Title
Modeling Flowsheet Data to Support Secondary Use

Permalink
https://escholarship.org/uc/item/7hw2f453

Journal
CIN Computers Informatics Nursing, 35(9)

ISSN
1538-2931

Authors

Westra, Bonnie L
Christie, Beverly
Johnson, Steven G
et al.

Publication Date
2017-09-01

DOI
10.1097/cin.0000000000000350

Peer reviewed

Modeling Flowsheet Data to Support Secondary Use

Bonnie L. Westra, PhD, RN, FAAN, FACMI[1,2], **Beverly Christie, DNP, RN**[3], **Steven G. Johnson, PhD, MS**[1], **Lisiane Pruinelli, PhD, MSN, RN**[1], **Anne LaFlamme, DNP, RN**[3], **Suzan G. Sherman, PhD, RN**[3], **Jung In Park, PhD, BS, RN**[1], **Connie W. Delaney, PhD, RN, FAAN, FACMI**[1,2], **Grace Gao, DNP, RN**[1], and **Stuart Speedie, PhD, FACMI**[2]

[1]University of Minnesota, School of Nursing, Minneapolis, MN, USA

[2]University of Minnesota, Institute for Health Informatics, Minneapolis, MN, USA

[3]Fairview Health Services & University of Minnesota Health, Minneapolis, MN, USA

Abstract

The purpose of this study was to create information models from flowsheet data using a data-driven consensus based method. Electronic health records contain a large volume of data about patient assessments and interventions captured in flowsheets that measure the same "thing," but the names of these observations often differ, according to who performs documentation or the location of the service (e.g., pulse rate in an intensive care, the emergency department, or a surgical unit documented by a nurse or therapist or captured by automated monitoring). Flowsheet data are challenging for secondary use due to the existence of multiple semantically equivalent measures representing the same concepts. Ten information models were created in this study: five related to quality measures (falls, pressure ulcers, venous thrombosis embolism, genitourinary system including catheter associated urinary tract infection, and pain management) and five high volume physiological systems: cardiac, gastrointestinal, musculoskeletal, respiratory, and expanded vital signs/anthropometrics. The value of the information models is that flowsheet data can be extracted and mapped for semantically comparable flowsheet measures from a clinical data repository regardless of the time frame, discipline, or setting in which documentation occurred. The 10 information models simplify the representation of the content in flowsheet data, reducing 1,552 source measures to 557 concepts. The amount of representational reduction ranges from 3% for Falls to 78% for the Respiratory System. The information models provide a foundation for including nursing and interprofessional assessments and interventions in common data models, to support research within and across health systems.

Corresponding Author: Bonnie L. Westra, PhD, RN, FAAN, FACMI, University of Minnesota, School of Nursing, 308 Harvard St SE, WDH 5-140, Minneapolis, MN 55434, 612-625-4470, Westr006@umn.edu.

Conflicts of Interest

The authors have no conflict of interest to report.

INTRODUCTION

Flowsheets are templated documentation forms in electronic health records (EHR) used by interprofessional health care clinicians, and if standardized, provide a rich source of data for secondary use such as quality improvement and research. Flowsheets are organized like a spreadsheet and include structured or semi-structured data for rapid documentation and visualization of assessments, interventions, and other types of data for a variety of health professions including nursing, physical, occupational, and speech therapy, social work, nutritionists, and others. The types of information captured in flowsheets are called flowsheet measures. Examples of measures are "Heart Rate," "Pain Rating," and "Pressure Ulcer Location." The inclusion of flowsheet data in clinical data repositories (CDRs) when combined with other data like patient demographics, laboratory results, and medical diagnoses can increase our understanding of factors contributing to outcomes, such as prevention of patient falls and pressure ulcers, or best methods of pain management. However, flowsheet data are not standardized within and across health systems. The purpose of this study was to create data-driven information models from EHR flowsheets to support secondary use of the data for quality improvement and research. Data extracted from EHRs can be mapped to these information models to describe care for a patient population using standards that are more useful for researchers both within and across healthcare organizations. These models enable analysis of how interprofessional care is related to patient outcomes.

