NCIN national cancer intelligence network

KING'S College LONDON

Data Quality and Completeness Report:

Lung Cancer and Mesothelioma Site Specific Clinical Reference Group

Sharma P Riaz
Victoria H Coupland
Karen M Linklater
Henrik Møller
Margreet Lüchtenborg

THAMES CANCER REGISTRY

Contents

1. Introduction

The National Cancer Intelligence Network lung cancer and mesothelioma Site Specific Clinical Reference Group covers neoplasms of the trachea, bronchus and lung as well as mesothelioma. Thames Cancer Registry investigates these cancers using data from the National Cancer Data Repository (NCDR). The NCDR contains information from the eight English cancer registries on all patients diagnosed with cancer in their respective catchment areas.

It is important to analyse the quality of the data as large proportions of missing or poor quality information will lead to potentially inaccurate conclusions being drawn. It also means that some more detailed analysis on specific subgroups would be difficult. It is vital to record the quality of these data to ensure improvements can be made. An annual report will help drive and measure any improvements.

This report explores the data quality and completeness of the lung cancer and mesothelioma datasets as derived from the NCDR. It reports on data on patients diagnosed over the ten-year period 1999-2008, while focussing on the most recent diagnosis year (2008).

2. Methods

Data were extracted from the NCDR on all cases of lung cancer (ICD-10 C33-C34) and mesothelioma (ICD10-C45) diagnosed in 1999-2008. There were 32,967 malignant neoplasms of the trachea, bronchus and lung and 2,075 mesothelioma registrations in 2008.

Data quality

The quality of the dataset was investigated for lung cancer and mesothelioma at cancer registry level (Table 1). The graphs and accompanying text will refer to each registry by their code.

Table 1: List of the eight English cancer registries.

Cancer registry code	**Cancer registry name**
ECRIC	Eastern Cancer Registration and Information Centre
NWCIS	North West Cancer Intelligence Service
NYCRIS	Northern & Yorkshire Cancer Registry and Information Service
Oxford	Oxford Cancer Intelligence Unit
SWCIS	South West Cancer Intelligence Service
Thames	Thames Cancer Registry
Trent	Trent Cancer Registry
WMCIU	West Midlands Cancer Intelligence Unit

The data quality measures investigated are listed below:

Death certificate only registrations

Many registrations for rapidly fatal cancers are initiated by a patient’s death certificate. These registrations are followed up in hospital systems or in the Hospital Episode Statistics (HES) dataset. Many cases are found and their details are updated to form a complete registration. However, some cases may not have been seen in a hospital and therefore further details cannot be found. These will remain death certificate only (DCO) registrations. These registrations have limited information and their date of diagnosis is the same as their date of death. They therefore have to be excluded from some analyses.

Basis of diagnosis

The basis of diagnosis is recorded for each cancer registration. Three groups were defined as follows: microscopically verified (cytology, histology of primary tumour and histology of metastases), clinically verified (clinical opinion, clinical investigation and death certificate) and not known (specific tumour markers, not known and missing).

Anatomical site

The unknown anatomical site group included patients with an ICD10 four digit code of Cxx.8 (overlapping lesion of [specific] cancer) and Cxx.9 ([specific] cancer, unspecified). See Appendix 1 for a full list of codes. Large proportions of patients with an unspecified anatomical site will limit our ability to analyse these cancers by specific subgroups.

Morphology

Morphology was classified as known (valid morphology codes) and not known (morphology codes: 8000, 8001 and missing). Large proportions of patients with an unknown morphology code will limit our ability to analyse these cancers by specific morphology subgroups.

Linked HES records

If a registration has no linked HES records, this could indicate that the matching has not been successful for that patient and as a result their treatment information may not have been included in our dataset. Also, the subset of HES data received by the cancer registries only includes patients with a diagnosis of cancer. Patients may have had surgery for their cancer, but have no cancer diagnosis coded in HES. Therefore, their surgery would not be linked to their cancer registration record. However, it could also mean that the patient has had no inpatient hospital activity. This will be important to consider in any future treatment analyses.

Ethnicity

Ethnicity has historically been poorly recorded in cancer registry datasets. Since 1995 it has been mandatory to collect ethnicity information within hospitals and therefore the NCDR includes ethnicity from the HES dataset. Large proportions of patients with a missing ethnicity code will make studies focussing on ethnicity less robust.

Stage variables

Stage is an important indicator of the prognosis and will influence the treatment that patients receive. The NCDR records TNM stage information. T describes the size of the tumour, N whether regional lymph nodes are involved and M describes distant metastasis. There are three types of TNM stage in the NCDR: pathological TNM (t_path, n_path, m_path, tnm_path), clinical TNM (t_clin, n_clin, m_clin, tnm_clin) and integrated TNM (t_int, n_int, m_int, tnm_int). The NCDR also includes the field "mets", which records if a patient has distant metastases or not and the field "nodes_postive", which records the number of nodes that were found to be positive. Each of these variables were analysed separately, with the proportion of registrations with a valid known or missing code calculated. For the individual T, N, M and "mets" fields a value of X was recorded as valid not known. In the "nodes-positive" field a value of 99 or 999 was defined as valid not known.

