National Cancer Data Repository: The next step; Merging the Registries Data

James Thomas

Database Manager, Research Group

Northern and Yorkshire Cancer Registry and
Information Service



Collaborators



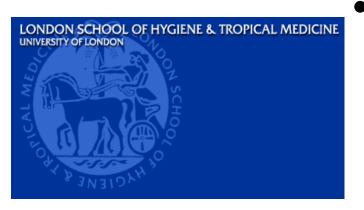


- Thames Cancer Registry
- David Forman, Eva Morris -NYCRIS



Michel Coleman, Libby Ellis, Laura Woods, Bernard Rachet – London School Hygiene and Tropical Medicine





Background

- Current National Cancer Data Repository
 - 1st stage to link ONS minimum cancer dataset with HES extract (containing all episodes for patients with cancer ICD code identified in HES DIAG fields)
 - 8.5 million tumours in ONS dataset, 340,000 per year
 - ONS tumours from 1971 2004
 - 34 million in patient HES episodes 01/04/1997 30/06/2007
 - ~80% Linkage rate for ONS > HES

Background

- De-duplication of national dataset. Identifying both duplicate registrations and multiple tumours for patients.
- Additional data items added to this dataset, including IMD, Gridlink, trust and cancer network information
- Production of algorithms to determine final ethnicity from HES data.
- Charlson co-morbidity score developed.
- Length of stay and No of hospital episodes calculated on a tumour basis
- Thames Cancer Registry have been responsible for the distribution of these data and it is being utilised numerous groups

Linking the Registries Data

- Q. Why? The ONS minimum cancer dataset already exists and all registries feed in to this
- A. Cancer registries put a lot of effort into collecting information that is outwith the minimum dataset
 - Detailed information on treatment of tumours
 - Place of Death
 - Higher levels of staging grading information
 - Patient names which will aid in the identification of ethnicities
 - Causes of death registry collected
 - Information on metastatic cancers
 - Nodes involved/Positive
 - Size of Tumour

Linking the Registries Data

Data requests were made to each of the 8 English Cancer Registries, including

- Malignant Neoplasms, InSitu tumours and benign CNS tumours
- 1990 2006 all tumours including extra regionals
- Patient details; names, NHSNO, Postcode, ethnicity, DCO, Death information
- Tumour details; site, type, size, grade, basis of diagnosis, nodes, mets and all collected staging data
- Treatment details, surg/CT/RT markers, OPCS coding, hospital and dates of treatment

Merging process

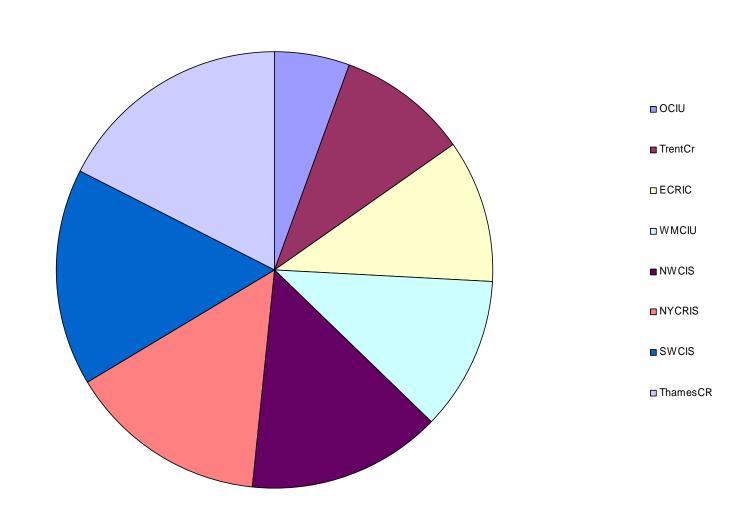
- 8 Cancer Registries
 - Differing levels of data collection
 - Similar but all slightly different data structures
- Needed to be combined into 1 single tumour table and 1 single treatment table
 - Data manipulation of table structures 25 tables
- Recoded individual fields to give a uniform dataset for most fields
 - ~300 separate fields checked and recoded where necessary
- For Colorectal tumours combined differing stage coding TNMPath/TNMClin/Dukes to give a final uniform stage
- Similar approach will be applied to other major cancer sites in collaboration with lead registries

