through this tool is de-identified and is organized by demographics, diagnoses, procedures, medications, lab results, and cohort analysis.
- Honest Broker is a data extraction tool that can use manual input or queries from i2b2 or TriNetX to extract data for an identified set of patients. The extractions may contain de-identified or identified data. While Honest Broker contains standard data tables that can be used for extraction, individuals also have the option to have custom tables created based on the needs of their studies.

4.2.2 Current CRDW Processes

Currently, the BMI department does not have a repeatable, standardized process for raising awareness of the CRDW and related tools. Individuals learn about the CRDW in a myriad of ways, such as through information on the CTSI website, conversations with other research team members, and contact with the BMI department. Upon request, BMI representatives have presented CRDW information at meetings for departments and/or research teams.

To obtain access to the CRDW tools, an individual joins the CTSI and confirms that he/she has current Collaborative Institutional Training Initiative (CITI) training for Human Subjects Research. Once these items are complete and a potential research project has been identified, the individual completes a feasibility or data release agreement identifying the principal investigator, a description of the research, and other relevant information such as the intended outcome of the

project (e.g., grant submission and abstract). Currently, access to these tools requires that the principal investigator be a member of the MCW faculty or have an adjunct faculty appointment with MCW. In addition, individuals must have an MCW domain account.

Once the agreement is completed and access is granted, the individual may begin using independently i2b2, TriNetX, and Honest Broker. In addition, the BMI Business Analyst contacts the individual to offer one-on-one or small group training. The purpose of this voluntary training is to review the data contained within the CRDW; discuss the proposed project and potential to complete the project within the CRDW; and demonstrate use of the query and extraction tools based on the project needs. Individuals are also provided contact information for the BMI Business Analyst and encouraged to contact the analyst for assistance as needed. All activity within the CRDW tools is tracked and documented using audit trails. These audit trails have been in use since the inception of the CRDW tools.

On an annual basis, the CTSI BMI department provides information to NCATS regarding achievement of program goals. Currently, this information includes number of individuals granted access, number of training sessions completed, and general information about outcomes achieved (such as the number of grant submissions completed). To provide this information, the CTSI BMI department tracks all CRDW access and training requests in a system called RISE. These requests have been tracked in RISE since 2016.

4.3 Study Overview

This research study was divided into four components. The following section provides a brief introduction to these components, a listing of the research questions and hypotheses related to

each component, a description of the target population for this study, a description of the roles and responsibilities within the study team, and an overview of the data processing and analysis performed.

4.3.1 Overview of Study Components

In the first component of this study, the awareness presentation intervention, a representative from the BMI department delivered 20-minute presentations to members of the CTSI research community. The purpose of these presentations was to raise awareness of the CRDW and its potential benefits with the goal of influencing the participants to move to the next stage in the DOI process (from knowledge to persuasion). To assess the knowledge obtained, participants completed pre- and post-assessments that measured their knowledge of the CRDW, as well as their self-reported likelihood to use the CRDW. To encourage participants to answer honestly and without fear of judgment, no identifying information (e.g., name, role/position, or department) was captured during the assessments. Increased accuracy of knowledge from pre- to post-assessment, as well as an increase in the self-reported likelihood to use would indicate progress in the DOI process to adopt the CRDW.

In the second component, the existing CRDW training process was assessed. Trainees completed pre- and post-assessments during the training session to assess the knowledge obtained and self-reported likelihood to use. In addition, trainees identified the potential outcomes for which the CRDW could be used (e.g., poster presentation, grant submission, and manuscript). To encourage trainees to answer honestly and without fear of judgment, no identifying information (e.g., name, role/position, or department) was captured during the

assessments. To assess progress from the knowledge stage of the DOI theory to the persuasion stage, the study used an increase in the accuracy of knowledge as well as an increase in self-reported likelihood to use as measurements. Additionally, the study captured the number of potential outcomes of CRDW use identified by trainees. An increase in the number of potential outcomes identified was hypothesized to indicate progress in the DOI process to adopt the CRDW.

In the third component, data from several BMI information systems was used to objectively measure outcomes achieved from the first two components. First, based on the DOI theory, if the awareness intervention was sufficient, the number of individuals requesting CRDW access and training should increase as they move from obtaining awareness-knowledge to seeking out more information about the CRDW as part of the persuasion stage. Within this study, data from the request tracking system RISE were used to obtain objective measures of these counts. In addition, if the training provided sufficient information, the majority of trainees should perform the necessary CRDW activities within 30 days of training to achieve their intended research outcomes. Currently, no objective measure exists for these activities; as a result, majority was defined as "at least 50%." Data from the feasibility agreement, data release agreement, and CRDW tool audit trails were used to track trainee activity within the CRDW to objectively measure trainee completion of the activities related to their intended research outcomes.

During the fourth and final component, trainees participated in a 30-day post-training assessment. The purpose of this assessment was to explore what factors may have influenced their decisions to adopt or reject use of the CRDW, as well as their completion of CRDW activities related to their intended research outcomes. Because these assessments were linked to

specific trainees, the study researcher provided an informational letter to each trainee indicating the confidentiality of the results so as to encourage honest responses to assessment items.

4.3.2 Research Questions and Hypotheses

These four components answer the following research questions by testing the following hypotheses:

Q1: Is delivering a 20-minute presentation regarding CRDW basic functions and processes a sufficient intervention to raise awareness of the CRDW for clinical and translational research purposes?

H1: Awareness presentations will increase the average number of correct answers from the pre-assessment to the post-assessment.

H2: Awareness presentations will increase the average self-reported level of likelihood to use from the pre-assessment to the post-assessment.

H3: The growth rate of CRDW access requests for individuals and projects will increase during the awareness presentation intervention period in comparison to the growth rate from 2016-2017 for the same time period.

H4: The growth rate of CRDW training requests for research project teams will increase during the awareness presentation intervention period in comparison to the growth rate from 2016-2017 for the same time period.

Q2: Does existing CRDW training provide sufficient information for trainees to perform the CRDW activities related to their intended research outcomes?

H5: CRDW training will increase the average number of correct answers from the pre-assessment to the post-assessment.

H6: CRDW training will increase the average number of identified potential outcomes from the pre-assessment to the post-assessment.

H7: CRDW training will increase the average self-reported level of likelihood to use from the pre-assessment to the post-assessment.

H8: Within 30 days of receiving CRDW training, at least 50% of individuals will complete the CRDW activities required to achieve their intended research outcomes.

Q3: What potential factors influence trainees' use of the CRDW for clinical and translational research purposes?

4.3.3 Target Population

The target population for this study was members of the CTSI research community. For the purposes of this study, "research community" was defined as anyone who conducts research and/or participates as a research team member. This definition included, but was not limited to, principal investigators, research coordinators, fellows, medical residents, research administrators, and graduate students. The "CTSI" qualifier referred to anyone who was employed by one of the eight organizations participating in the CTSI. For example, a research fellow employed by BCW would be within the target population, while a research fellow employed by the Aurora Health Care system would be excluded from the target population. In addition, all participants must

read, speak, write, and comprehend the English language, as that is the language used within all the CRDW tools and data.

4.3.4 Study Roles and Responsibilities

The principal investigator for this study was Katie McCarthy (identified as "study researcher" throughout the remainder of this document). The study researcher was responsible for:

- Developing the study design and protocol.
- Developing the recruitment protocols.
- Creating the assessment tools for the CRDW awareness presentation intervention, CRDW training, and 30-day post-training follow-up.
- Writing the protocol submission and obtaining approval from the Institutional Review Board (IRB).
- Performing data collection and analysis.
- Entering and maintaining the study documentation and results.
- Leading and managing all aspects of the study.

The study team also included two representatives from the BMI department, who served as subject matter experts and participated in the execution of the study protocol. Kristen Osinski is the BMI Business Analyst and served as the primary BMI support for the study researcher. With direction from the study researcher, the BMI Business Analyst was responsible for developing and delivering the CRDW awareness presentations and CRDW training sessions. She also executed the recruitment protocols for the CRDW awareness presentations and 30-day

post-training follow-up assessments and assisted with data collection. Finally, she participated in review of all the assessments created by the study researcher.

Bradley (Brad) Taylor is the Chief Research Informatics Officer within the BMI department. He served as the MCW and CTSI sponsor for this study. In this role, he approved the participation of the BMI Business Analyst and ensured the study researcher had access to the BMI tools necessary to complete the study. In addition, he assisted with recruitment for the CRDW awareness presentations and delivered some of the CRDW awareness presentations.

4.3.5 Data Processing and Analysis

The study researcher performed data preparation, cleaning, and analysis using Microsoft Excel 2016, R version 3.5.1, and RStudio version 1.1.463. For all statistical tests, a level of significance of 0.05 was used.

4.4 Study Component #1: Awareness Intervention

During this component, a member of the BMI department (the BMI Business Analyst or Chief Research Informatics Officer) delivered a 20-minute presentation to a group of individuals from the target population. During the presentation, participants completed pre- and post-assessments to determine the knowledge obtained from the presentation content.

Should this presentation intervention prove sufficient for raising awareness of the CRDW, the BMI department plans on implementing this intervention as a standard process. Based on this plan, the minimum sample size for this component mirrored the 3-4 presentations per month that

could be sustained by the BMI department as a standard process. Because the study period was three months, the goal was completion of 9-12 presentations during the study period.

During refinement of the procedure and assessment tools for this component, an average of seven individuals participated in each presentation, and 82% of these individuals participated in the assessments. Using this information as a basis for determining minimum sample size, if group sizes are 6-8 individuals and 75% participate in the assessment, the potential number of completed assessments is 40-72 assessments. As a result, the minimum sample size for this component is 40 completed pre-assessments and associated post-assessments.

Based on the recruitment protocol developed by the study researcher, a BMI department representative (Business Analyst or Chief Research Informatics Officer) contacted leaders identified as primary contacts for each CTSI partner organization to determine interest in scheduling awareness presentations for groups within their organizations. When interest was expressed by a leader, the BMI representative worked with that leader to schedule an appropriate date, time, and location and obtained an estimate of the number of potential attendees. As part of the recruitment protocol, the BMI representative also obtained verbal permission to conduct pre- and post-assessments during the scheduled session. If permission was obtained, the BMI representative contacted the study researcher regarding the date and time of the session, location of the session, and number of potential attendees. The study researcher used this information to prepare the appropriate number of pre- and post-assessment copies for the session.

4.4.1 Research Procedure

At the start of the scheduled session, the study researcher explained the purpose of this study and pre- and post-assessments, as well as the voluntary nature of the study and the anonymity of the results. To individuals who consented to participate, the study researcher then provided the pre-assessment and an envelope to in which to store the completed pre-assessment. Once the pre-assessments were complete, the BMI representative delivered the presentation and answered any questions that were posed. At the end of the session, the study researcher provided each individual who completed the pre-assessment with a copy of the post-assessment and instructed the individual to store this post-assessment in the same envelope as the pre-assessment. Once the assessments were completed, the study researcher collected all the envelopes. Finally, the number of individuals attending the session and the number of pre- and post-assessment pairs completed were documented.

Only presentations that covered the standard CRDW awareness presentation and conformed to this protocol were included within this study component. For example, if a department requested a CRDW presentation that only described one of the CRDW tools, then this presentation was not the standard CRDW awareness presentation and was not counted as a presentation for the purpose of this study component.

4.4.2 Content Development

Based on factors identified by the TAM and D&M IS Success Model as important for technology adoption, the study researcher identified learning objectives for the CRDW awareness presentation intervention. Specifically, perceived usefulness and task compatibility

were identified as strong determinants; therefore, the presentation objectives focused on key elements related to research work that are supported by the CRDW. These objectives are listed below.

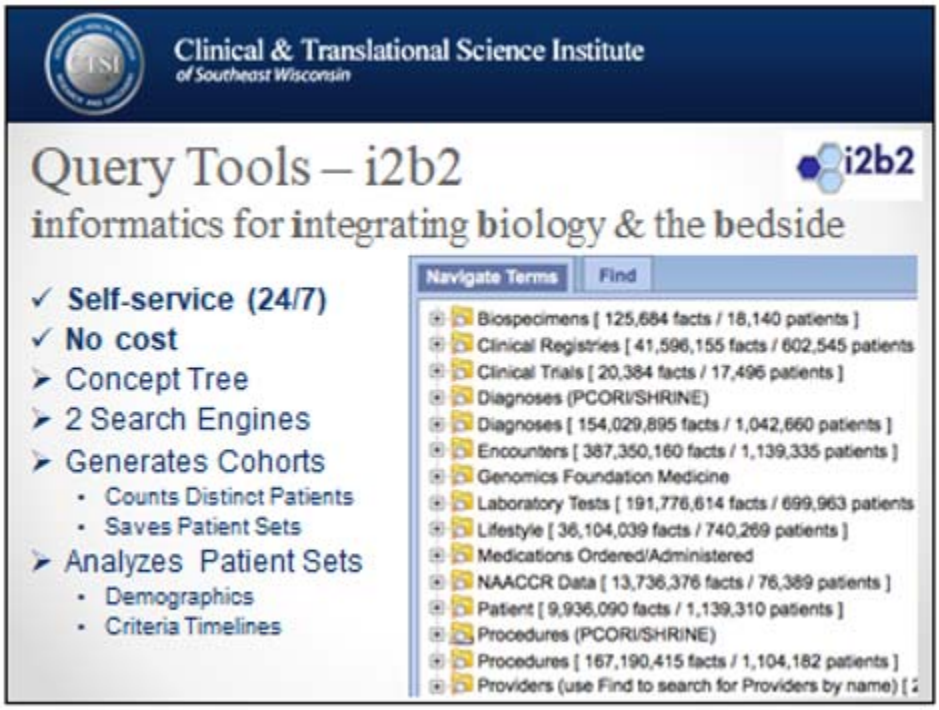
At the completion of this presentation, attendees should be able to:

- Identify the standard data elements available within the CRDW.
- Identify the situations where use of CRDW data requires Institutional Review Board (IRB) approval.
- Identify the BMI services that are provided free-of-charge related to the CRDW.

Based on these objectives and direction from the study researcher, the BMI Business Analyst developed the content of the awareness presentation. In addition, similar presentations have been delivered previously by the BMI department, so experience gained from these presentations was used to inform the structure and flow of this new presentation.

Below is an example of one of the content slides developed for the presentation.

Figure 3. Example Slide from Awareness Presentation Intervention



4.4.3 Assessment Tool Development

Using Bloom's taxonomy, the study researcher classified the presentation objectives as *factual* and *conceptual* knowledge at the *remember* cognitive level. Using these dimensions as a guide, the study researcher created an assessment tool with a question related to each objective. The goal of these questions was to determine if the learning objectives for the presentation were achieved. A final question was developed to determine the individual's self-reported likelihood to use the CRDW, with a follow-up question if "unlikely" or "very unlikely" was selected to determine the reason for this selection. The goal of this question was to obtain an indication of

the individual's attitude toward the CRDW (a key component of DOI theory and the TAM model) and to gain some qualitative information for anyone indicating a low likelihood of use.

4.4.4 Refinement of Procedure and Assessment Tool

To refine the awareness presentation content, assessment, and procedure, the study researcher and BMI Business Analyst delivered two presentations to different departments within the MCW research community. A total of 17 individuals participated in these presentations, with 14 individuals completing the assessments (82%). Three individuals did not participate as they arrived after the pre-assessments were completed. While the research procedure outlined above was followed, the data captured on the assessments was not retained (as the sole purpose of these sessions was to refine the content, tools, and procedure).

During this process, the study researcher identified the following improvements:

- The original presentation contained too much content to be covered in 20 minutes. The study researcher and BMI Business Analyst modified the content to meet the identified time constraints. A copy of the final presentation is provided in Appendix B: Awareness Presentation Content.
- Because the pre- and post-assessments contain the exact same questions, the study researcher copied pre-assessments on one color paper, while the post-assessments were copied on a different color. This process proved helpful in distinguishing between the two assessments when they were completed and placed within the same envelope. This process was continued during the research study.

- The study researcher reworded the three questions related to the presentation learning objectives to reduce ambiguity and better align with the presentation.
- For the question relating to likelihood to use the CRDW, several individuals completed the follow-up question even though they indicated a positive likelihood to use the CRDW. The study researcher changed the formatting of the likelihood question and follow-up question to more clearly indicate when the follow-up question should be completed.

In addition to these sessions, the study researcher had the assessment tool reviewed by an individual with expertise in survey design. This review resulted in some minor wording changes, but no substantial content changes. A copy of the revised assessment is provided in Appendix C: Awareness Presentation Assessment Tool.

4.4.5 Data Processing and Analysis Procedure

The purpose of study component #1 was to test the following hypotheses:

H1: Awareness presentations will increase the average number of correct answers from the pre-assessment to the post-assessment.

H2: Awareness presentations will increase the average self-reported level of likelihood to use from the pre-assessment to the post-assessment.

To test hypothesis H1, the answers from the first three questions on the awareness presentation assessments were analyzed. The first three questions are shown below.

Figure 4. First Three Questions from Awareness Presentation Assessment

1) Which of the following information is currently available in the standard tables within the clinical research data warehouse (CRDW) and can be queried using both i2b2 and TriNetX? Please check all that apply.

- ☐ Patient race and ethnicity
- ☐ Progress notes
- ☐ Lab tests
- ☐ Flowsheet data
- ☐ Diagnosis
- ☐ Medications

2) To use the CRDW to access de-identified information, you need to obtain IRB approval. Please select one answer.

- ☐ True
- ☐ False

3) Which of the following Biomedical Informatics (BMI) services do you think you will be charged for? Please check all that apply.

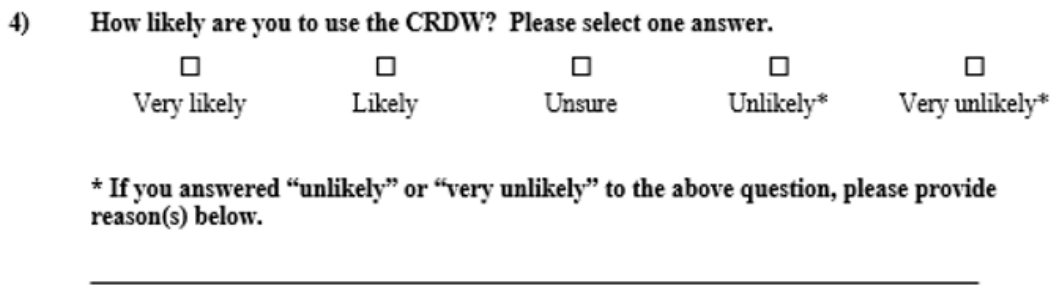
- ☐ One-on-one training for CRDW tools
- ☐ Creation of custom tables within Honest Broker
- ☐ General support from BMI analysts (such as answering questions)
- ☐ Support for creating queries with standard CRDW tools

Because question #1 had six possible answers, question #2 had one possible answer, and question #3 had four possible answers, the maximum number of potential correct answers for one assessment was eleven. Each assessment was reviewed to determine the number of correct answers. For example, if an individual correctly identified five answers for question #1, zero answers for question #2, and three answers for question #3, the individual's correct answer score was 8 (5 + 0 + 3).

The mean and standard deviation for the total number of correct answers were calculated for the pre- and post-assessments. Bar charts and boxplots were used to visualize the results from these questions as well as the overall number of correct answers from this section of the assessment. Then, the study researcher tested the normality of the data using Q-Q plots and the results of the Shapiro-Wilk normality tests. Based on the results of these tests, the study researcher performed the appropriate parametric or non-parametric hypothesis testing method.

To test hypothesis H2, the answers from the fourth question on the awareness presentation assessments were analyzed. The fourth question is shown below.

Figure 5. Fourth Question from Awareness Presentation Assessment



To process this data, each level of likelihood was assigned a numeric equivalent ("Very Likely" was 4, "Likely" was 3, "Unsure" was 2, "Unlikely" was 1, and "Very Unlikely" was 0). The mean and standard deviation for the level of likelihood were calculated for the pre- and post-assessments. Bar charts and boxplots were used to visualize the results from this question. Because this assessment data uses an ordinal scale of measurement, hypothesis testing was performed using a non-parametric method (Wilcoxon matched-pairs signed-ranks test).

4.5 Study Component #2: Assessment of CRDW Training

This study component assessed the existing CRDW training process to determine if sufficient information was provided for trainees to complete the CRDW activities related to their intended research outcomes. During this component, the BMI Business Analyst delivered one hour of CRDW training to one or more members of a research team. Because research teams often contain individuals with unique skills and work is delegated based on these skills, trainees were not required to have attended an awareness presentation prior to training. For example, a principal investigator may have attended an awareness presentation, identified the potential uses

for the CRDW in his/her research, and then selected one or more research team members to receive CRDW training who had not attended the awareness presentation.

Because individuals receiving CRDW training may not have participated in the awareness presentations, the CRDW training provided the same foundational knowledge covered in the awareness presentation, but added hands-on practice using the CRDW for a research project identified by the individuals receiving the training. During the training, pre- and post-assessments were completed to determine if sufficient information was provided.

Because one of the hypotheses to be tested within this study was whether or not the number of requests for CRDW training increases, the awareness presentation was the only recruitment used for this component. Due to the anonymity of the participants in the awareness presentations, this study component did not link those attending awareness presentations to those attending training sessions. Based on experience gained during the refinement of the procedure and assessment tool for this intervention, the goal was to have 75% of trainees complete a pre-assessment and associated post-assessment.

4.5.1 Research Procedure

This intervention began when one or more members of a research team entered a CRDW feasibility or data release agreement. Upon receipt of this agreement, the BMI Business Analyst followed the current CRDW training process and contacted the requestor to determine if the requestor (and associated research team members) wanted to schedule CRDW training.

If interest was expressed, the BMI Analyst completed the following tasks within the current CRDW training process:

- Worked with the requestor to schedule an appropriate date, time, and location and obtained the names of the individuals that would be attending the training.
- Used the list of names to ensure each trainee had access to the CRDW.
- Used the information from the agreement to determine the feasibility of completing the indicated research project with the CRDW and prepared for the training demonstration.

During the scheduled session, the BMI Analyst followed the study protocol and explained the purpose of this research study and the pre- and post-assessments, as well as indicated that participation in the study was voluntary and the results were anonymous. Trainees who consented to participate were provided the pre-assessment and an envelope in which to store the completed pre-assessment. Once the pre-assessments were complete, the BMI Business Analyst provided the training, allowed the trainees to try hands-on activities within the CRDW, and answered any questions that were posed. At the end of the session, each trainee who completed the pre-assessment was provided a copy of the post-assessment and asked to store this post-assessment in the same envelope as the pre-assessment. Once complete, all the envelopes were collected by BMI Business Analyst and then provided to the study researcher.

Only training sessions that covered the standard CRDW training content and conformed to this protocol were included within this study component. For example, if a researcher requested refresher training focusing on a specific area of confusion, then this training session did not cover the standard CRDW training content and was not counted as a training session for the purpose of this study component.

4.5.2 Content Development

The BMI Business Analyst developed the existing CRDW training and has used this training approach and content for approximately two years. As a result, no new content was developed. The study researcher reviewed the existing content to ensure it provided the same CRDW foundational knowledge as the awareness presentation. Based on this review, no revisions were identified. Because the hands-on activities are based on the trainee's potential research project, no standard content was available for this portion of the training.

