

Using Cancer Waiting Times Data as an Initial Notification for Cancer Registration

Christine Head, Rachael Cayley, Paul Edwards, Caroline Brook
Northern and Yorkshire Cancer Registry and Information Service (NYCRIS)
St James's Institute of Oncology, Level 6, Bexley Wing, St James's University Hospital, Beckett Street, Leeds, LS9 7TF

Background and Aims

In 2004 NYCRIS shared the results of an evaluation of Cancer Waiting Times (CWT) data in a presentation entitled "How Useful is the Cancer Waiting Times dataset for the Purposes of Cancer Registration?" which concluded that CWT data were a useful early source of notification for cancer registration. That conclusion was treated with a degree of caution by the Registry community. The work reported here aims to demonstrate that:

- The Cancer Waiting Times dataset is a useful notification for cancer registration.
- It is possible to automate the processing of the CWT dataset.
- Combining CWT data with other sources can reduce data abstraction from case notes.

Step 1: Are the data useable?

To enable CWT records for 2005 to 2007 to be matched against the NYCRIS database personal demographic data were obtained via NSTS and a diagnosis date was determined from the available dates in priority order as follows:

- 1 Clinical Intervention Date (First Diagnostic Test)
- 2 First Seen By Specialist Date (Cancer)
- 3 Date First Seen
- 4 Cancer Specialist Referral Date
- 5 Cancer Referral Decision Date
- 6 Referral Request Received Date
- 7 MDT Discussion Date
- 8 Decision to Treat Date
- 9 First Definitive Treatment Date
- 10 Death Date returned by NSTS

Using exact 3-character site matching 23% of the 242,132 records matched to a NYCRIS record with a diagnosis date within one month of the derived diagnosis date. When matching was based on CWT site groups and a difference in diagnosis date of up to one year the match rate improved to 31%. For 2005 and 2006 data (i.e. completed registration years) the matching exercise was repeated for CWT records with a Confirmed Cancer Status (n=58,387) and the match rates were 63% and 83% respectively.

Unmatched CWT records for 2005 to 2007 which had a NYCRIS postcode and cancer diagnosis (excluding non-melanoma skin, ill-defined and secondary sites) were registered and followed up via hospital case notes (n=4,525). 81% of registrations made from the CWT data extracts have been followed-up (n=3,685): 76% of these cases have proved to be new registrations (n=2,807), with only 13% found to be non-malignant (see Figure 1).

Methods

- To assess the quality of the CWT dataset using extracts from the National CWT database for 2005 to 2007. These data were compared with NYCRIS registered cases for the equivalent diagnosis years. Where a match for a CWT patient could not be found the record was registered and followed up via hospital case notes.
- To use a 6-month extract of 2008 CWT data to develop an automated notification process.
- To load the CWT data as the initial data source for all 2008 registrations and then complete registrations for 2008 diagnosed breast cancers without access to case notes by combining CWT data with additional notifications or links to Trust systems.

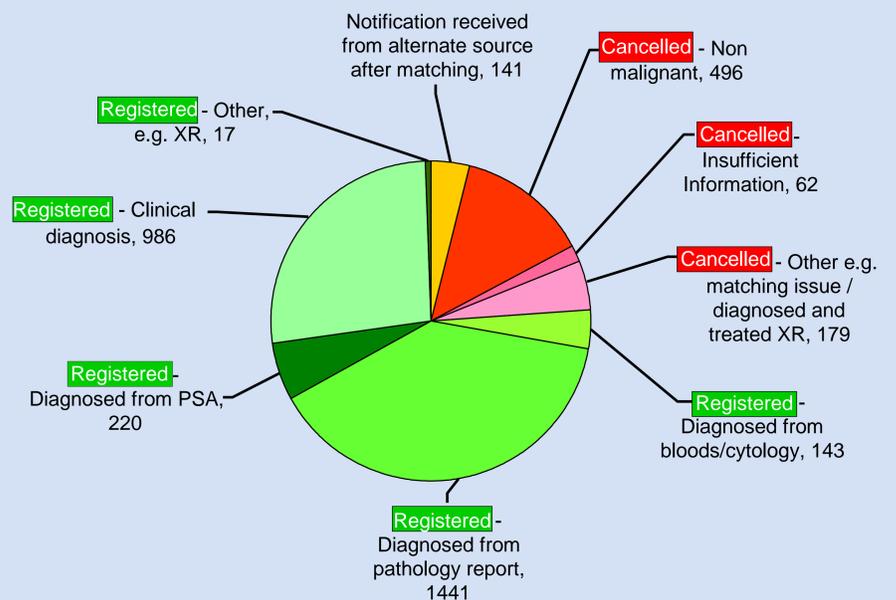
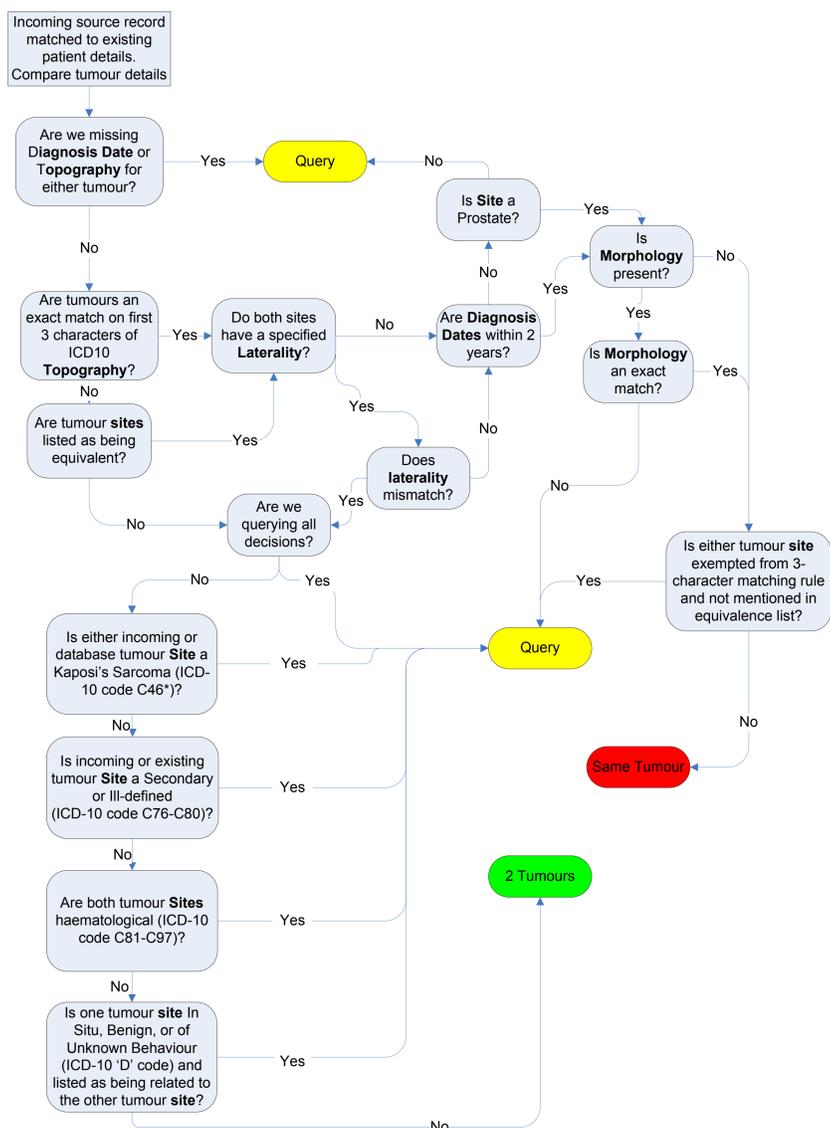


