Single-Source Tumor Documentation – Reusing Oncology Data for Different Purposes

Markus Riesa, Hans-Ulrich Prokoschb, Matthias W. Beckmannb,c, Thomas Bürkleb

aChair of Medical Informatics, University Erlangen-Nuremberg, bComprehensive Cancer Center Erlangen-Nuremberg, University Hospital Erlangen, cDepartment of Obstetrics and Gynecology, University Hospital Erlangen, Germany

Keywords
- Tumor documentation · Cancer · Single-source · Information technology · Dataset

Summary
We present a path towards single-source tumor documentation established at the Comprehensive Cancer Center Erlangen-Nürnberg (CCC-EN). Our goal was to derive data for cancer quality assurance and certification, cancer registry documentation and cancer research directly from routine care documentation. Therefore, clinical documentation activities were analyzed and a cancer data superset, containing these required elements, was developed. This superset was then split into appropriate clinical documentation packages, and the existing information technology infrastructure was analyzed and adapted to accommodate those documentation packages. A clinical documentation package is the amount of cancer-relevant data that can be captured within a clinical encounter. This grouping of data enables integration into existing clinical documentation workflows. We present examples in which single-source tumor documentation has been successfully established at the CCC-EN. The resulting cancer documentation reference model is described and its transferability to other institutions discussed.

Schlüsselwörter
- Tumordokumentation · Krebs · Single-Source · Informationstechnologie · Datensatz

Zusammenfassung
Introduction

Cancer is the cause of 12% of illness-related deaths worldwide [1]. According to Eschenbach and Buetow, we are currently in the midst of an explosion of knowledge about cancer as a disease process. In their vision paper from 2006 on cancer informatics [2], they cite the US National Cancer Institute (NCI), which had challenged the oncology community to eliminate the suffering and death due to cancer by 2015, and admit that it will be impossible to achieve this ambitious goal without better interconnectivity and coordination across the cancer enterprise. The authors proposed the Cancer Biomedical Informatics Grid (caBIG™) as an interoperable biomedical informatics infrastructure to overcome these obstacles. One of their key issues focused on ‘Medical or clinical informatics enabling the management, analysis and dissemination of clinical and public health data, including the use of informatics infrastructure and applications such as clinical trial management systems, electronic health records, and cancer registries’. However, the caBIG™ tools mainly provided support for cancer research, bioinformatics, and molecular research, and the close relationship between clinical care documentation, cancer registries and translational cancer research received less attention [3]. On the other hand, organizations such as the International Agency for Research on Cancer (IARC – WHO), the National Cancer Data Base (USA), the National Cancer Data Strategy (Australia) and the Working Group of German Tumor Centers (Germany) have emphasized the importance of cancer data registration and have tried to standardize cancer registry documentation [4–6]. National cancer plans for optimized care have been set up [7–9], but often this results in additional paper- or information technology (IT)-based documentation efforts, e.g. for quality assurance, registry documentation and/or cancer research [10]. In Germany, the National Cancer Plan has identified the multitude of independent and uncoordinated documentation requirements as a major barrier for efficient cancer documentation. The existence of parallel independent documentation systems leads to a tremendous workload, and thus hinders acceptance among physicians. Incompatible documentation systems overstrain hospitals and clinicians, and non-comparable datasets also lead to varying results and inadequate conclusions. Thus, harmonization of normative regulations, the definition of 1 nationwide minimal cancer dataset and a data-sparring homogeneous cancer documentation were fundamental recommendations within the cross-sectional field of action for ‘cancer documentation’ [11].

When the Comprehensive Cancer Center Erlangen-Nuremberg (CCC-EN) was founded in 2008 as 1 of 11 oncology centers of excellence in Germany, those recommendations were defined as general principles of all IT developments, and a comprehensive single-source approach fulfilling the requirements of cancer center certification, cancer registry documentation and cancer research, at the same time, was declared as our ultimate goal. We therefore developed a reference model for single-source tumor documentation, comprising a generic clinical documentation workflow and a cancer data superset. This superset was divided into clinical documentation packages in which every single documentation package could be documented at the clinical encounter, i.e. where such data originate, and within the respective clinical system that is already in routine use at the respective organizational unit of the hospital. Within the overall IT infrastructure, the different systems were interfaced to provide data items for research purposes, and, in particular, a complete cancer data superset for the cancer registry database.

