



Published in final edited form as:

J Nurs Adm. 2013 June ; 43(6): 355–360. doi:10.1097/NNA.0b013e3182942c3c.

Conducting Research Using the Electronic Health Record Across Multi-Hospital Systems: Semantic Harmonization Implications for Administrators

Dr. Kathryn H. Bowles, PhD, RN [Professor and Ralston House Endowed Term Chair in Gerontological Nursing, Director of the Center for Integrative Science in Aging],
School of Nursing, University of Pennsylvania, Philadelphia, PA

Dr Sheryl Potashnik, PhD [Research Project Manager],
School of Nursing, University of Pennsylvania, Philadelphia, PA

Dr. Sarah J. Ratcliffe, PhD [Associate Professor of Biostatistics and Epidemiology, Senior Scholar, Clinical Epidemiology and Biostatistics],
Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Ms Melissa Rosenberg, RN BSN [Research Assistant],
School of Nursing, University of Pennsylvania, Philadelphia, PA

Ms Nai-Wei Shih, IMBA [Research Admin Coordinator],
School of Nursing, University of Pennsylvania, Philadelphia, PA

Mr. Maxim Topaz, MA, RN [Doctoral Student],
School of Nursing, University of Pennsylvania, Philadelphia, PA

Dr John H. Holmes, PhD [Associate Professor of Medical Informatics in Epidemiology, Senior Scholar Center for Clinical Epidemiology and Biostatistics], and
Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Dr Mary D. Naylor, PhD, RN [Marian S Ware Professor in Gerontology, Director of the NewCourtland Center for Transitions and Health]
School of Nursing, University of Pennsylvania, Philadelphia, PA

Abstract

Administrators play a major role in choosing and managing the use of the electronic health record (EHR). The documentation policies and EHR changes enacted or approved by administrators affect the ability to use clinical data for research. This article illustrates the challenges that can be avoided through awareness of the consequences of customization, variations in documentation policies and quality, and user interface features. Solutions are posed that assist administrators in avoiding these challenges and promoting data harmonization for research and quality improvement.

Corresponding Author: Dr. Bowles, University of Pennsylvania School of Nursing, Room 340 Fagin Hall, 418 Curie Boulevard, Philadelphia, Pennsylvania 19104 (bowles@nursing.upenn.edu)..

Conflicts: There are no conflicts to disclose regarding this manuscript.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Using data elements from the electronic health record (EHR) for purposes beyond clinical documentation, billing, and administration is a rapidly growing practice (1). The increasing number of EHR installations and several recent national policy initiatives have supported this trend. Meaningful use of electronic health information, mandated by the Health Information Technology for Economic and Clinical Health (HITECH) Act, involves using EHR and related technology to improve quality, safety and efficiency of patient care; engage patients and families; improve care coordination; and ensure adequate privacy and security for personal health information (2). Further, the Institute of Medicine proposed a *learning health system* in which we use patient and health care information for research and continuous improvement in health and healthcare (3).

However, we are in the beginning stages of seeing researchers actually use EHR data for studies. Secondary use of data meant for clinical documentation, billing, or administrative purposes presents several challenges including issues around the collection from original sources, storage, aggregation, linkage and transmission of health data (4). In addition, to increase the likelihood of generalizable results, researchers often seek samples from a variety of sites. This requires the merging of data from multiple sources and semantic harmonization may be an issue. Semantic harmonization is the process of combining data from heterogeneous sources into a single clinical system (5). It can occur in one health system or across multiple systems.

The purpose of this article is to describe the challenges, solutions, and implications related to semantic harmonization while conducting research using EHR data from 4 hospitals. The implications and importance of this topic is critical to nurse managers and hospital administrators because they develop and approve documentation and data-use policies and provide strategic guidance to the informatics nurse leaders.

Background

Our research team is developing a clinical decision support (CDS) system for discharge planning that will assist clinicians to identify patients in need of post-acute referral. Based on seminal patient characteristics, the system will recommend the most appropriate site of care (i.e. home care, skilled nursing facility, etc.) (National Institute of Nursing Research 2RO1-007674). A critical role of CDS is to provide healthcare providers, patients, and caregivers with general and person-specific information, intelligently filtered and organized, to enhance health and healthcare (6). Examples of CDS interventions might include alerts/reminders, clinical guidelines, order sets, patient data reports and dashboards, documentation templates, diagnostic support, and other tools. Most often, these CDS interventions arise from electronic health record (HER) data that is calculated, synthesized, or analyzed to provide meaningful information to the user at a critical decision point (6,7).