Results

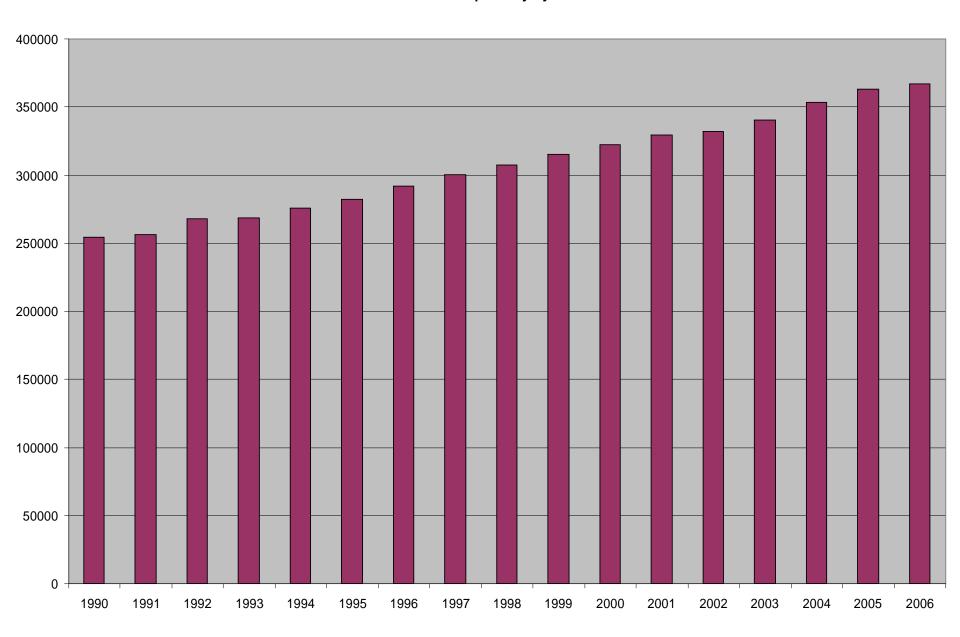
- Merged dataset now exists containing 5,281,243 tumours
 - 1990 2006 from all 8 ECRs
- 9,235,731 rows of treatment information
 - covering 4,352,745 tumours
 - 2,470,741tumours have an OPCS coded treatment