4.5.3 Assessment Tool Development

Trainees participating in this component may or may not have attended a CRDW awareness presentation. For example, a research team manager may have attended an awareness presentation and encouraged his/her research staff (who did not attend the presentation) to obtain CRDW training. For this reason, the CRDW training provided the same foundational knowledge as the awareness presentation to ensure that all trainees obtained the same base knowledge.

Because the CRDW training covered the same foundational knowledge as the awareness presentation, the same three learning objective questions from the presentation assessment were included on the training assessment tool. The goal of these questions was to determine if the learning objectives for the training were achieved.

Because DOI theory indicates that individuals in the persuasion stage of technology adoption will mentally consider ways to apply the technology to their current and future needs, the study researcher developed a new question to identify the trainee's intended outcomes in using of the

CRDW. Currently, the feasibility and data release agreement forms for the CRDW include a question that asks the individual to indicate his/her intended outcomes in using the CRDW. The options that the individual may select are protocol, poster presentation, grant submission, abstract, manuscript, and other. As a result, these same intended outcome options were used for the new assessment question developed by the study researcher. The goal of this question was to identify a trainee's intended use of the CRDW before and after training.

Finally, the same question from the awareness presentation assessment related to an individual's self-reported likelihood to use the CRDW was included. The goal of this question was to obtain an indication of the individual's attitude toward the CRDW (a key component of DOI theory and the TAM model) and gain some qualitative information for anyone indicating a low likelihood of use.

4.5.4 Refinement of Procedure and Assessment Tool

To refine the training assessment and procedure, the BMI Business Analyst delivered two training sessions to different research project teams within the MCW research community and provided the results to the study researcher. A total of four individuals participated in these presentations, with four individuals completing the assessments (100%). While the research procedure outlined above was followed, the data captured on the assessments was not retained (as the sole purpose of these sessions was to refine the tool and procedure).

During this process, the study researcher identified the following improvements:

- Because the pre- and post-assessments contain the exact same questions, the study researcher copied pre-assessments on one color paper, while the post-assessments were copied on a different color. This process proved helpful in distinguishing between the two assessments when they were completed and placed within the same envelope. This process was continued during the research study.
- The study researcher reworded the three questions related to the training learning objectives to reduce ambiguity and better align with the training. The changes made were aligned with the changes identified during the awareness presentation refinement process.
- For the question relating to likelihood to use the CRDW, several individuals completed the follow-up question even though they indicated a positive likelihood to use the CRDW. The study researcher changed the formatting of the likelihood question and follow-up question to more clearly indicate when the follow-up question should be completed. The changes made were aligned with the changes identified during the awareness presentation refinement process.

In addition to these sessions, the study researcher had the assessment tool reviewed by an individual with expertise in survey design. This review resulted in some minor wording changes, but no substantial content changes. A copy of the revised assessment is provided in Appendix D: CRDW Training Assessment Tool.

4.5.5 Data Processing and Analysis Procedure

The purpose of study component #2 was to test the following hypotheses:

H5: CRDW training will increase the average number of correct answers from the pre-assessment to the post-assessment.

H6: CRDW training will increase the average number of identified potential outcomes from the pre-assessment to the post-assessment.

H7: CRDW training will increase the average self-reported level of likelihood to use from the pre-assessment to the post-assessment.

To test hypothesis H5, the answers from the first three questions on the training assessments were analyzed. The first three questions are shown below.

Figure 6. First Three Questions from CRDW Training Assessment

1) Which of the following information is currently available in the standard tables within the clinical research data warehouse (CRDW) and can be queried using both i2b2 and TriNetX? Please check all that apply.

- ☐ Patient race and ethnicity
- ☐ Progress notes
- ☐ Lab tests
- ☐ Flowsheet data
- ☐ Diagnosis
- ☐ Medications

2) To use the CRDW to access de-identified information, you need to obtain IRB approval. Please select one answer.

- ☐ True
- ☐ False

3) Which of the following Biomedical Informatics (BMI) services do you think you will be charged for? Please check all that apply.

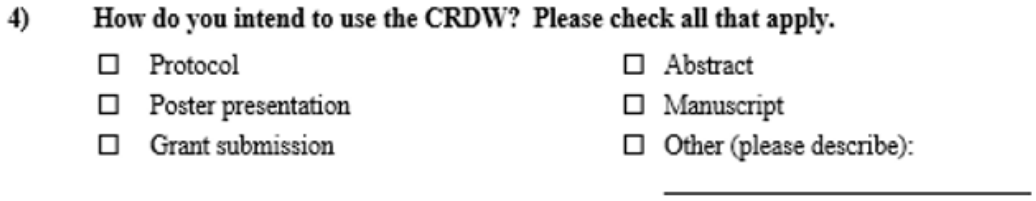
- ☐ One-on-one training for CRDW tools
- ☐ Creation of custom tables within Honest Broker
- ☐ General support from BMI analysts (such as answering questions)
- ☐ Support for creating queries with standard CRDW tools

Because question #1 had six possible answers, question #2 had one possible answer, and question #3 had four possible answers, the maximum number of potential correct answers for one assessment was eleven. Each assessment was reviewed to determine the number of correct answers. For example, if an individual correctly identified five answers for question #1, zero answers for question #2, and three answers for question #3, the individual's correct answer score was 8 (5 + 0 + 3).

The mean and standard deviation for the total number of correct answers were calculated for the pre- and post-assessments. Bar charts and boxplots were used to visualize the results from these questions as well as the overall number of correct answers from this section of the assessment. Then, the study researcher tested the normality of the data using Q-Q plots and the results of the Shapiro-Wilk normality tests. Based on the results of these tests, the study researcher performed the appropriate parametric or non-parametric hypothesis testing method.

To test hypothesis H6, the answers from the fourth question on the training assessments were analyzed. The fourth question is shown below.

Figure 7. Fourth Question from CRDW Training Assessment



A total of six potential outcomes could be selected by the trainee. The mean and standard deviation for the number of potential outcomes were calculated for the pre- and post-assessments. Bar charts and boxplots were used to visualize the results from this question.

Then, the study researcher tested the normality of the data using Q-Q plots and the results of the Shapiro-Wilk normality tests. Based on the results of these tests, the study researcher performed the appropriate parametric or non-parametric hypothesis testing method.

To test hypothesis H7, the answers from the fifth question on the training assessments were analyzed. The fifth question is shown below.

Figure 8. Fifth Question from CRDW Training Assessment

4) How likely are you to use the CRDW? Please select one answer.

☐ Very likely ☐ Likely ☐ Unsure ☐ Unlikely* ☐ Very unlikely*

*** If you answered "unlikely" or "very unlikely" to the above question, please provide reason(s) below.**

_____

To process this data, each level of likelihood was assigned a numeric equivalent ("Very Likely" was 4, "Likely" was 3, "Unsure" was 2, "Unlikely" was 1, and "Very Unlikely" was 0). The mean and standard deviation for the level of likelihood were calculated for the pre- and post-assessments. Bar charts and boxplots were used to visualize the results from this question. Because this assessment data uses an ordinal scale of measurement, hypothesis testing was performed using a non-parametric method (Wilcoxon matched-pairs signed-ranks test).

4.6 Study Component #3: Objective Measurement of Outcomes

The third study component involved obtaining data from several BMI information systems to objectively measure outcomes achieved from the first two components. As this component involves review of existing data, no recruitment procedures or goals were established for this component.

4.6.1 Research Procedure

The third component involved the study researcher obtaining data from several BMI information systems to objectively measure outcomes achieved from the first two components. Specifically, the following measures were obtained:

- The number of CRDW access requests by individual and research project.
- The number of CRDW training requests by research project team.
- The number of trainees performing the CRDW activities related to their intended research outcomes.

Currently, all requests for access and training for the CRDW are tracked within a system known as RISE. These data have been tracked since 2016. To determine if the awareness intervention was sufficient, the study researcher obtained RISE data for the number of requests for CRDW access and training during the study time period. The study researcher used this data to determine if the intervention had increased the number of requests in comparison to prior years.

To obtain the measurement of trainees performing the CRDW activities related to their intended research outcomes, the first step was reviewing the list of trainees who completed training during the study period. The study researcher linked the trainees to their intended outcomes (e.g., protocol, poster presentation, and grant submission) identified in the completed feasibility or data release agreements. In addition, the study researcher requested CRDW audit trail data for each of the trainees. The study researcher reviewed the audit trail data to determine if trainees completed the appropriate actions within the CRDW to achieve their intended research outcomes. For example, if the intended outcome was a grant submission, minimally, the

researcher would likely use the CRDW for a potential cohort of patients that could be included in the study. For this use case, completion of a query within the CRDW was required to achieve the intended research outcome. Individuals who completed the necessary CRDW actions within 30 days of receiving training were identified as successfully completing the CRDW activities. This timeline aligned with the 30-day post-training assessment; as a result, no CRDW activities more than 30 days post-training were included as part of this study.

4.6.2 Refinement of Procedure

To determine if the objective measures for CRDW access and training requests were feasible, the study researcher obtained a sample of data from the RISE system for 2016, 2017, and 2018. Review of this sample confirmed that the necessary data for the applicable timeframes were available in RISE and could be used to obtain the number of requests and growth rates for analysis purposes. Because the existing request documentation provided the necessary information for this study and to maintain consistency with historical data, no changes were made to this tracking for this study.

To determine if the objective measure for CRDW activities was feasible, the study researcher obtained a sample of data for feasibility agreements, data release agreements, and audit trails from 2016, 2017, and 2018. Review of this sample confirmed that the necessary data for the applicable timeframes were available and could be used to obtain the number of trainees that completed the CRDW activities related to their intended research outcomes. Because the existing request documentation provided the necessary information for this study and to maintain

consistency with historical data, no changes were made to the CRDW feasibility agreements, data release agreements, and audit trails for this study.

4.6.3 Data Processing and Analysis Procedure – Access Requests

The purpose of measuring CRDW access requests was to test the following hypothesis:

H3: The growth rate of CRDW access requests for individuals and projects will increase during the awareness presentation intervention period in comparison to the growth rate from 2016-2017 for the same time period.

To test hypothesis H3, the CRDW access requests for the three-month study period (July 30 through October 30) in 2016, 2017, and 2018 were exported from the RISE system into a Microsoft Excel file. Within this data file, the key fields included the date of the request, the name and email of the principal investigator involved in the request, the project ID and name associated to the request, the name of the individual for whom access should be granted, and the type of tool for which access was granted. Individuals could request access to one or both types of CRDW tools, and each type of tool was documented as a separate request. As a result, an individual could have one or two records within the access dataset. For example, if an individual requested access to the cohort tools (i2b2 and TriNetX), one record would exist for that individual's request. If the individual also requested access to Honest Broker, a second record would exist for that individual.

To process this data, the study researcher first reviewed the file to identify any blanks, duplicated requests, or obvious errors (such as records that were outside of the study period). If any issues

were found, they were documented and resolved by the researcher. Next, the researcher identified the number of unique individuals requesting access. If an individual had two requests (one for the cohort tools and one for Honest Broker), these records were combined to avoid double-counting the individuals. The researcher also identified the number of unique projects associated to the access requests. If a project ID was associated to more than one request (indicating multiple individuals requested access as part of that project), these requests were combined to avoid over counting an individual project.

Once the study researcher identified unique records, the number of access requests by individual and by project were counted for each year (2016, 2017, and 2018). The growth rates for 2016-2017 and 2017-2018 were calculated using the formula:

$$Growth\ Rate\ (\%) = \frac{Year\ 2\ Count - Year\ 1\ Count}{Year\ 1\ Count}\ x\ 100$$

The study researcher compared the calculated growth rates for 2016-2017 and 2017-2018 to determine if the hypothesized growth occurred.

4.6.4 Data Processing and Analysis Procedure – Training Requests

The purpose of measuring CRDW training requests was to test the following hypothesis:

H4: The growth rate of CRDW training requests for research project teams will increase during the awareness presentation intervention period in comparison to the growth rate from 2016-2017 for the same time period.

To test hypothesis H4, CRDW training requests for the three-month study period (July 30 through October 30) in 2016, 2017, and 2018 were exported from the RISE system into a

Microsoft Excel file. Within this data file, the key fields included the date of the request, the name and email of the principal investigator involved in the request, and the project ID and name associated to the request. Each request within this dataset could involve one or more individuals from a project team. In addition, a project team may request multiple training sessions. As a result, the dataset could include multiple records for the same project.

To process this data, the study researcher first reviewed the file to identify any blanks or obvious errors (such as records that were outside of the study period). If any issues were found, they were documented and resolved by the researcher. Next, the researcher identified the number of unique projects that had requested training. If a project ID was associated to more than one request, these requests were combined to avoid over counting an individual project.

Once the study researcher identified unique records, the number of training requests by project were counted for each year (2016, 2017, and 2018). The growth rates for 2016-2017 and 2017-2018 were calculated using the formula:

$$Growth\ Rate\ (\%) = \frac{Year\ 2\ Count - Year\ 1\ Count}{Year\ 1\ Count}\ x\ 100$$

The study researcher compared the calculated growth rates for 2016-2017 and 2017-2018 to determine if the hypothesized growth occurred.

4.6.5 Data Processing and Analysis Procedure – Outcomes

The purpose of measuring CRDW trainee completion of CRDW activities related to their intended research outcomes was to test the following hypothesis:

H8: Within 30 days of receiving CRDW training, at least 50% of individuals will complete the CRDW activities required to achieve their intended research outcomes.

To test hypothesis H8, a mapping table was created linking the intended research outcomes identified on research teams' Data Release Agreements with the appropriate actions within the CRDW. Based on the expertise of the BMI Business Analyst, the following mapping table was created.

Table 10. Mapping Table for Intended Research Outcomes and CRDW Actions

Intended Outcome	CRDW Actions
Protocol	Query (in i2b2 and/or TriNetX)
Poster Presentation	Query (in i2b2 and/or TriNetX) and Extract (from Honest Broker)
Grant Submission	Query (in i2b2 and/or TriNetX)
Abstract	Query (in i2b2 and/or TriNetX) and Extract (from Honest Broker)
Manuscript	Query (in i2b2 and/or TriNetX) and Extract (from Honest Broker)

"Other" was also an option as an intended outcome. The BMI Business Analyst reviewed any agreements in which this option was selected and determined the appropriate CRDW actions based on other information provided within the agreement.

With the mapping complete, the next step was to obtain audit trail data from i2b2, TriNetX, and Honest Broker for each of the trainees who participated in study component #2 (assessment of CRDW training). For each trainee, a listing of his/her activities within 30 days of his/her

training was aggregated from the separate CRDW tool audit trails, and then the listing was compared to the appropriate CRDW actions based on intended research outcome. Trainees that completed the mapped CRDW actions within 30 days of his/her training were identified as successfully completing the CRDW activities.

4.7 Study Component #4: 30-day Post-training Assessment

The fourth and final component was designed to explore what factors may have influenced individuals' decisions to adopt or reject use of the CRDW, as well as their achievement of outcomes. The study researcher contacted volunteers identified at the end of the CRDW training sessions and asked them to complete a 30-day post-training assessment. This assessment gathered quantitative and qualitative information about individuals' perceptions of CRDW use, which was used to explore potential reasons for CRDW adoption or rejection, as well as completion or lack of completion of the CRDW activities related to their intended research outcomes.

Recruitment for this study component occurred at the end of the CRDW training sessions. Following the recruitment protocol, at the end of each CRDW training session, the BMI Business Analyst asked each trainee if he/she was willing to participate in a 30-day post-training assessment. If the trainee declined, the BMI Business Analyst documented the trainee's name and date of training and indicated that the trainee would not participate. This documentation ensured the trainee was not inadvertently contacted for the last component of this study and that a complete list of trainees was available at the end of the study. If the trainee consented, the BMI Business Analyst documented the trainee's name and date of training and indicated the

trainee's preferred follow-up method (in-person interview, phone interview, or email with an electronic survey) and contact information. The goal was to have 10% of trainees complete the 30-day post-training assessment. This goal was based on previous response rates for similar interview and survey data collection procedures.

4.7.1 Research Procedure

During the study, the study researcher accessed the list of trainees. For trainees who consented to be contacted for a 30-day post-training assessment, the study researcher contacted each trainee based on the identified preferred follow-up method.

For individuals who indicated a preference for an electronic survey, a link to the survey was sent to the individual via email. The first email contact occurred within 30-38 days post-training. If the survey was not completed within one week after the first contact, one additional follow-up email was sent to remind the trainee to complete the survey.

For individuals who indicated a preference for an in-person or phone interview, the study researcher contacted the trainee using his/her preferred contact method to schedule an appropriate date, time, and location (if needed). The first contact occurred within 30-38 days post-training. If needed based on lack of response, a second contact occurred one week after the initial contact. For these interviews, the survey researcher asked the same questions that were provided on the electronic survey. Trainees were provided the response scales for each question to ensure consistency in responses.

All responses completed within 30-45 days post-training were included in this study component. In addition, responses to the interviews and electronic surveys were combined into a single dataset for review and analysis.

4.7.2 Assessment Tool Development

The goal of the 30-day post-training assessment was to gather quantitative and qualitative information about individuals' perceptions of CRDW use, which could be used to explore potential reasons for achievement or lack of achievement of outcomes. As a result, the study researcher developed an assessment tool to gather information about key components the technology adoption frameworks identify as strong determinants of use.

In the first section of the assessment tool, the individual was asked if he/she used the CRDW. If the answer to this question was yes, further questions asked about the activities performed, the perceptions of ease of use of the CRDW tools, and how the individual used the data (or plans to use the data). The goal of this section was to document actual use, perceived ease of use, and perceived usefulness of the CRDW.

The second section of the assessment tool determined if the individual requested support from the BMI department regarding the CRDW. Because support was considered an important factor in technology adoption, the goal of this question was to understand if individuals were using the available support.

The third section of the assessment tools asked for individual perceptions of the potential benefits or barriers to using the CRDW. Previous technology adoption framework research

provided example survey items to measure factors influencing technology adoption. The study researcher adapted these items to reference the CRDW and related tools. The goal of this section was to gather information that could be used to explore potential reasons for completion or lack of completion of the CRDW activities related to their intended research outcomes.

The fourth section addressed an individual's self-reported likelihood to use the CRDW, which was similar to the question that was included in the first two study component assessments. The purpose of this question was to obtain an indication of the individual's attitude toward the CRDW (a key component of DOI theory and the TAM model) and gain some qualitative information for anyone indicating a low likelihood of use.

The final section allowed the user to provide any additional comments about his/her experience with the CRDW. The goal of this section was to capture information that individuals wanted to provide but were not covered in the previous sections.

The study researcher documented this assessment tool in a Microsoft Word template that could be used to record the results from in-person and phone interviews. Additionally, the tool was built in an electronic survey tool (REDCap) for the emailed surveys.

4.7.3 Refinement of Procedure and Assessment Tool

The study researcher had the content of the assessment tool reviewed by an individual with expertise in survey design. This review resulted in some minor wording changes, as well as a format change to the third section. The written and electronic forms for the assessment were also reviewed with the BMI Business Analyst for content and flow. This review resulted in some

minor wording changes. No issues with survey flow were identified. A copy of the revised assessment is provided in Appendix E: 30-day Post-training Assessment Tool.

Because the study researcher has conducted similar research procedures for other projects, no additional refinement activities were performed for this research procedure.

4.7.4 Data Processing and Analysis Procedure

The purpose of study component #4 was to investigate potential factors that influence trainees' use of the CRDW for clinical and translational research purposes, specifically within 30-45 days of their original training.

The first section of the assessment gathered information about individual use of the CRDW. The questions in this section are shown below.

Figure 9. 30-day Post-training Assessment – Section 1 Questions

Questions	
Since receiving clinical research data warehouse (CRDW) training, have you used any of the CRDW tools (i2b2, TriNetX, or Honest Broker)?	○ Yes ○ No

If the answer to the previous question is Yes, the following additional questions appear:

What activities did you complete using the CRDW tools? (Please check all that apply.)	☐ Obtain a count of the number of patients in a particular cohort. ☐ Analyze a patient set in order to see a demographic distribution. ☐ Extract data tables for a patient set.

For each CRDW tool listed below, rate its ease of use.	Very Easy to Use	Easy to Use	Difficult to Use	Very Difficult to Use
i2b2	○	○	○	○
TriNetX	○	○	○	○
Honest Broker	○	○	○	○

How have you used (or intend to use) the data (e.g., patient count or data tables) obtained form the CRDW? (Please check all that apply.)	☐ Used, or plan to use, for cohort discovery ☐ Used, or plan to use, for hypothesis development ☐ Used, or plan to use, for a prospective study ☐ Used, or plan to use, for a retrospective study ☐ Other (please specify)

To process and analyze the data from the use of the CRDW question, the study researcher obtained a count of each response and calculated the percentage of participating trainees indicating they had used the CRDW since their training sessions.

If the trainee indicated that he/she had used the CRDW, then he/she would see the subsequent three questions. For the first and third subsequent question, trainees could select more than one option. As a result, the study researcher counted the number of responses for each option and calculated the percentage of participating trainees indicating each option.

For the second subsequent question in this section, the number of options for answers were collapsed to provide directional information due to the relatively small sample size in relation to the number of options provided as answers. The options for rating of the ease of use of CRDW tools included "Very easy to use," "Easy to use," "Difficult to use," and "Very difficult to use." These ease of use ratings were collapsed into "Easy to use" (for "Very easy to use" and "Easy to use") and "Difficult to use" (for "Difficult to use" and "Very difficult to use") to provide a clearer summary of assessment answers. The study researcher counted the responses for each rating and then determined any directional trends within the responses.

The second section of the assessment gathered information about requests for CRDW support. The questions in this section are shown below.

Figure 10. 30-day Post-training Assessment – Section 2 Questions

Did you contact the Biomedical Informatics (BMI) department for support with the CRDW?	○ Yes ○ No

If the answer to the previous question is Yes, the following additional question appears:

Who did you contact for support within the BMI department?	_____

To process and analyze the data from the CRDW support question, the study researcher obtained a count of each response and calculated the percentage of participating trainees indicating they had requested CRDW support since training. For the trainees indicating that they had requested support, they would see the subsequent question. The study researcher reviewed these responses to determine which individuals were contacted for CRDW support.

The third section of the assessment obtained individual perceptions of the potential benefits and barriers to using the CRDW. The questions in this section are shown below.

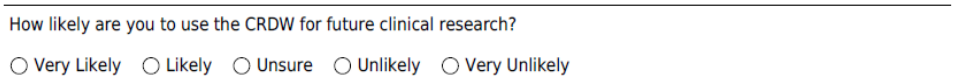
Figure 11. 30-day Post-training Assessment – Section 3 Questions

For the statements below, indicate your level of agreement.	Strongly Agree	Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree	Strongly Disagree
The training I received was sufficient for me to use the CRDW effectively.	○	○	○	○	○
I lacked the necessary support to use the CRDW effectively.	○	○	○	○	○
Overall, I found the CRDW easy to use.	○	○	○	○	○
I lacked the time to effectively use the CRDW.	○	○	○	○	○
The data I needed was not available within the CRDW.	○	○	○	○	○
Using the CRDW was beneficial to my work.	○	○	○	○	○

To analyze the data from this section of the assessment, the study researcher collapsed the number of options for answers to provide directional information due to the relatively small sample size in relation to the number of options provided as answers. The options included "Strongly Agree," "Somewhat Agree," "Neither Agree nor Disagree," "Somewhat Disagree," and "Strongly Disagree." These options were collapsed into "Agree" (for "Strongly Agree" and "Somewhat Agree"), "Neutral" (for "Neither Agree nor Disagree"), and "Disagree" (for "Somewhat Disagree" and "Strongly Disagree") to provide a clearer summary of assessment answers. The study researcher counted the responses for each rating and then determined any directional trends within the responses. The results were also visualized using bar charts organized by positively and negatively worded statements.