Our study used de-identified data from variables gathered through approximately 1200 nursing admission assessments and ongoing documentation throughout the hospital stay. We used the electronic data elements to create case studies of older adults describing their health status including their socio-demographic, functional, cognitive, physical, mental, emotional, home environment and social support descriptors. Examples of data elements include assessments of caregiver characteristics; medications and medication management; medical conditions, activities of daily living/other functional parameters; and depression, fall, and pressure ulcer risk scores. Interprofessional teams of doctors, nurses, social workers, and physical therapists will evaluate the case studies and based on patient characteristics, determine the need for post-acute referral, and if so, which site is most appropriate to meet their needs. Four separate hospitals, all using an EHR from the same vendor, provided the

data elements used in the case studies. The study was approved by the institutional review board (IRB) of the University of Pennsylvania and the hospital IRB when required.

Challenges to data harmonization

The realities of doing research across different institutions using the EHR, even when working with sites that have an EHR from the same vendor, are daunting. Numerous issues and questions emerged while conducting this type of research. There are differing versions of the EHR, customizations, variations in documentation policies and quality, and user interface features to contend with. We will provide descriptions and examples of these challenges, followed by a discussion of solutions and implications.

Version control

Although the EHR at each hospital was from the same vendor, among the 4 hospitals there were 3 different versions of the software in use. Unfortunately, when designing the study there was no discussion of or foresight to ask about various versions because it was assumed that there were commonalities with the same vendor across sites. Differing versions of the software presented a challenge because hospitals using earlier versions did not collect the data elements and related pick list options (the standard list of responses attached/belonging to a particular question) the same way as is present in more recent versions.

Customization

Commonly, sites make specific customized changes to the EHR based on practicing clinicians' or clinical managers' requests. These customizations may change the standard variable names, pick list values or both, and can occur in multiple ways such as:

Adding local detailed information to existing choices—Data on Supplemental Digital Content 1, <http://links.lww.com/JONA/A226>, illustrates the addition of information about local agencies that patients are discharged to. The original data elements listed in the study ontology column describe generic discharge destinations, such as home with home health services or inpatient rehabilitation facility. Looking across the columns we see the pick list choices no longer match the standard EHR choices. While this may or may not be of value locally, it is not useful information when doing cross-site research and reporting.

Reversing the wording of pick list options—One hospital reversed the terms for documenting patient orientation. Instead of documenting *disoriented to* person, place, time, and situation, they changed the word to *oriented to*. This minor change created the opposite meaning of the orientation data element compared to that of the other 3 sites.

Adding or subtracting standard pick list options—Several sites customized pick list option responses by adding or deleting the standardized responses that existed in the EHR product. For example, Table 1 illustrates this problem of adding types of agencies or additional categories of people. In this example some of the pick list choices no longer match the question of *lives with* (a person) but rather is now a place (state veterans home). Numerous other examples include fields such as differing nutrition risk and diet tolerance.

Changing variable or table names—When working with multiple databases to access or merge data, it is important that items are a semantic match, meaning the naming conventions are the same. Changing variable or database table names makes writing a query to access or merge the data challenging. In the event that variable or database table names are changed from version to version by the company producing the EHR, this represents something under vendor control and administrators should ask about backward compatibility

(See Table, Supplemental Digital Content 2, example of variable or table names, <http://links.lww.com/JONA/A227>). These variations in table names affected the ability to use queries written at 1 site at other sites in an attempt to access the same variables. Queries in this study had to be re-written.

Substituting different evidence based tools than what the product comes with

—One of the sites inserted a different evidence based tool for assessing fall risk than the other 3 sites. Although they substituted another credible/evidence-based alternative, doing this is not recommended because it made the components of fall risk assessment non-comparable between sites. These customized changes jeopardize the opportunity to have standardized measures useful for research, quality improvement and benchmarking.

Documentation Policies

Variation in unit or hospital documentation policies can affect the quality and content of clinical EHR data. For example, in our study, 3 hospitals have a policy that fall risk is assessed on every shift. This policy assures that documentation of fall risk will be present in the database during a given time period. The remaining site does not mandate documentation of this data element so its presence is highly variable making it difficult to find. This issue resulted in missing data for 1 site which could have been used to evaluate opportunities for improvement of this nurse-sensitive outcome.

