The HSCC Clinical Data Warehouse: The South Carolina Biomedical Research Tool of the Future

November 2013
About Health Sciences South Carolina

Health Sciences South Carolina (HSSC) was created in 2004 as the nation’s first statewide biomedical research collaboration. HSSC members include the six largest health systems in South Carolina as well as the state’s three senior research universities. HSSC was created by its founding member institutions in the recognition that the state’s public health and economic well-being would be greatly advanced if health care organizations exchanged competitiveness for cooperation.

HSSC works on behalf of the entire state of South Carolina to make good health possible. Patient-centered health care collaboration is the hallmark of each HSSC initiative. This includes the development and implementation of the South Carolina Clinical Data Warehouse, one of the nation’s foremost health care information management tools, which HSSC designed to improve the efficiency and effectiveness of statewide patient care and biomedical research and which serves as a national model to improve health care.

Why South Carolina Needs a Clinical Data Warehouse

The state of health care in South Carolina is no secret; United Health Foundation ranks South Carolina 46th in the nation in overall health. The South Carolina health care crisis is due to many factors, including historical, educational and socioeconomic realities, as well as a significant percentage of the population lacking quality access to health care and health care insurance. Major health care problems in the state include but are not limited to high incident rates statewide of stroke, diabetes, cardiac heart disease, tobacco-related illness, and infant mortality. Another statewide health care issue is higher incident rates of certain cancers, such as cervical and prostate cancer.

With these issues in mind, HSSC asked its member hospital systems to list the most important services HSSC could offer to assist in addressing and solving the state’s major health care needs. The answer was clear:

1. Create and implement a statewide system that collects population health management and provides clinical and comparative effectiveness data,
2. Establish and implement statewide tools for collecting and analyzing data for clinical and research purposes, and
3. Serve as a convener of the brightest minds and unique resources to implement statewide health care research demonstrations and projects.

With the Clinical Data Warehouse, HSSC has created a single information tool that provides the above services statewide. The Clinical Data Warehouse links electronic patient records from across the state and enables clinicians and researchers to follow patient conditions in real-time. It also allows biomedical researchers to study specific health disorders across a much broader, de-identified (anonymous) and aggregated patient population base. Finally, the Clinical Data Warehouse will serve to recruit the biomedical industry to the state and boost the state’s economy, as the infrastructure for this health care information tool is unique nationally.
The Clinical Data Warehouse: A Brief History

Throughout the 1980s and early 1990s, hospital electronic data were mostly related to patient billing and scheduling. Enormous computerized systems were developed to organize and facilitate patient appointments as well as generate and process bills and insurance claims. At the turn of the 21st century, health care providers began to invest in the creation of computer database systems that could store and communicate important patient data between and among various components of a hospital for clinical purposes. One example of how this works is when a patient’s lab result is sent electronically to another area of a hospital for analysis by a clinician.

All major health care providers now use electronic systems for storing and retrieving clinical data, yet there are numerous electronic systems from which health care providers may choose. For example, a provider can choose one electronic system for lab results and another to track medications. The management and operation of these systems for any single health care provider is very complex and requires a team of individuals to operate and maintain, including getting the different systems “to communicate” with one another. And while operating the systems at any one provider is a challenge in itself, an even greater challenge is getting electronic systems at different providers “to communicate” with one another.

In 2006, Health Sciences South Carolina leadership proposed the idea to its member institutions of the creation of a unique health care information tool that could integrate the data systems from each member hospital and then communicate that information to member universities: a comprehensive, statewide patient database for research. Such a statewide database would function as an innovative “learning system” across the state and become a critical tool in developing a South Carolina Rapid Learning Health System. A Rapid Learning Health System uses advances in information technology to collect and compile through clinical research the evidence needed to deliver the most effective and up-to-date care for patients.

Leaders at HSSC member institutions were excited by the prospect of clinical researchers being able to study statewide patient populations in real time and to retrieve detailed health care data beyond their own patient databases. Researchers need access to large pools of patient information in order to develop and test their scientific hypotheses, so the creation of a single, integrated statewide data system would create an enormous opportunity for an increased number of biomedical research projects, including large-scale projects. Such a statewide tool could eventually lead to reduced health care costs and improved treatments for disease, as well as economic development opportunities through partnerships with the biomedical industry and the development of new medical technologies.

