



Aalborg Universitet

AALBORG  
UNIVERSITY

## Terminology-based recording of clinical data for multiple purposes within oncology

Brønnum, Dorthe; Højen, Anne Randorff; Gøeg, Kirstine Rosenbeck; Elberg, Pia Britt

*Published in:*  
Exploring Complexity in Health: An Interdisciplinary Systems Approach

*DOI (link to publication from Publisher):*  
[10.3233/978-1-61499-678-1-267](https://doi.org/10.3233/978-1-61499-678-1-267)

*Creative Commons License*  
CC BY-NC 4.0

*Publication date:*  
2016

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

### *Citation for published version (APA):*

Brønnum, D., Højen, A. R., Gøeg, K. R., & Elberg, P. B. (2016). Terminology-based recording of clinical data for multiple purposes within oncology. In *Exploring Complexity in Health: An Interdisciplinary Systems Approach: Proceedings of MIE2016 at HEC2016, 28 August-2 September 2016, Munich, Germany* (pp. 267-271). IOS Press. <https://doi.org/10.3233/978-1-61499-678-1-267>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

### **Take down policy**

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.

# Terminology-Based Recording of Clinical Data for Multiple Purposes Within Oncology

Dorthe BRØNNUM<sup>a,1</sup>, Anne Randorff HØJEN<sup>b</sup>, Kirstine Rosenbeck GØEG<sup>b</sup> and  
Pia Britt ELBERG<sup>b</sup>

<sup>a</sup> Department of Oncology, Aalborg University Hospital, Denmark

<sup>b</sup> Department of Health Science and Technology, Aalborg University, Denmark

**Abstract.** Introduction: Collecting clinical data once for the use in both electronic health record (EHR) and registries requires semantic interoperability. This paper presents the results of a systematic semantic analysis of similarities and differences in clinical documentation across regional EHR and a national oncology registry to assess options for an integration of recording templates. Methods: A comparison of current clinical information in EHR and the national registry was carried out, using SNOMED CT as frame of reference to find exact-, similar- and non-match. Results: Exact match was found for 9 out of 19 items from the registry and EHR, relating to clinical history, observations and findings at the examination and tumor control. Similar match concerned clinical findings of more common side effects to therapy whether present or absent. Both EHR and the registry had information with no compared match. Conclusion: Clinical documentation during a follow-up in head and neck cancer contains a core set of items recorded in both EHR and registry, representing clinical history, observations and more common side effects and tumor evaluation. These core items could be the point of departure for integration or re-design of EHR-systems.

**Keywords.** Clinical informatics, hospital information systems, computerized medical record systems, research infrastructure and EHR data reuse.

## 1. Introduction

Clinical documentation is required for a wide range of purposes: patient care, administration, public health, quality and research. This information is typically managed in separate incompatible systems such as regional electronic health record (EHR), national registries and temporary research databases. Besides requiring time-consuming multiple registrations during a patient encounter, this infrastructure also impairs the use of data across patient encounters.

Collecting clinical data once and reusing data for different purposes requires semantic interoperability [1]. Ries, Krumm and Anderson [2-4] have presented their work of identifying, structuring and harmonizing clinical data in EHR for secondary use in quality management, registries and research.

This paper presents the results of a systematic semantic analysis of similarities and differences in current documentation in EHR and a national oncology registry for quality and research purposes (DAHANCA - Danish Head and Neck Cancer Group).

---

<sup>1</sup> Corresponding Author: dsb@rn.dk, Department of Oncology, Aalborg University Hospital, Hobrovej 18-22, 9000 Aalborg, Denmark;

The aim of this analysis was to identify semantic similarities and differences in the clinical content of information across EHR and a national registry to assess options for an integration of recording templates. Integrated templates would minimize parallel documentation and support semantic interoperability.

## 2. Methods

In this study, a semantic comparison of clinical information in EHR and registry was chosen because of the well-known granularity differences between the two systems, with the EHR containing detailed clinical information and the registry with classifications of more general results. In an integration situation, identifying areas of similar information and not only exact matched information, would allow finely grained information to be drawn from the EHR along with classification of the information and communication to the registry.

### 2.1. Material

The scope of this analysis was set to include the clinical situation of an outpatient follow-up, where information on clinical history, morbidity, side effects and tumor control was documented in EHR and registry for quality and research purposes. This context was suitable for further analysis due to similar clinical procedure across regional centers, a delimited timespan of each visit and noticed similarities and differences in clinical information for EHR and registry.

Consultant oncologists at Danish centers for head and neck cancer therapy were asked to collect clinical notes representative of the documentation in a follow-up patient visit. Four out of six centers contributed with clinical notes, and all with documentation in free text. A template from the national clinical quality and research database from a follow-up visit in head and neck cancer was retrieved, consisting of 19 highly structured items with a total of 56 outcomes (DAHANCA registry).