Figure 1: Results of hospital case note follow-up for 2005 - 2007 CWT initiated registrations

Step 2: Can we automate tumour matching?

Specific tumour matching rules had to be developed to handle records containing 3-character ICD10 codes, and no morphology data, as provided by the CWT dataset. This was achieved by developing groups of equivalent tumour sites which could be considered a match. The incoming site code was transformed to a full ICD10 code by adding the NOS 4th digit 9. For sites where the NOS option was not available a dummy 'X' code was used e.g. C21 Anus/Anal Canal.



Step 3: Can we make a notification?

For NYCRIS to use a record as a notification it must contain patient demographics, tumour site and diagnosis date, and preferably a hospital. These data items are all available in the CWT dataset. Referral, Care Plan and limited treatment details are also available.

An Access database was developed to pre-process the CWT dataset into a format that could be loaded into the cancer registration database (PRAXIS). Only records with a Confirmed Cancer Status were included in the load file. A diagnosis date was derived for each record (as defined in Step 1). Basis of diagnosis was set to Clinical Investigation if a Clinical Intervention Date and Organisation Code (Provider First Diagnostic test) were present, otherwise it was set to Clinical. Missing organisation codes were set to a string that would subsequently fail validation so that a user could enter a "best guess" hospital.

To enable the data to be automatically processed in PRAXIS, major restructuring of the dataset was required to create a separate import record for each organisation mentioned in the original CWT record. A subset of data items were common to every import record: patient demographics, tumour site details, cancer status, source and type of referral, MDT details and planned treatment type. Each CWT record could generate up to 6 import records as follows:

- Organisation Code (Provider First Seen)
- Organisation Code (Provider First Cancer Specialist)
- Organisation Code (Provider First Diagnostic Test)
- Organisation Code (Provider Decision To Treat)
- Organisation Code (Provider First Treatment)
- Referring Organisation Code

As source and type of referral were present on each import record, merge rules were defined such that a maximum of three referral events were retained in PRAXIS: first seen, first seen cancer specialist and referring organisation.

The organisation codes used in the CWT dataset should have been national hospital codes. However, some Trust codes were present, as were codes no longer in use. A mapping table was built to prevent these codes failing subsequent validations. Also, all the postcodes within the CWT dataset were compared with the PRAXIS postcode table ahead of processing so that any new postcodes could be added.

Although the intention was to load the 2008 CWT data prior to processing other data sources this was not possible; the automated process is not due to be implemented until late summer 2009. The following results have been obtained loading test data files onto a snapshot of the PRAXIS database taken in January 2009.

Month	Import CWT Data	Records Loaded	Failed Validation	Passed Validation	New Patient / New Tumour	Existing Patient / New Tumour	Existing Patient / Existing Tumour	Manual Match Required	% Automatically Processed
Feb	10926	26	10900	432	109	8946	1413	87%	
Mar	10064	42	10022	572	121	8035	1294	87%	
Apr	12269	34	12235	1216	215	9605	1199	90%	

Conclusion

The Cancer Waiting Times dataset is a useful notification for cancer registration and it is possible to process it automatically. We have yet to determine whether combining CWT data with other sources can reduce data abstraction from case notes.