Methods

Current clinical documentation processes at Erlangen University Hospital were analyzed, and 13 medical specialty executives and 11 external cancer documentation experts were interviewed using a semi-structured interview guide covering oncology care processes, cancer registration, IT support and future use of data. These processes led to the identification of clinical documentation steps, which were then chronologically ordered and assigned to organizational units and roles. We identified recurring documentation steps within the history of individual patients and the association of these recurrences with particular types of care (e.g. hospital inpatient, hospital outpatient or ambulatory care). In addition, the current medium of documentation (e.g. paper or electronic) and the respective documentation system, if data were acquired electronically, were identified.

We matched these documentation scenarios with cancer documentation requirements from the viewpoint of cancer center certification, cancer quality assurance programs, and cancer research, as well as cancer registry documentation. The latter was defined as the joint set of basis tumor documentation of the Working Group of German Tumor centers (Arbeitsgemeinschaft Deutscher Tumorzentren, ADT), the dataset of the Society of Epidemiological Cancer Registries (Gesellschaft der epidemiologischen Krebsregister in Deutschland, GEKID) and the Bavarian cancer registry law (BCRL). We also identified 16 different European and German quality assurance and certification initiatives; within the CCC-EN context we focused on the German national quality insurance program and the 10 certificates of the German Cancer Association (Deutsche Krebsgesellschaft, DKG) [12].

Within a hospital-wide interdisciplinary working group comprising IT staff, medical staff and members of the cancer registry, we thus defined a cancer data superset based on the analysis described above. Currently, however, although no specific research data elements have been added, we have confirmed that the cancer data superset meets basic requirements, e.g. for clinical trials documentation, and have added data items to document a patient’s informed consent and assignment to clinical studies. In this process we realized that, currently, none of the existing German cancer datasets (e.g. ADT or GEKID) meets clinical documentation reality. Although the datasets group data elements, e.g. into primary visit data or primary therapeutic procedure, these data are not completely accessible, e.g. in a primary visit. Clinical and pathological TNM staging for example are determined in different steps during patient care. The latter is only available when a tissue sample has been examined in the pathology department and, at present, is documented in a freetext pathology report. However, both items are required for the ADT primary visit data.

Similar to the approach of Altmann et al. [13], we divided the cancer data superset into clinical documentation packages according to the previously identified clinical documentation workflow and the respective
clinical event for which those data are typically recorded. In a final step, we assigned each documentation package to the respective computerized information system at the CCC-EN and identified which IT interfaces needed to be implemented.

Results

The generic clinical documentation workflow, depicted in Figure 1, illustrates that 4 staff groups are involved in tumor documentation: administrative staff in admissions, medical staff from different medical departments at various points of time, staff of quality management responsible for certification activities, and cancer registry staff. The documentation workflow starts with patient registration and administrative data documentation (1.1), which is followed by obtaining the patient’s informed consent for data reuse in the cancer registry and biomaterials storage (1.2). Next, the clinical department responsible for the cancer entity (e.g. urology for prostate, bladder and renal cancer) collects anamnestic and initial assessment data (2.1), organizes the tumor conference (3.4), and, depending on the tumor staging, performs surgical procedures (4.1) (with the exception of non-solid tumors or primary radiooncology treatment) or organizes other therapeutic alternatives (e.g. immunotherapy) (4.4). Later, the responsible clinical department organizes follow-up care (5.1). The pathology department (3.1), radiology department, nuclear medicine (3.2) and laboratory (3.3) are involved at various times in the diagnostic and staging process. Departments for radiooncology (4.2) and oncology (4.3) are important providers for non-surgical therapy. Cross-sectional supportive measures are provided from departments such as psycho-oncology, social services and palliative medicine including pain therapy (4.5–4.10). In addition, we included data reuse for cancer research and biobanking, for quality assurance and for cancer registry purposes.