BACKGROUND

Modeling and representation of flowsheet data have been addressed in a few studies where researchers have discussed the relevance of this data for quality improvement and information retrieval.[1,2] Other investigators have operationalized flowsheet data in ontologies for inclusion in data repositories, [3,4] but investigators have reported data harmonization problems and, consequently, limitations for multi-site studies. In one study, a data-driven ontological approach was used to create a pressure ulcer information model.[5] Additional information models are needed as well as a process for mapping multiple types of flowsheet measures to the identified concepts.

While flowsheet data is a rich source of information, secondary use is limited in CDRs due to multiple challenges in normalizing that data.[6] These challenges include: (1) the massive amount of data in flowsheets, (2) many unique measures for semantically equivalent concepts that may have different names and are not linked through an information architecture in the EHR (i.e., "Heart Rate" and "Pulse"), and (3) local customization of the value sets (i.e., the set of allowable answers for a flowsheet measure) within and across EHR implementations. In a pilot study of 199,665 encounters from one CDR, investigators noted that 34% of the data was documented in flowsheets which was twice the size of the next largest data type – orders and procedures (17%).[7] There are a variety of reasons for multiple semantically equivalent flowsheet measures: multiple EHR builders add new flowsheet measures without reusing existing ones; or, requests for customization for slight variations in names or value sets by discipline, programs, or settings (i.e., emergency department or intensive care units) contribute to duplication. Additionally, insufficient tracking and

mapping during EHR software upgrades may result in deprecation of some flowsheet measures while new ones are created that represent essentially the same concepts. The result is semantically equivalent flowsheet measures that are stored with different flowsheet identification (ID) numbers. For secondary use, this means that all semantically equivalent flowsheet measures must be linked for valid conclusions about quality measures or research for a population receiving care over time, in different settings, or by different disciplines. Thus, information models are needed to map semantically equivalent concepts from flowsheet data.[7]

METHOD

Purpose and Design

The purpose of this study was to create data-driven information models from EHR flowsheets to support secondary use of the data for quality improvement and research. The study is a retrospective observational study using an iterative consensus-based approach to identify concepts from multiple resources, but only those concepts supported by actual patient data were included in information models. Concepts represent assessment questions and interventions performed about a clinical topic, such as pain. The concepts are logically organized into a hierarchical model and used to map semantically equivalent flowsheet measures to concepts.

Data Source

The University of Minnesota (UMN) maintains a CDR that includes EHR data from one health system composed of seven hospitals, over 40 clinics in a midwestern state. The CDR is maintained under the auspices of UMN's Clinical and Translational Science Institute (CTSI). The CDR has more than 2.5 million patients and more than 4 billion rows of data that include patient encounters, demographics, medical diagnoses, procedures, laboratory results, medications, notes, and flowsheet measures. The flowsheet data represents more than 34% of all rows contained in this CDR. After approval by the Institutional Review Board, a de-identified subset of 199,665 encounters representing 66,660 patients who received care between October 20, 2010 and December 27, 2013 was provided in a secure data shelter. The scope of the project included development of 10 information models. Initial topics were five clinical quality measures from a pilot study. Topics were later expanded to include review of systems, building on the proposed model for flowsheet data by Warren et al.[1] The five quality measures were falls, pressure ulcers, venous thrombosis embolism (VTE), genitourinary system including catheter associated urinary tract infection (CAUTI), and pain management. The five high volume physiological systems were cardiac, gastrointestinal, musculoskeletal, respiratory, and expanded vital signs/anthropometrics.

Process

In the health system's EHR, flowsheet measures are organized into templates and groups. Templates represent the screen where data were documented and contain groups of individual flowsheet measures that are logically related for a specific topic. Groups consist of a set of closely related measures that are collectively used in one or more templates. Examples of templates are "Emergency Department" or "Adult Patient Admission."

Examples of groups are assessments for "Skin" or "Musculoskeletal System." Flowsheet measures can be included in many groups and templates. There are semantically equivalent flowsheet measures which display within different groups or templates but have different names. Flowsheet measures represent assessments, interventions, or other phenomenon.