Quality of documentation

We have found several issues affecting the quality of documentation. First, the thoroughness of individual clinicians' documentation varies greatly (8-10). Clinicians may not value documentation or recognize that their data are being used for other purposes. Although most clinicians are very thorough and document each data element consistently, others are accepting of missing data or too busy to be consistently thorough. Clinicians often complain that documentation takes too long and the information is not used to support patient care or quality improvement (11). The 2nd issue is exacerbated when clinicians take shortcuts and enter free text rather than use pick list choices. This may result in redundant information, completely different responses that often don't relate to the question asked, or variations in spelling and abbreviation that are cumbersome to condense and code. This practice negates the value of structured, electronic data. Third, education about how to interpret the meaning/intention of the EHR content is critical to assure reliable data entry among the various users. Unclear questions within the EHR will result in varied interpretations and therefore unreliable data and information.

User Interface

The design of the EHR user interface may also present a challenge. Depending on how the EHR presents the data to the user, important data elements could be missed. At one of our sites, an assessment of activities of daily living was hidden under other information and had to be brought forward by the clinician to present itself for documentation. If a clinician did not know this, or did not take the effort to make the item visible, it could be forgotten and never collected.

Further, although part of the system was designed to notify the clinician that their documentation was incomplete, we identified that this function was limited. It did not tell clinicians *what* was incomplete. Staff had to scroll through screens to find the missing item(s). Requiring this extra effort in a busy clinical setting is likely to result in non-compliance.

Solutions

The challenges to data harmonization noted above required a huge amount of work to overcome. The 1st step was to work with the multiple hospital staff at each site, including informatics nurses, database architects, data analysts, and nurse leaders to determine where data elements were housed, how they were named, and when and how they were measured. Placing each variable and its pick list choices on a spreadsheet enabled us to see matches and differences across all sites. Our task was to harmonize and merge the data from all 4 sites into a single large database. To achieve that, table names, variable names, and pick list choices had to be congruent. Our solution was creation of our own ontology, or framework of standard terms, to map the disparate terms to (12). Ontology development requires consensus on the concept definitions and nomenclature standards (13) so the team held group meetings where word-for-word we reviewed the terms and came to consensus on how to map each term to the ontology. We called our ontology the gold standard. Next a unique identifier code was created for every data element in the data set mapping the term to the gold standard term (See Figure, Supplemental Digital Content 3, <http://links.lww.com/JONA/A228>).

Version control

The ontology and subsequent mapping was our solution to harmonize data differences resulting from the various versions and customizations that occurred. Little can be done prospectively about version control, but being aware of it up front may have led us to choose different partners who were on the same software version. Also, mappings supplied by the vendors may provide backward compatibility among versions.

Customization

In contrast to version control, customization can be avoided. Users must be aware of the consequences of making unique changes to an EHR product sold by a leading information system vendor. These changes complicate or negate the ability to do research or quality measurement across systems. In fact, some sites also admitted that customization creates problems internally when they want to upgrade to a new version as all of their changes are lost and must be rebuilt. EHR customization could prevent users from upgrading and therefore miss out on potentially improved products. Although customization is always tempting to increase adoption of the EHR, when data are merged across systems, problems arise with mismatched terms and redundancy. Users must consider and be educated about these issues as we validate whether the customization is really necessary and creates added value worth these risks.

Documentation policies

The particular EHR vendor used by the 4 sites provides a venue for users to collaborate via a user's group. Participation in the user's group is highly recommended to gain insight into the care processes and experience of other sites. This communication among system users may decrease the variation in documentation policies that impact how the information system is used. Through collaborative problem solving, policies may be drafted that benefit research and quality improvement efforts that lead to meaningful use. Participation in our study helped the sites gain valuable insights they could share with others. Further, administrators and nurse leaders should consider the documentation policies in place and determine the frequency and requirements for important assessments regarding fall risk, pain, catheter use, depression, functional status, and pressure ulcer risk.

Quality of documentation

The challenges to documentation quality were addressed by education, monitoring, and positive reinforcement. We held grand round presentations, used voice over PowerPoint and storyboard presentations to educate the nurses about the study. We provided pens with the slogan, *Nursing Documentation Makes a Difference*. We collaborated with unit committee leaders to serve as champions to monitor and encourage high quality documentation and recognition that nursing data was being put to use for improved patient care. For every step along the way we maintained a constant stream of communication via on-site visits, email communications, telephone and web conferencing. These efforts supported the study operations but are equally critical in general operations to assure high quality documentation for quality measurement. Administrators and nurse leaders should assess the evaluation plans that measure the completeness and accuracy of EHR documentation and determine the actions taken for improvement or maintenance.

User Interface

The impact of the user interface design on documentation workflow became evident during the study. Workflow is always a critical element in health information technology implementation and research (14,15). The way certain data elements were presented on the screen or the timing in the workflow of their appearance can affect the quality of the data. We addressed this issue by altering the screen configuration so that buried items were brought forward so that the clinician was aware of them. The sites we worked with were unaware that, because of the EHR design, a patient could be discharged from their care without having their function for activities of daily living measured. Administrators and nurse leaders should request a review of the screen placement of critical data elements while considering how it prompts the user to ask questions and complete the data entry. They should also request periodic reports on the completeness of documentation. Education was critical to identify where the data elements were located and how to interpret and answer them.