The set of the data items required for or resulting from each documentation step comprised 286 data items, which were ordered into 20 clinical documentation packages (see fig. 2). A clinical documentation package is the amount of cancer-relevant data that can be captured within a certain clinical activity or encountered by 1 staff group and within 1 clinical documentation system (e.g. EMR (the electronic medical record), RIS (radiology information system), LIS (laboratory information system), pathology (pathology documentation system), or surgery (the surgical theatre documentation system)). We identified 3 different types of datasets: those that are required once only for each patient, such as patient administrative data; those that have to be documented for each cancer entity of a patient (1 patient may have several cancer occurrences simultaneously or at different times); and those that are required repeatedly, e.g. diagnostics, treatments and follow-ups. Some of these documentation packages can also be documented in outpatient or ambulatory settings.

The cancer data superset should be captured only once at its origin and reused wherever needed. The superset is based on the combined datasets of the ADT, GEKID and BCRL plus additional clinical data elements (see table 1). 161 elements (56%) come from the ADT dataset. 89% of the ADT elements have been used, omitting for example data collected from autopsy. Interestingly, 19% of the GEKID and 24% of the BCRL dataset elements are not part of the ADT, prompting us to add 64 elements from GEKID and 45 from BCRL for mandatory complete coverage of those datasets. None of the German standard datasets included documentation of data from supportive activities, such as psycho-oncology, pain care or palliative care. Thus, to complete the cancer data superset, we added another 83 elements from those areas, which are mostly also required as indicators within cancer organ center certifications (e.g. as elements in the dotted documentation packages in Figure 3).

The main goal of single-source tumor documentation is data reuse. Consequently, data should be captured in structured, numeric or date format for further computing purposes. Of the 286 data elements, 83% fulfill this requirement; 14% are unstructured free text data elements containing, for example, name or address, which also had to be accepted. To cross-check our cancer data superset with similar international activities, we performed an automated mapping of its data items to the NCI thesaurus and found appropriate NCI concepts for 98% of the 286 data elements.

Currently, a stepwise process is underway to implement all those documentation items within the IT systems of Erlangen University Hospital. For the Urology Clinic (prostate, kidney and bladder cancer patients) the complete set of clinical documentation packages has already been implemented, support-

Table 1. A Cancer data superset and the different proportions of the ADT, GEKID and BCRL datasets used to develop it. B The internal structure of this superset: this is important to enable data reuse

<table>
<thead>
<tr>
<th>A: Cancer data superset and relationship to subsets</th>
<th>ADT</th>
<th>GEKID</th>
<th>BCRL</th>
<th>Added data elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of data elements</td>
<td>161</td>
<td>64</td>
<td>45</td>
<td>83</td>
</tr>
<tr>
<td>Proportion used for cancer data superset, %</td>
<td>89</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Proportion of cancer data superset, %</td>
<td>56</td>
<td>22</td>
<td>16</td>
<td>29</td>
</tr>
<tr>
<td>Part of ADT but not of, %</td>
<td>–</td>
<td>68</td>
<td>79</td>
<td>–</td>
</tr>
<tr>
<td>Part of GEKID but not of, %</td>
<td>19</td>
<td>–</td>
<td>50</td>
<td>–</td>
</tr>
<tr>
<td>Part of BCRL but not of, %</td>
<td>24</td>
<td>29</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B: Structure of cancer data superset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of data element</td>
</tr>
<tr>
<td>Structured options or lists</td>
</tr>
<tr>
<td>Data fields</td>
</tr>
<tr>
<td>Numeric fields</td>
</tr>
<tr>
<td>Names, addresses, etc.</td>
</tr>
<tr>
<td>Free text</td>
</tr>
<tr>
<td>Sum</td>
</tr>
</tbody>
</table>
Thus, the respective clinical IT applications (such as the EMR, the surgery documentation system, and RIS) and non-clinical IT systems (such as the cancer registry system or research databases) were interfaced to realize single-source documentation. In our example data from the urology department, such as the cancer diagnosis, ICD-O (International Classification of Disease for Oncology) morphology code, and the ECOG and clinical TNM classification are reused with a national biobanking project portal of the German Prostate Cancer Consortium [15, 16], for the required quality indicators for the German Cancer Association (DKG) certification, and for the Erlangen Cancer Registry database (see fig. 3).

