

Transferring HIS Data to Population-Based Cancer Registries – Concept and First Implementations

Bernhard BREIL¹, Martin DUGAS

*Department of Medical Informatics and Biomathematics, University of Münster,
Germany*

Abstract. Cancer is the second leading cause of death worldwide and in focus of epidemiological research. In Germany the cancer registration law stipulates an electronic report to the population-based cancer registry (PBCR). In this context the Comprehensive Cancer Centre Münster (CCCM) required a new concept to support the obligation to register cancer diseases. We analysed Hospital Information System (HIS) data structures related to cancer documentation and PBCR documents. Our main idea was to export available data items from the HIS and to convert them into the import format of the PBCR. We analysed HIS data and developed an XML-based converter to support an electronic reporting procedure. Using available HIS data can avoid redundant data entry and supports information workflow within the CCCM. HIS data can provide a secondary use beyond clinical routine in form of reporting, quality assurance and clinical research.

Keywords. HIS, interoperability, population-based cancer registries, routine data

1. Introduction

Cancer is the second leading cause of death worldwide [1]. Therefore, cancer data is in focus of many studies and population-based cancer registries (PBCR) have been used in epidemiological research [2]. The main task of cancer registries is to collect, analyse and interpret frequencies of cancer diseases to monitor incidence and trends of a defined population [1]. To support this data collection in Germany legal stipulations, derived from the cancer registration law in effect since April 2005 [2], postulate electronic reports of basic cancer data. For this reason the regional PBCR provides an electronic tool to support the reporting procedure.

At the moment relevant data is transferred to the cancer registration centres by this tool with separate, manual data entry which is slow and error-prone. Concerning that physicians spend about 25% of their working time for documentation [3] it is important to re-use these data. Some data for PBCR are already electronically available in the Hospital Information System (HIS). The establishment of the Comprehensive Cancer Centre (CCC) [4] at the university hospital Münster demanded a new concept of transferring tumour data to cancer registration centres with less redundant manual data

¹ Corresponding Author: Department of Medical Informatics and Biomathematics, University of Münster, Domagkstraße 9, 48149 Münster, Germany; E-mail: bernhard.breil@ukmuenster.de.

input. Hence, our main goal is to analyse which data is already available in the HIS and provide a concept, how manual input can be reduced by a new optimised workflow using exported HIS data.

2. Methods

We analysed the current reporting process regarding PBCRs at the university hospital of Münster, Germany, a tertiary referral centre.

We compared the data needed for the epidemiological reports with HIS data to identify items which can be electronically transferred. Thus, we assessed available input formats of the PBCR. We compared catalogues and coding systems to assess semantic interoperability. This is important to assure that all cases can be interpreted by the PBCR. Based on this analysis we designed a concept to improve this reporting procedure.

3. Results

3.1. Analysis

Currently reports for PBCR and routine documentation for patient care are separate tasks. After identifying the cancer patients in the HIS, PBCR data are documented manually with a different software which allows sending an electronic report to the regional PBCR. This tool has an import function for a defined input format.