In 2008, HSSC assembled a team of health information technology and computer science experts, along with physicians, researchers and administrators at HSSC member hospitals to begin organization and careful construction of the Clinical Data Warehouse. Critical funding for the project was provided by The Duke Endowment, the State of South Carolina (through the South Carolina SmartState™ Program), and HSSC member institutions themselves. During this time, infrastructure pieces were acquired and installed throughout the state; also, a team of
computer programmers began writing thousands of lines of computer programming code. Simultaneously, the HSSC member institution Chief Information officers and clinicians met regularly to discuss policies related to the general organization of the Clinical Data Warehouse, including especially data use privacy and security. A Governance Committee was developed and ultimately formed, which includes representatives from each HSSC member institution and the purpose of which is to administer policies for the day-to-day operations of the Clinical Data Warehouse, including policies in the Data Use Agreement; HIPAA; and the Federal Policy for the Protection of Human Subjects legislation, or the Common Rule, especially with respect to impact on the protection, privacy and security of patients and patient data.

Because such a project had never been attempted, the Clinical Data Warehouse developmental stage took several years. In addition, HSSC member hospitals were necessarily focused on a number of major technology issues, including system-wide implementation of electronic health records, meeting the requirements of the 2009 Health Information Technology for Economic and Clinical Health (HITECH Act), and adoption of Meaningful Use standards in the Centers for Medicare and Medicaid Services Incentive Program. All precautions were taken by HSSC and its member institutions to ensure adherence to patient protection, federal regulations and member institution missions in the design of the Clinical Data Warehouse.

During the development process and prior to the launch of the Clinical Data Warehouse in September 2013, HSSC engaged two private expert IT firms to conduct external evaluations and testing to ensure the system met appropriate data and health care industry standards. Both evaluations were very positive and provided HSSC with suggestions of additional tools that could strengthen further the Clinical Data Warehouse and ensure security. HSSC contracted with Oracle and Recombinant by Deloitte to implement these tools, and also implemented an open source interface platform developed by the National Institutes of Health in conjunction with Harvard called i2b2 (Informatics for Integrating Biology and the Bedside). The development team spent Summer 2013 loading historical data from the three initial provider participants (Greenville Health System, Medical University of South Carolina and Palmetto Health) into the database, and again conducted end-to-end testing of the system. In September 2013, the Clinical Data Warehouse “went live” with an initial number of patient data types (demographics, diagnoses, and procedures) available to researchers. Additional data types, such as medications and laboratory results, will be added to the system in 2013 and 2014. HSSC plans to “onboard” a fourth provider system, Spartanburg Regional Health System, onto the Clinical Data Warehouse during the 2014 calendar year.

At present, HSSC is training clinical researchers throughout the state on how to use the Clinical Data Warehouse for the purposes of data analysis, comparative effectiveness research and the design of biomedical research studies. The Clinical Data Warehouse is also central to a proposed major patient-centered research collaboration with three North Carolina research institutions. Over time, HSSC will continue to expand the research capacity of the Clinical Data Warehouse and consider novel ways that this unique learning system can impact improved health outcomes and lead the state toward becoming a Rapid Learning Health System.
How the Clinical Data Warehouse Solution Works

As stated above, health care providers can choose from a variety of software systems to store and use electronic health records. The health care industry has developed standard formats and terms that are used to exchange electronic information across systems. Without this level of standardization, software system engineers would constantly face the daunting task of interpreting and translating the information, resulting in quality and inefficiency issues.

In the health care industry, the most common interoperability standard is known as Health Level Seven (HL7). HL7 was developed in the late 1980s; like any language standard, human or computer, variations and permutations develop quickly. In other words, while health care software systems tend to “speak” HL7 universally, each independent HL7 system is uniquely modified and customized (think “colloquial” vocabulary and “accents”) so that it becomes difficult for the clean exchange of information across multiple HL7 systems. HSSC resolved this problem by assembling a team of coders who manually mapped institutional-specific code to a common terminology across all participating Clinical Data Warehouse institutions.