### 2.2. Identifying information structures and common semantic in clinical notes in EHR

An initial analysis of the information within the four clinical notes was performed to identify different statement types. The four notes and the Danish national guideline for a clinical examination in a follow-up visit in head and neck cancer were compared. Headlines in the text, replicate information and side effects/morbidity were used to create a list of statement types under which all sentences in the clinical notes were structured.

### 2.3. Semantic comparison of EHR and Registry data

Clinical data in EHR and registry were analyzed by comparing statement types representative of free text in notes and registry items. This comparison was done using SNOMED CT as a tool, an international terminology also translated by the Danish health authorities. Similarities and differences was found in adherence to published methods for analysis of clinical content [5]. In an iterative process both the initial statement types and registry items/outcomes were compared to SNOMED CT, finding

terms representative of the expressions in the text and registry outcomes. Existing guidelines were used in the mapping to SNOMED CT [6], by means of which clinical terms were found for more than 90% of the outcomes in the registry.

Similarities and differences in EHR and registry were defined as follows:

- Exact match: where two terms are identical.
- Similar match: where two terms are closely linked by the relationship in SNOMED CT. For example, “Disturbance of salivary secretion” is a close relationship to “Dry mouth” thus similar match was found.
- Non-match: No match, nor exact or similar, between terms representing the expression in text and terms representing the high structured data in the registry or reverse.

### 3. Results

Semantic similarities and differences of existing documentation were identified as exact match, similar match and non-match.

Exact match was found for 9 out of 19 registry items, with identical information in registry and at least one of the clinical notes. Exact match was found for clinical history, observations, clinical findings at the examination and tumor evaluation.

Similar match was found for 6 out of 19 registry items, where information in the registry was closely linked to the information in EHR by terminology relations. Similar match was related to clinical findings of more common side effects whether present or absent. All six similar matches also had exact match. (Table 1)

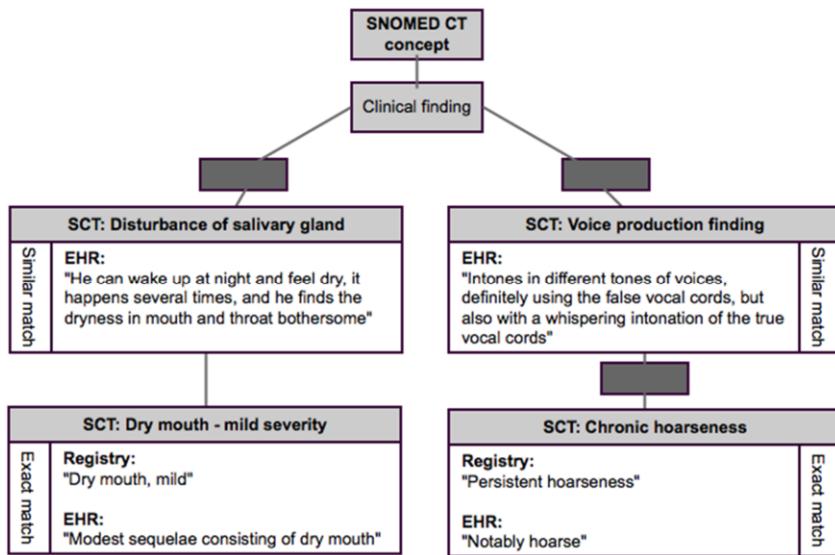
**Table 1.** Exact and similar matches in comparison of clinical data in EHR and registry

| Context                           | Exact match  | Similar match   |
|-----------------------------------|--|---|
| Clinical history and observations | Smoker – Nonsmoker<br>Weight                                   |   |
|                                   | Dysphagia §#<br>Dry mouth §#<br>Hoarseness §                   | Swallowing finding §<br>Disturbance of salivary gland §<br>Voice production finding § |
| Clinical examination findings     | Atrophic mucosa §<br>Fibrosis of skin #<br>Oedema of larynx §# | Mucosal finding §<br>Larynx finding §   |
| Tumor evaluation                  | Status of tumor finding §                                      | Sub-conclusions, tumor finding§   |

§ Both exact and similar match were found.

# Common morbidity according to Danish national guideline for follow-up in head and neck cancer

Two examples of exact and similar match are illustrated in Figure 1. “Dry mouth” has exact match in EHR and registry and similar match by close relation to “Disturbance of salivary gland”. “Chronic hoarseness” has exact match in EHR and registry and similar match by close relation to “Voice production finding”.



