TOR VERGATA
UNIVERSITÀ DEGLI STUDI DI ROMA

DOTTORATO DI RICERCA
IN SCIENZE INFERMIERISTICHE E SANITA' PUBBLICA

XXXIII CICLO - Anno Accademico 2019/2020

Dottorando: Dott.ssa Mariangela Vanalli

Tutor: Dott. Fabio D'Agostino e Dott. Davide Ausili

Titolo tesi: Impact, standards, and analysis of a clinical nursing information system in a hospital setting

ABSTRACT

Background

Nursing documentation is a knowledge source of patient and provable evidence demonstrating how decisions are made and decision outcomes are recorded. Recently, introduction of new technologies has resulted in changes in health care organizations and practices. Documentation is the most important of these changes in transition from a paper-based health record to an electronic health record (EHR). Using modern technologies is unavoidable in the knowledge era, EHR is the new technology enabling health system evolution. In Italy, a clinical decision support system (END-CDSS) called professional assessment instrument (PAI) has been implemented in one large hospital in Rome. The PAI is an electronic health record for documenting nursing care according to the structure of the nursing process. The use of the nursing process has shown an improvement in both patient outcomes and documentation accuracy that it should always be assessed. Electronic nursing documentation should include the nursing process phases; furthermore, these nursing care elements should be expressed using a standardized nursing taxonomy. The documentation of nursing care in EHRs with standardized nursing terminology provides a means for producing consistent data about patient needs assessment, nursing diagnoses, planning, delivering interventions and patient outcomes evaluation to share, compare, and merge with other data across systems.

The Clinical Care Classification (CCC) system is one of the few standard nursing terminologies recognized by the American Nurses Association (ANA) and is a royalty-free terminology, but an Italian version of it is still not available. The CCC system developed to describe nursing care and useful for interoperability of nursing concepts in EHR. Interoperability of health information systems is crucial: healthcare providers and researchers working with data must understand the importance of data mining.

Nurses play a key role in data collection and generation of patient information. Nursing records that are based on a combination of structured data entry and unstructured data entry should ensure that the free text data are available for reuse whereas single free text data such as clinical narrative notes would not allow comparison with other standardized languages. However, narrative notes present a challenge because they can take on many forms as it reflects nurse's perception of the patient condition and can include a variety of highly telegraphic terms and many abbreviations. This variability makes it difficult to extract useful information from nursing notes, due to their non-standardization.

An automated text analysis can extract specific data and convert unstructured data into structured data from a corpus comprised of a significant amount of narrative data, and it can be run using the natural language processing (NLP) and data mining techniques. Standardized nursing terminologies and non-standardized nursing terms assisted by NLP techniques can be the source of data for comparing terms to determine their semantic equivalence through a cross-mapping strategy with String Metric-assisted Assessment of Semantic Heterogeneity (SMASH). The accuracy of this SMASH strategy could allow to highlight and validate semantic differences between non-standardized nursing terms and other standardized languages, underlying the transformation process and the results connected to the transition from non-standardized nursing terms to standardized uniform nursing data.

Objective

The objectives of this doctoral research were 1) to evaluate whether a nursing information system called PAI can improve the accuracy of nursing documentation in an Italian hospital unit; and 2) to describe the translation process of the CCC into Italian to achieve a meaning-equivalent and cross-culturally appropriate version and to preliminarily test its clinical validity; and 3) to identify unstructured free-text nursing activities recorded by nurses in EHRs and then map these nursing activities into standard nursing activities using the SMASH method and NLP techniques.

Methods

Three studies were conducted to achieve the three objectives. For the first objective, a quasi-experimental longitudinal design was conducted, collecting data from the PAI using the D-Catch instrument. For the second objective, a translation with cross-cultural adaptation was performed in four phases: forward-translation, back-translation, review, and dissemination. Subsequently a pilot cross-mapping study between nursing activities in free-text nursing documentation and the CCC interventions was conducted.

For the third objective, a cross mapping study with three phases was conducted: a) text summarisation component with NLP techniques, b) a consensus analysis by four experts to detect the category of word stems of unstructured free-text nursing notes and c) cross-mapping method with SMASH.

Results

The first study showed that the PAI implementation significantly improved ($p < .001$) the chronologically descriptive accuracy and diagnostic accuracy of the nursing documentation after periods of four months, eight months, and one year. The second study found that all elements of the CCC system were translated into Italian. Semantic and conceptual equivalences were achieved. In the pilot cross-mapping study, 77.8% of the free-text nursing activities were mapped into the CCC interventions, where 40.9% and 36.9% were fully or partially mapped, respectively.

The third study described the results of the three phases. In the first phase, a total of 8491 tokens were extracted from free-text nursing notes and analysed with the text mining application R@1-1; subsequently, the identification of the tokens, 1087 word stems, were calculated with the application of cut off set in the tokenizer, semantic, stop word removal and stemming phases. In the second phase, of 1087 word stems extracted with NLP techniques, only those belonging to the macro area "nursing activities" were considered. A total of 548 word stems, were categorized in nursing activities by four experts with 100% agreement. In the third phase, these 548 word stems that were unstructured free-text nursing activities, were mapped to PAI standard nursing activities and lexical similarities (with Jaro-Winker distance [JWD] and Levenshtein distance [LD]) were the metrics used for the SMASH. Of 548 concepts searched for automated lexical mapping (with JWD), 451 (82.3%) unstructured free-text nursing activities were mapped to PAI standard nursing activities (with <13 LD), 47 (8.7%) were partial mapped (with <13 LD), while 50 (9.0%) were not mapped (with >13 LD). This automated mapping yielded a recall of 0.95%, a precision of 0.94%, an accuracy of 0.91% and an F-measure of 0.96. The F-measure (ranging from 0 to 1) indicates good reliability of this cross-mapping automated procedure. The area under the ROC curve of 0.97 (95% CI 0.96-0.98) shows that the automated procedure is accurate and performs to the cross-mapping method.

Discussion and conclusion

This doctoral program represented the first Italian study that showed a relationship between a clinical nursing information system that uses an SNT and nursing documentation accuracy, the first study that described the translation and cross-cultural adaptation of the CCC system in Italian, and the first study was applied text analytics with NLP techniques and cross-mapping strategy with SMASH to process free-text nursing data recorded in the EHRs. The first study showed that an electronic nursing documentation using END-CDSS can support nurses in the nursing process by improving their clinical reasoning skills. The second study outlined the translation process of the CCC and its cross-cultural adaptation in the Italian context. The CCC was designed to act as a clinical interface terminology and, thus, it could be an ideal language for the integration of nursing care data implementation into EHRs, in view of the expanding digitization of clinical documentation.

The third study showed how the NLP techniques and the SMASH method were a feasible approach to extract data related to nursing concepts that were not recorded through structured data entry and to analyse the semantic heterogeneity between unstructured free-text nursing activities and PAI standard nursing activities.

Keywords: Cross-mapping, professional assessment instrument, standardized nursing terminology, nursing documentation accuracy, nursing process, electronic nursing documentation, Clinical Care Classification system, cross-cultural comparison, nursing minimum data set, natural language processing.