As data is collected from participating institutions for the Clinical Data Warehouse, the varying formats and “dialects” are filtered through an Interface Engine and stored in an Operational Data Store. The Interface Engine is responsible for receiving transactions from the sources systems and transforms that data into a format that can be accepted by the Operational Data Store. The Operational Data Store is a real-time transactional database. The secure Operational Data Store will contain all updates to patient clinical information (laboratory results, medications, patient diagnoses, etc.) as well as personal identifying markers (name, contact information, governmental ID numbers, etc.).

At the same time information is filtered through the Interface Engine and into the Operational Data Store, the demographic information for each patient is also fed into a Master Patient Index, which is a Clinical Data Warehouse Solution tool that resolves duplicate information across participating HSSC member institutions into one “unique identifier” for each patient. The Master Patient Index helps resolve cases where a patient might, for example, register as Joanne Smith at one provider but as Joanne Smith-Jones at another provider.

On a scheduled basis, patient health information from the Operational Data Store and Master Patient Index is consolidated and synthesized and extracted into the Data Warehouse. Information within the Data Warehouse remains identified.

Data is extracted from the Data Warehouse to specific data marts and patient registries. The primary data mart for clinical research is called i2b2 (Informatics for Integrating Biology and the Bedside). Patient information in i2b2 is de-identified; that is, an individual’s personal identifying markers (name, contact information, governmental ID numbers) are removed.

There are two ways to access information that is securely managed within the Data Warehouse Solution: (1) authorized researchers and clinicians can use the i2b2 tool to gather general population health data to design research projects; or (2) authorized researchers and clinicians can be provided identified information through a thorough justification and approval process, described in “The Safety and Privacy of Personal Medical Information” (see page 9).
this latter process, an HSSC-designated individual known as the “Honest Broker” collects identified information and securely provides it to the approved researcher.

Clinical Data Warehouse Solution Overview

Figure 1.0. This figure shows the Clinical Data Warehouse infrastructure. Patient information is fed in real time from the provider systems through the Interface Engine and into the Operational Data Store and Master Patient Index, then synthesized and consolidated and placed in the Data Warehouse for facilitation of delivery to specific data marts and registries. De-identified information is drawn from i2b2 in order to design prospective research studies. Identified information and detailed de-identified data can be drawn directly from the Data Warehouse through a thorough approval process.

Figure 2.0. A screen shot from the i2b2 query interface. i2b2 is the open source program that “strips away,” or de-identifies, data and does not permit a researcher to access patient HIPAA personal identifiers. Researchers and clinicians then use i2b2 to conduct de-identified and aggregated data searches of patient populations in the Clinical Data Warehouse. Such queries assist researchers and clinicians with designing research projects.
How the Clinical Data Warehouse Will Be Used (Practical Examples)

As stated above, researchers require large pools of patient information to develop and test scientific theories. Until recently, researchers had no simple way to study patient populations within and especially beyond their associated provider—and even then, studying a patient population in real time was unthinkable. General advances in technology have greatly improved this situation for clinical researchers, but most researchers remain restricted to accessing information within their own provider patient base unless they collaborate with researchers at other systems. But the Clinical Data Warehouse provides clinical researchers at HSSC member institutions with an integrated learning tool whereby patient information across the entire state can now be surveyed and tracked in real time.

Like other scientists, clinical researchers engage in rigorous survey work before proceeding with a research project that may take years to complete. Imagine that a researcher is interested in studying a rare condition such as Sickle Cell Disease. Nationally there are an estimated 90,000 individuals diagnosed with Sickle Cell Disease, less than one percent of the population. Conducting queries of the Sickle Cell Disease population within a single South Carolina provider system yields a small population sample from which to build a potential research patient cohort. However, with the Clinical Data Warehouse, a researcher can triple or quadruple previous sample sizes, expanding queries to include more than 3 million patients across the state. Researchers in South Carolina now have a better chance of determining the potential success of a given research project, and easier ways to build patient cohorts.