De-identified flowsheet data were extracted and summarized in two spreadsheets. The first spreadsheet, entitled "Documentation Context," showed the relationship of flowsheet measures within the templates and groups in which they were found. The second spreadsheet entitled "Summarized Measures" included a count for the frequency unique flowsheet measures were documented across all templates and groups and included the data type (i.e., numeric, text, date, choice list) and the set of values (answers documented). The researchers used these two spreadsheets to develop multiple information models.

Creation of Information Models

Each investigator selected a clinical topic for creating an information model, identified concepts from the spreadsheets, research, evidence-based practice guidelines, Web sites that include clinical data models, and other resources such as textbooks. Investigators used these concepts and related synonyms to search for concept terms in templates, groups, and flowsheet measures. Any flowsheet measures that had fewer than 10 observations were eliminated; 10 was used as a cut point to eliminate measures that were part of the model build or flowsheets measures that were designed but not used. Investigators created the information models in spreadsheets that organized the concepts in a hierarchical manner with manual mapping of one more many flowsheet IDs to the concepts and added value sets from choice list measures. While resources exist to find concepts and synonyms for terms as well as display information in a hierarchical manner, such as Mind Mapper (Irvine, CA), these resources do not automate the mapping process resulting in a need for manual mapping.

Validation of Information Models

Investigators presented the information models and mappings to flowsheet IDs for consensus validation during weekly team meetings. Through this review process, rules were refined to ensure consistency in mapping flowsheet data to clinical concepts and information models. Each model was reviewed by a second investigator to affirm mappings of flowsheet IDs to concepts, identify any flowsheet measures that may have been missed, and present findings to the research team for validation.

RESULTS

The flowsheet data consisted of 153,049 data points for 14,564 measures (each measure is one type of row) in 2,972 groups in 562 templates. There were 10 information models created. The number of flowsheet measures mapped to an information mode ranged from 59 to 309 (see Table 1). As shown in Table 1, the left column represents the name of the information model and in the second column, the number of unique flowsheet measures associated with concepts in the information model. The right hand columns demonstrate that the number of concepts to which the flowsheet measures were mapped and the organization

of concepts into classes sets of closely related measures. The information models simplified the representation of the content in flowsheet data from a total of 1,552 flowsheet measures to 557 concepts within the 10 information models. Figure 1 demonstrates how a concept in the information model is associated with multiple flowsheet measures that are semantically equivalent (for example, "Genitourinary Conditions" is mapped to three different flowsheet measures as indicated by the three unique IDs in the ID column). The amount of reduction ranged from 3% for Falls to 78% for the Respiratory System. Figure 2 is a depiction of an information model using Unified Modeling Language (UML®, Needham, MA) showing the Genitourinary Information Model concepts and the relationships between these groups developed in Microsoft® Office Visio® (Redmond, WA).

All 10 of the information models are available in the "Supplemental Digital Content (SDC 1). High Level Clinical Information Models from Flowsheet Data." The high level information models include classes (groups of concepts) and the concepts in the information models; they do not include the data type, values, or mappings to specific flowsheet IDs (these are available upon request from the primary author).

DISCUSSION

In this study, 10 clinical information models were created from EHR flowsheet data using a data-driven consensus based approach to support secondary use of the data. The information models encompass data related to five quality measures required for reporting to the Centers for Medicare and Medicaid Services or the Joint Commission. Additional information models include review of systems. The data-driven approach by Harris et al. [5] was used and a unique aspect of this study was the inclusion of multiple information models and extension of this process to include mapping flowsheet data to concepts in the information model for replication across health care systems.