The fourth section of the assessment obtained an individual's self-reported likelihood to use the CRDW. The question in this section is shown below.

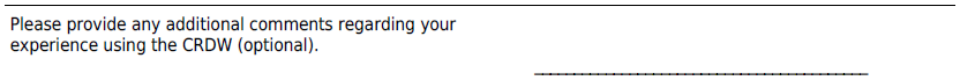
Figure 12. 30-day Post-training Assessment – Section 4 Question



To analyze the data from this section of the assessment, the study researcher collapsed the number of options for answers to provide directional information due to the relatively small sample size in relation to the number of options provided as answers. The options included "Very Likely," "Likely," "Unsure," "Unlikely," and "Very Unlikely." These options were collapsed into "Likely" (for "Very Likely" and "Likely"), "Unsure" (for "Unsure"), and "Unlikely" (for "Unlikely" and "Very Unlikely") to provide a clearer summary of assessment answers. The study researcher counted the responses for each rating and then determined any directional trends within the responses.

The fifth and final section of the assessment allowed the individual to provide any additional comments about his/her experience with the CRDW. The question in this section is shown below.

Figure 13. 30-day Post-training Assessment – Section 5 Question



The study researcher reviewed any responses from this section and included them as part of the analysis for the appropriate section. For example, if the individual included a comment that he/she was unable to use the CRDW since training, the study researcher reviewed this comment in the context of the other assessment answers. Due to the relatively small sample size for this study component, no thematic analysis was performed based on the comments from this section of the assessment.

4.8 Record Retention

For the pre- and post-assessments, as well as the 30-day post-training assessment, the study researcher entered the results (answers to each assessment or follow-up question) into the REDCap system hosted at Medical College of Wisconsin. REDCap (Research Electronic Data Capture) is a secure, web-based application designed to support data capture for research studies, providing 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources.[125,126] In addition, the study researcher entered the results into a Microsoft Excel spreadsheet stored on an encrypted hard drive. This second copy served as confirmation of the

REDCap data entry and provided a backup should the REDCap data be lost or compromised. The study researcher retained the original paper records (such as completed pre- and post-assessments) in a secure location.

Additional electronic documentation generated during the study was stored in an MCW box.com folder shared by the BMI Business Analyst and the study researcher. This additional documentation could include, but was not limited to, data reports for the objective measurements study component and backups of any Microsoft Excel spreadsheets used for data capture and/or analysis. For backup purposes, the study researcher regularly copied these files to an encrypted external hard drive owned by the researcher.

4.9 Researcher Role and Experience

For this project, the study researcher's role was as principal investigator and evaluator for each of the study components. Specifically, this included the creation of assessment tools, review of data obtained from related CRDW data sources (such as audit trails and the RISE system), and completion of the 30-day post-training assessments. These activities were feasible for the study researcher to complete based on her background and professional experience.

The study researcher for this project has been a member of the CTSI since 2016 and has served as an advisor for several CTSI Quality and Efficiency module projects, including two projects for the BMI department. These experiences have allowed the study researcher to understand the CTSI structure, learn about the BMI department, and become acquainted with key individuals within the CTSI.

In addition, the study researcher has more than twelve years of professional teaching experience in higher education, including four years in a program director position. During this time, the study researcher has developed more than 20 courses, redesigned the curricula for two degree programs, and gained valuable insights by working with a professional instructional designer and completing coursework in adult education. As a result, the study researcher has deep experience developing course learning objectives and assessments and knowledge of adult education principles.

Finally, the study researcher spent more than a decade in information-technology related roles leading, managing, and supporting individuals and systems within three of the CTSI partner organizations (BCW, Froedtert Hospital, and MCW). This experience includes more than seven years implementing, supporting, and maintaining EHR-related software, interfaces, security, and infrastructure. Experience gained leading enterprise EHR implementation projects has provided the study researcher with critical, firsthand experience with technology adoption at both individual and organization levels.

4.10 Institutional Review Board (IRB)

This study was approved by the Medical College of Wisconsin/Froedtert Hospital Institutional Review Board (IRB) in accordance with 45 CFR 46.111 by expedited review, Categories 5 and 7, as a minimal risk study (see Appendix A: Institutional Review Board Approval Letter). As part of existing IRB reliance master agreements, this approval was granted for the Medical College of Wisconsin – Milwaukee campus, Milwaukee School of Engineering, and University of Wisconsin-Milwaukee.

4.11 Timeline and Budget

The study period timeline was 32 weeks, with most of the data collection and analysis for this study occurring from August 2018 through November 2018. This timeline was chosen for several reasons. First, many new researchers, residents, and students began working at CTSI partner institutions during this timeframe. Having the interventions occur during this time period provided the greatest opportunity to involve a diverse cohort of individuals. Second, this timeframe allowed the study to be completed prior to the submission of the CTSA renewal proposal, which is due in April 2019. Results from this study may be included within this renewal packet as evidence of goal achievement. Finally, this timeline allowed sufficient time to assess the outcomes of the awareness presentation and training, but was short enough to minimize negative effects should these components result in ineffective or detrimental results.

Below is the high-level timeline for this study.

Table 11. High-level Timeline for the Study

Task	Start Date	End Date	April 2018	May 2018	June 2018	July 2018	August 2018	September 2018	October 2018	November 2018	December 2018	January 2019	February 2019	March 2019
Dissertation proposal	**4/6/2018**	**4/30/2018**												
Draft of proposal to Dr. Luo	4/6/2018	4/6/2018	X											
Proposal to committee	4/16/2018	4/16/2018	X											
Meetings with committee members (as needed)	4/16/2018	4/27/2018	X											
Proposal defense presentation	4/30/2018	4/30/2018	X											
Dissertation project	**5/1/2018**	**12/31/2018**												
Obtain IRB approval	5/1/2018	7/26/2018		X	X	X								
Contact leaders to schedule awareness presentations	7/26/2018	10/14/2018				X	X	X	X					
Conduct awareness presentations	7/30/2018	10/30/2018				X	X	X	X					
Conduct training sessions	7/30/2018	10/30/2018				X	X	X	X					
Conduct 30-day follow-up assessments	8/30/2018	11/30/2018					X	X	X	X				
Data analysis	7/30/2018	12/31/2018				X	X	X	X	X	X			
Dissertation	**1/21/2019**	**3/8/2019**												
Final draft of dissertation to Dr. Luo	1/21/2019	1/21/2019										X		
Draft of dissertation to committee	2/11/2019	2/11/2019											X	
Meetings with committee members (as needed)	2/11/2019	2/25/2019											X	
Dissertation defense presentation	2/25/2019	3/8/2019											X	X
Communication to committee	**6/15/2018**	**1/14/2019**												
Update to committee	6/15/2018	6/15/2018			X									
Update to committee	7/31/2018	7/31/2018				X								
Update to committee	8/31/2018	8/31/2018					X							
Update to committee	9/30/2018	9/30/2018						X						
Update to committee	10/31/2018	10/31/2018							X					
Update to committee	11/30/2018	11/30/2018								X				
Update to committee	1/14/2019	1/14/2019										X		

From a budget perspective, the efforts of Bradley Taylor and Kristen Osinski from the BMI department, as well as the use of REDCap software, were supported by the National Center for Advancing Translational Sciences, National Institutes of Health, Award Number UL1TR001436. The efforts of Dr. Luo (UWM advisor) and committee members (Dr. Cisler, Dr. Kate, Dr. Patrick, and Dr. Payne) were provided at no cost to the project. The efforts of Katie McCarthy were not monetarily compensated. The study participants were not compensated as part of this study. As a result, no project budget allocation was needed for resources or participants.

This study involved procurement of office supplies such as pens, paper, envelopes, and printer ink. The estimate for these costs was $500 and was provided by the study researcher. No other costs were associated to this project.

4.12 Study Methodology Limitations

Although this study has several strengths with the potential to make an important contribution to literature regarding data warehouse adoption and use for clinical and translational research purposes, there are potential methodology limitations that need to be noted.

First, although the CTSA hubs share common CRDW tools, such as i2b2, each CTSA hub has variation in data warehouse implementation. For example, the Honest Broker tool used at the CTSI is unique to that CTSI. As a result, further work would be needed to determine if these study components could be generalized for use at other CTSA hubs, especially in relation to the objective measures used.

Second, the 30-day post-training assessment population is based on volunteers identified at the completion of CRDW training sessions. A potential exists that volunteer bias could exist within this cohort. For example, only those individuals who feel confident in their use of the CRDW may volunteer, which may cause results showing more activity than the overall trainee population. Additionally, knowing that a 30-day post-training assessment exists may change the behaviors of all trainees in relation to their use of the CRDW. For instance, trainees may rush to complete their work within 30 days of training even if this would not be their normal behavior.

Finally, the 30-day post-training assessment focuses on individuals receiving CRDW training and does not attempt to contact individuals who received access to the CRDW but did not receive training. While focusing on individuals receiving training is reasonable based on the desire to explore the link between training and completion of CRDW activities, it is possible that choosing this population could result in bias (such as individuals who receive training are

generally less comfortable with technology than those who choose not to receive training). As a result of these limitations, care should be used when interpreting and attempting to generalize the results of this study.

CHAPTER 5: RESULTS

Within this chapter, the results of each of the study components are presented, including the acceptance or rejection of each of the study's hypotheses.

5.1 Study Procedures and Timeline

After receiving IRB approval on July 26, 2018, data collection for this study occurred according to the project timeline. Throughout the study, no deviations from the procedures described in the previous chapter were encountered.

5.2 Study Component #1: Awareness Intervention

During the study period, seven presentations were held with a total of 54 participants attending. Based on informal introductions occurring prior to the presentations, participants included representatives from four of the eight CTSI partner organizations (CHW, MCW, MSOE, and UWM) and held a variety of research-related roles (including investigators, academic faculty, research coordinators, research administrators, laboratory management and staff, fellows, medical students, and graduate students).

Of the 54 participants, a total of 50 participants attempted the pre- and post-assessments. Of these 50 assessment pairs, one was found to be incomplete and was removed from the sample. As a result, a total of 49 participants (91%) successfully completed the presentation pre- and post-assessments, which exceeded the minimum sample size of 40 completed assessment pairs originally proposed for this study.

The purpose of study component #1 was to test the following hypotheses:

H1: Awareness presentations will increase the average number of correct answers from the pre-assessment to the post-assessment.

H2: Awareness presentations will increase the average self-reported level of likelihood to use from the pre-assessment to the post-assessment.

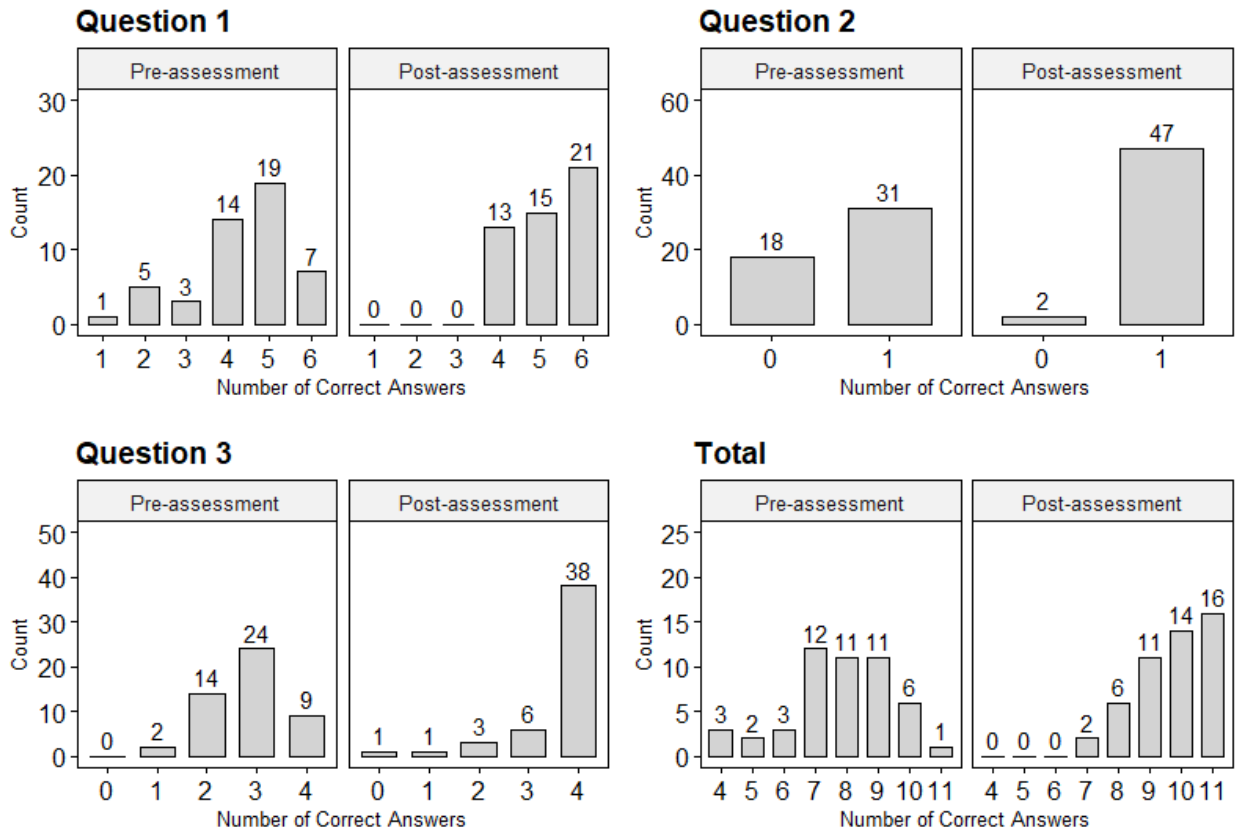
Below is a summary of the results for these hypotheses.

5.2.1 Hypothesis H1 Results – Correct Answers

To test hypothesis H1, the answers from the first three questions on the awareness presentation assessments were analyzed. The maximum number of potential correct answers for one assessment was eleven. Within the dataset, the total number of correct answers ranged from four to eleven. In the pre-assessments, the mean (standard deviation) was 7.8 (1.6) total correct answers. In the post-assessments, the mean (standard deviation) was 9.7 (1.2) total correct answers.

The figure below shows the frequency of the number of correct answers for the pre-assessment and post-assessment.

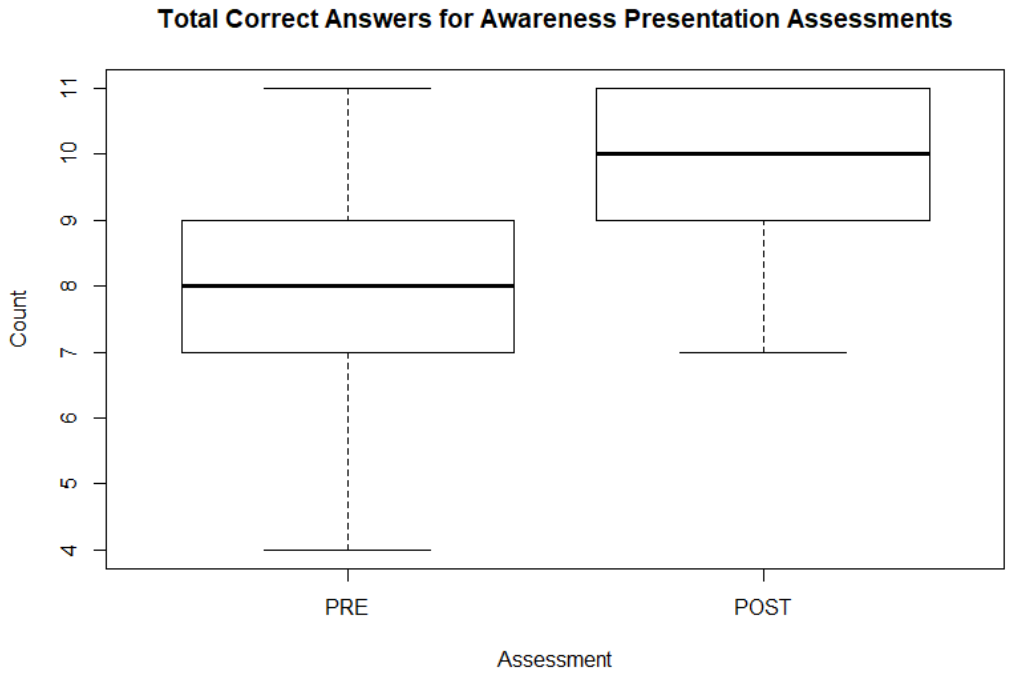
Figure 14. Correct Answers for Awareness Presentation Assessments



As shown in the figure above, the number of correct answers for the individual questions 1, 2, and 3 increased from the pre-assessment (before the awareness presentation) to the post-assessment (after the awareness presentation). The total number of correct answers for this section also increased from the pre-assessment to the post-assessment. This supports hypothesis H1.

To visualize the change in the median number of total correct answers from pre-assessment to post-assessment, a boxplot was created (and is shown below).

Figure 15. Boxplot of Total Correct Answers for Awareness Presentation Assessments



As shown in the figure above, the median number of total correct answers from the pre-assessments was 8 correct answers. The median number of total correct answers from the post-assessments was 10 correct answers, which is an increase of two correct answers from the pre-assessment. This supports hypothesis H1.

Next, the study researcher tested the normality of the data using Q-Q plots and the results of the Shapiro-Wilk normality tests. Based on the results of these tests, normality could not be assumed for this data; therefore, hypothesis testing was performed using a non-parametric method (Wilcoxon matched-pairs signed-ranks test). The Wilcoxon matched-pairs signed-ranks test indicated that the post-assessment scores were statistically significantly higher than

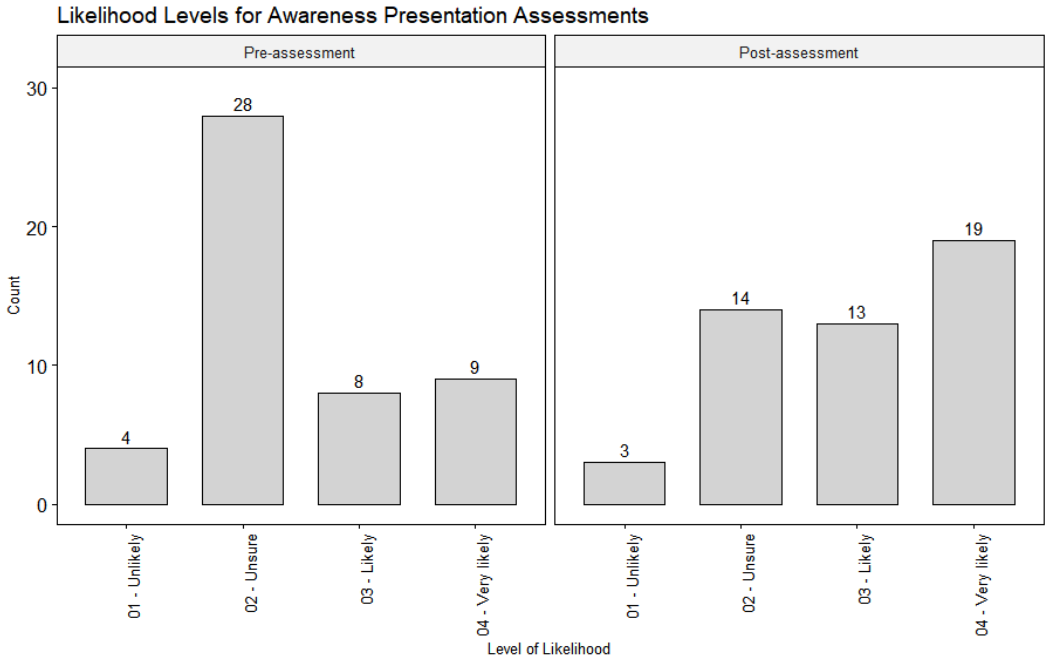
pre-assessment scores (Wilcoxon T(37) = 241, p < 0.000, one-tailed test). As a result, hypothesis H1 was accepted.

5.2.2 Hypothesis H2 Results – Likelihood to Use

To test hypothesis H2, the answers from the fourth question on the awareness presentation assessments were analyzed. The possible options for this question were "Very likely" to "Very unlikely." Within the dataset, answers ranged from "Very likely" to "Unlikely." No participants selected "Very unlikely" in response to this question. In the pre-assessments, the mean (standard deviation) was a 2.4 (0.9) level of likelihood. In the post-assessments, the mean (standard deviation) was a 3.0 (1.0) level of likelihood.

The figure below shows the frequency of the levels of likelihood for the pre-assessment and post-assessment.

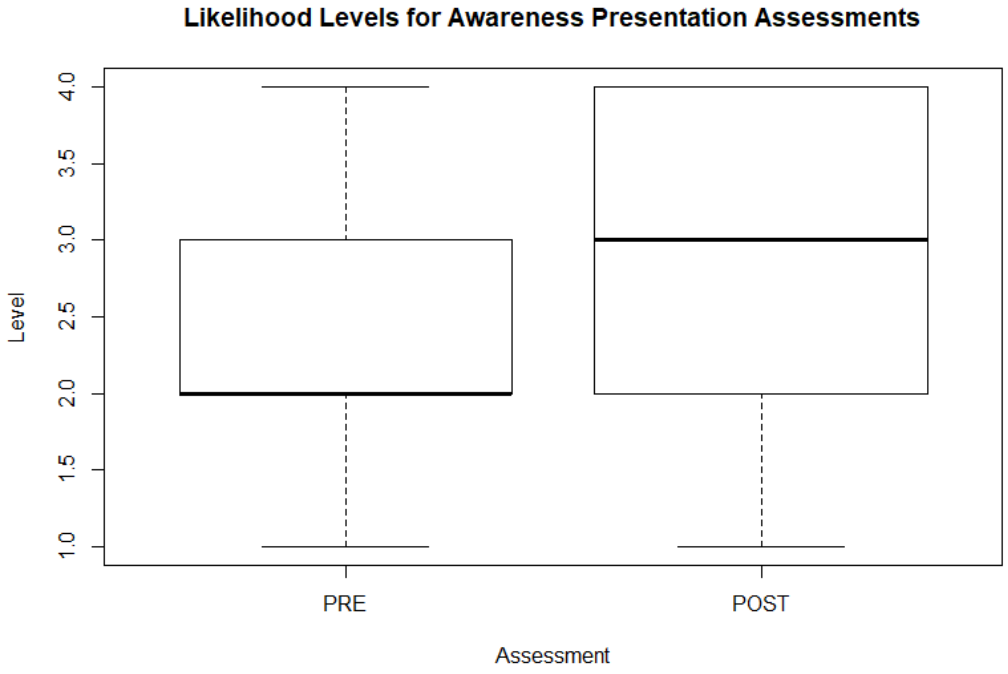
Figure 16. Levels of Likelihood for Awareness Presentation Assessments



As shown in the figure above, the majority of participants indicated they were unsure if they would use the CRDW in the pre-assessments (prior to the awareness presentations). In the post-assessments (after the awareness presentations), the majority of participants indicated that they were very likely or likely to use the CRDW, which indicates an increase in level of likelihood from the pre-assessments to the post-assessments. This supports hypothesis H2.