This format was specified in an XML schema definition (EKRRNW.xsd), which can be imported into the database of the PBCR. Overall the PBCR specified about 50 attributes, 17 of which are mandatory, additional twelve are recommended. Figure 1 shows an extract of this file. Identity, diagnosis, therapy and reporting institution are mandatory items; data about death, pathological findings and code systems are optional.

```
- <xs:element name="MeldungDaten">
  - <xs:complexType>
    - <xs:sequence>
      + <xs:element name="Identitaet"/></xs:element>
      + <xs:element name="Tod" minOccurs="0"/></xs:element>
      + <xs:element name="Diagnose"/></xs:element>
      + <xs:element name="Therapie"/></xs:element>
      + <xs:element name="Codes" minOccurs="0"/></xs:element>
      + <xs:element name="Pathologie" minOccurs="0"/></xs:element>
      + <xs:element name="Meldestelle" type="xs:positiveInteger"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

Figure 1. Extract from the PBCR import schema (EKRRNW.xsd). A report (MeldungDaten) consists of the following items: identity (Identitaet), death (Tod), diagnosis (Diagnose), therapy (Therapie), codes (Codes), pathology (Pathologie) and reporting institution (Meldestelle).

We compared PBCR data items and HIS data; Table 1 provides an overview. Most data items are available because they are required for billing and quality assurance (QA). Therefore, diagnoses are already coded with the International Classification of

Diseases (ICD); diagnoses and surgery dates are documented within the HIS. Results of pathological findings as well as information about cause of death are not available for all patients in our system. When we started the analysis, the information, whether the patient is informed about the reporting and whether he permitted a further contact was only documented on paper. Because this information is required for the electronic report we added these fields in our HIS forms to make all mandatory items available in the HIS.

Table 1. Availability of PBCR items in HIS specified by the population-based cancer registry North Rhine-Westphalia, Germany

PBCR Data Items	Content	Available in HIS
identity	name, first name, birth name, date of birth, street, postal code, city, nationality, sex	nearly completely available
death	cause of death, day of death, place of death	optional: not available
diagnosis	tumour diagnosis date, diagnosis text, certainty, ICD code, TNM code, location, grading	ICD code and date available TNM, location and grading partly available
therapy	aim of therapy, surgery, chemotherapy, radiation	ICPM code and date available (required for billing and QA)
codes	ICD version, topography, morphology	ICD version available, topography and morphology partly available
pathology	pathological finding, date of finding	optional: in some cases available
patient information	patient is informed, contacting permitted	now available
reporting institution	ID of reporting institution	fixed number for university clinic

Up to now the TNM and the ICD-O Classifications are only used in some departments, however, they are, although not required in the specification, important for epidemiological research.

3.2. Concept and Implementation

The analysis showed that HIS data can cover all required PBCR attributes. Consequently, we designed a report to extract required cancer data patient-centric from the HIS as a comma-separated file (the reporting tool did not enable a direct XML export). This export function is limited to two physicians, who are responsible for data transfer to the PBCR. The next step was to convert this file into the format specified by the PBCR. We developed a tool to convert these CSV files into XML files compliant with the schema. Special attention was given to missing attributes and date items.

Figure 2 shows the workflow from HIS export to the final import in the PBCR database in four steps. First, data are exported using the report. Then the resulting CSV file was converted into an appropriate XML format. This XML file was processed by the tool EPIDem [2]. EPIDem allows the transfer of the data to the PBCR. For safety reasons EPIDem is located in a separated network, defined by a virtual LAN (VLAN)

of the Association of Physicians called “KV Safenet”. A prototype was tested successfully with 200 patients and an upcoming routine operation is planned soon.

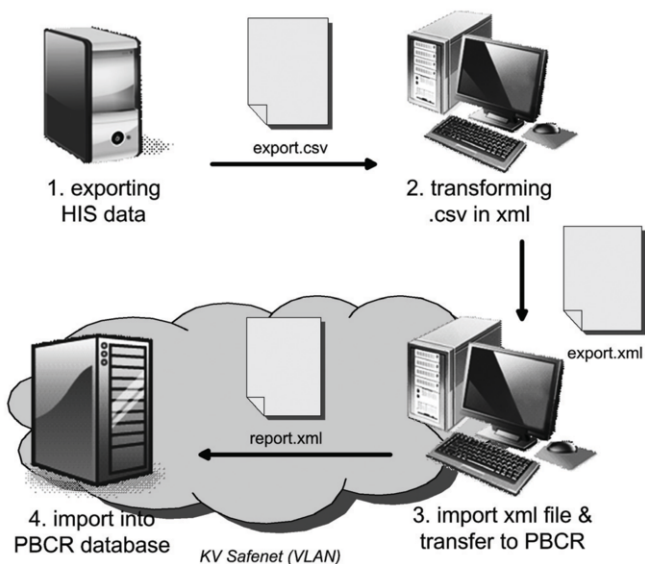


Figure 2. Overview of transferring HIS data to cancer registries

4. Discussion

While comparing medical records with data of cancer registries in the UK Pascoe et al. showed that electronic codes for cancer have a poor level of completeness and correctness in medical records [5]. In Germany ICD codes are being monitored closely by QA and healthcare insurances. Therefore we assume that HIS data quality exceeds the data described by Pascoe et al. This should be clarified in further studies.