The information models are intended to support data delivery to researchers when EHR data are needed over time, across units or settings, and documented by numerous disciplines. Examples of current projects that use flowsheet data are: relating the impact of compliance with Surviving Sepsis Guidelines to patient outcomes, discovering factors associated with unintended Intensive Care Unit admission after elective surgery, or discovering factors associated with CAUTI. In addition to supporting research within a single organization, the information models derived in this study can enable the extension of common data models for comparison across settings such as those used by the Patient-Centered Outcomes Research Institute (PCORI) and National Center for Advancing Translational Science (NCATS). [8,9]

Results of this study build upon and expand previous research to standardized flowsheet data for secondary use. Warren proposed a model for organizing flowsheet data in i2b2 (Informatics for Integrating Biology & the Bedside, Boston, MA). The i2b2 tool is widely used by academic health centers to query their CDRs in comparable ways across systems; however, the proposed configuration for flowsheet data is not included yet. The information models developed in this study build on Warren's proposed model for organizing flowsheet data in i2b2.[7,10] The method used by Harris et al.[5] informed how information models were

created in this study. There are some differences, however. The current study builds on their pressure ulcer model and adds models for nine additional clinical areas.

Ideally, EHR software vendors would use common information models with nationally recognized data standards for flowsheet data; however, this is not yet the case. The chaos in flowsheet data exists in the most modern EHRs and is not unique to implementation in any specific health setting. When vendors do not have common information models and support customization of their systems, the result is inconsistent data within and across systems. The University of Minnesota has hosted the Nursing Knowledge Big Data Science Conference for the past 4 years to support a national action plan for identifying, standardizing, implementing, and effectively using sharable and comparable nurse-sensitive data.[11] Representatives from practice, industry, academia, and professional and governmental organizations attend this think-tank type summit and collaborate throughout the year via the 10 virtual working groups to achieve the vision of sharable and comparable nurse-sensitive data to support interoperability, quality improvement, and research. Considerable effort has gone into standardizing documentation that supports billing; this same effort has not supported standardization of nursing documentation such as flowsheets to demonstrate the value of nursing care.[12] Implementation of a national action plan that supports sharing common information models and data standards across vendors and health systems is essential and this study provides a foundation for such effort.

FUTURE RESEARCH

Future research is needed to increase the generalizability of findings from this study. A second phase is in process for validation of the information models by other organizations. Once this process is completed, concepts will be mapped to standardized terminologies. Consistent with 2016 Interoperability Standards[13] and the American Nurses Association's position on use of Nursing Terminologies,[14] assessments and outcomes will be coded with Logical Observation Identifiers Names and Codes (LOINC) and Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT) will be used to code the value sets for assessments, problems, and interventions. Both LOINC and SNOMED-CT terminologies are evolving for physiological concepts,[15] such as those in our information model which will be used and expanded as a result of our work. Research is needed to create additional information models as this study addressed only a portion of flowsheet data. Furthermore, a method for continuously updating the information models as the flowsheet data change is also needed. Finally, studies are needed that use these information models across settings to demonstrate the value of standardized models for flowsheet data used in research. The authors plan to test one or more of the information models across two or more PCORI funded sites to expand their common data model to address a research question such as the most common non-pharmacological interventions and their association with reducing pain.

LIMITATIONS

There are several limitations to this study. The information models were created using data from a single organization; external validation from other organizations and clinical experts is needed. The generalizability of the information models may be limited until conceptual

definitions and data standards are included. There may be missing concepts in information models due to using a subset of data from one CDR rather than all flowsheet data in the CDR.

A manual process was used to create information models. This can lead to errors. An open software tool is in development, FloMap, to automate rules for finding and mapping flowsheet measures to information model concepts. This tool was developed by one of the research team (S. Johnson, Minneapolis, MN) but is not yet available since it is in beta testing to improve the usability and decrease the effort for validating, creating, and maintaining information models.

CONCLUSION

The purpose of this study was to create information models from flowsheet data using a data-driven consensus based method. Flowsheet data are challenging for secondary use due to the existence of multiple semantically equivalent measures representing the same concepts. Ten information models were created in this study: five related to quality measures (falls, pressure ulcers, venous thrombosis embolism (VTE), genitourinary system including catheter associated urinary tract infection (CAUTI), and pain management), and five high volume physiological systems: cardiac, gastrointestinal, musculoskeletal, respiratory, and expanded vital signs/anthropometrics. The value of the information models is that flowsheet data can be mapped and extracted for semantically comparable flowsheet measures from a CDR regardless of the time frame, discipline, or setting in which documentation occurred. The 10 information models simplify the representation of the content in flowsheet data, reducing 1,552 source measures to 557 concepts. The amount of representational reduction ranges from 3% for Falls to 78% for the Respiratory System. The information models provide a foundation for including nursing and interprofessional assessments and interventions in common data models, such as PCORI, to support research within and across health systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Warren JJ, Manos EL, Connolly DW, Waitman LR. Ambient findability: Developing a flowsheet ontology for i2B2. Nurs Inform. 2012; 2012:432.