Implications

The lessons learned in this multi-site study are generalizable to clinical information consumers. Take home messages include; 1) avoid local customization; 2) participate and set policies within the wider user's group; 3) educate the nurses to value their contribution to quality documentation and make them aware their data is being used for larger purposes; and 4) critically review the design of the system for workflow issues that may impact quality. Since users of clinical information systems are usually unaware of database principles regarding the naming, storage, and retrieval of patient information (16), the experiences described above provide administrators and nurse leaders with several real life examples of how day-to-day documentation practices can impact research and process improvement. Due to a lack of standardization between and among clinical information systems world-wide, many complex processes must be undertaken to resolve discrepancies (4,13,17). Forethought about these issues is critical to simplify and avoid these costly errors.

In addition, we make 4 more recommendations. One, to meet meaningful use criteria, most health systems are building data warehouses to facilitate the storage, retrieval, and transfer of essential clinical data as needed to answer research and quality questions (18). It is imperative that data collected by nurses are also housed in the warehouses to assure inclusion with other essential EHR data elements to tell the complete story of a patient's health care needs and effectiveness of interventions. Second, administrators must appoint representatives to health system information technology committees to inform decisions about which data elements are important for measurement of nursing effectiveness. Critical

review of the content, and understanding the location and meaning of data elements in the nursing documentation system provides essential information to inform these decisions. Next, administrators are highly encouraged to seek the purchase of information systems that use standardized languages mapped to a reference terminology such as SNOMED CT (19) or the 3M Healthcare Data Dictionary (20). To date, most vendors have not widely incorporated the standardized languages for nursing documentation such as the Omaha System (21), Clinical Care Classification (22) or the North American Nursing Diagnoses Association (23). Use of a standardized language already mapped to the information system reference terminology will eliminate the costly and time consuming step of building the ontology and completing the mapping required for our project. Using standards is a global recommendation to facilitate data sharing and interoperability (24). Thus we recommend that EHR developers utilize the recommended standardized languages for nursing documentation and that locally we adhere to the standards to avoid the challenges experienced and discussed herein. Finally, we suggest the addition of a guiding principle to the Guiding Principles for the Nurse Executive to Enhance Clinical Outcomes by Leveraging Technology (25). Nurse executives clearly recognize that their organizations are part of a larger community of health systems, and this guiding principal should apply to how information technology is managed in terms of changes made to the EHR and the documentation policies created and applied. We want to avoid creating silos of data customized specifically for local applications. The costs of harmonizing such data after the fact are high in terms of time and labor and the lengthy process impedes our ability to use the EHR as a rich data source to do research and quality improvement.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The project described was supported by Award Number R01NR007674 from the National Institute of Nursing Research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Nursing Research or the National Institutes of Health.

The authors thank the nursing leadership and information systems personnel at the 4 hospitals for their support and perseverance in completing the data retrieval process.

References

1. Hayrinen K, Saranto K, Nykanen P. Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *Int J Med Inform.* 2008; 77(5):291–304. [PubMed: 17951106]
2. Office of the National Coordinator for Health Information Technology (ONC), Department of Health and Human Services. Health Information Technology: Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology, 2014 Edition; Revisions to the Permanent Certification Program for Health Information Technology. Federal Register. 2012; 77(171):54163–54292. [PubMed: 22946139]
3. Institute of Medicine. Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary. Grossmann, C.; Powers, B.; McGinnis, JM., editors. National Academies Press; Washington, DC: 2011.
4. 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 Inform Assoc.* 2007; 14(1):1–9. [PubMed: 17077452]
5. Neumann, EK. Semantic harmonization: the power and limitation of ontologies. In: Alterovitz, G.; Ramoni, M., editors. Knowledge-Based Bioinformatics: From Analysis to Interpretation. John Wiley & Sons, Ltd; Chichester, UK: 2010.