Results

Breakdown of Tumours Supplied by the Separate Cancer Registries



Tumours in Repository by Year



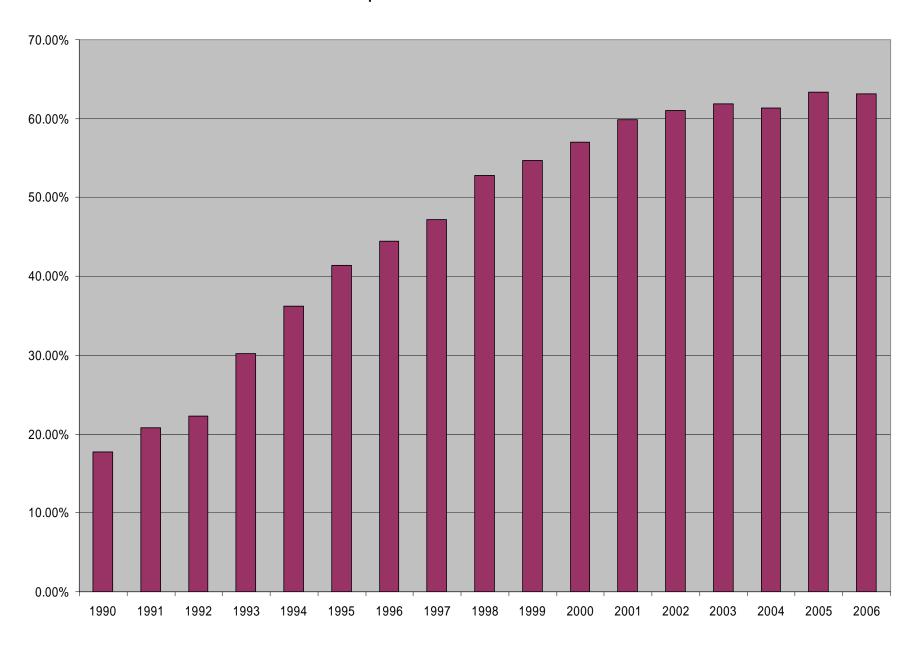
Data Completeness

- NHS number is supplied for 98.2% of the registry tumours
- Staging Information is available for 60.3% of the major tumour sites, ~50% in ONS file
 - NOTE, even for these sites stage is coded in many different ways

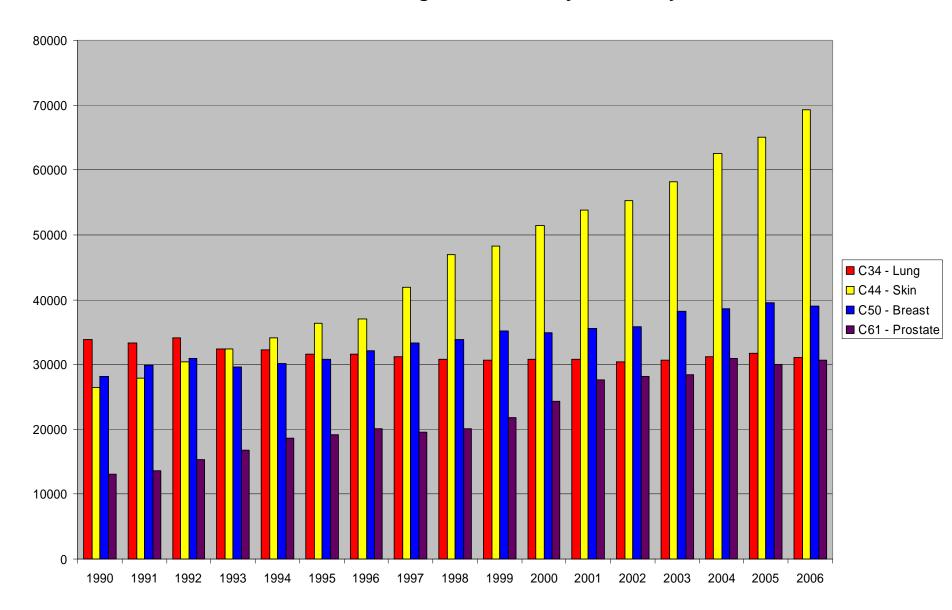
Data Completeness

Field	Overall %	C18-20 %	C50 %
	Complete	Complete	Complete
Gender	100%		
Postcode	99.9%		
Tumour site	100%		
Stage	25.0%	52.3%	54.2%
Nodes sampled	8.3%	22.3%	31.2%
Nodes positive	18.5%	29.7%	33.5%
Tumour size	11.2%	18.4%	49.7%

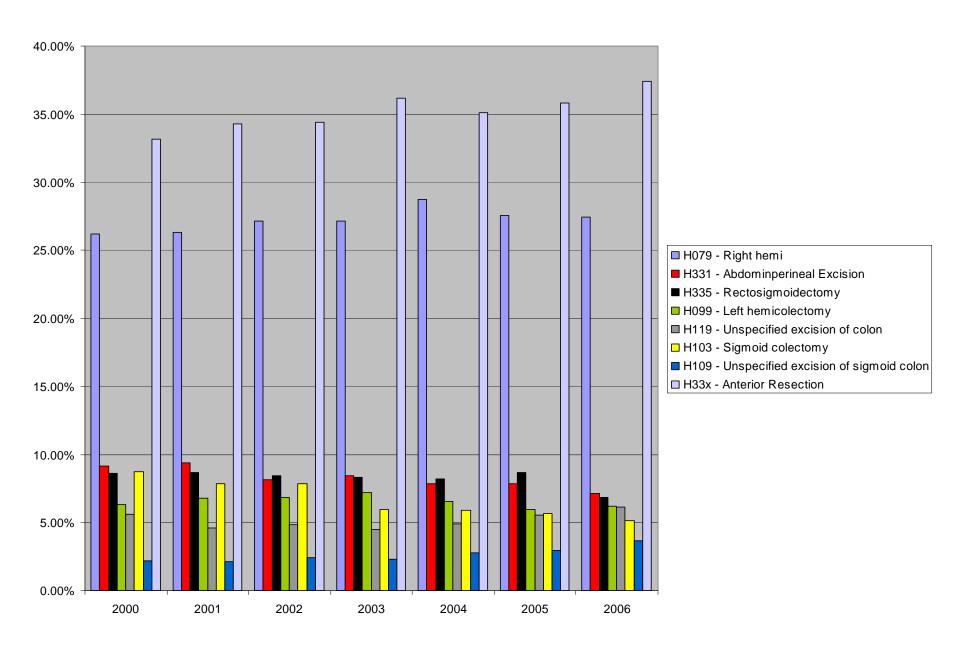
Data Completeness - Tumour size C50 Breast