**Figure 1.** Exact match in “Dry mouth” and similar match by close relation to “Disturbance of salivary gland”. Exact match in “Chronic hoarseness” and similar match by close relation to “Voice production finding”. Dark gray boxes illustrate the hierarchical tree of SNOMED CT, but are not relevant in this example.

There was no match between text in clinical notes and registry items concerning clinical summary, general findings and procedures for the examination. In reverse order there was no match between registry items and EHR related to infrequent clinical findings. (Table 2). Below are two examples of non-matched data, both related to a clinical examination of the mouth and tongue:

- Only present in EHR: “No suspicious findings by inspection and palpation.”
- Only present in the registry: “Susceptible to caries” (not present, mild, moderate or severe grade)

**Table 2.** Non-matches in comparison of clinical data in EHR and registry.

| Context             | Text in EHR with no match to registry items                                      | Registry items with no match to text in EHR   |
|---------------------|--|---|
| Clinical summary    | Reason for visit<br>Diagnosis and given treatment                                | Status for follow-up  |
| General findings    | State of nutrition<br>Health and performance<br>Pain<br>Social status            |   |
| Procedures          | Procedures for examination<br>Procedures for laryngoscopia<br>Partial evaluation |   |
| Clinical findings # |  | Gastrointestinal tube<br>Perichondritis<br>Susceptible to caries<br>Neurological symptom<br>Osteoradionekrosis<br>Tracheostomy/laryngectomy<br>Other side effects or new cancer |

# Infrequent clinical findings according to Danish national guideline for follow-up in head and neck cancer

## 4. Discussion

Comparison of similarities and differences in clinical documentation with SNOMED CT demonstrates the possibility to use and reuse clinical data in EHR and registry. The results concerning similar match revealed the need for flexible documentation. Both highly structured items and semi-structured documentation would be necessary to fulfill the different purposes of documentation in both EHR and registry, as emphasized by Rosenbloom [7].

The four clinical notes used for comparison represented typical examples of documentation in a follow up visit in head and neck cancer. Three patient visits were uncomplicated, whereas the fourth required intervention due to positive findings related to morbidity. Clinical notes from more complicated visits, like the one mentioned or with tumor relapse could have added valuable information to the content in EHR.

Current clinical documentation formed the basis for comparing information in EHR and registry. A re-design of EHR would make it possible to adjust the content and possibly improve the results, but a re-design would also require negotiations among clinical experts to reach consensus of a clinical meaningful representation of core data items within the SNOMED CT structure, to allow for re-design of regional EHR.

## 5. Conclusion

The study demonstrated that a semantic comparison can inform of the contents of EHR compared to a quality and research registry, as it takes into account the different granulation levels. Using semantic comparison, we showed, that follow-up in head and neck cancer contains a core set of items recorded in both EHR and registry, representing the clinical history, observations and more common side effects whether present or absent and tumor evaluation. These core items could be the point of departure for integration or re-design of EHR-systems.

## References

- [1] J. J. Cimino, Desiderata for controlled medical vocabularies in the twenty-first century, *Methods Inf. Med.* **37** (1998), 394-403.
- [2] M. Ries, H.U. Prokosch, M.W. Beckmann and T. Burkle, Single-source tumor documentation - reusing oncology data for different purposes, *Onkologie* **36** (2013), 136-141.
- [3] R. Krumm, A. Semjonow, J. Tio, H. Duhme, T. Burkle, J. Haier, M. Dugas and B. Breil, The need for harmonized structured documentation and chances of secondary use - results of a systematic analysis with automated form comparison for prostate and breast cancer, *J. Biomed. Inform.* **51** (2014), 86-99.
- [4] H. V. Anderson, W. S. Weintraub, M. J. Radford, M. S. Kremers, M. T. Roe, R. E. Shaw, D. M. Pinchotti and J. E. Tcheng, Standardized cardiovascular data for clinical research, registries, and patient care: a report from the Data Standards Workgroup of the National Cardiovascular Research Infrastructure project, *J. Am. Coll. Cardiol.* **61** (2013), 1835-1846.
- [5] K. R. Gøeg, R. Chen, A. R. Højern, and P. Elberg, Content analysis of physical examination templates in electronic health records using SNOMED CT, *Int. J. Med. Inf.* **83** (2014), 736-749.
- [6] A. R. Højern and K. R. Gøeg, Snomed CT implementation. Mapping guidelines facilitating reuse of data, *Methods Inf. Med.* **51** (2012), 529-538.
- [7] S. T. Rosenbloom, J. C. Denny, H. Xu, N. Lorenzi, W. W. Stead, and K. B. Johnson, Data from clinical notes: A perspective on the tension between structure and flexible documentation, *J. Am. Med. Inform. Assoc.* **18** (2011), 181-186.