While it is true that certain health care research studies—especially clinical trials and the development of biomedical technologies—require years of regulatory approval, there are ways that the Clinical Data Warehouse can be used to positively impact health care outcomes in the near-future. Because of the Clinical Data Warehouse, researchers can determine how new guidelines and industry practices impact patient populations across broad geographical areas. For instance, one early pilot project being developed is a study to determine whether narrow-spectrum antibiotics as a first-line treatment for Community Acquired Pneumonia in children, as recommended by the Pediatric infectious Diseases Society and the Infectious Diseases Society of America, have the intended effect to decrease the incidence of multi-drug resistant organisms. With the Clinical Data Warehouse, researchers will be able to perform historical comparisons prior to and after the implementation of these new recommended practices.

Also, the Clinical Data Warehouse is already attracting the interest of outside health care providers and organizations and increasing partnership opportunities for HSSC members. Another Clinical Data Warehouse pilot project is being developed with major health care organizations in Massachusetts to increase surgery safety. And just recently, HSSC member institutions partnered with three North Carolina academic health care institutions, Duke University, University of North Carolina and Wake Forest University, on a major information collaboration that would not have been possible before the existence of the South Carolina Clinical Data Warehouse. Finally, the biomedical industry is expressing an interest in how the Clinical Data Warehouse might be a useful tool for future clinical studies and FDA approvals.
How the Clinical Data Warehouse is Organized and Operated

The Clinical Data Warehouse is a complex biomedical information tool and requires administrative input and oversight from health care leaders across South Carolina. The Clinical Data Warehouse is overseen by the Health Sciences South Carolina Clinical Data Warehouse Governance Committee. Not only does the Governance Committee oversee the daily operation of the Clinical Data Warehouse, but it holds as a foremost responsibility the protection of patient privacy and security of patient data. The Governance Committee is composed of voting members appointed by the HSSC Policy Steering Committee (one each per member) as well as several HSSC administrators including the Chief Executive Officer, the Chief Medical Officer and the Chief Medical Information Officer.

The Governing Committee operates according to the terms of the Data Collaboration Agreement, which must be signed by all HSSC member institutions that participate in the Clinical Data Warehouse. The purpose of the Governance Committee is (a) to establish and administer policies for the Clinical Data Warehouse’s daily operation, (b) to promote policies in accordance with HIPAA (federal), the Common Rule (federal), applicable state regulations, and the Data Collaboration Agreement (HSSC), (c) to assist in issue or dispute resolution, and (d) to support and aid any and all other activities related to the Clinical Data Warehouse.

The Governance Committee is supported and advised by three formal Work Groups:

- **Data Quality & Stewardship Work Group**: provides oversight for the HSSC CDW data request process; identifies new use cases and defines data requirements; reviews requests to expand the database; helps identify issues and challenges associated with data access and quality; advises on issues associated with data quality; assists member institutions with understanding and resolving data issues associated with the Clinical Data Warehouse;

- **Security Work Group**: oversees security classifications of access to data; provides information on best security practices; advises on encryption, security policies, and security scans; and engages in security audit procedures;

- **Institutional Review Board/Data Use/Privacy Work Group**: advises on IRB and Common Rule requirements; HIPAA and other privacy standards; makes recommend-dations for the Data Use Agreement, policies and standards for research, and Treatment Payments & Health Care Operations (TPO); recommends access of data for queries not requiring a separate IRB approval; monitors data auditing activity; recommends governance for research, develops feedback mechanisms to and from local IRBs, privacy, and compliance offices; and reports on regulatory and compliance issues, audits, complaints, and disputes.