To visualize the change in the median level of likelihood to use from pre-assessment to post-assessment, a boxplot was created (and is shown below).

Figure 17. Boxplot of Levels of Likelihood for Awareness Presentation Assessments



As shown in the figure above, the median level of likelihood was 2 (which is the numeric value of the "Unsure" option) in the pre-assessments. In the post-assessments, the median level of likelihood was 3 (which is the numeric equivalent of the "Likely" option). This increase supports hypothesis H2.

The Wilcoxon matched-pairs signed-ranks test indicated that the post-assessment likelihood levels were statistically significantly higher than pre-assessment likelihood levels (Wilcoxon $T(9.5) = 83$, $p < 0.000$, one-tailed test). As a result, hypothesis H2 was accepted.

5.3 Study Component #2: Assessment of CRDW Training

During the study period, eight training sessions were held with a total of 17 trainees attending. Due to the anonymity of the awareness presentation participants, it is unknown if any of these trainees attended the awareness presentations prior to receiving training. Of the 17 trainees, all successfully completed the training pre- and post-assessments, which exceeds the 75% participation rate originally proposed for this study.

The purpose of study component #2 was to test the following hypotheses:

H5: CRDW training will increase the average number of correct answers from the pre-assessment to the post-assessment.

H6: CRDW training will increase the average number of identified potential outcomes from the pre-assessment to the post-assessment.

H7: CRDW training will increase the average self-reported level of likelihood to use from the pre-assessment to the post-assessment.

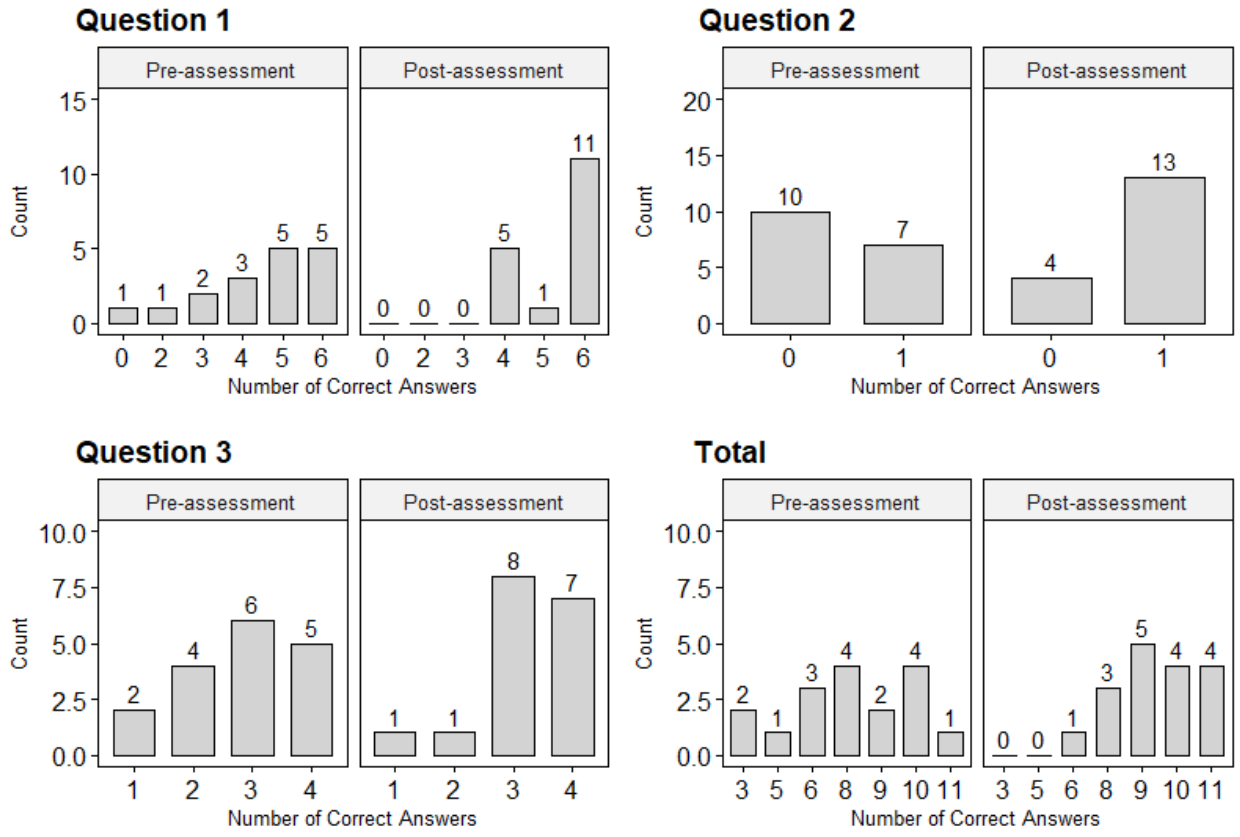
Below is a summary of the results for these hypotheses.

5.3.1 Hypothesis H5 Results – Correct Answers

To test hypothesis H5, the answers from the first three questions on the training assessments were analyzed. The maximum number of potential correct answers for one assessment was eleven. Within the dataset, the total number of correct answers ranged from three to eleven. In the pre-assessments, the mean (standard deviation) was 7.6 (2.4) total correct answers. In the post-assessments, the mean (standard deviation) was 9.3 (1.4) total correct answers.

The figure below shows the frequency of the number of correct answers for the pre-assessment and post-assessment.

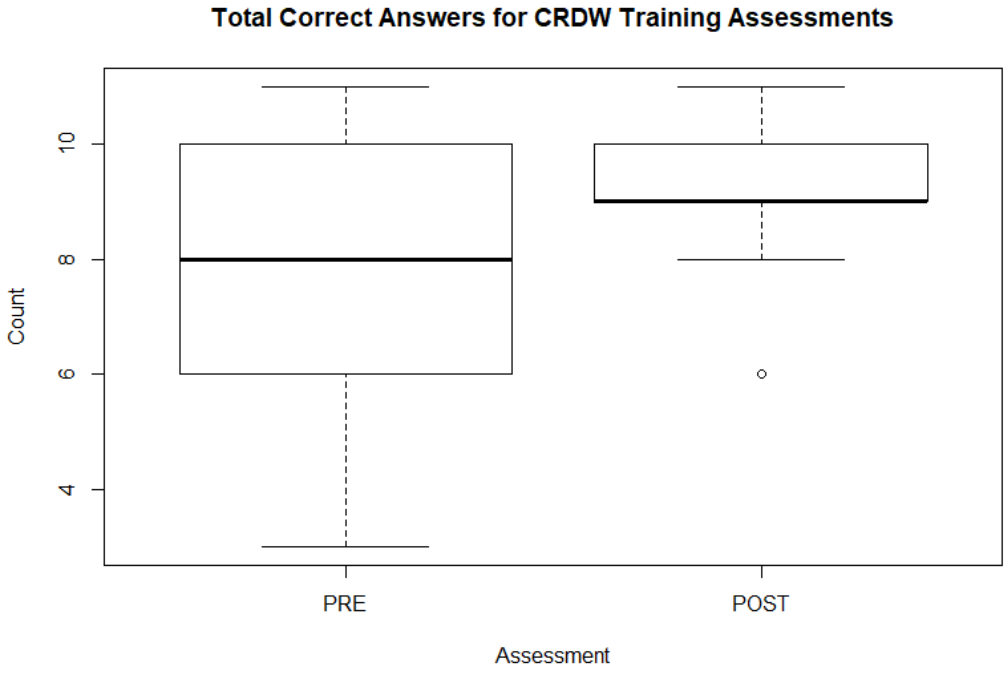
Figure 18. Correct Answers for CRDW Training Assessments



As shown in the figure above, the number of correct answers for the individual questions 1, 2, and 3 increased from the pre-assessment (before the training session) to the post-assessment (after the training session). The total number of correct answers for this section also increased from the pre-assessment to the post-assessment. This supports hypothesis H5.

To visualize the change in the median number of total correct answers from pre-assessment to post-assessment, a boxplot was created (and is shown below).

Figure 19. Boxplot of Total Correct Answers for CRDW Training Assessments



As shown in the figure above, the median number of total correct answers from the pre-assessments was 8 correct answers. The median number of total correct answers from the post-assessments was 9 correct answers, which is an increase of one correct answer from the pre-assessment. This supports hypothesis H5.

Next, the study researcher tested the normality of the data using Q-Q plots and the results of the Shapiro-Wilk normality tests. Based on the results of these tests, normality could not be assumed for this data; therefore, hypothesis testing was performed using a non-parametric method (Wilcoxon matched-pairs signed-ranks test). The Wilcoxon matched-pairs signed-ranks test indicated that the post-assessment scores were statistically significantly higher than pre-assessment scores (Wilcoxon $T(8.5) = 21$, $p = 0.005$, one-tailed test). As a result, hypothesis H5 was accepted.

5.3.2 Hypothesis H6 Results – Potential Outcomes

To test hypothesis H6, the answers from the fourth question on the training assessments were analyzed. A total of six potential outcomes could be selected by the trainee. Within the dataset, the number of potential outcomes selected ranged from one outcome to five outcomes. In the pre-assessments, the mean (standard deviation) was 2.4 (1.4) potential outcomes. In the post-assessments, the mean (standard deviation) was 2.2 (1.3) potential outcomes.

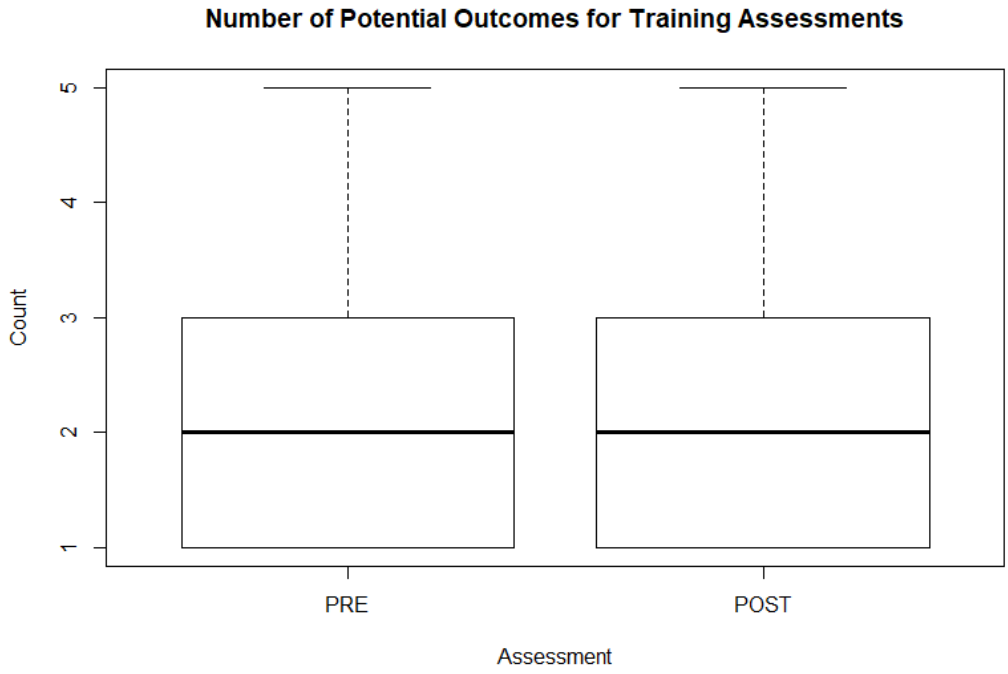
The figure below shows the frequency of the number of potential outcomes for the pre-assessment and post-assessment.

Figure 20. Number of Potential Outcomes for Training Assessments

Based on this figure, the majority of trainees indicated one potential outcome during the pre-assessment (before the training session). In the post-assessment (after the training session), the majority of trainees indicated one potential outcome. As a result, an increase in potential outcomes identified did not occur. This does not support hypothesis H6.

To visualize the change in the median number of potential outcomes from pre-assessment to post-assessment, a boxplot was created (and is shown below).

Figure 21. Boxplot of Number of Potential Outcomes for Training Assessments



As shown above, the boxplots for the pre-assessment and post-assessment show no difference in the number of potential outcomes identified, which does not support hypothesis H6.

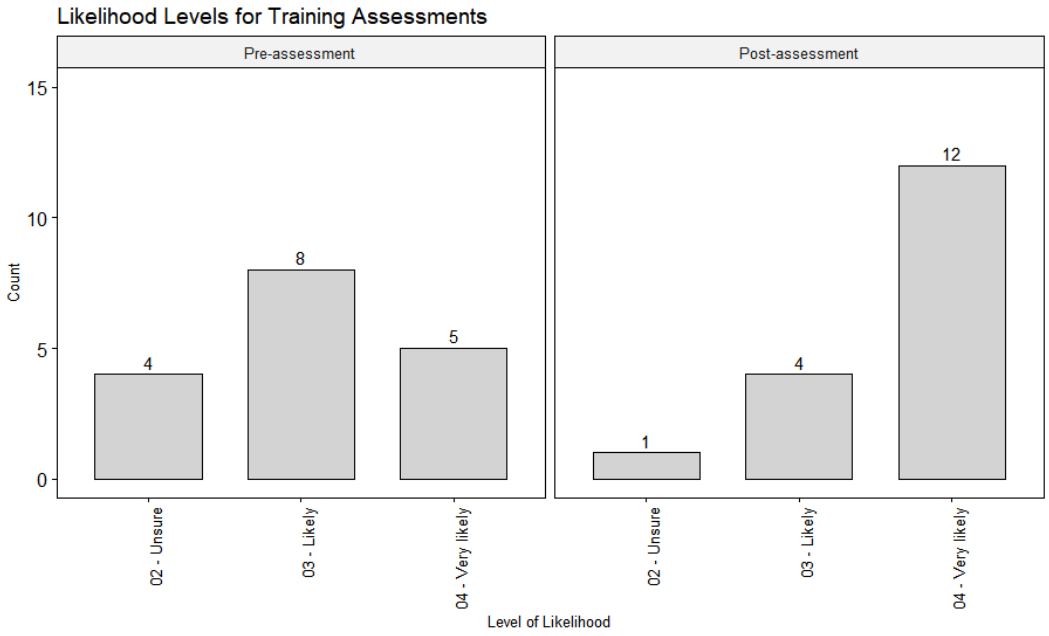
Next, the study researcher tested the normality of the data using Q-Q plots and the results of the Shapiro-Wilk normality tests. Based on the results of these tests, normality could not be assumed for this data; therefore, hypothesis testing was performed using a non-parametric method (Wilcoxon matched-pairs signed-ranks test). The Wilcoxon matched-pairs signed-ranks test indicated that the post-assessment potential outcomes were not statistically significantly different than pre-assessment potential outcomes (Wilcoxon $T(9) = 0$, $p = 0.824$, two-tailed test). As a result, hypothesis H6 was rejected.

5.3.3 Hypothesis H7 Results – Likelihood to Use

To test hypothesis H7, the answers from the fifth question on the training assessments were analyzed. The possible options for this question were "Very likely" to "Very unlikely." Within the dataset, answers ranged from "Very likely" to "Unsure." No participants selected "Unlikely" or "Very unlikely" in response to this question. In the pre-assessments, the mean (standard deviation) was a 3.0 (0.7) level of likelihood. In the post-assessments, the mean (standard deviation) was a 3.6 (0.6) level of likelihood.

The figure below shows the frequency of the levels of likelihood for the pre-assessment and post-assessment.

Figure 22. Levels of Likelihood for Training Assessments

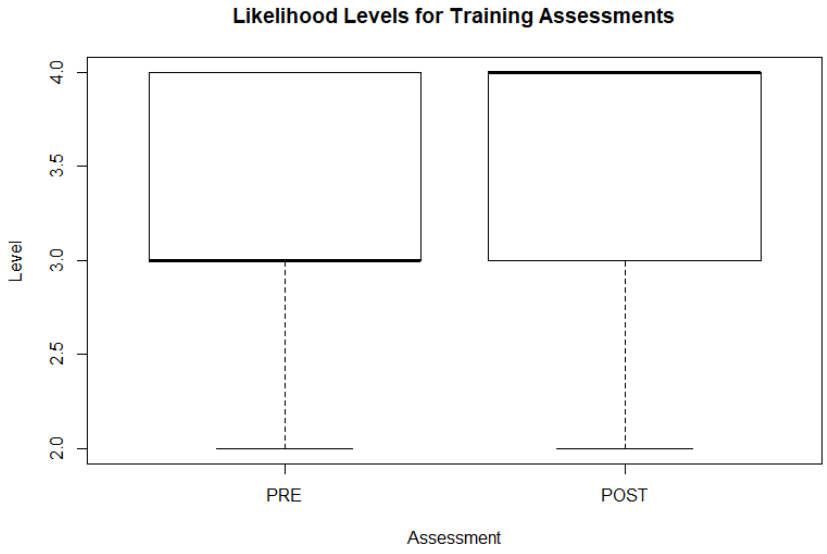


As shown in the figure above, the majority of participants indicated they were likely to use the CRDW in the pre-assessments (prior to the training sessions). In the post-assessments (after the training sessions), the majority of participants indicated that they were very likely to use the

CRDW, which indicates an increase in level of likelihood from the pre-assessments to the post-assessments. This supports hypothesis H7.

To visualize the change in the median level of likelihood to use from pre-assessment to post-assessment, a boxplot was created (and is shown below).

Figure 23. Boxplot of Levels of Likelihood for Training Assessments



As shown in the figure above, the median level of likelihood was 3 (which is the numeric value of the "Likely" option) in the pre-assessments. In the post-assessments, the median level of likelihood was 4 (which is the numeric equivalent of the "Very Likely" option). This increase supports hypothesis H7.

The Wilcoxon matched-pairs signed-ranks test indicated that the post-assessment likelihood levels were statistically significantly higher than pre-assessment likelihood levels (Wilcoxon $T(3.5) = 8$, $p = 0.011$, one-tailed test). As a result, hypothesis H7 was accepted.

5.4 Study Component #3: Objective Measurement of Outcomes

Study component #3 involved the objective measurement of CRDW access requests, CRDW training requests, and trainee completion of CRDW activities related to their intended research outcomes. Below is a summary of the results for these measurements and related hypotheses.

5.4.1 Hypothesis H3 Results – Access Requests

The purpose of measuring CRDW access requests was to test the following hypothesis:

H3: The growth rate of CRDW access requests for individuals and projects will increase during the awareness presentation intervention period in comparison to the growth rate from 2016-2017 for the same time period.

The following table summarizes the number of individuals requesting access to the CRDW during the study period in 2016, 2017, and 2018, as well as the growth rate calculated year-to-year.

Table 12. CRDW Access Requests by Individual in 2016, 2017, and 2018

2016	2017	2018	Growth Rate (2016-2017)	Growth Rate (2017-2018)
26	61	39	135%	-36%

Based on this analysis, the growth rate of access requests by individual decreased by 36% during the study period compared to the previous year.

The following table summarizes the number of projects associated to CRDW access requests during the study period in 2016, 2017, and 2018, as well as the growth rate calculated year-to-year.

Table 13. CRDW Access Requests by Project in 2016, 2017, and 2018

2016	2017	2018	Growth Rate (2016-2017)	Growth Rate (2017-2018)
15	31	18	107%	-42%

Based on this analysis, the growth rate of access requests by project decreased by 42% during the study period compared to the previous year.

Because the growth rate for individual and project access requests decreased during the study period instead of increasing as hypothesized, hypothesis H3 was rejected.

5.4.2 Hypothesis H4 Results – Training Requests

The purpose of measuring CRDW training requests was to test the following hypothesis:

H4: The growth rate of CRDW training requests for research project teams will increase during the awareness presentation intervention period in comparison to the growth rate from 2016-2017 for the same time period.

The following table summarizes the number of projects associated to CRDW training requests during the study period in 2016, 2017, and 2018, as well as the growth rate calculated year-to-year.

Table 14. CRDW Training Requests by Project in 2016, 2017, and 2018

2016	2017	2018	Growth Rate (2016-2017)	Growth Rate (2017-2018)
8	20	9	150%	-55%

Note that nine training sessions were documented in RISE during the study period in 2018, but only eight training sessions were part of the CRDW training assessment (study component #2) that occurred during this same period. This discrepancy was the result of a refresher training session that was held during the study period for an individual that previously received CRDW training. Although this training session did not meet the criteria to be included in study component #2 (assessment of CRDW training), this training session was included in the RISE records as a training request (as RISE training requests include all types of CRDW training provided).

Based on this analysis, the growth rate of training requests by project decreased by 55% during the study period compared to the previous year. Because the growth rate for project team training requests decreased during the study period instead of increasing as hypothesized, hypothesis H4 was rejected.

5.4.3 Hypothesis H8 Results – Outcomes

The purpose of measuring trainee completion of CRDW activities related to their intended research outcomes was to test the following hypothesis:

H8: Within 30 days of receiving CRDW training, at least 50% of individuals will complete the CRDW activities required to achieve their intended research outcomes.

Of the 17 trainees involved in study component #2 (assessment of CRDW training), four individuals (24%) completed the appropriate CRDW actions mapped to their identified intended research outcomes. Because less than 50% of individuals receiving training completed the CRDW activities required to achieve their identified research intended outcomes, hypothesis H8 was rejected.

5.5 Study Component #4: 30-day Post-training Assessment

The purpose of study component #4 was to investigate potential factors that influence trainees' use of the CRDW for clinical and translational research purposes, specifically within 30-45 days of their original training. Of the 17 trainees that participated in study component #2 (assessment of CRDW training), 16 trainees (94%) consented to be contacted for a 30-day post-training assessment. Of the consenting trainees, a total of ten participants successfully completed the assessment within 30-45 days of their original training, resulting in a 59% participation rate. This participation rate exceeds the 10% rate originally proposed for this study component.

Additionally, of the trainees who completed the 30-day post-training assessment, two trainees (20%) completed the CRDW activities related to their intended research outcomes and eight

trainees (80%) did not complete the CRDW activities related to their intended research outcomes in study component #3 (objective measurement of outcomes). This achievement ratio is similar to the ratio for all the trainees who participated in this study.

Although the participation rate was higher than originally proposed, the sample size is relatively small in relation to the number of options provided as answers. For example, the options for rating of the ease of use of CRDW tools included "Very easy to use," "Easy to use," "Difficult to use," and "Very difficult to use." As a result, the number of options were collapsed to provide directional information as previously described.

5.5.1 Section #1: Use of the CRDW Tools

In the first section of the assessment, individuals were asked if they used the CRDW. If so, further questions asked about the activities performed, the ease of use of the CRDW tools, and how the data has been (or will be) used. Six of the ten trainees (60%) indicated that they had used the CRDW tools since receiving training and, therefore answered three additional questions related to CRDW use.