In total, based on the described workflow model and cancer data superset, we have implemented 12 different documentation scenarios at the CCC-EN so far. Our goal is a hospital-wide implementation of this concept for all cancer locations.
Discussion

The lack of communication and information interchange between systems, the lack of simplified standards and the lack of implementation and adherence to available standards were recently identified by Chahal as the major challenges for establishing efficient IT infrastructures for cancer research [17]. Lau et al. [18] emphasized the value of EMR data for conducting epidemiological and outcome research, while Ochs and Casagrande in 2008 [19] proposed an integrated information architecture for comprehensive cancer centers and also stressed the need for methods to integrate research systems with other (clinical) enterprise systems. Although this problem has been partially addressed (e.g. in the NCI-founded caBIG™) [20] and the National Health Service (NHS)-funded Cancer Grid initiative [21], to our knowledge no single-source reference architecture based on 1 standardized dataset exists. We could also find no publications on a comprehensive cancer center IT implementation process in which such a single-source approach has been strictly followed to support clinical care, quality assurance, cancer registration and translational research.

Developments like the NCI thesaurus can be seen as a first step, but have the limitation that they do not supply a dataset for tumor documentation [22], but rather a full collection of potential data elements. The NCI thesaurus has never been mapped on real clinical processes to define which elements should be captured in which specific clinical encounter.

Although many German publications describe the goal to integrate single-source tumor documentation directly in the clinical care process [23–28], the recommendations of the German Cancer Plan are often misinterpreted as ‘the need to establish 1 standardized cancer documentation system for the whole cancer care process’ [11]. In our opinion, however, the IT landscape in university hospitals is extremely heterogeneous and subsets of cancer data are already documented in various systems. Therefore, our approach concentrated on: defining an institution-wide cancer data superset (based on the German datasets of ADT, GEKID and BCRL, enriched by additional data elements to meet the requirements of German quality assurance initiatives); mapping of subsets (documentation packages) of this superset on particular clinical encounters and organizational units; the documentation of those items within the already available clinical systems, and the comprehensive integration of all such systems with communication interfaces.

This approach can potentially be adapted to any other institution by: (1) Analysis of clinical workflows in cancer care and development of an (optimized) clinical documentation workflow; (2) analysis of national cancer documentation requirements in detail with the goal of developing a detailed cancer data superset covering the essential requirements for cancer care, quality assurance, cancer registration and cancer research; (3) grouping the data elements of this cancer data superset into clinical documentation packages, which can be completed in the respective clinical encounter; and (4) analysis of the existing IT infrastructure, defining appropriate documentation systems for each clinical documentation package, and implementing an interface infrastructure for cooperative data processing.

The cancer data superset presented in this paper was developed based on the German datasets of the ADT, GEKID and BCRL, enriched by additional data elements to meet the requirements of German quality assurance initiatives. In other countries this cancer data superset must be developed or adapted according to national requirements, such as the
National Cancer Data Base (USA) [4] or the National Cancer Data Strategy (Australia) [5]. Existing supersets like ours can potentially be adjusted to the requirements in other institutions, since we believe that especially the clinical documentation requirements are similar.

Correct grouping of the data elements into clinical documentation packages depends heavily on the analysis quality in step 1. We found that in most cases it is not required to implement the complete documentation packages from scratch, because some of the required data elements are already documented in an electronic system and only the remainder needs to be implemented.

To summarize, single-source tumor documentation can only be established when IT cooperates closely with medical staff and the cancer registry. Implementation is a complex and long-lasting project involving at least half of hospital’s clinical disciplines. A complex network of different IT applications is required for tumor documentation. Data reuse is possible if clinical data are documented in high quality [29]. To that end medical staff must be motivated with additional benefits, such as pre-populated physician letters, patient summaries, or follow up reminders directly within the patient care process, but also with possibility of reuse of such data in research projects.

Acknowledgements

The authors would like to thanks Marcus Martin, Dr. Sabrina Petsch, Stefan Schick and Prof. Dr. Wallich for their continuous support during the development of the cancer data superset and the practical implementation phase. Parts of the described projects have been funded by German Cancer Aid.

Disclosure Statement

There is no conflict of interest.

References

7 Does the UK really have an effective cancer plan? [27].