3. Results

3.1.1 *Quality of the lung cancer dataset, England, 2008*

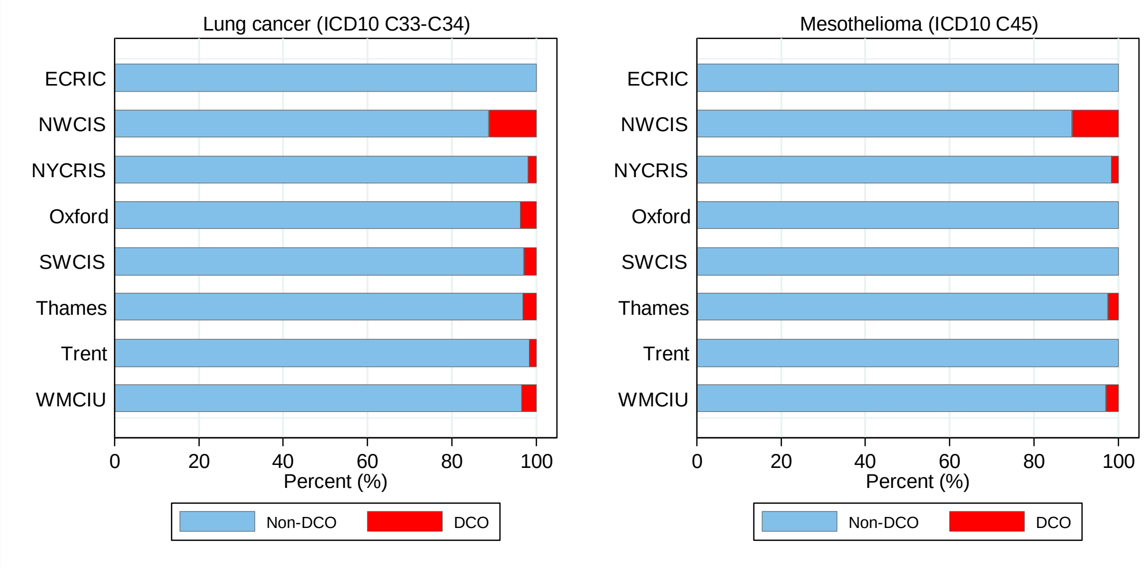
	England		ECRIC		NWCIS		NYCRIS		Oxford		SWCIS		Thames		Trent		WMCIU	
Total number of registrations	(n=32967)		(n=3,423)		(n=5,221)		(n=5,697)		(n=1,254)		(n=4,316)		(n=6,007)		(n=3,720)		(n=3,329)	
	Number	(%)	Number	(%)	Number	(%)	Number	(%)	Number	(%)	Number	(%)	Number	(%)	Number	(%)	Number	(%)
Death certificate only																		
Death certificate only	1247	3.78	0	0.00	588	11.26	115	2.02	47	3.75	124	2.87	191	3.18	65	1.75	117	3.51
Non-DCO registrations	31720	96.22	3423	100.00	4633	88.74	5582	97.98	1207	96.25	4192	97.13	5816	96.82	3655	98.25	3212	96.49
Anatomical site (excluding DCO registrations)																		
Known anatomical site	20732	65.36	3025	88.37	2758	59.53	3984	71.37	532	44.08	2592	61.83	3382	58.15	2461	67.33	1998	62.20
Basis of diagnosis (excluding DCO registrations)																		
Microscopically verified	22199	69.98	2,370	69.24	2,949	63.65	3,836	68.72	864	71.58	2,930	69.90	4,279	73.57	2,579	70.56	2,392	74.47
Clinically verified	9279	29.25	1040	30.38	1670	36.05	1690	30.28	312	25.85	1185	28.57	1,487	25.57	1076	29.44	819	25.50
Ethnicity (excluding DCO registrations)																		
Known	27941	88.09	2925	85.45	4248	91.69	4899	87.76	1060	87.82	3616	86.26	5018	86.28	3333	91.19	2842	88.48
No linked record in Hospital Episode Statistics (excluding DCO registrations)																		
Linked	29247	92.2	3096	90.45	4413	95.25	5126	91.83	1015	91.55	3842	91.65	5271	90.63	3431	93.87	2963	92.25
Morphology (excluding DCO registrations)																		
Known	28503	89.85	3287	95.88	4248	91.69	3930	70.4	1133	93.79	3865	92.18	5635	96.89	3431	93.87	2979	92.75
Valid known stage (excluding DCO registrations)																		
Clinical																		
T	3179	10.02	0	0	348	7.52	0	0	1	0.08	446	10.66	1580	27.18	0	0	804	25.05
N	3480	10.97	0	0	335	7.24	0	0	1	0.08	459	10.97	1535	26.39	0	0	1150	35.81
M	3321	10.47	0	0	349	7.54	0	0	1	0.08	648	15.48	1153	19.85	0	0	1169	36.4
TNM	1872	5.9	0	0	14	0.3	2	0.04	1	0.08	485	11.59	0	0	0	0	1,370	42.69
Pathological																		
T	2137	6.74	0	0	979	21.14	0	0	0	0	300	7.17	417	7.18	0	0	441	13.74
N	2131	6.72	0	0	942	20.35	0	0	45	3.73	328	7.84	364	6.27	0	0	452	14.09
M	1704	5.37	0	0	697	15.05	0	0	89	7.37	508	12.11	53	0.91	0	0	357	11.12
TNM	1320	4.16	0	0	54	1.17	0	0	89	7.37	487	12	0	0	0	0	690	21.5
Intergrated																		
T	3621	11.42	2,450	71.64	0	0	0	0	0	0	0	0	0	0	0	0	1171	36.46
N	3455	10.89	2,001	58.51	0	0	0	0	0	0	0	0	0	0	0	0	1454	45.28
M	3143	9.19	1,778	51.99	0	0	0	0	0	0	0	0	0	0	0	0	1365	42.51
TNM	4571	14.41	2,761	80.73	0	0	0	0	0	0	0	0	0	0	0	0	1810	56.4