For PBCR intended for epidemiological research it is important that cancer registry data is complete [6]. Using electronic HIS reports with ICD codes and including the criteria of the patient’s agreement can support completeness of the report data.

At the moment the specified schema definition is a proprietary solution of the PBCR and only valid for Germany. The XML file has the advantage that it is easy to create and also readable by humans and computers. Nevertheless, concerning system interoperability an internationally standardised format would be advantageous. Regarding a study conducted in the UK there are still variations in relevant data items as well as in their definition [7]. Therefore, the use of the clinical document architecture (CDA) should be analysed to provide a solution beyond Germany.

Other important aspects are privacy and data integrity of patient data. The legal basis of the export results from the cancer registration law. The PBCR creates pseudonyms of permanent storable identity information and makes other items anonymous. Therefore, no data was transferred in plaintext while exporting data to the PBCR. This was approved by the responsible data protection officer.

There is evidence that physicians spend 2–3 hours of a working day on documentation [3]. Thus, it is important to re-use this documentation for quality assurance and other reporting duties. Our analysis shows that there are still attributes

which are not documented within the HIS. But the value of cancer registries and their activities depend heavily on the quality of their data [8, 9]. It is important to complete these reports by updating HIS documentation to monitor recent changes in survival rates as soon as possible [10].

HIS data structure is important for secondary use. Besides an export to a cancer registry it can be used for quality assurance as well as for clinical research [11]. Structured documentation combined with standardised coding systems can help to avoid multiple data input for different purposes and contributes to the “single source” approach [12].

5. Conclusion

Routine HIS data can support reporting to cancer registries. Using electronic exports of HIS data avoids double input of already available data, which is particularly important given the high documentation workload of physicians.

Acknowledgements. Dr. Rainer Fricke and Dr. Volker Krieg (epidemiological cancer registry North Rhine-Westphalia, director: Professor Dr. Hans Werner Hense) provided data structures and support, Professor Dr. Jörg Haier (managing director of CCCM) provided HIS reports.

References

- [1] Parkin, D.M. (2006) The evolution of the population-based cancer registry. *Nature Reviews Cancer* 6:603–612.
- [2] Cancer registry North Rhine-Westphalia, <http://www.krebsregister.nrw.de>.
- [3] Ammenwerth, E., Spötl, H.-P. (2009) The time needed for clinical documentation versus direct patient care – A work-sampling analysis of physicians’ activities. *Methods of Information in Medicine* 48:84–91.
- [4] Comprehensive Cancer Centre Münster, <http://www.cccm.uni-muenster.de/>.
- [5] Pascoe, S.W., Neal, R.D., Heywood, P.L., Allgar, V.L., Miles, J.N., Stefoski-Mikeljevic, J. (2008) Identifying patients with a cancer diagnosis using general practice medical records and cancer registry data. *Family Practice* 25:215–220.
- [6] Parkin, D.M., Bray, F. (2009) Evaluation of data quality in the cancer registry: Principles and methods. Part II, *European Journal of Cancer* 45(5):756–764.
- [7] Edwards, D., Bell, J. (2000) Cancer registries – Future development and uses in Britain. *Journal of Public Health Medicine* 22:216–219.
- [8] Bray, F., Parkin, D.M. (2008) Evaluation of data quality in the cancer registry: Principles and methods. Part I: Comparability, validity and timeliness. *European Journal of Cancer* 45(5):747–755.
- [9] Larsen, I.K., Småstuen, M., Johannesen, T.B., Langmark, F., Parkin, D.M., Bray, F., Møller, B. (2009) Data quality at the Cancer Registry of Norway: An overview of comparability, completeness, validity and timeliness. *European Journal of Cancer* 45(7):1218–1231.
- [10] Gondos, A., Bray, F., Brewster, D.H., Coebergh, J.W., Hakulinen, T., Janssen-Heijnen, M.L., Kurtinaitis, J., Brenner, H. (2008) EUNICE Survival Working Group. Recent trends in cancer survival across Europe between 2000 and 2004: A model-based period analysis from 12 cancer registries. *European Journal of Cancer* 44:1463–1475.
- [11] Breil, B., Semjonow, A., Dugas, M. (2009) HIS-based electronic documentation can significantly reduce the time from biopsy to final report for prostate cancer and supports quality management as well as clinical research. *BMC Medical Informatics and Decision Making* 9:5.
- [12] Kush, R., Alschuler, L., Ruggeri, R., Cassells, S., Gupta, N., Bain, L., Claise, K., Shah, M., Nahm, M. (2007) Implementing single source: The STARBRITE proof-of-concept study. *Journal of the American Medical Informatics Association* 14:662–673.