The first question asked trainees to identify the activities they completed using the CRDW tools. Six trainees (100%) had obtained a count of patients within a particular cohort. One trainee also analyzed the demographic distribution and extracted data tables for a patient set.

The second question asked trainees to rate the ease of use of each of the CRDW tools. Based on the responses to this question, the majority of trainees found i2b2 and Honest Broker easy to use, while trainees were evenly split on the ease of use of TriNetX.

The third question asked trainees how they have used or plan to use the data obtained from the CRDW. Trainees were allowed to select one or more of five possible plans. The most common selection was using the data for cohort discovery (three trainees), closely followed by using the data for prospective studies (two trainees) and using the data for retrospective studies (two trainees).

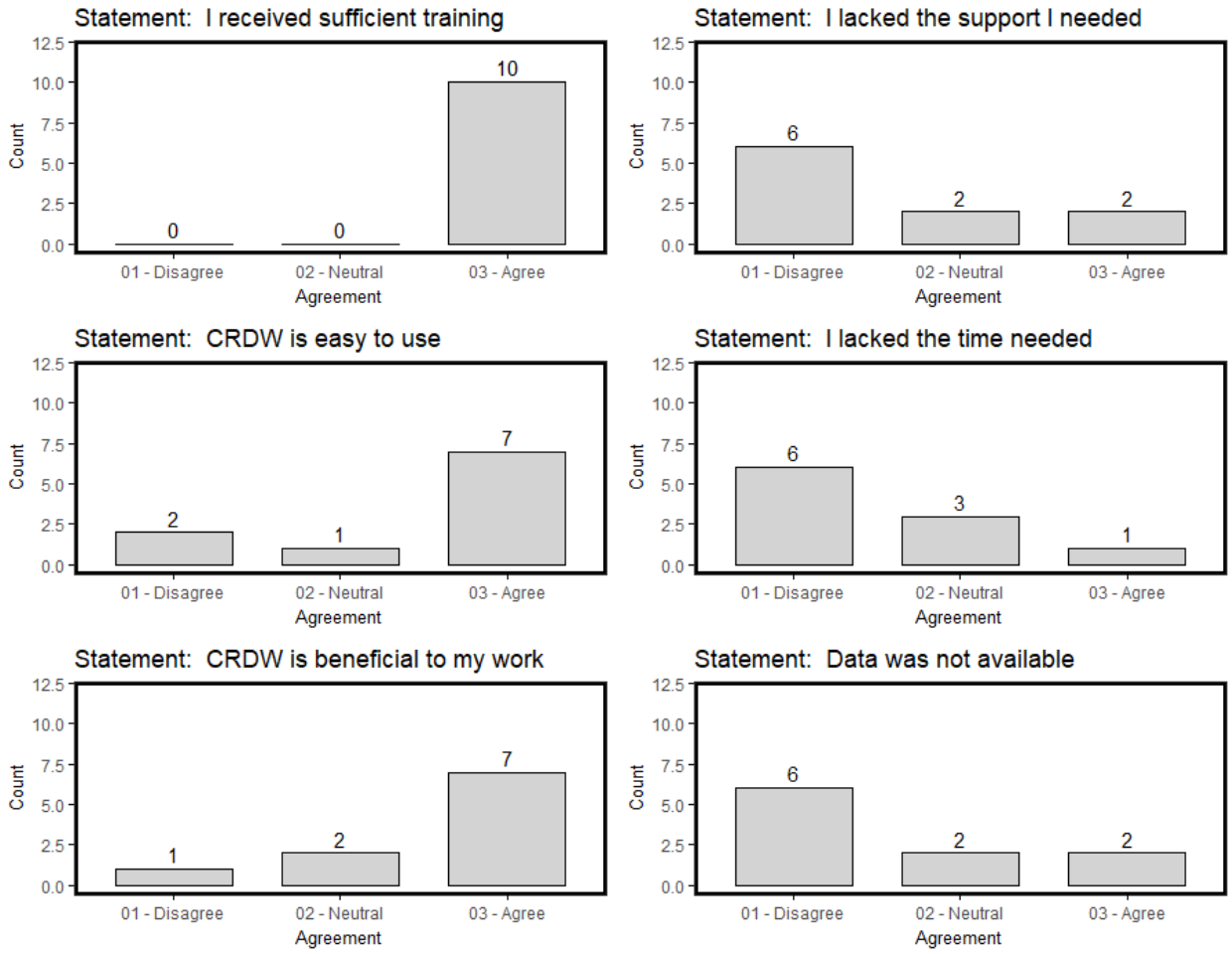
5.5.2 Section #2: Requesting Support

In the second section of the assessment, trainees were asked if they contacted the BMI department to receive CRDW support. Four trainees (40%) indicated that they had requested support from the BMI department. Of these trainees, all contacted the BMI Business Analyst to obtain this support. Two of the four trainees also provided information within the comments section of the assessment, identifying the BMI Business Analyst as being an effective resource and phenomenal in providing education and support.

5.5.3 Section #3: Benefits and Barriers to CRDW Use

In the third section of the assessment, trainees were asked about their perceptions of potential benefits and barriers to using the CRDW. The figure below shows the level of agreement for each of these statements. The left column of graphs shows the results for the three positively phrased statements; the right column shows the results for the three negatively phrased statements.

Figure 24. Responses to Benefits and Barriers Statements



5.5.4 Section #4: Likelihood to Use the CRDW

In the fourth section of the assessment, trainees were asked about their level of likelihood to use the CRDW for future clinical research. Eight trainees (80%) reported being likely to use the CRDW again.

One trainee (10%) reported being unsure of using the CRDW in the future. In the comments section of the assessment, this trainee identified that he/she had not yet had an opportunity to access the data within the CRDW, but felt the CRDW would be beneficial in the future. This

trainee expressed a concern that receiving training and not using it immediately may present some problems, but the trainee also identified that he/she knew how to reach out for support.

One trainee (10%) reported being unlikely to use the CRDW in the future. In this trainee's assessment responses, he/she identified lack of support, difficulty using the Honest Broker tool, and lack of availability of needed data within the CRDW as potential barriers. The trainee also somewhat disagreed with the statement that the CRDW was beneficial for his/her work. The trainee did not request support from the BMI department. The trainee did not provide any further comments within his/her assessment responses.

CHAPTER 6: DISCUSSION

The purpose of this study was to assess and measure the impact of raising awareness and providing training on the completion of CRDW activities related to intended research outcomes. This study also explored potential factors that influenced the completion of CRDW activities. Specifically, three research questions were examined:

Q1: Is delivering a 20-minute presentation regarding CRDW basic functions and processes a sufficient intervention to raise awareness of the CRDW for clinical and translational research purposes?

Q2: Does existing CRDW training provide sufficient information for trainees to perform the CRDW activities related to their intended research outcomes?

Q3: What potential factors influence trainees' use of the CRDW for clinical and translational research purposes?

Four study components were used to investigate these questions. The previous chapter presented detailed results from each of these components. This chapter will provide an overall summary of the findings, the interpretation of the findings by research question, the limitations of the study, and the implications of the findings for future research.

6.1 Summary of Study Findings

The following table presents the eight hypotheses tested in this study and indicates if the hypotheses were supported based on the findings from each study component.

Table 15. Assessment of Study Hypotheses

Hypothesis	Hypothesis Supported
Research Question #1	
H1: Awareness presentations will increase the average number of correct answers from the pre-assessment to the post-assessment.	Yes
H2: Awareness presentations will increase the average self-reported level of likelihood to use from the pre-assessment to the post-assessment.	Yes
H3: The growth rate of CRDW access requests for individuals and projects will increase during the awareness presentation intervention period in comparison to the growth rate from 2016-2017 for the same time period.	No
H4: The growth rate of CRDW training requests for research project teams will increase during the awareness presentation intervention period in comparison to the growth rate from 2016-2017 for the same time period.	No
Research Question #2	
H5: CRDW training will increase the average number of correct answers from the pre-assessment to the post-assessment.	Yes
H6: CRDW training will increase the average number of identified potential outcomes from the pre-assessment to the post-assessment.	No
H7: CRDW training will increase the average self-reported level of likelihood to use from the pre-assessment to the post-assessment.	Yes
H8: Within 30 days of receiving CRDW training, at least 50% of individuals will complete the CRDW activities required to achieve their intended research outcomes.	No

6.2 Research Question #1: Raising Awareness of the CRDW

The approach of this study mirrored the five innovation-decision stages identified in the DOI theory. As a result, research question #1 (Q1) was used to determine if a specific intervention (presentation) to raise awareness was successful in increasing knowledge of the CRDW, as well as moving individuals into the persuasion stage, where individuals actively seek out information regarding the CRDW.

Based on the results for research hypotheses H1 and H2, the presentation did increase participants' knowledge of the CRDW, as well as their self-reported level of likelihood to use the CRDW, both of which would indicate a successful awareness intervention. In reviewing the results of hypotheses H3 and H4, the number of CRDW access and training requests did not increase as anticipated, but instead decreased during the study period when compared to previous years. As a result, while the knowledge stage was reached, the persuasion stage was not reached during the study period.

One possible explanation is that the awareness intervention may have been successful (as indicated by the increase in knowledge and likelihood to use), but the timeline for the study did not provide adequate time for individuals to complete the steps necessary to request access and/or training. For example, completion of the Data Release Agreement (the first step in CRDW access) requires that the principal investigator for the intended project be MCW faculty. Individuals from other CTSI partner sites who attended the awareness presentations may have needed time to find an MCW collaborator and, therefore, may not have requested access during the study period.

Another possible explanation is that participants may have proactively attended the CRDW presentations to understand the availability of CRDW tools for future research that may not occur for months or years. While participants were asked about their likelihood to use the CRDW before and after the presentation and the likelihood of use increased, they were not asked about their potential timelines for using the tools. As a result, the BMI department may see an influx of access and/or training requests after this study as grant submission deadlines approach. Fortunately, because the schedule of presentations has been documented, requests could be monitored to determine if there is a spike after the study period.

One further explanation is that the awareness presentation intervention provided sufficient information for participants to identify that the CRDW tools and/or data would not meet the needs of their current research projects. In this scenario, an individual would not have immediately requested access or training for the CRDW, but may request access and/or training for future research projects. This explanation is supported by the increase in the likelihood to use level identified in the post-assessments, but the lack of immediate increase in access and training requests for the CRDW.

Limited historical data and lack of a standardized awareness process may also have affected the results for hypotheses H3 and H4. The requests for CRDW access and training have been tracked since 2016. As a result, only two growth rates exist for comparison (2016 to 2017 and 2017 to 2018). Because a significant spike in requests was tracked in 2017, the 2016-2017 growth rates significantly increased. To have a higher 2017-2018 growth rate, an even more significant spike in requests would need to occur in 2018. If the large number of requests in 2017 was an anomaly, comparing the growth rate from 2017-2018 to the growth rate from

2016-2017 may be misleading. Without additional historical data, no conclusion can be drawn about a "typical" year-to-year growth rate. Additionally, without a standardized process for raising awareness, no documentation exists of activities that may have affected the number of 2017 requests. Fortunately, as requests continue to be tracked, a retrospective review could be completed to review the number of requests year-to-year to determine typical growth rates. As a result of this study, the BMI department also has documentation of when CRDW awareness presentations and training sessions occurred, and they could use this information to review changes in request rates after the study period.

A final possible explanation is that the awareness presentation content was insufficient to move individuals from the knowledge stage to the persuasion stage. While the content of the presentation was focused on factors found to be important in technology adoption, these factors and the approach to delivering the content were not confirmed with the target population prior to performing this study. For example, although the presentation focused on the data available (which was identified as important for perceived usefulness and task compatibility), the actual content may have focused too much on specific data elements available and not on stories of how the available data had been used for research purposes (which may have been the focus desired by the target population). As a result, the increase in knowledge and likelihood to use may have been insufficient to cause immediate requests for access and training.

6.3 Research Question #2: Achievement of Outcomes Post-training

For research question #2 (Q2), the goal was to determine if the CRDW training provided sufficient information for trainees to complete the CRDW activities related to their intended research outcomes. Based on the results for research hypotheses H5 and H7, the training does increase knowledge of the CRDW and self-reported likelihood to use the CRDW. Interestingly, training did not significantly change the number of potential outcomes identified by trainees (research hypothesis H6). In addition, only 24% of the trainees completed the CRDW actions related to their intended research outcomes within 30 days of completing training (research hypothesis H8). As a result, while the persuasion stage was reached, the decision stage was not reached during the study period.

For research hypothesis H6, trainees were asked to identify their intended outcomes from using the CRDW. The intended outcome options included protocol, poster presentation, grant submission, abstract, manuscript, and other. Trainees were allowed to select one or more (up to a maximum of six) of these intended outcomes. Within this study, no significant change in the number of intended outcomes existed between the pre-assessments and the post-assessments. One explanation is that trainees discovered (during training) that the CRDW would not be useful for their intended research projects. This explanation is contradicted, though, by the likelihood to use indicated by the trainees, which increased from pre-assessment to post-assessment.

To investigate this contradiction further, the intended outcomes selected in the pre-assessment and post-assessment by individual trainee were compared. In addition, for each individual trainee, the likelihood to use value was compared between the pre- and post-assessment. Finally,

a numeric change in level of likelihood to use was calculated. For example, if a trainee indicated he/she was unsure about using the CRDW in the pre-assessment, and then indicated being very likely to use the CRDW in the post-assessment, this was a two-level change in likelihood to use (since the likelihood to use level increased by two). Below are the results of this investigation.

Table 16. Comparison of Intended Outcomes and Likelihood to Use

Change in Likelihood to Use	**Number of Trainees**	**Intended Outcomes**		
		Increased	**Stayed the Same**	**Decreased**
Increased by two levels	3		1 trainee	2 trainees
Increased by one level	5	1 trainee	3 trainees	1 trainee
No change in likelihood to use	8	2 trainees	6 trainees	
Decreased by one level	1		1 trainee	
Totals	17	3 trainees	11 trainees	3 trainees

Within these results, an interesting pattern occurs. For trainees indicating a significant change in the likelihood to use the CRDW (two levels), the majority of trainees indicated fewer intended outcomes of CRDW use after training. For trainees indicating a small change (one level) or no change in likelihood to use, the majority of trainees indicated the same number of intended outcomes in the pre- and post-assessment. One explanation is that trainees with a significant change in likelihood to use were able to focus on a specific task they wanted to complete with the CRDW, so they reduced their number of intended outcomes as a result.

Further research is needed to determine if this pattern continues to emerge with a larger group of trainees. Additionally, future research should include questions to obtain qualitative information about the selection of intended outcomes, such as asking trainees about why they selected

specific outcomes. The goal of this qualitative information would be to understand, more fully, the link between outcomes and likelihood to use. Also, because the training pre- and post-assessments were de-identified for this study, the study researcher was unable to correlate these findings with a trainee's completion of activities within the CRDW. For example, the study researcher could not determine if trainees with a significant change in likelihood and a lower number of intended outcomes completed the CRDW activities related to their intended research outcomes at a greater rate than other trainees. Future research should identify the trainee in the pre- and post-assessments so that these results can be correlated with the results of reviewing audit trails of CRDW activity.

For research hypothesis H8, only 24% of the trainees completed the CRDW actions related to their intended outcomes within 30 days of completing training instead of the hypothesized 50% or greater. One of the most likely explanations for these results is the 30-day window for post-training follow-up was insufficient based on the intended outcomes identified in the feasibility and/or data release agreements by the particular trainees involved. Within these agreements, trainees could select one intended outcome of requesting access to the CRDW tools. Of the 17 trainees who participated, ten trainees (59%) identified creation of a manuscript as the intended outcome within their feasibility/data release agreements and seven (41%) identified "other" intended outcomes. For those identifying other intended outcomes, most involved innovative quality improvement projects or investigation of new care models, which are significant undertakings. For all but one trainee, the outcomes identified were linked to completion of a query (typically to identify an appropriate patient cohort) and then a data extraction (either to obtain data about specific patients or obtain identified patient information used for completion of chart review). While most trainees (82%) completed the query activities within 30 days, only

four completed the data extraction within that timeline. Considering this timeline in the context of the intended outcomes, this timing is logical; teams may initially query to determine if a sufficient patient population exists to perform their studies, but then may need time to refine their research hypotheses and protocols before performing the data extraction. These activities can be time-consuming and may not fit within a 30-day window.

This explanation seems to be supported by the results from two of the components within this study. First, within the 30-day post-training assessment, four of the ten respondents identified a neutral or agreement response to the "I lacked the time to effectively use the CRDW" statement, which may indicate that 30 days is insufficient time to begin working with CRDW tools. In addition, one of the comments from this assessment identified that a trainee had not yet had the time to use the CRDW for his/her work. Second, when reviewing one of the early trainee's achievement of intended outcomes using audit trail data, the trainee completed the needed query within the 30 days post-training, but did not complete the extraction until three months later (which happened to be within the timeline of the audit trails obtained, but outside of the 30-day window for that particular trainee). Fortunately, because the training dates for all of the participating trainees are known, CRDW activity could be reviewed at a later date to determine if a longer study period would change the results for the completion of CRDW activities.

Further evidence for this explanation may be the lack of change in the number of potential outcomes identified by participants before and after training as part of the pre- and post-assessments. In creating the training assessment, a base assumption was that trainees would already have been planning to use the CRDW in every way possible to support their research, implying that trainees would query and extract all data from the CRDW. As a result, the training

assessment measured only whether or not new research outcomes could be envisioned after training.

A better approach may have been to gauge the trainee's planned activities within the CRDW before and after training. For example, a trainee may have identified using the CRDW for creation of a study protocol and grant submission. Prior to starting training, this trainee may have intended to use the CRDW only to obtain cohort information. However, during training, the trainee discovered the potential to extract some of the needed data from the CRDW. While the potential outcome of using the CRDW would not change (still being a study protocol and grant submission), the intended activities to be performed in the CRDW may change.

If trainees' planned activities within the CRDW changed as a result of training, further evidence exists for the 30-day window being too short to assess outcome achievement. In the above example, if the trainee discovered the potential to extract information from the CRDW, the trainee may also have had to reconsider his/her data gathering plans, which could require more than 30 days to complete.

6.4 Research Question #3: Factors Influencing CRDW Use

For research question #3 (Q3), the goal was to explore potential factors that influence trainees' use of the CRDW for clinical and translational research purposes, specifically within 30-45 days of their original training.

Based on the findings for the previous research questions, time may be the most significant factor in CRDW use. From the initial time to find an MCW collaborator to the time needed to

refine data gathering processes, trainees may "feel" that their existing processes require less time than this "new" technology.

Ease of use may also be a factor (and may be correlated to the "time" factor). Within the 30-day post-training follow-up assessment, the ease of use ratings by tool and the responses to the statement "Overall, I found the CRDW easy to use" were somewhat mixed. Of the ten trainees completing the assessment, most agreed that the CRDW was easy to use. Of the six trainees that had used the tool since training, three identified TriNetX, two identified i2b2, and one identified Honest Broker as difficult to use. However, caution was exercised in using the ease of use ratings to determine which tool(s) may be challenging. In this section of the assessment, a "Did not use" option was originally planned, but was mistakenly omitted from the final assessment provided to all of the trainees. Although only one trainee identified this issue in his/her comments on the assessment, it is unclear if the six trainees who responded to this question rated the ease of use based on experience or simply selected a rating to be allowed to complete the remainder of the assessment. As a result, further investigation is required to determine if ease of use is a significant factor.

The type of CRDW support available may also be a factor. Currently, support is available by contacting the BMI Business Analyst. Trainees that pursued this support provided positive comments within the 30-day post-training assessment. Interestingly, two respondents who reported a lack of support did not contact the BMI department. These respondents may have preferred another form of support (such as the ability to search online information about the CRDW or view tutorial videos). As a future step, follow-up could occur with these trainees to determine if additional types of support would be a significant factor in use of the CRDW.

Finally, the data available within the CRDW may be a factor. Of the ten trainees responding to the 30-day post-training assessment, four either agreed with or provided a neutral response to the statement "The data I needed was not available within the CRDW." Should the CRDW lack data needed for research, this may also explain why most trainees did not perform a data extraction from the CRDW and, therefore, did not achieve their intended outcomes as defined in study component #3 (objective measurement of outcomes).

From the 30-day post-training assessment responses, training is not seen as a barrier to CRDW use, as all trainees felt that they received sufficient training.

6.5 Implications of the Findings

The purposes of this study were to identify interventions that increase research team awareness of data warehouses, to assess outcomes of data warehouse training, and to define objective measures for determining if research teams complete the data warehouse activities related to their intended research outcomes. The following section discusses the implications of the study findings on achieving these purposes.

The findings of this study indicate that an awareness presentation intervention can increase research team knowledge of the CRDW, as well as their self-reported likelihood to use. This intervention alone does not appear to result in increased CRDW use within a 90-day period. Several implications can be derived from this finding. First, this finding serves as an initial baseline that can be used for future research timelines. Second, awareness interventions need to be tailored to the target population. Failure to check the awareness presentation content and delivery with representatives from the target population may have reduced the effectiveness of

the intervention. Finally, interventions should include an assessment item to understand the timeline that individuals have for pursing adoption of new technology. For example, individuals who participated in this study's presentations may have done so planning for research that they do not intend to pursue for several months. Understanding these timelines is critical to refining two of the initial objective measurements within this study: the number of access CRDW requests and the number of CRDW training requests.

This study also found that a majority of trainees did not complete the CRDW activities related to their intended research outcomes within 30 days of receiving CRDW training. Two key implications can be identified from this finding. First, the definition of "outcome achievement" needs to be refined with assistance from the target population. Within this study, the definition of "outcome achievement" was derived based on the experience of the key training and support individual for the CRDW. While this definition is a sufficient baseline, the next step should be working with individuals from the target population to obtain their feedback on this definition and what they consider "outcome achievement" related to data warehouse use. A second implication from this finding is that more research is needed to understand the timeline for trainees to complete activities post-training, as the 30-day window appears to be too short based on initial findings. Additionally, the appropriate window may vary depending on the intended research outcomes identified by the trainee.

Similar to other research using the conceptual frameworks that served as the basis of this study, time, ease of use, type of support, and availability of data appear to be potential factors influencing CRDW use and achievement of outcomes. Future work is needed to understand the significance and correlation of these factors. For example, future work could focus on the

creation of a predictive model with a target variable of actual CRDW use and predictor variables that are correlated to actual use such as time, ease of use, type of support, and availability of data. Part of the model creation would be a calculation of the significance of each of these predictors within the model. The significance of the predictive factors could then inform research to influence these predictors in order to increase actual use of the CRDW.

From a research protocol perspective, this study provides new protocols that can serve as a foundation for objective measurement of technology use and adoption. One of the limitations noted in research related to the technology adoption frameworks is the lack of objective measures for actual technology use, because most studies have focused on measurements of intention to use or likelihood to use. Within this study, both the likelihood to use and actual use were measured using techniques that are common regardless of technology platform and can be used retrospectively or prospectively (as done in this study). For example, while CTSA hubs may have chosen different data warehouse implementations, most will have some form of request process and audit trails. The process used in this study can be easily adapted based on a CTSA hub's specific technology implementation and used to review past and current activities within the data warehouse.