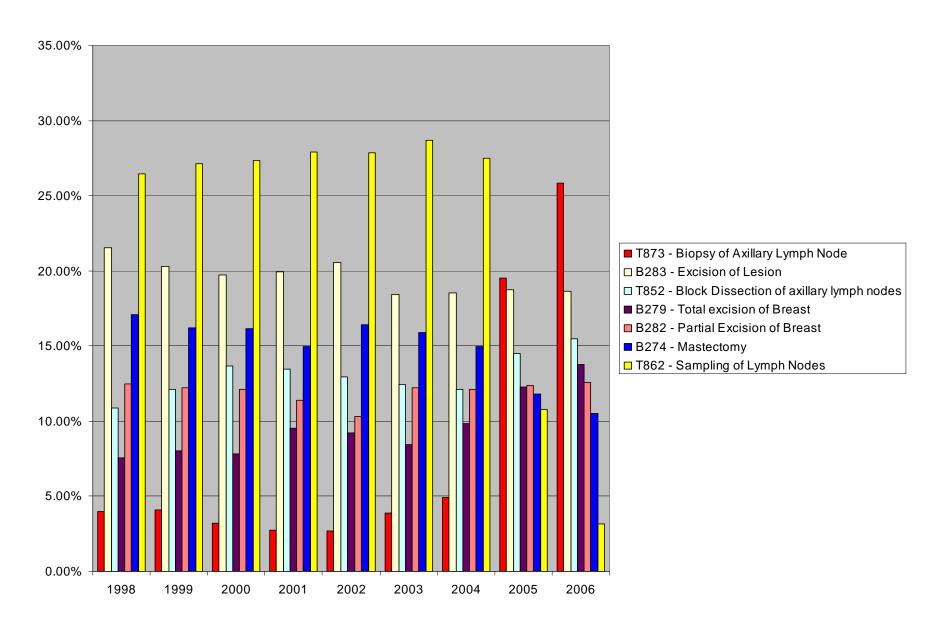
Total Tumours Registered for Major Sites By Year



Breakdown of Common Colorectal Treatments by Year



Breakdown of Common Breast Cancer Treatments by Year



Next Steps

- ONS2004/HES is available on request and has been distributed
- Merged registry dataset is being linked to HES
- All merged registry data should be available during the summer months; exact date tbc
- New request will be issued to registries for 2007 data, post ONS submission
- Much speedier turn around of data, ~ 2-3 months

Next Steps

- Linkage factors for ONS/Registry data have been supplied for linkage to GPRD
- Colorectal Screening data has been linked for colorectal patients:- retrospective and prospective enriching registry/repository data
- Working to collaborate with NATCANSAT to push forward linking in Out Patient HES and a Rapid HES feed

Conclusion

- NCDR will aim to provide a central data resource
 - Increasing Efficiency
 - Decreasing duplication of effort
- This resource is vast and needs to be utilised
- Average lifetime of 75 years is about 27,000 days
- If we can increase survival of breast cancer for example it will the equivalent of a whole lifetime
 - That's per year