**Who Owns the Clinical Data Warehouse?**

It is important to note that the South Carolina Clinical Data Warehouse is not a single database, but an integration of unique databases of participating HSSC member institutions. HSSC is a steward of information and data records that are owned respectively by its member institutions and by patients themselves. HSSC takes this stewardship role extremely seriously (see next section) and conforms to all federal, state and institutional requirements with respect to the maintenance, distribution and use of information and data in its care.
The Safety and Privacy of Personal Medical Information

The safety of patients and the security of patient health information and data is paramount to HSSC. Enormous precautions have been taken to ensure that the Clinical Data Warehouse conforms to governmental and industry security standards. Clemson University serves as the host institution of the Clinical Data Warehouse’s applications, databases and host systems. These infrastructure components are overseen by the Clemson Office of Information Security and Privacy as well as the Data Quality Stewardship Advisory Group plus the Data Review Request and Review Committee. Clemson has particular expertise in securing health care data, as it also serves as the Medicaid eligibility and claims processor for the SC Department of Health and Human Services.

One important fact to keep in mind is that researchers cannot access any of a patient’s HIPAA identifiers when querying health information and data with the Clinical Data Warehouse i2b2 interface tool. The Clinical Data Warehouse system engages in a computing process called “de-identification” of data; that is, the majority of an individual’s personal identifying markers (name, contact information, governmental ID numbers, etc.) are detached from the health record. Also, only authorized users (clinical researchers approved by each HSSC member institution) are permitted to access de-identified data. These authorized users must undergo special training that informs them of the restrictions on data use and the requirements of data protection; they must also agree to the terms of a strict Data Usage Agreement.

After surveying patient population information to design a research study, a researcher must complete a thorough approval process to receive detailed data. This includes a consultation with HSSC biostatisticians and an “honest broker” who help the researcher consider additional statistical questions about the potential study. Following this consultation, a request for a limited data set must be approved by a Data Review Committee before being sent to an Institutional Review Board (IRB), the committee that approves and monitors the use of patient data and oversees the general welfare of patients in research studies at each institution from which data originates. (HSSC has organized a collaborative electronic IRB, or eIRB, which permits researchers to receive inter-institutional permission for Warehouse-related research projects in one process.) If the data request is approved, the honest broker creates and reviews the limited data set, then submits the information to the researcher.

Again, ensuring the safety of patients and the security of patient health information is a primary and vital part of HSSC’s mission and has been a guiding principle in all stages of the design, implementation and operation of the Clinical Data Warehouse.
The Future of the Clinical Data Warehouse

Just two decades ago, the idea of a statewide, integrated database system that allows clinical researchers immediate access to broad patient population data would have been considered a technological pipedream in the health care industry—not just because the technology needed to be created, but also because it is more common for health care providers to compete rather than collaborate. One of the greatest strengths of the Clinical Data Warehouse, beyond its technological innovation, is that it is grounded in a unique spirit of cooperation between South Carolina’s major health care providers and research intensive academic institutions. Nearly one decade ago, the leaders at all HSSC member institutions understood that working together was the best possible way to improve the health and well-being of all South Carolinians.

Today, the Clinical Data Warehouse represents a new technology—a statewide health care learning system—that can immediately and positively impact the lives of patients throughout South Carolina. Researchers now have the ability to analyze health conditions across the entire state in real time. This increased knowledge will lead to improved judgments about the kinds of projects researchers engage in; it will also help the health care community determine whether regulations and industry standards are working as intended. These are projects and valuations that are precisely relevant to South Carolinians today.

Like any technology, the Clinical Data Warehouse has the potential to expand in a myriad of ways to improve health outcomes and the economy of our state. Each expanded application of this learning system tool will further establish a Rapid Learning Health System across our state. Such a system relies on information technology to deliver timely methods of care for patients; HSSC is confident that the Clinical Data Warehouse is a centerpiece to developing such a system. And no matter where the Clinical Data Warehouse of tomorrow is headed, the overall well-being of all South Carolinians will always be its foundational core.

The Clinical Data Warehouse is a learning system tool that has the ability to improve each and every one of our lives.

Contact Information

For media inquiries and general queries about HSSC and the South Carolina Clinical Data Warehouse, please contact:

Arik Bjorn, HSSC Academic Program Manager
bjorn@maibox.sc.edu

For technical queries about the South Carolina Clinical Data Warehouse, please contact:

Rick Larsen, HSSC Clinical and Informatics Program Director
larser@musc.edu