Finally, the participation rates within this study were significantly higher than originally anticipated. Several factors seemed to influence these rates. First, the assessment tools were short and simple to complete. For example, both the awareness presentation and training assessments were a page in length and included five questions or less. Second, assessments occurred either during meetings or during a known timeframe (such as 30 days post-training). As a result, participants did not need to find additional time in their schedules and/or could plan

time within their schedules to complete the assessment. Finally, the background and reason for each assessment were clearly described before participants agreed to complete the assessments. This information proved important in participation decisions, as several individuals directly expressed a desire to help with PhD research and/or assist in improving CRDW processes. In future work, these factors should be considered when trying to assess potential participation rates.

6.6 Study Finding Limitations

Although this study has several strengths and contributes to closing the gaps identified within the literature corpus and practical implementation, several limitations of the findings need to be noted.

First, the study period, especially the 30-day window for post-training follow-up, may have been too short to completely assess CRDW use. Within this study, identifying that 30 days may be insufficient is an important finding, as no baseline previously existed. An important next step in exploring this topic should be understanding the "typical" timeline for use of the CRDW within research projects.

Second, although this study exceeded the proposed sample sizes and participation rates, the number of participants was still fairly small and did not include representation from all CTSI partner institutions. This study was a first step in gathering information about CRDW use, but further work is needed to fully reflect the experience of all CTSI partners as well as other CTSA hubs.

Finally, improvements to the research methodology were identified. For example, a question related to planned activities within the CRDW would have assisted in understanding of changes to trainees' plans for using the CRDW within their research. Additionally, refinement of the assessment item related CRDW tool ease of use is needed to gather more accurate information about perceived ease of use. As a result, future research should incorporate these lessons learned to gain further knowledge.

6.7 Future Research

Based on the findings from this study, several areas of future work have been identified.

- Future research needs to focus on refining the objective measurement of "outcome achievement." This research may include focus groups with the target population to refine the current definition, retrospective studies of audit trail data to obtain a timeline and baseline of outcome achievement, and replication of this process focusing on use of different technology platforms.
- Research also needs to occur to better understand the timing of technology adoption. This research may include interviews with individuals after awareness presentations to understand how likelihood to use is linked to the timeline for requesting access and/or training, studies identifying the typical time between access request and training request, and review of CRDW activity to determine the typical activity pattern that appears within audit trails (e.g., do most users access the tool once or multiple times and how much time elapses between uses).
- Further research is needed to understand the significance and correlation of the factors influencing CRDW use. This research may include extending this study to all CTSI partners

and/or replicating this study with other CTSA hubs and providing more detailed assessments/interviews specifically targeting significance and correlation.

6.8 Conclusion

Although significant research is being conducted about and with data warehouses for current and future clinical and translational science purposes, critical gaps exist within the literature corpus real-world implementation and practice. This study focused on three of these gaps: the need to identify interventions that increase research team awareness of data warehouses, the need to assess outcomes of data warehouse training, and the need to define objective measures for data warehouse activities related to research outcome achievement.

To address these gaps, a study was conducted with the CTSI of Southeast Wisconsin focusing on use of the CRDW. The study included an intervention to raise research team awareness of the CRDW, assessments of the awareness intervention and existing CRDW training, and definition and use of objective measures for completion of CRDW activities related to their intended research outcomes after CRDW training. Four study components were included within the study, and these components were based on four conceptual frameworks related to adoption of technology and assessment of learning objectives: Diffusion of Innovation theory, Technology Acceptance Model, DeLone and McLean Information Systems Success Model, and Bloom's Taxonomy of Educational Objectives.

Key study findings suggest that the awareness intervention and training were successful in increasing research team knowledge of the CRDW and self-reported likelihood to use, but this increase did not result in a subsequent increase in CRDW access requests or training requests

within the study period. In addition, only 24% of trainees completed the CRDW activities related to their intended research outcomes within 30 days of training. The time needed for adopting the technology, the ease of use of the CRDW tools, the types of CRDW support available, and the data available within the CRDW may all be factors influencing these outcomes.

Although several study limitations have been noted, this study provides important findings regarding attainment of technology knowledge and its links to actual technology use, the correlation between self-reported likelihood to use and actual technology use, the timeline of technology adoption, and foundational protocols for objective measurement of technology use and adoption. Future research extending these findings should focus on refining the objective measurement of "outcome achievement," understanding the timing of technology adoption, and measuring the significance and correlation of factors influencing CRDW use.

CHAPTER 7: REFERENCES

1. Zerhouni E. The NIH roadmap. *Science*. 2003;302(5642):63-72. doi:10.1126/science.1091867.

2. Zerhouni EA. Clinical research at a crossroads: the NIH roadmap. *J Investig Med*. 2006;54(4):171-173. doi:10.2310/6650.2006.X0016.

3. National Center for Advancing Translational Sciences (NCATS). About the CTSA program. https://ncats.nih.gov/ctsa/about. Accessed April 2, 2018.

4. National Center for Advancing Translational Sciences (NCATS). CTSA program hubs. https://ncats.nih.gov/ctsa/about/hubs. Accessed April 2, 2018.

5. Turisco F, Keogh D, Stubbs C, Glaser J, Crowley WF. Current status of integrating information technologies into the clinical research enterprise within US academic health centers: strategic value and opportunities for investment. *J Investig Med*. 2005;53(8):425-433. doi:10.2310/6650.2005.53806.

6. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Informatics Assoc*. 2013;20(1):144-151. doi:10.1136/amiajnl-2011-000681.

7. Almabhouh A, Saleh AR, Ahmad A. A framework of data warehouse systems success: an empirical study. *Int J Electron Commun Comput Eng*. 2012;3(3):664-677. https://www.researchgate.net/profile/Alaaeddin_Almabhouh/publication/262342311_A_Framework_of_Data_Warehouse_Systems_Success_An_Empirical_Study/links/0deec5375bc1237f46000000.pdf. Accessed March 29, 2018.

8. MacKenzie SL, Wyatt MC, Schuff R, Tenenbaum JD, Anderson N. Practices and perspectives on building integrated data repositories: results from a 2010 CTSA survey. *J Am Med Informatics Assoc*. 2012;19(e1):e119-e124. doi:10.1136/amiajnl-2011-000508.

9. Rogers EM. *Diffusion of Innovations*. 5th ed. New York, NY: Free Press; 2003.

10. Moses H, Matheson DHM, Cairns-Smith S, George BP, Palisch C, Dorsey ER. The anatomy of medical research: US and international comparisons. *JAMA*. 2015;313(2):174-189. doi:10.1001/jama.2014.15939.

11. Squires D, Anderson C, Squires CD, Anderson C. U.S. health care from a global perspective: spending, use of services, prices, and health in 13 countries. *Commonw Fund*. 2015. http://johngarven.com/blog/wp-content/uploads/2017/07/Spending-Use-of-Services-Prices-and-Health-in-13-Countries-The-Commonwealth-Fund.pdf. Accessed April 5, 2018.

12. Sarkar IN. Biomedical informatics and translational medicine. *J Transl Med*. 2010;8:(22) 1-12. doi:10.1186/1479-5876-8-22.

13. Kulikowski CA, Shortliffe EH, Currie LM, et al. AMIA board white paper: definition of biomedical informatics and specification of core competencies for graduate education in the discipline. *J Am Med Informatics Assoc*. 2012;19(6):931-938. doi:10.1136/amiajnl-2012-001053.

14. National Institutes of Health (NIH). History. https://www.nih.gov/about-nih/who-we-are/history. Accessed April 2, 2018.

15. National Institutes of Health (NIH). A short history of the National Institutes of Health. https://history.nih.gov/exhibits/history/index.html. Accessed April 2, 2018.

16. National Institutes of Health (NIH). NIH organization. https://www.nih.gov/about-nih/what-we-do/nih-almanac/nih-organization. Accessed April 2, 2018.

17. Katz IT, Wright AA. Scientific drought, golden eggs, and global leadership - why Trump's NIH funding cuts would be a disaster. *N Engl J Med*. 2017;376(18):1701-1704. doi:10.1056/NEJMp1703734.

18. Johnson JA. NIH funding: FY1994-FY2016. *Congr Res Serv*. 2016:1-5. https://fas.org/sgp/crs/misc/R43341.pdf. Accessed March 30, 2018.

19. Sung NS, William F. Crowley J, Genel M, et al. Central challenges facing the national clinical research enterprise. *JAMA*. 2003;289(10):1278-1287. doi:10.1001/jama.289.10.1278.

20. Woolf SH. The meaning of translational research and why it matters. *JAMA*. 2008;299(2):211-213. doi:10.1001/jama.2007.26.

21. Dzau VJ, Fineberg H V. Restore the US lead in biomedical research. *JAMA*. 2015;313(2):143-144. doi:10.1001/jama.2014.17660.

22. Balas EA, Chapman WW. Road map for diffusion of innovation in health care. *Health Aff*. 2018;37(2):198-204. doi:10.1377/hlthaff.2017.1155.

23. Payne PRO, Embi PJ, Niland J. Foundational biomedical informatics research in the clinical and translational science era: a call to action. *J Am Med Informatics Assoc*. 2010;17(6):615-616. doi:10.1136/jamia.2010.005165.

24. Bernstam E V., Hersh WR, Johnson SB, et al. Synergies and distinctions between computational disciplines in biomedical research: perspective from the Clinical and Translational Science Award program. *Acad Med*. 2009;84(7):964-970. doi:10.1097/ACM.0b013e3181a8144d.

25. Katzan IL, Rudick RA. Time to integrate clinical and research informatics. *Sci Transl Med*. 2012;4(162):(162fs41) 1-4. doi:10.1126/scitranslmed.3004032.

26. Henry J, Pylypchuk Y, Searcy T, Patel V. Adoption of electronic health record systems among U.S. non-federal acute care hospitals: 2008-2015. *ONC Data Br*. 2016;(35).

https://www.healthit.gov/sites/default/files/briefs/2015_hospital_adoption_db_v17.pdf. Accessed March 30, 2018.

27. Zhang NJ, Seblega B, Wan T, Unruh L, Agiro A, Miao L. Health information technology adoption in U.S. acute care hospitals. *J Med Syst*. 2013;37(2):9907 (1-9). doi:10.1007/s10916-012-9907-2.

28. Blumenthal D. Wiring the health system - origins and provisions of a new federal program. *N Engl J Med*. 2011;365(24):2323-2329. doi:10.1056/NEJMsr1110507.

29. Blumenthal D. Implementation of the federal health information technology initiative. *N Engl J Med*. 2011;365(25):2426-2431. doi:10.1056/NEJMsr1112158.

30. Adler-Milstein J, Jha AK. HITECH act drove large gains in hospital electronic health record adoption. *Health Aff*. 2017;36(8):1416-1422. doi:10.1377/hlthaff.2016.1651.

31. Kahn MG, Kaplan D, Sokol RJ, DiLaura RP. Configuration challenges: implementing translational research policies in electronic medical records. *Acad Med*. 2007;82(7):661-669. doi:10.1097/ACM.0b013e318065be8d.

32. Murphy SN, Weber G, Mendis M, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Informatics Assoc*. 2010;17(2):124-130. doi:10.1136/jamia.2009.000893.

33. Bayley KB, Belnap T, Savitz L, Masica AL, Shah N, Fleming NS. Challenges in using electronic health record data for CER: experience of 4 learning organizations and solutions applied. *Med Care*. 2013;51:S80-S86. doi:10.1097/MLR.0b013e31829b1d48.

34. Hripcsak G, Duke JD, Shah NH, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform*. 2015;216:574-578. http://www.ncbi.nlm.nih.gov/pubmed/26262116. Accessed April 3, 2018.

35. Observational Health Data Sciences and Informatics. OHDSI - Observational Health Data Sciences and Informatics. https://ohdsi.org/. Accessed April 3, 2018.

36. Beresniak A, Schmidt A, Proeve J, et al. Cost-benefit assessment of using electronic health records data for clinical research versus current practices: Contribution of the electronic health records for clinical research (EHR4CR) European project. *Contemp Clin Trials*. 2016;46:85-91. doi:10.1016/J.CCT.2015.11.011.

37. De Moor G, Sundgren M, Kalra D, et al. Using electronic health records for clinical research: the case of the EHR4CR project. *J Biomed Inform*. 2015;53:162-173. doi:10.1016/J.JBI.2014.10.006.

38. The European Institute for Innovation through Health Data (i~HD). About i~HD. http://www.i-hd.eu/index.cfm/about/about-i-hd/. Accessed April 3, 2018.

39. Weber SC, Seto T, Olson C, Kenkare P, Kurian AW, Das AK. Oncoshare: lessons learned from building an integrated multi-institutional database for comparative effectiveness research. *AMIA Annu Symp proceedings AMIA Symp*. 2012;2012:970-978. http://www.ncbi.nlm.nih.gov/pubmed/23304372. Accessed April 3, 2018.

40. Stanford Medicine. The oncoshare project. http://med.stanford.edu/oncoshare.html. Accessed April 3, 2018.

41. Sharda R, Delen D, Turban E. *Business Intelligence and Analytics: Systems for Decision Support*. 10th ed. Pearson; 2014.

42. Huser V, Kayaalp M, Dodd ZA, Cimino JJ. Piloting a deceased subject integrated data repository and protecting privacy of relatives. *AMIA . Annu Symp proceedings AMIA Symp*. 2014;2014:719-728. http://www.ncbi.nlm.nih.gov/pubmed/25954378. Accessed April 3, 2018.

43. Pakhomov S, Weston SA, Jacobsen SJ, Chute CG, Meverden R, Roger VL. Electronic medical records for clinical research: application to the identification of heart failure. *Am J Manag Care*. 2007;13(6):281-288. http://www.ncbi.nlm.nih.gov/pubmed/17567225. Accessed April 3, 2018.

44. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. *BMJ*. 2009;338:(b81) 1-9. http://www.ncbi.nlm.nih.gov/pubmed/19174434. Accessed April 3, 2018.

45. Arts DGT, de Keizer NF, Scheffer G-J. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. *J Am Med Informatics Assoc*. 2002;9(6):600-611. doi:10.1197/jamia.M1087.

46. Liaw ST, Rahimi A, Ray P, et al. Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. *Int J Med Inform*. 2013;82(1):10-24. doi:10.1016/J.IJMEDINF.2012.10.001.

47. Johnson SG, Speedie S, Simon G, Kumar V, Westra BL. A data quality ontology for the secondary use of EHR data. *AMIA Annu Symp proceedings AMIA Symp*. 2015;2015:1937-1946. http://www.ncbi.nlm.nih.gov/pubmed/26958293. Accessed April 3, 2018.

48. Kahn MG, Brown JS, Chun AT, et al. Transparent reporting of data quality in distributed data networks. *EGEMS (Washington, DC)*. 2015;3(1):1052. doi:10.13063/2327-9214.1052.

49. Cook JA, Collins GS. The rise of big clinical databases. *Br J Surg*. 2015;102(2):e93-e101. doi:10.1002/bjs.9723.

50. Richesson RL, Horvath MM, Rusincovitch SA. Clinical research informatics and electronic health record data. *Yearb Med Inform*. 2014;9(1):215-223. doi:10.15265/IY-2014-0009.

51. Botsis T, Hartvigsen G, Chen F, Weng C. Secondary use of EHR: data quality issues and informatics opportunities. *AMIA Jt Summits Transl Sci proceedings AMIA Jt Summits Transl Sci*. 2010;2010:1-5. http://www.ncbi.nlm.nih.gov/pubmed/21347133. Accessed April 3, 2018.

52. Sáez C, Martínez-Miranda J, Robles M, García-Gómez JM. Organizing data quality assessment of shifting biomedical data. *Stud Health Technol Inform*. 2012;180:721-725. http://www.ncbi.nlm.nih.gov/pubmed/22874286. Accessed April 3, 2018.

53. Richesson RL, Hammond WE, Nahm M, et al. Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH health care systems collaboratory. *J Am Med Informatics Assoc*. 2013;20(e2):e226-e231. doi:10.1136/amiajnl-2013-001926.

54. Danciu I, Cowan JD, Basford M, et al. Secondary use of clinical data: the Vanderbilt approach. *J Biomed Inform*. 2014;52:28-35. doi:10.1016/J.JBI.2014.02.003.

55. Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care*. 2013;51:S30-S37. doi:10.1097/MLR.0b013e31829b1dbd.

56. Weiskopf NG, Rusanov A, Weng C. Sick patients have more data: the non-random completeness of electronic health records. *AMIA Annu Symp proceedings AMIA Symp*. 2013;2013:1472-1477. http://www.ncbi.nlm.nih.gov/pubmed/24551421. Accessed April 3, 2018.

57. Köpcke F, Trinczek B, Majeed RW, et al. Evaluation of data completeness in the electronic health record for the purpose of patient recruitment into clinical trials: a retrospective analysis of element presence. *BMC Med Inform Decis Mak*. 2013;13(1):(37) 1-8. doi:10.1186/1472-6947-13-37.

58. Coorevits P, Sundgren M, Klein GO, et al. Electronic health records: new opportunities for clinical research. *J Intern Med*. 2013;274(6):547-560. doi:10.1111/joim.12119.

59. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. *J Biomed Inform*. 2013;46(5):830-836. doi:10.1016/J.JBI.2013.06.010.

60. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012;13(6):395-405. doi:10.1038/nrg3208.

61. Safran C, Bloomrosen M, Hammond WE, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association white paper. *J Am Med Informatics Assoc*. 2007;14(1):1-9. doi:10.1197/jamia.M2273.

62. National Institutes of Health (NIH). The 21st Century Cures Act. https://www.nih.gov/research-training/medical-research-initiatives/cures. Accessed April 3, 2018.

63. Rindfleisch TC, Brutlag DL. Directions for clinical research and genomic research into the next decade: implications for informatics. *J Am Med Informatics Assoc*. 1998;5(5):404-411. doi:10.1136/jamia.1998.0050404.

64. Marsolo K, Spooner SA. Clinical genomics in the world of the electronic health record. *Genet Med*. 2013;15(10):786-791. doi:10.1038/gim.2013.88.

65. Canuel V, Rance B, Avillach P, Degoulet P, Burgun A. Translational research platforms integrating clinical and omics data: a review of publicly available solutions. *Brief Bioinform*. 2015;16(2):280-290. doi:10.1093/bib/bbu006.

66. Kohane IS. Ten things we have to do to achieve precision medicine. *Science*. 2015;349(6243):37-38. doi:10.1126/science.aab1328.

67. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med*. 2015;372(9):793-795. doi:10.1056/NEJMp1500523.

68. Shah SH, Arnett D, Houser SR, et al. Opportunities for the cardiovascular community in the precision medicine initiative. *Circulation*. 2016;133(2):226-231. doi:10.1161/CIRCULATIONAHA.115.019475.

69. Hudson KL, Collins FS. The 21st Century Cures Act - a view from the NIH. *N Engl J Med*. 2017;376(2):111-113. doi:10.1056/NEJMp1615745.

70. National Institutes of Health (NIH). Cancer Moonshot[SM]. https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative. Accessed April 3, 2018.

71. National Institutes of Health (NIH). Regenerative medicine innovation project. https://www.nih.gov/rmi. Accessed April 3, 2018.

72. Mirnezami R, Nicholson J, Darzi A. Preparing for precision medicine. *N Engl J Med*. 2012;366(6):489-491. doi:10.1056/NEJMp1114866.

73. Tenenbaum JD, Avillach P, Benham-Hutchins M, et al. An informatics research agenda to support precision medicine: seven key areas. *J Am Med Informatics Assoc*. 2016;23(4):791-795. doi:10.1093/jamia/ocv213.

74. Nambiar R, Bhardwaj R, Sethi A, Vargheese R. A look at challenges and opportunities of big data analytics in healthcare. In: *2013 IEEE International Conference on Big Data*. IEEE; 2013:17-22. doi:10.1109/BigData.2013.6691753.

75. Simpao AF, Ahumada LM, Gálvez JA, Rehman MA. A review of analytics and clinical informatics in health care. *J Med Syst*. 2014;38(4):(45) 1-7. doi:10.1007/s10916-014-0045-x.

76. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff*.

2014;33(7):1123-1131. doi:10.1377/hlthaff.2014.0041.

77. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Heal Inf Sci Syst*. 2014;2:(3) 1-10. doi:10.1186/2047-2501-2-3.

78. Adam NR, Wieder R, Ghosh D. Data science, learning, and applications to biomedical and health sciences. *Ann N Y Acad Sci*. 2017;1387(1):5-11. doi:10.1111/nyas.13309.

79. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. *MIS Q*. 2003;27(3):425-478. doi:10.2307/30036540.

80. Hameed MA, Counsell S, Swift S. A conceptual model for the process of IT innovation adoption in organizations. *J Eng Technol Manag*. 2012;29(3):358-390. doi:10.1016/j.jengtecman.2012.03.007.

81. Greenhalgh T, Robert G, Bate P, Macfarlane F, Kyriakidou O. *Diffusion of Innovations in Health Service Organisations: A Systematic Literature Review*. Malden, MA: Blackwell Publishing Ltd; 2005.

82. Dorr DA, Cohen DJ, Adler-Milstein J. Data-driven diffussion of innovations: successes and challenges in 3 large-scale innovative delivery models. *Health Aff*. 2018;37(2):257-265. doi:10.1377/hlthaff.2017.1133.

83. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. *Manage Sci*. 1989;35(8):982-1003. doi:10.1287/mnsc.35.8.982.

84. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. *Manage Sci*. 2000;46(2):186-204. doi:10.1287/mnsc.46.2.186.11926.

85. Chau PYK, Hu PJ-H. Investigating healthcare professionals' decisions to accept telemedicine technology: an empirical test of competing theories. *Inf Manag*. 2002;39(4):297-311. doi:10.1016/S0378-7206(01)00098-2.

86. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q*. 1989;13(3):319-340. doi:10.2307/249008.

87. Morton ME, Wiedenbeck S. A framework for predicting EHR adoption attitudes: a physician survey. *Perspect Heal Inf Manag*. 2009;6(Fall):1-19. http://www.ncbi.nlm.nih.gov/pubmed/20169013. Accessed April 1, 2018.

88. Chismar WG, Wiley-Patton S. Does the extended technology acceptance model apply to physicians. In: *Proceedings of the 36th Annual Hawaii International Conference on System Sciences, 2003*. IEEE; 2003. doi:10.1109/HICSS.2003.1174354.

89. Holden RJ, Karsh B-T. The technology acceptance model: its past and its future in health care. *J Biomed Inform*. 2010;43(1):159-172. doi:10.1016/J.JBI.2009.07.002.

90. Hu PJ, Chau PYK, Sheng ORL, Tam KY. Examining the technology acceptance model using physician acceptance of telemedicine technology. *J Manag Inf Syst*. 1999;16(2):91-112. doi:10.2307/40398433.

91. Yi MY, Jackson JD, Park JS, Probst JC. Understanding information technology acceptance by individual professionals: Toward an integrative view. *Inf Manag*. 2006;43(3):350-363. doi:10.1016/J.IM.2005.08.006.

92. Wu J-H, Wang S-C, Lin L-M. Mobile computing acceptance factors in the healthcare industry: a structural equation model. *Int J Med Inform*. 2007;76(1):66-77. doi:10.1016/J.IJMEDINF.2006.06.006.