2. Kim H, Harris MR, Savova GK, Chute CG. The first step toward data reuse: Disambiguating concept representation of the locally developed ICU nursing flowsheets. Comput Inform Nurs. 2008; 26(5):282–289. DOI: 10.1097/01.NCN.0000304839.59831.28 [PubMed: 18769183]

3. Chow M, Beene M, O'Brien A, et al. A nursing information model process for interoperability. J Am Med Inform Assoc. 2015; 22(3):608–614. DOI: 10.1093/jamia/ocu026 [PubMed: 25656515]

4. Kim H, Choi J, Secalag L, Dibsie L, Boxwala A, Ohno-Machado L. Building an ontology for pressure ulcer risk assessment to allow data sharing and comparisons across hospitals. AMIA Annu Symp Proc. 2010; 2010:382–386. [PubMed: 21347005]

5. Harris MR, Langford LH, Miller H, Hook M, Dykes PC, Matney SA. Harmonizing and extending standards from a domain-specific and bottom-up approach: An example from development through use in clinical applications. J Am Med Inform Assoc. 2015; 22(3):545–552. DOI: 10.1093/jamia/ocu020 [PubMed: 25670750]

6. Waitman LR, Warren JJ, Manos EL, Connolly DW. Expressing observations from electronic medical record flowsheets in an i2b2 based clinical data repository to support research and quality improvement. AMIA Annu Symp Proc. 2011; 2011:1454–1463. [PubMed: 22195209]

7. Johnson SG, Byrne MD, Christie B, et al. Modeling flowsheet data for clinical research. AMIA Jt Summits Transl Sci Proc. 2015; 2015:77–81. [PubMed: 26306244]

8. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc. 2014; 21(4):578–582. DOI: 10.1136/amiajnl-2014-002747 [PubMed: 24821743]

9. FitzHenry F, Resnic F, Robbins S, et al. Creating a common data model for comparative effectiveness with the observational medical outcomes partnership. Applied Clinical Informatics. 2015; 6(3):536–547. [PubMed: 26448797]

10. Westra BL, Christie B, Johnson SG, et al. Expanding interprofessional EHR data in I2B2. AMIA Joint Summits on Translational Science proceedings AMIA Summit on Translational Science. 2016

11. Delaney CW, Pruinelli L, Alexander S, Westra BL. 2016 nursing knowledge big data science initiative. Comput Inform Nurs. 2016; 34(9):384–386. DOI: 10.1097/CIN.0000000000000288 [PubMed: 27606577]

12. Welton JM. What's a nurse's value? making cents of care. Nurs Econ. 2016; 34(2):57, 81. [PubMed: 27265945]

13. Office of the National Coordinator for Health Information Technology. [Accessed April 21, 2016] 2016 interoperability standards advisory. https://www.healthit.gov/sites/default/files/2015interoperabilitystandardsadvisory01232015final_for_public_comment.pdf. Updated 2016

14. ANA Board of Directors. [Accessed April 21, 2016] Inclusion of recognized terminologies within EHRs and other health information technology solutions. http://www.nursingworld.org/MainMenuCategories/Policy-Advocacy/Positions-and-Resolutions/ANAPositionStatements/Position-Statements-Alphabetically/Inclusion-of-Recognized-Terminologies-within-EHRs.html. Updated 2015

15. Matney SA, Settergren T, Carrington JM, Richesson RL, Sheide A, Westra BL. Standardizing physiologic assessment data to enable big data analytics. Western Journal of Nursing Research. 2016:1–15. First published on line July 18, 2015.

