2007

Development of a specialist medical vocabulary for radiation oncology

Andrew Alexis Miller
University of Wollongong, amiller@uow.edu.au

Ping Yu
University of Wollongong, ping@uow.edu.au

Publication Details

Andrew, M. & Yu, P. (2007). Development of a specialist medical vocabulary for radiation oncology. In K. Kuhn, J. Warren & T. Leong (Eds.), World Congress on Medical Informatics (pp. P342). Netherlands and HISA Ltd Australia: IOS Press.

Development of a specialist medical vocabulary for radiation oncology

Abstract

Controlled medical terminologies that have been developed to describe terms utilized in the field of Radiation Oncology, include SNOMED-CT and DICOM-RT. However, a literature review has failed to provide evidence that the coverage and the level of granularity of these nomenclatures have satisfied the needs of radiation oncologists. Indeed most investigations conclude that the coverage is generally unsatisfactory. Furthermore, there is no evidence that an objective specification of the specialist medical terms used in Radiation Oncology has been developed. We report the development of a Specialist Medical Vocabu-lary for Radiation Oncology using an objective and systematic method of discovery of data elements published in the Radia-tion Oncology literature. The importance of the data elements to radiation oncologists is judged according to the criterion that a submitted report has been deemed worthy of publica-tion. Within the time period of discovery, 97 articles were retrieved and, during the analysis of 80 articles, 622 individual data elements and 2392 instances of use were found. Infrequent data elements comprised the majority of individual data elements (54%), and frequently used data elements were a minority (27 individual data elements with 10 or more in-stances of use). However these 10 data elements comprised 49.5% of the total data elements found.

Keywords

vocabulary, oncology, development, radiation, medical, specialist

Disciplines

Physical Sciences and Mathematics

Publication Details

Andrew, M. & Yu, P. (2007). Development of a specialist medical vocabulary for radiation oncology. In K. Kuhn, J. Warren & T. Leong (Eds.), World Congress on Medical Informatics (pp. P342). Netherlands and HISA Ltd Australia: IOS Press.

MEDINFO 2007
HISA Ltd Australia

Development of a Specialist Medical Vocabulary for Radiation Oncology

A. Andrew Miller[a], Ping Yu[b]

[a]*Department of Radiation Oncology, Illawarra Cancer Care Centre, The Wollongong Hospital, Wollongong, Australia*
[b]*Health Informatics Research Centre, School of Information Systems and Technology, University of Wollongong, Australia*

Abstract

Controlled specialist medical vocabularies (SMV) include some terms used in Radiation Oncology (RO), although no objective specification of the specialist medical terms used in RO has been published.

We are developing a Specialist Medical Vocabulary for Radiation Oncology (SMV-RO) using an objective and systematic method of discovery of data elements. The importance of any data element to radiation oncologists is judged according to the criterion that it was included in a report deemed worthy of publication in the RO literature.

From a defined period, 97 articles were retrieved. Analysis of 80 articles found 622 individual data elements and 2392 instances of use. Infrequent data elements comprised the majority of individual data elements (54%), and frequently used data elements were a minority (27 individual data elements with 10 or more instances of use). However these 10 data elements comprised 49.5% of the total data elements found.

Keywords:

vocabulary, controlled, radiation oncology, clinical, specialist medical vocabulary, oncology information system

Introduction

Radiation oncologists are medical specialists who treat cancer patients with high energy x-rays and work with software designed to minimize human error in the delivery of radiation to patients which also includes an electronic medical record module (the OIS-EMR).

Specialist Medical Vocabularies (SMV) list standardized terms devoid of relationships [1], but are the first step in defining ontologies. Any SMV should undergo frequent review to addressing inadequacies [2] as clinically unused SMVs appear to decay over periods of only 4-5 years [3].

Clinical data in a SMV format informs clinical decision making [4] and clinical outcomes, but must be stored in matching data structures in the OIS-EMR. This will enable successful data retrieval with its standardized meaning intact, which is a requirement for successful implementation of a national electronic health record (EHR) which derives its information for patient interaction.

The oncology data is a small proportion of a patient's EHR. A SMV for Radiation Oncology (SMV-RO) has not been developed, although there is some overlap with general terminologies (MeSH thesaurus, SNOMED), and specific collections (NCI's Common Data Elements dictionary for oncology trials [5]).

Research aims

We aim to document the SMV used in the domain of RO.

Research methodology

The SMV-RO should be determined by systematic extraction of terms from a corpus of text which is carefully sampled to be maximally representative.

The corpus that informs the practice and knowledge of RO is the published literature. The importance and relevance of the data sets used in the literature is determined by its acceptance for publication after peer review.

Content analysis can analyze and classify written communications in conjunction with domain specific knowledge. A methodology of "*text deconstruction*" was developed and applied by a radiation oncologist (AAM) to find and list the data elements used in a report's description of data collection, measurement and analysis. Aggregation of these data elements formed the SMV-RO. The initial manual analysis establishes a gold standard within a well defined corpus of the literature, against which to later compare automated tools.

A focused PubMed search was undertaken for articles describing the use of radiotherapy in the month of February 2006 (search "Major MeSH Heading [Radiotherapy] AND 2006/02[pdat]"). Other specialist RO journals, general oncology journals and general medical journals were also surveyed.

Results

The data collection and analysis for the first 97 articles are presented, which revealed 622 data elements used in the analysis or description of study groups. Seventeen articles did not include patient data and were excluded.