93. Wu J-H, Shen W-S, Lin L-M, Greenes RA, Bates DW. Testing the technology acceptance model for evaluating healthcare professionals' intention to use an adverse event reporting system. *Int J Qual Heal Care*. 2007;20(2):123-129. doi:10.1093/intqhc/mzm074.

94. Aggelidis VP, Chatzoglou PD. Using a modified technology acceptance model in hospitals. *Int J Med Inform*. 2009;78(2):115-126. doi:10.1016/J.IJMEDINF.2008.06.006.

95. Kowitlawakul Y. The technology acceptance model: predicting nurses' intention to use telemedicine technology (eICU). *CIN Comput Informatics, Nurs*. 2011;29(7):411-418. doi:10.1097/NCN.0b013e3181f9dd4a.

96. Gagnon M-P, Ghandour EK, Talla PK, et al. Electronic health record acceptance by physicians: testing an integrated theoretical model. *J Biomed Inform*. 2014;48:17-27. doi:10.1016/J.JBI.2013.10.010.

97. Abdekhoda M, Ahmadi M, Gohari M, Noruzi A. The effects of organizational contextual factors on physicians' attitude toward adoption of electronic medical records. *J Biomed Inform*. 2015;53:174-179. doi:10.1016/J.JBI.2014.10.008.

98. Moore GC, Benbasat I. Development of an instrument to measure the perceptions of adopting an information technology innovation. *Inf Syst Res*. 1991;2(3):192-222. doi:10.1287/isre.2.3.192.

99. Chau PYK, Hu PJ-H. Information technology acceptance by individual professionals: a model comparison approach. *Decis Sci*. 2001;32(4):699-719. doi:10.1111/j.1540-5915.2001.tb00978.x.

100. King WR, He J. A meta-analysis of the technology acceptance model. *Inf Manag*. 2006;43(6):740-755. doi:10.1016/J.IM.2006.05.003.

101. Zhang H, Cocosila M, Archer N. Factors of adoption of mobile information technology by homecare nurses. *CIN Comput Informatics, Nurs*. 2010;28(1):49-56. doi:10.1097/NCN.0b013e3181c0474a.

102. Pai F-Y, Huang K-I. Applying the technology acceptance model to the introduction of healthcare information systems. *Technol Forecast Soc Change*. 2011;78(4):650-660.

doi:10.1016/J.TECHFORE.2010.11.007.

103. Keil M, Beranek PM, Konsynski BR. Usefulness and ease of use: field study evidence regarding task considerations. *Decis Support Syst*. 1995;13(1):75-91. doi:10.1016/0167-9236(94)E0032-M.

104. Turner M, Kitchenham B, Brereton P, Charters S, Budgen D. Does the technology acceptance model predict actual use? A systematic literature review. *Inf Softw Technol*. 2010;52(5):463-479. doi:10.1016/J.INFSOF.2009.11.005.

105. DeLone WH, McLean ER. Information systems success: the quest for the dependent variable. *Inf Syst Res*. 1992;3(1):60-95. doi:10.1287/isre.3.1.60.

106. Delone WH, McLean ER. The DeLone and McLean model of information systems success: a ten-year update. *J Manag Inf Syst*. 2003;19(4):9-30. doi:10.1080/07421222.2003.11045748.

107. Petter S, DeLone W, McLean ER. Information systems success: the quest for the independent variables. *J Manag Inf Syst*. 2013;29(4):7-62. doi:10.2753/MIS0742-1222290401.

108. Petter S, Delone W, Mclean E. Measuring information systems success: models, dimensions, measures, and interrelationships. *Eur J Inf Syst*. 2008;17(3):236-263. doi:10.1057/ejis.2008.15.

109. Wixom BH, Watson HJ. An empirical investigation of the factors affecting data warehousing success. *MIS Q*. 2001;25(1):17-41. doi:10.2307/3250957.

110. Hwang M, Xu H. The effect of implementation factors on data warehousing success: an exploratory study. *J Information, Inf Technol Organ*. 2007;2:1-14. https://digitalcommons.butler.edu/cob_papers/77. Accessed March 29, 2018.

111. Spruit M, Sacu C. DWCMM: The data warehouse capability maturity model. *J Univers Comput Sci*. 2015;21(11):1508-1534. doi:10.3217/jucs-021-11-1508.

112. Wears RL, Berg M. Computer technology and clinical work: still waiting for godot. *JAMA*. 2005;293(10):1261-1263. doi:10.1001/jama.293.10.1261.

113. Shin B. An exploratory investigation of system success factors in data warehousing. *J Assoc Inf Syst*. 2003;4:141-168. http://web.a.ebscohost.com.ezproxy.lib.uwm.edu/ehost/detail/detail?vid=4&sid=420a399d-915d-4bf5-b9d3-674027399133%40sessionmgr4008&bdata=JkF1dGhUeXBlPWlwLHVpZCZzaXRlPWVob3N0LWxpdmUmc2NvcGU9c2l0ZQ%3D%3D#AN=16701845&db=buh. Accessed March 30, 2018.

114. Chenoweth T, Corral K, Demirkan H. Seven key interventions for data warehouse success. *Commun ACM*. 2006;49(1):114-119. doi:10.1145/1107458.1107464.

115. Farrokhi V, Pokoradi L. Organizational and technical factors for implementing business intelligence. *Ann Oradea Univ Fascicle Manag Technol Eng*. 2013;XII (XXII)(1):75-78. https://doaj.org/article/24ab4adce83b4731a2d8c70335bcab47. Accessed March 29, 2018.

116. Cato P, Golzer P, Demmelhuber W. An investigation into the implementation factors affecting the success of big data systems. In: *2015 11th International Conference on Innovations in Information Technology (IIT)*. IEEE; 2015:134-139. doi:10.1109/INNOVATIONS.2015.7381528.

117. Yeoh W, Popovič A. Extending the understanding of critical success factors for implementing business intelligence systems. *J Assoc Inf Sci Technol*. 2016;67(1):134-147. doi:10.1002/asi.23366.

118. Lorenzi NM, Riley RT, Blyth AJC, Southon G, Dixon BJ. Antecedents of the people and organizational aspects of medical informatics: review of the literature. *J Am Med Informatics Assoc*. 1997;4(2):79-93. doi:10.1136/jamia.1997.0040079.

119. Krathwohl DR. A revision of Bloom's taxonomy: an overview. *Theory Pract*. 2002;41(4):212-218. http://web.a.ebscohost.com.ezproxy.lib.uwm.edu/ehost/detail/detail?vid=0&sid=1131bc05-7d6f-4f80-a337-47d29137b079%40sessionmgr4009&bdata=JkF1dGhUeXBlPWlwLHVpZCZzaXRlPWVob3N0LWxpdmUmc2NvcGU9c2l0ZQ%3D%3D#db=aph&AN=8550701. Accessed April 3, 2018.

120. Su WM, Osisek PJ, Starnes B. Applying the revised Bloom's taxonomy to a medical-surgical nursing lesson. *Nurse Educ*. 2004;29(3):116-120. http://ovidsp.tx.ovid.com.ezproxy.lib.uwm.edu/sp-3.28.0a/ovidweb.cgi?QS2=434f4e1a73d37e8c6114a63cc85fea09b53aa2625be8650d2c61e829aa0e44b26c043c586cf610efd3c659130c11289a503dfc6eb2043a19286edec3d2956d5594a9df5ab05771b4926c0d1d2edabafb5da0225c89d80212f5cb09. Accessed April 4, 2018.

121. McNeil RC. A program evaluation model: using Bloom's taxonomy to identify outcome indicators in outcomes-based program evaluations. *J Adult Educ*. 2011;40(2):24-29. http://web.b.ebscohost.com.ezproxy.lib.uwm.edu/ehost/detail/detail?vid=0&sid=b7672a6c-4879-4959-924a-1f9401c15bc6%40sessionmgr104&bdata=JkF1dGhUeXBlPWlwLHVpZCZzaXRlPWVob3N0LWxpdmUmc2NvcGU9c2l0ZQ%3D%3D#AN=85651355&db=ehh. Accessed April 4, 2018.

122. Clinical & Translational Science Institute of Southeast Wisconsin. About CTSI. https://ctsi.mcw.edu/about/history/. Accessed April 2, 2018.

123. Clinical & Translational Science Institute of Southeast Wisconsin. Clinical & Translational Research Informatics. https://ctri.mcw.edu/. Accessed April 2, 2018.

124. Clinical & Translational Science Institute of Southeast Wisconsin Biomedical Informatics. Clinical Research Data Warehouse. https://ctri.mcw.edu/cda/i2b2-cohort-discovery-tool/.

Accessed April 2, 2018.

125. Clinical and Translational Science Institute (CTSI) of Southeast Wisconsin. REDCap. https://redcap.mcw.edu/. Accessed December 31, 2018.

126. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381. doi:10.1016/J.JBI.2008.08.010.

APPENDIX A: INSTITUTIONAL REVIEW BOARD APPROVAL LETTER

The following pages show the institutional review board (IRB) approval letter for this project.

Medical College of Wisconsin /
Froedtert Hospital
Institutional Review Board

To: Reza Shaker, MD
Bradley Taylor
CC: Kristen Osinski
Jake Luo, PhD
Katie McCarthy

Date: 7/26/2018

Re: Project Title: THE ASSESSMENT OF TECHNOLOGY ADOPTION INTERVENTIONS AND OUTCOME ACHIEVEMENT RELATED TO THE USE OF A CLINICAL RESEARCH DATA WAREHOUSE
PRO ID: PRO00032267

IRB Approval Date: 7/25/2018

IRB Expiration Date: 7/24/2019

The MCW/FH Institutional Review Board #5 has granted approval for the above-referenced submission in accordance with 45 CFR 46.111 by expedited review, Categories 5 and 7.

Approval has been granted for the following institutions:

Medical College of Wisconsin - Milwaukee Campus
Milwaukee School of Engineering†
UW-Milwaukee†

The IRB has granted approval of an alteration of the informed consent requirements at 45 CFR 46.116 for the first, second, and fourth components of this project. However, you must use the IRB-approved consent language.

The IRB has granted approval of a waiver of informed consent requirements at 45 CFR 46.116 for the third component.

The items listed below were submitted and reviewed when the IRB approved this submission. Research must be conducted according to the IRB-approved documents listed below:

LET-PRO00032667 Approval Letter for Med Resident Fellow
LET-PRO00032667 Approval Letter for Grad Students
LET-PRO00032667 Approval Letter for Med Students
SUR-PRO00032267 Comp 4 Follow-up Assessment
PRO-PRO00032267 Protocol Summary
ICF-PRO00032267 Comp 1 Informational Letter
ICF-PRO00032267 Comp 2 Informational Letter
ICF-PRO00032267 Comp 4 Informational Letter
SUR-PRO00032267 Comp 2 Training Assessment
SUR-PRO00032267 Comp 1 Awareness Assessment

Any and all proposed changes to this submission must be reviewed and approved by the IRB prior to implementation. When it is necessary to eliminate hazards to subjects, changes may be made first. This should be followed promptly via a Reportable Event for a protocol deviation and Amendment.

In accordance with federal regulations, continuing approval for this submission is required prior to 7/24/2019. The Continuing Progress Report (CPR) must be received by the IRB with enough time to allow for review and approval prior to the expiration date. Failure to submit the CPR in a timely manner may result in the expiration of IRB approval.

A Final Report must be submitted to the IRB within 30 days of when all project activities and data analysis have been completed.

All Unanticipated Problems Involving Risks to Subjects or Others (UPIRSO) must be reported promptly to the MCW/FH IRB according to the IRB Standard Operating Procedures.

If your project involves the use of any Froedtert Health resource such as, space, staff services, supplies/equipment or any ancillary services - lab, pharmacy, radiology, protected health/billing information or specimen requests, OCRICC approval is required before beginning any research activity at those sites.

If you have any questions, please contact the IRB Coordinator II for this IRB Committee, Dana Hirn Mueller, at 414-955-8601 or dhirn@mcw.edu.

Sincerely,

Kathryn Gaudreau
David Clark, PhD
IRB Chair
MCW/FH Institutional Review Board #5

APPENDIX B: AWARENESS PRESENTATION CONTENT

The following pages show the content covered during the awareness presentation (study component #1) for this research study. The content and slides were developed by the BMI Business Analyst, Kristen Osinski, and reviewed by the study researcher.

Clinical & Translational Science Institute
of Southeast Wisconsin

Biomedical Informatics
Clinical Research Data Warehouse

- Introduction
- CRDW Content, Tools, Process
- Access to CRDW
- Questions

Kristen Osinski, MS
Biomedical Informatics Business Analyst

Clinical & Translational Science Institute
of Southeast Wisconsin

Regional consortium of 8 Milwaukee-area Health Care and Academic institutions partnering to advance health through research and discovery

Funded through NIH NCATS CTSA program

Clinical & Translational Science Institute
of Southeast Wisconsin

Greater Plains Collaborative (GPC) funded by PCORI CDRN program
Regional consortium of academic medical centers across many states committed to a shared vision of improving health care delivery

Clinical & Translational Science Institute
of Southeast Wisconsin

Definition of Biomedical Informatics

Biomedical informatics (BMI) is the interdisciplinary field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving and decision making, motivated by efforts to improve human health.

https://www.amia.org/biomedical-informatics-core-competencies

CTSI's Biomedical Informatics team (ctri.mcw.edu)
Aggregates and transforms health system data from various source systems and formats to research-ready state
- ✓ Supports data harmonization & site interoperability
- ✓ Develops and implements data sharing solutions
- ✓ Leverages cross-institutional relationships to expand knowledge sharing and scientific collaboration

Clinical & Translational Science Institute
of Southeast Wisconsin

Biomedical Informatics

Role of Business Analyst

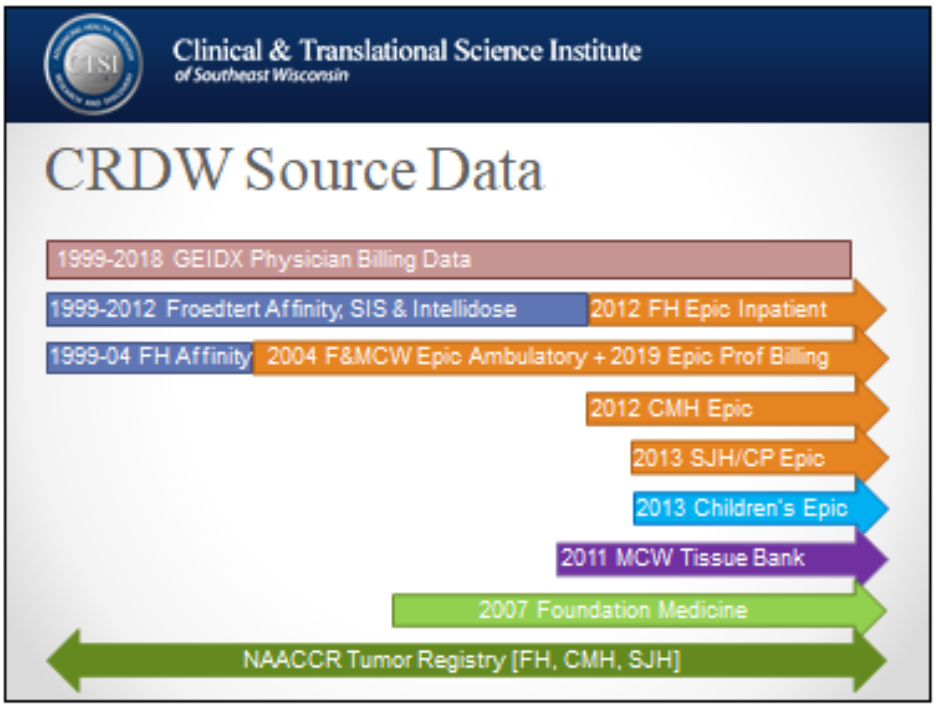
- Customer intake & consultations
- Facilitate access to CRDW
- Train users (1-on-1 or small team)
  - ✓ Provide project-specific training
  - ✓ Develop user guide
- Support users navigating CRDW tools
  - ✓ Find cohort characterization criteria
  - ✓ Design efficient queries
- Translate data from clinical/financial sources to programmers who transform it for research tools

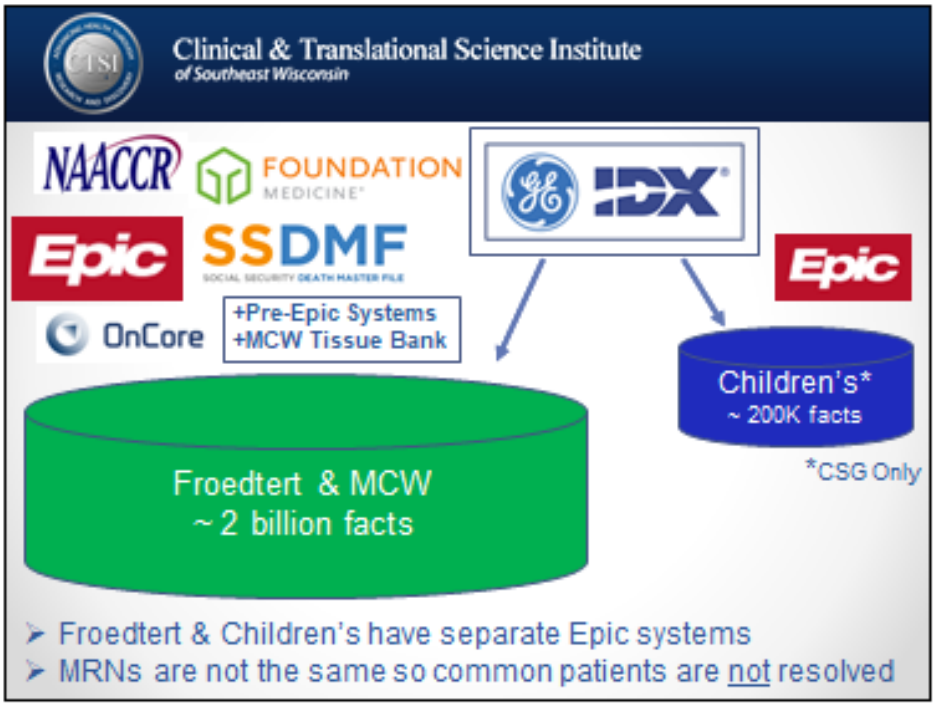
Clinical & Translational Science Institute
of Southeast Wisconsin

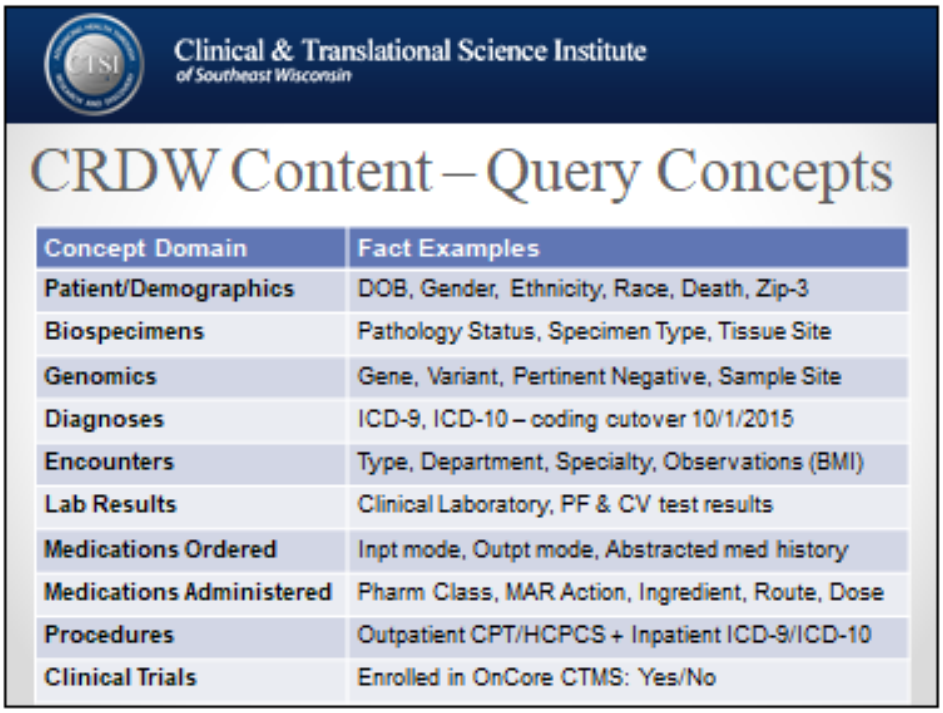
CRDW Content

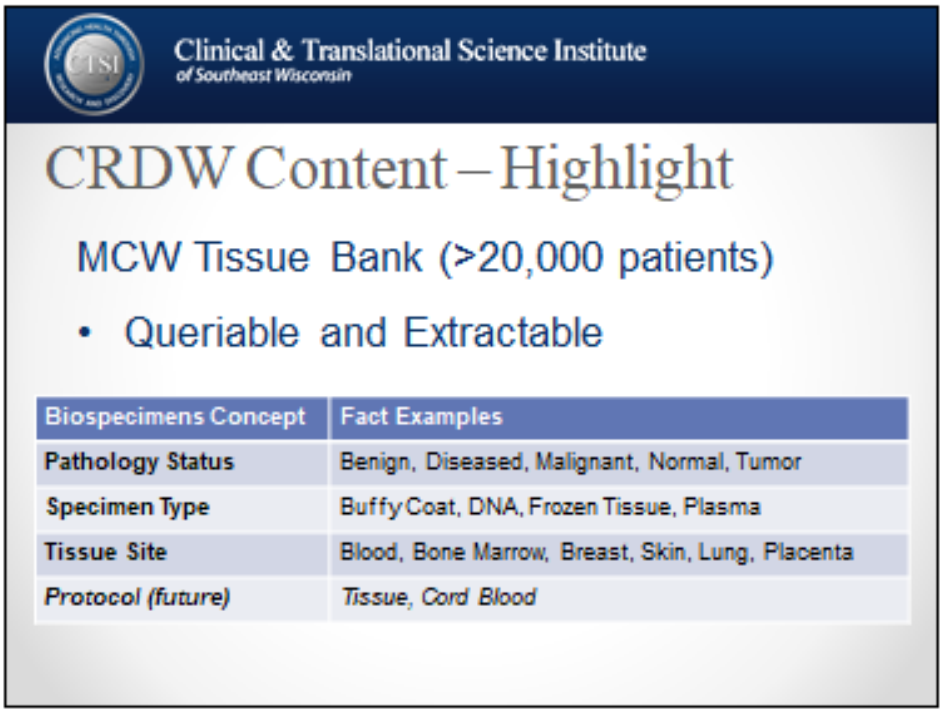
Data transformation

Clinical Use Context	Research Use Context
Individual patient	Aggregated patient cohorts
Real-time access (supports workflow)	Delayed access (supports 2ndary use)
Implicit (patient chart is pre-programmed to enable rapid human inference)	Explicit (total re-assembly of data w/date-tagging, standardization, and coding)
Identified	De-identified and Identified
Treatment, Payment, Operations	Clinical & Translational Research / Quality Improvement (ops research)