Question	Type	Information Model Concept Name	Value Type	ID
	Class	Genitourinary Assessment		
	Question	Genitourinary WDL	8	606179
1	Value	WDL		
1	Value	WDL Except		
	Question	Genitourinary Conditions	8	674599, 675571, 605415
5	Value	acute renal failure		
5	Value	benign prostatic hyperplasia		
5	Value	burning		
5	Value	chronic kidney disease		
5	Value	decreased output		
5	Value	prostatitis		
5	Value	pyelonephritis		
5	Value	recent urinary tract infection		
5	Value	recurrent urinary tract infection		
5	Value	renal calculi		
5	Value	retention		
5	Value	urgency		
	Question	Bladder Symptoms	8	676482
28	Value	bladder distended		
28	Value	bladder fullness		
28	Value	bladder pressure		
28	Value	bladder spasm		
28	Value	flank pain		
28	Value	other (see comments)		
28	Value	urinary retention		
	Question	Voiding Symptoms	8	674075
36	Value	burning		
36	Value	decreased output		
36	Value	denies		
36	Value	excessive odor		
36	Value	frequency		

Figure 1.
Partial example of a high level Genitourinary clinical Information Model from flowsheet data.
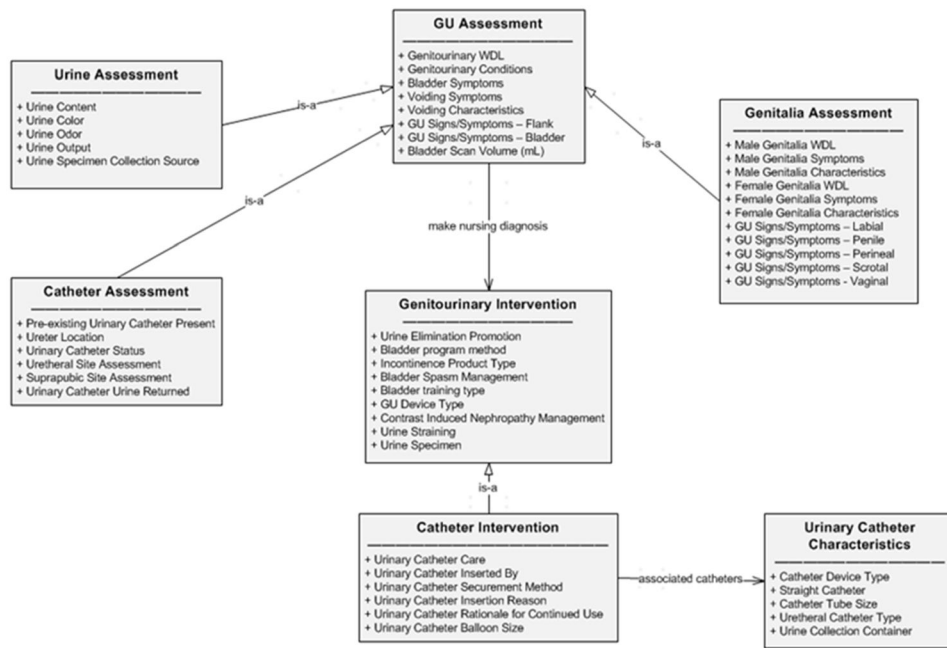
Figure 2.
UML model for Genitourinary clinical information model developed in Microsoft® Office Visio®.

Table 1

Information Models Derived from Flowsheet Data

Information Model Name	Number Flowsheet Measures	Number Information Model Concepts/Classes	
		Concepts	Classes
Cardiovascular System	241	84	8
Falls	59 *	57	4
Gastrointestinal System	60	28	3
Genitourinary System (including CAUTI)	79	38	3
Musculoskeletal System	276	72	9
Pain	309	80	12
Pressure Ulcers	104	56	6
Respiratory System	272	61	12
Venous Thrombosis Embolism (VTE)	67	16	8
Expanded Vital Signs/Anthropometrics	85	48	10

*Observations for Falls is underreported as assessment questions from multiple groups are integrated throughout the flowsheets and responses to these assessments trigger recommendations for interventions to prevent falls. This method prevents duplicate data entry, but also makes it challenging to track actual numbers of flowsheet measures that are included.