6. Greenes, RA. *Clinical Decision Support: The Road Ahead*. 1st ed.. Elsevier Academic Press; Waltham, MA: 2007.
7. Osheroff, JA.; Teich, JM.; Levick, D., et al. *Improving Outcomes with Clinical Decision Support: An Implementer's Guide*. 2nd ed.. Healthcare Information and Management Systems Society; Chicago, IL: 2012.
8. Carroll AE, Tarczy-Hornoch P, O'Reilly E, Christakis DA. Resident documentation discrepancies in a neonatal intensive care unit. *Pediatrics*. 2003; 111(5 Pt 1):976–980. [PubMed: 12728074]
9. Jefferies D, Johnson M, Griffiths R. A meta-study of the essentials of quality nursing documentation. *Int J Nurs Pract*. 2010; 16(2):112–124. [PubMed: 20487056]
10. Samuels JG, Fetzer S. Pain management documentation quality as a reflection of nurses' clinical judgment. *J Nurs Care Qual*. 2009; 24(3):223–231. [PubMed: 19525763]
11. Jones, SS.; Koppel, R.; Ridgely, MS.; Palen, TE.; Wu, S.; Harrison, MI. *Guide to Reducing Unintended Consequences of Electronic Health Records*. Agency for Healthcare Research and Reform; Rockville, MD: 2011.
12. Yu AC. Methods in biomedical ontology. *J Biomed Inform*. 2006; 39(3):252–266. [PubMed: 16387553]
13. Schuurman N, Leszczynski A. A method to map heterogeneity between near but non-equivalent semantic attributes in multiple health data registries. *Health Informatics J*. 2008; 14(1):39–57. [PubMed: 18258674]
14. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. *Risk Manag Healthc Policy*. 2011; 4:47–55. [PubMed: 22312227]
15. Russ AL, Saleem JJ, Justice CF, Woodward-Hagg H, Woodbridge PA, Doebbeling BN. Electronic health information in use: Characteristics that support employee workflow and patient care. *Health Informatics J*. 2010; 16(4):287–305. [PubMed: 21216808]
16. Branson A, Hauer T, McClatchey R, Rogulin D, Shamdasani J. A data model for integrating heterogeneous medical data in the health-e-child project. *Stud Health Technol Inform*. 2008; 138:13–23. [PubMed: 18560104]
17. Shi, Y.; Liu, X.; Xu, Y.; Ji, Z. Semantic-based data integration model applied to heterogeneous medical information system; The 2nd International Conference on Computer and Automation Engineering (ICCAE); Singapore. 2010; Singapore: Nanyang Technological University; p. 624-628.
18. Chaudhuri S, Dayal U. An overview of data warehousing and OLAP technology. *ACM SIGMOD record*. 1997; 26(1):65–74.
19. International Health Terminology Standards Development Organisation. SNOMED CT. 2012. Available at: <http://www.ihtsdo.org/snomed-ct/> [Accessed November 12, 2012]
20. Healthcare Informatics. The Vocabulary of Interoperability: An Interview with Dr. Hon Pak, former U.S. Army CIO. *Healthcare Informatics*; Jun-Jul. 2012 p. 53-54.
21. Martin, KS. *The Omaha System: A Key to Practice, Documentation, and Information Management*. 2nd ed.. Elsevier Saunders; St Louis, MO: 2005.
22. Saba VK. Nursing classifications: Home health care classification system (HHCC): An overview. *Online J Issues Nurs*. 2002; 7(3):9. [PubMed: 12410634]
23. NANDA International. NANDA-I, NIC and NOC for Safe Patient Care. Available at: <http://www.nanda.org/NNN.aspx> [Accessed November 8, 2012]
24. World Health Organization. Workshop on semantic interoperability prerequisites for efficient e-health systems: Initial considerations. Available at: <http://www.who.int/classifications/terminology/prerequisites.pdf>. Updated January 31, 2005 [Accessed November 8, 2012]
25. American Association of Nurse Executives. AONE Guiding Principles for the nurse executive to enhance clinical outcomes by leveraging technology. Mar 6. 2009 Available at: http://www.aone.org/resources/PDFs/AONE_GP_Leveraging_Technology.pdf2013

Table 1

Standard Pick List Options.

Data Element	Study Ontology	Hospital A	Hospital B	Hospital C	Hospital D
Lives With	-alone -child(ren) -domestic partner -facility resident -friend(s) -grandparent(s) -other relative(s) -parent(s) -sibling(s) -spouse -homeless	-alone -children -spouse -significant other -parents -other relative (specify) -friend -grandparent(s)	-Alone -Children -Spouse -Significant Other -Parent -Other Relative -Domestic Partner -Nursing Home -Mental Health Facility -State Veterans Home -Assisted Living Facility -Group/Boarding Home Facility -Homeless	-alone -children -spouse -significant other -parents -other relative (specify) -domestic partner -friend -grandparent(s)	-alone -child(ren) -dependent child(ren) -domestic partner -facility resident -friend(s) -grandparent(s) -other relative(s) -parent(s) -sibling(s) -significant other -spouse