Clinical & Translational Science Institute
of Southeast Wisconsin

CRDW Tools – HIPAA Context

- **Query (i2b2 & TriNetX)**
  - ✓ Always de-identified
  - ✓ Unlimited use

- **Data Extraction (Honest Broker)**

De-identified data tables

- Unlimited use, no IRB protocol required

Identified data tables

- Limited access, requires IRB approval

Clinical & Translational Science Institute
of Southeast Wisconsin

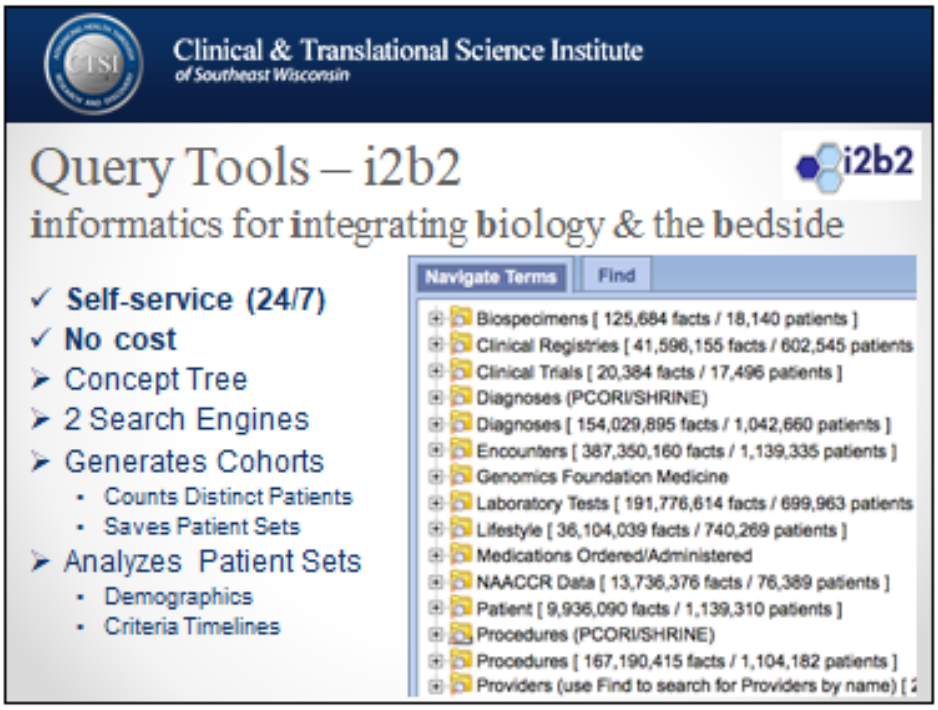
Query Tools – i2b2 & TriNetX

Use Cases

- ✓ Hypothesis Development/Feasibility
- ✓ Cohort Counts for grant submission
- ✓ Cohort Definition for data analysis

Benefits

- ✶ **Minutes** of query time vs. hours/weeks of manual chart review to find eligible subjects
- ✶ No HIPAA risk because data is anonymous



Clinical & Translational Science Institute
of Southeast Wisconsin

Data Extract Tools – Honest Broker

Use Cases

- ✓De-identified retrospective analysis
- ✓Identified chart data collection
- ✓Identified prospective study recruitment

Benefits

- ✶ **Minutes** of data extraction time vs. hours, weeks or months of manual chart review to gather research variables
- ✶ HIPAA-compliant, IRB-approved

Clinical & Translational Science Institute
of Southeast Wisconsin

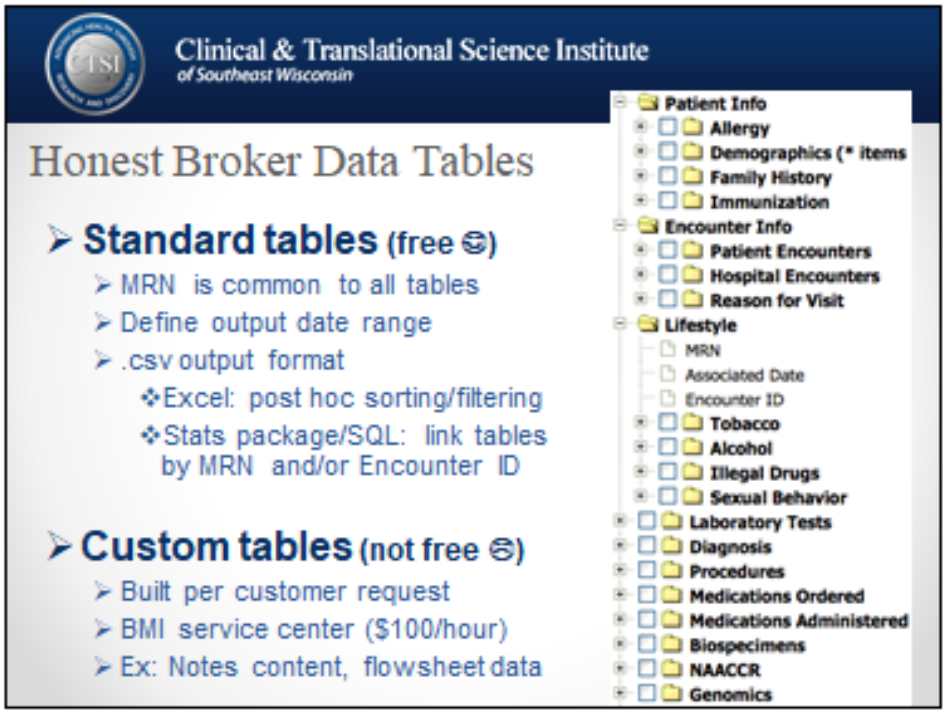
Data Extract Tools – Honest Broker

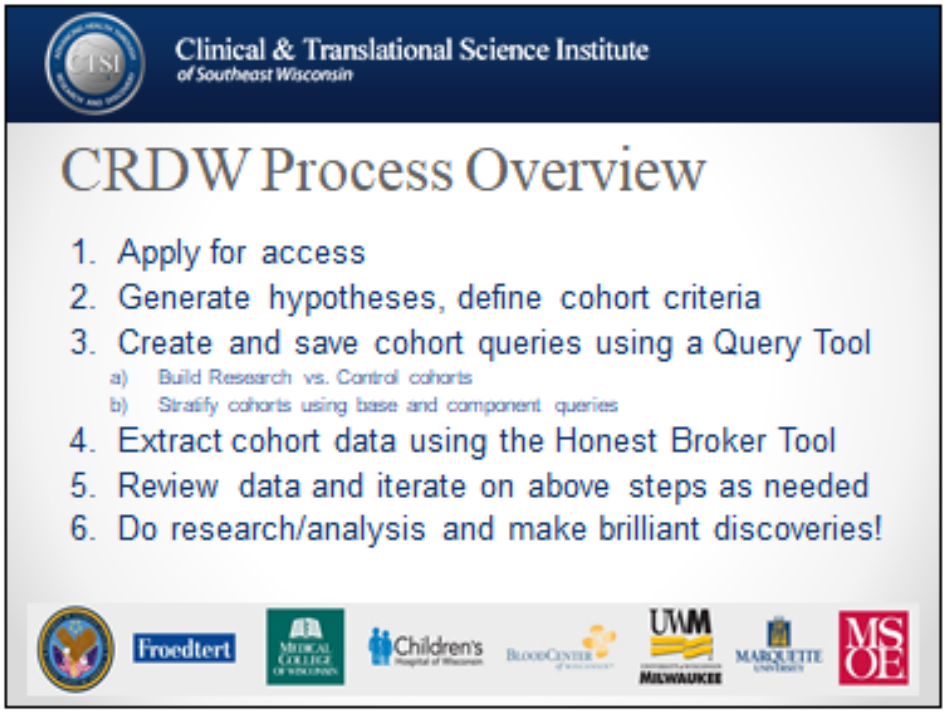
Query Tool Integration

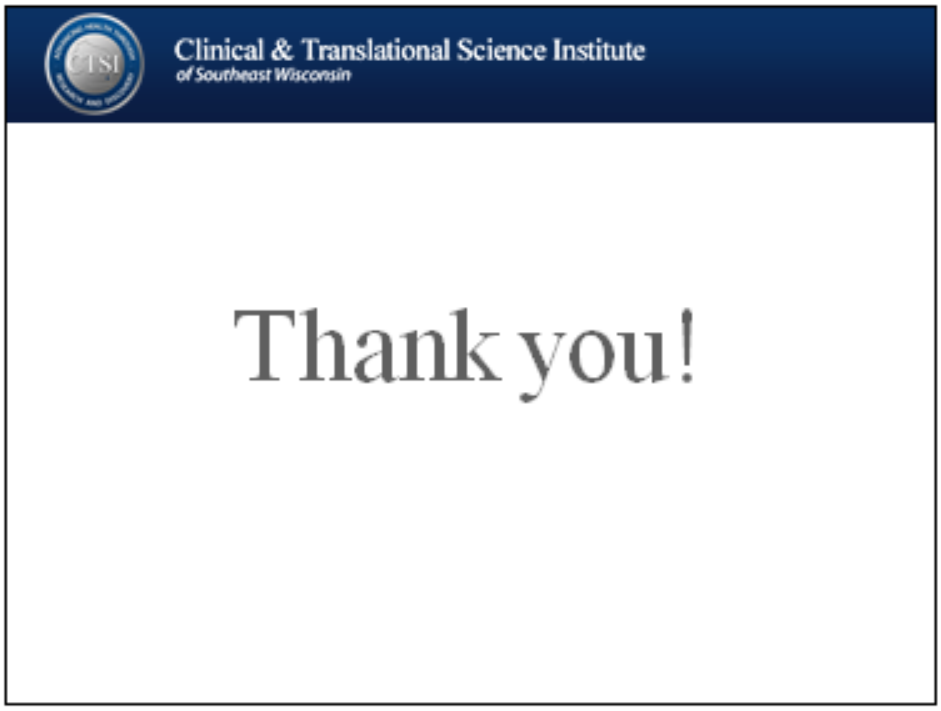
- ➢ i2b2 – Auto link to HB
- ➢ TriNetX – Export IDs to HB

Self-Service Data Extraction

- ➢ De-identified (cohorts > 10)
- ➢ Identified (any cohort size)
- ➢ .csv files delivered via email







APPENDIX C: AWARENESS PRESENTATION ASSESSMENT TOOL

The following page shows the assessment tool used before and after the awareness presentation (study component #1).

Biomedical Informatics CRDW Presentation – Assessment

1) Which of the following information is currently available in the standard tables within the clinical research data warehouse (CRDW) and can be queried using both i2b2 and TriNetX? Please check all that apply.

- ☐ Patient race and ethnicity
- ☐ Progress notes
- ☐ Lab tests
- ☐ Flowsheet data
- ☐ Diagnosis
- ☐ Medications

2) To use the CRDW to access de-identified information, you need to obtain IRB approval. Please select one answer.

- ☐ True
- ☐ False

3) Which of the following Biomedical Informatics (BMI) services do you think you will be charged for? Please check all that apply.

- ☐ One-on-one training for CRDW tools
- ☐ General support from BMI analysts (such as answering questions)
- ☐ Creation of custom tables within Honest Broker
- ☐ Support for creating queries with standard CRDW tools

4) How likely are you to use the CRDW? Please select one answer.

☐ Very likely ☐ Likely ☐ Unsure ☐ Unlikely* ☐ Very unlikely*

*** If you answered "unlikely" or "very unlikely" to the above question, please provide reason(s) below.**

_____

APPENDIX D: CRDW TRAINING ASSESSMENT TOOL

The following page shows the assessment tool used before and after the one-on-one/small group CRDW training (study component #2).

Biomedical Informatics CRDW Training – Assessment

1) Which of the following information is currently available in the standard tables within the clinical research data warehouse (CRDW) and can be queried using both i2b2 and TriNetX? Please check all that apply.

- ☐ Patient race and ethnicity
- ☐ Progress notes
- ☐ Lab tests
- ☐ Flowsheet data
- ☐ Diagnosis
- ☐ Medications

2) To use the CRDW to access de-identified information, you are required to obtain IRB approval. Please select one answer.

- ☐ True
- ☐ False

3) Which of the following Biomedical Informatics (BMI) services do you think you will be charged for? Please check all that apply.

- ☐ One-on-one training for CRDW tools
- ☐ General support from BMI analysts (such as answering questions)
- ☐ Creation of custom tables within Honest Broker
- ☐ Support for creating queries with standard CRDW tools

4) How do you intend to use the CRDW? Please check all that apply.

- ☐ Protocol
- ☐ Poster presentation
- ☐ Grant submission
- ☐ Abstract
- ☐ Manuscript
- ☐ Other (please describe): _____

5) How likely are you to use the CRDW? Please select one answer.

☐ Very likely ☐ Likely ☐ Unsure ☐ Unlikely* ☐ Very unlikely*

*** If you answered "unlikely" or "very unlikely" to the above question, please provide reason(s) below.**

_____

APPENDIX E: 30-DAY POST-TRAINING ASSESSMENT TOOL

The following pages show the assessment tool used 30 days after training was received (study component #4). Although participants had the option of completing this assessment via electronic survey or phone or in-person interview, the content of this assessment is shown as formatted for the electronic survey in REDCap (with section numbers and additional information added to describe the flow of the survey questions).

Section #1: Use of the CRDW Tools

Questions	
Since receiving clinical research data warehouse (CRDW) training, have you used any of the CRDW tools (i2b2, TriNetX, or Honest Broker)?	◯ Yes ◯ No

If the answer to the previous question is Yes, the following additional questions appear:

What activities did you complete using the CRDW tools? (Please check all that apply.)	☐ Obtain a count of the number of patients in a particular cohort. ☐ Analyze a patient set in order to see a demographic distribution. ☐ Extract data tables for a patient set.

For each CRDW tool listed below, rate its ease of use.				
	Very Easy to Use	Easy to Use	Difficult to Use	Very Difficult to Use
i2b2	◯	◯	◯	◯
TriNetX	◯	◯	◯	◯
Honest Broker	◯	◯	◯	◯

How have you used (or intend to use) the data (e.g., patient count or data tables) obtained form the CRDW? (Please check all that apply.)	☐ Used, or plan to use, for cohort discovery ☐ Used, or plan to use, for hypothesis development ☐ Used, or plan to use, for a prospective study ☐ Used, or plan to use, for a retrospective study ☐ Other (please specify)

Section #2: Requesting Support

Did you contact the Biomedical Informatics (BMI) department for support with the CRDW?	◯ Yes ◯ No

If the answer to the previous question is Yes, the following additional question appears:

Who did you contact for support within the BMI department?	_____

Section #3: Benefits and Barriers to CRDW Use

For the statements below, indicate your level of agreement.					
	Strongly Agree	Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree	Strongly Disagree
The training I received was sufficient for me to use the CRDW effectively.	◯	◯	◯	◯	◯
I lacked the necessary support to use the CRDW effectively.	◯	◯	◯	◯	◯
Overall, I found the CRDW easy to use.	◯	◯	◯	◯	◯
I lacked the time to effectively use the CRDW.	◯	◯	◯	◯	◯
The data I needed was not available within the CRDW.	◯	◯	◯	◯	◯
Using the CRDW was beneficial to my work.	◯	◯	◯	◯	◯

Section #4: Likelihood to Use the CRDW

How likely are you to use the CRDW for future clinical research?

◯ Very Likely ◯ Likely ◯ Unsure ◯ Unlikely ◯ Very Unlikely

Section #5: Comments

Please provide any additional comments regarding your experience using the CRDW (optional).	_____

CURRICULUM VITAE

Katie A. McCarthy, MSMI, CQIA(ASQ), CQSE(ASQ)

Milwaukee School of Engineering (MSOE)
1025 North Broadway, Milwaukee, WI 53202
Phone: 414-277-2316
Email: mccarthk@msoe.edu

Academic Degrees

Ph.D. in Health Sciences — **Candidate (tentative: May 2019)**
University of Wisconsin-Milwaukee — Milwaukee, WI
Area of concentration: Clinical and Health Informatics

M.S. Medical Informatics — **May 2005**
Milwaukee School of Engineering (MSOE) — Milwaukee, WI
Medical College of Wisconsin (MCW)

- Thesis: Discovering Helix – Implementing a Genome-enabled Laboratory Information System
- Independent research: Bayesian networks applied to mammography diagnosis

B.S. Computer Engineering — **May 1999**
B.S. Technical Communication — **May 1999**
Milwaukee School of Engineering (MSOE) — Milwaukee, WI

- Fred Loock Outstanding Student Award
- Alumni Student Achievement Award
- Who's Who Among Colleges and Universities

Professional Experience

Milwaukee School of Engineering (MSOE) — **November 2005 to present**
Milwaukee, WI

Assistant Professor, Rader School of Business — 2013 to present
Adjunct Lecturer, Rader School of Business — 2005 to 2013

- Served as Program Director for the M.S. Medical Informatics (MSMI) program, jointly offered by MSOE and MCW
- Served as specialty coordinator for Bachelor of Business Administration (BBA) Information Technology Systems (ITS) coordinator
- Served as academic advisor for MSMI, BBA-ITS, and B.S. Management Information Systems (MIS) students
- Led curriculum redesign for the MSMI program, and co-led the curriculum creation for the BBA program

- Served as faculty advisor for the Quality and Efficiency module of the Clinical and Translational Science Institute (CTSI) of Southeast Wisconsin
- Served as faculty advisor for students participating in the Gus Ramirez Lean Scholars Program through Froedtert and the Medical College of Wisconsin
- Served as co-designer and instructor for the Rader School of Business Information Technology summer program for high school students
- Taught undergraduate and graduate courses for the following degree programs: MSMI, Master of Business Administration (MBA), M.S. Nursing (MSN), M.S. Engineering Management (MSEM), BBA, and B.S. Industrial Engineering
- Courses taught in the following areas: informatics, analytics, data integration and management, system development and implementation, quality improvement and management, Six Sigma, Lean, design of experiments, statistics, healthcare economics

Froedtert and The Medical College of Wisconsin **2006 to 2014**
Milwaukee, WI

Information Technology (IT) Project Manager, Epic EMR Program 2011 to 2014
- Responsible for managing large Epic electronic medical record (EMR) projects, involving large cross-functional teams and budgets of approximately $3 million.
- Projects managed included data conversion into Epic from two other EMRs, implementation of more than 100 interfaces and negotiation of contracts with the numerous vendors involved, implementation of anesthesia device integration with Epic, redesign of Epic security for approximately 12,000 users, and design and validation of a new framework for managing order transmittal for Epic.

IT Manager, Core Infrastructure Administration 2009 to 2011
- Responsible for management of approximately 50 employees that comprised the telecommunications, data network, database, integration/interfaces, security and disaster recovery, and Epic EMR core systems teams within the IT department.
- Responsible for operating budgets totaling over $5 million.

Clinical Informatics Team Lead – Orders and Results 2007 to 2009
- Responsible for leading team to design, implement, and maintain/support enterprise ordering and resulting processes within the Epic electronic medical record, including processes for laboratory, radiology, cardiology, medication, procedure, and referral ordering.
- Responsible for analysis and design of interfaces to support laboratory, cardiology, and medication processes.
- Responsible for daily team leadership, including team structure development, process improvement, staff mentoring and training, and resource allocation.

Other positions held:
Clinical Informatics Systems Analyst 2006 to 2007
IT Systems Analyst 2006

BloodCenter of Wisconsin (BCW) **2001 to 2006**
Milwaukee, WI

Positions held:
Diagnostic Laboratories Quality Engineer
Diagnostic Laboratories Lead Quality Specialist
Information Services (IS) Quality/Validation Analyst

- Responsible for daily oversight of the Diagnostic Laboratories Quality department and the ongoing training and mentoring of the employees within the department, including development of the new employee training program for the department.
- Responsible for creation and implementation of quality practices, management, and tracking systems for individual departments, the Hematopoietic Progenitor Cell (HPC) Program, and the Diagnostic Laboratories business unit (such as document control, training, project management, resource tracking, change management, project life cycle, validation, equipment tracking, and quality control and monitoring).
- Responsible for testing/validation of computer systems and applications, including the lab information system (Cerner) and the blood operating system (LifeTrak).
- Responsible for daily quality activities, such as document creation and review; staff training; creation and tracking of metrics; identification and implementation of corrective actions; preparation of regulatory agency packets; facilitation of departmental, internal, and external audits; and completion of audit responses.
- Responsible for creation and delivery of training such as Change Management, Advanced Microsoft Word Skills, Quality Basics, Project Administration, Effective Writing, and Statistical Techniques (Basic and Intermediate).

Plexus Technology Group **1999 to 2001**
Neenah, WI

Embedded Software Design Engineer

- Responsible for the analysis, design, implementation, and testing of software for medical devices, safety systems, and datacom/telecom products.
- Member of the Recruitment/Interview Team, the New Employee Training Committee, and the Software Diagnostics and Testing Focus Group.

Publications and Presentations

- *Impact of Awareness and Training on the Adoption of Clinical Research Data Warehouses.* Katie McCarthy. American Medical Informatics Association (AMIA) Informatics Educators Forum (IEF). June 2019 (accepted).
- *Merging Farms: Combining Two Instances of Epic*. Bradley Howard, Katie McCarthy, Kathy Patrino, and Michael Sura. Epic Users' Group Meeting. September 2014.
- Featured in *Going from Epic to Epic*: *One Health System's Unique Journey*. Healthcare Informatics. December 2013.
- *Cutover Planning and Implementation: Keys to a Successful Go-live.* Katie McCarthy and Nanda Kothinti. Epic Project Managers Advisory Council. April 2013.
- *Do Computer-Aided Diagnosis Systems in Mammography Need to be Trained to Individual Observers?* Charles Kahn, Katie McCarthy, and Elizabeth Burnside. Radiological Society of North America (RSNA) Annual Meeting. November 2005.

Professional Affiliations

- American Medical Informatics Association (AMIA) — 2002 to present
- American Society for Quality (ASQ) — 2001 to present
  Certified Quality Improvement Associate (CQIA) — 2003 to present
  Certified Software Quality Engineer (CQSE) — 2002 to present
- Association for Computing Machinery (ACM) — 2000 to present
- Healthcare Information and Management Systems Society (HIMSS) — 2006 to present

Honors

- MSOE Falk Engineering Educator Award (finalist) — 2014, 2015
- MSOE Outstanding Mentor Award (nominee) — 2000, 2002

Community and Professional Involvement

- MSOE Nursing Advisory Committee, Member — 2017 to present
- BloodCenter of Wisconsin (BCW) — 2004 to present
  Lecturer for Specialist in Blood Banking (SBB) program
- Third Coast Consortium for Biomedical and Health Informatics — 2015 to 2017
  Co-organizer
- HIMSS Dairyland Regional Networking event — 2014 to 2017
  Co-organizer

Professional Development

- Massachusetts Institute of Technology (MIT) 2016
  Professional development certificate: Big Data and Social Analytics
- Blackboard Learning Management System 2015
  Webinar: The Ins and Outs of Designing Exemplary Courses
- Massachusetts Institute of Technology (MIT) 2016
  Professional development course: Tackling the Challenges of Big Data
- ACM Computer-Human Interaction (CHI) conference 2014
- Epic User Group Meeting (UGM) 2009, 2013, 2014
- Epic Project Managers Advisory Council 2013
- ASQ World Conference for Quality and Improvement 2006
- Cerner Healthcare Conference 2004
- MSOE Quality Engineer training 2004
- ASQ Quality 101 training 2003
- ASQ Software Quality Engineering training 2002