A total of 2392 instances of data element were catalogued (mean of 3.8 instances/data element). The data elements were unevenly distributed as is seen in Table 1.

Table 1 - Distribution of data elements

Frequency of occurrence	Individual data elements (count)	Total number of occurrences (sum)	% of data elements (count)	% of occurrences (sum)
5n1	523	886	84.1	37.0
10n6	44	331	7.1	13.8
15n11	25	308	4.0	12.9
20n16	9	163	1.4	6.8
25n21	3	69	0.5	2.9
30n26	6	168	1.0	7.0
35n31	3	100	0.5	4.2
40n36	5	187	0.8	7.8
45n41	2	83	0.3	3.5
50n46	2	97	0.3	4.1
55n51	0	0	0.0	0.0
60n56	0	0	0.0	0.0
Total	622	2392	100	100

The commonest data elements used include: Diagnosis (n=65), Date of Birth (n=50), Histology (n=47), Surgery Decision (n=42), Radiotherapy Decision (n=41), and Date of Death (n=38). Data elements seen once were commonest (345 occurrences), but while one to five occurrences produced 37% of the total elements found, the remaining 15.9% of data elements produced 64% of the total elements found.

Inferred data elements were common. The most frequent inferred data element is the date of event which has preeminence in the analysis of outcome in oncology reports.

Data elements were categorized into physical and temporal separations within the oncology treatment and management process as shown in Table 2.

Table 2 - Distribution of data elements within the Corpus

	Individual data elements	Total frequency of	Entries per data element (mean)
PATIENT	305	1044	3.4
Disease	240	764	3.2
Diagnosis	35	429	12.3
THERAPY	258	1232	4.8
Surgery	15	151	10.1
Radiotherapy	111	453	4.1
Drugs	20	106	5.3
Chemotherapy	71	335	4.7
Other Therapy	1	2	2
OUTCOMES	60	181	3.0
Recurrence	14	99	7.1
Response	8	37	4.6
Patient Milieu	**341**	**576**	**1.7**

Many data elements defined the tumour's characteristics (e.g., oncogenes expressed). The 'Patient Milieu' group comprises patient assessments of anatomy (imaging), physiology (blood tests), psychology (Quality of Life items), and clinical status (signs & symptoms) undertaken initially, while on treatment and after treatment, and constituted 54.8% of the total, and were used infrequently (mean 1.69 entries/data element).

Discussion

Assessments form a very large part of this SMV, although the usage frequency is low. The correlation of biological factors with clinical outcome may improve cancer therapy and indicates that authors should itemize their data sets (data elements and choices) to permit similar data collection. Other departments should be able to include new assessments into the OIS-EMR to permit medical staff to assess and record data in identical formats for later analysis. The Common Data Elements approach of the NCI is a relevant example and easily reviewed for potential terms [5].

The lack of relationships between terms within this specialist medical vocabulary were apparent. The data elements discovered always related to an event related to a patient, a patient's diagnosis, or the patient's treatment for their diagnosis. When the vocabulary is used to define an ontology and OIS-EMR, the storage of these terms should allow for reconstruction of the relationships present at collection. For example, an assessment of an entity such as depression could be measured as a psychological factor as well as part of a Quality of Life Assessment, an eligibility factor for an intervention, or a biological factor modifying the response to therapy, a prognostic factor predicting outcome, or an outcome of therapy.

To cover the complete range of assessment terms is impossible by the method we have used. The dynamic nature of oncological knowledge argues for the establishment of a modern data set with periodic updates [8], ideally through the submission of data sets by the authors of the published reports as part of the submission for publication.

Conclusion

While the vocabulary is a starting point that has successfully demonstrated that specified terms used in Radiation Oncology can be derived from the Radiation Oncology literature, further work remains to reflect the complexity of this vocabulary by the inclusion of concepts and relationships, in particular any ontology must explicitly quantify the time of measurement [6]. Future work should map this vocabulary into other systematic nomenclatures (SNOMED, DICOM, and ultimately the UMLS), as well as seek software methods to automate discovery of new terms.

The nature of many terms already discovered argues that the OIS-EMR should possess an assessment tool that can assign assessments to the correct setting, that are scalable to include many data elements that are used infrequently and are able to be drawn from established nomenclatures.

References

[1] P. Fabry, R. Baud, P. Ruch, C. Despont-Gros, and C. Lovis, "Methodology to ease the construction of a terminology of problems," *International Journal of Medical Informatics*, vol. 75, pp. 624-632, 2006.

[2] D. E. Oliver, "Synchronization of Diverging Versions of a Controlled Medical Terminology," *Proceedings of the 1998 AMIA Annual Fall Symposium*, pp. 850–854, 1998.

[3] M. F. Chiang, D. S. Casper, J. J. Cimino, and J. Starren, "Representation of ophthalmology concepts by electronic systems Adequacy of controlled medical terminologies," *Ophthalmology*, vol. 112, pp. 175-183, 2005.

[4] R. Serban, A. ten Teije, F. van Harmelen, M. Marcos, and C. Polo-Conde, "Extraction and use of linguistic patterns for modelling medical guidelines," *Artificial Intelligence in Medicine*, vol. 39, pp. 137-149, 2007.

[5] NCI, "CDE Browser: Data Element Search. http://cdebrowser.nci.nih.gov/CDEBrowser/," National Cancer Institute.

[6] Y. Lee, K. Supekar, and J. Geller, "Ontology integration: experience with medical terminologies," *Comput Biol Med*, vol. 36, pp. 893-919, 2006.