3.1.2 Quality of the mesothelioma dataset, England, 2008

Total number of registrations	England		ECRIC		NWCIS		NYCRIS		Oxford		SWCIS		Thames		Trent		WMCIU	
	(n=2075)		(n=264)		(n=266)		(n=310)		(n=87)		(n=361)		(n=443)		(n=170)		(n=174)	
	Number	(%)	Number	(%)	Number	(%)	Number	(%)	Number	(%)	Number	(%)	Number	(%)	Number	(%)	Number	(%)
Death certificate only																		
Death certificate only	50	0.00	0	0.00	29	10.90	5	1.16	0	0.00	0	0.00	11	2.48	0	0.00	5	2.87
Non-DCO registrations	2025	0.00	264	100.00	237	89.10	305	98.39	87	100.00	361	100.00	432	97.52	170	100.00	169	97.13
Anatomical site (excluding DCO registrations)																		
Known anatomical site	1694	89.78	255	96.59	237	100.00	274	89.84	64	73.56	274	75.90	432	100.00	143	84.12	139	82.25
Basis of diagnosis (excluding DCO registrations)																		
Microscopically verified	1754	86.62	238	90.15	179	75.53	274	89.84	77	88.51	286	79.22	382	88.43	155	91.18	163	96.45
Clinically verified	200	9.88	26	9.85	58	24.47	30	9.84	5	5.75	13	3.60	48	11.11	15	8.82	5	2.96
Ethnicity (excluding DCO registrations)																		
Known	1794	88.59	235	89.02	221	93.25	269	88.20	77	88.51	303	83.93	373	86.34	159	93.53	157	92.90
No linked record in Hospital Episode Statistics (excluding DCO registrations)																		
Linked	1848	91.26	241	91.29	227	95.78	279	91.48	77	88.51	318	88.09	386	89.35	160	94.12	160	94.67
Morphology (excluding DCO registrations)																		
Known	2019	99.75	264	100.00	237	100.00	305	100.00	87	100.00	361	100.00	432	100.00	167	98.24	167	99.41
Valid known stage (excluding DCO registrations)																		
Clinical																		
T	19	0.94	0	0.00	5	2.11	0	0.00	0	0.00	0	0.00	14	3.24	0	0.00	0	0.00
N	13	0.64	0	0.00	5	2.11	0	0.00	0	0.00	1	0.28	7	1.62	0	0.00	0	0.00
M	12	0.59	0	0.00	5	2.11	0	0.00	0	0.00	2	0.55	5	1.16	0	0.00	0	0.00
TNM	2	0.10	0	0.00	0	0.00	0	0.00	0	0.00	2	0.55	0	0.00	0	0.00	0	0.00
Pathological																		
T	39	1.93	0	0.00	35	14.77	0	0.00	0	0.00	0	0.00	3	0.69	0	0.00	1	0.59
N	37	1.83	0	0.00	32	13.50	0	0.00	1	1.15	0	0.00	3	0.69	0	0.00	1	0.59
M	40	1.98	0	0.00	33	13.92	0	0.00	1	1.15	5	1.39	1	0.23	0	0.00	0	0.00
TNM	7	0.35	0	0.00	1	0.42	0	0.00	1	1.15	5	1.39	0	0.00	0	0.00	0	0.00
Intergrated																		
T	17	0.84	16	6.06	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.59
N	18	0.89	17	6.44	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.59
M	16	0.79	16	6.06	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
TNM	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00

3.2 Death certificate only

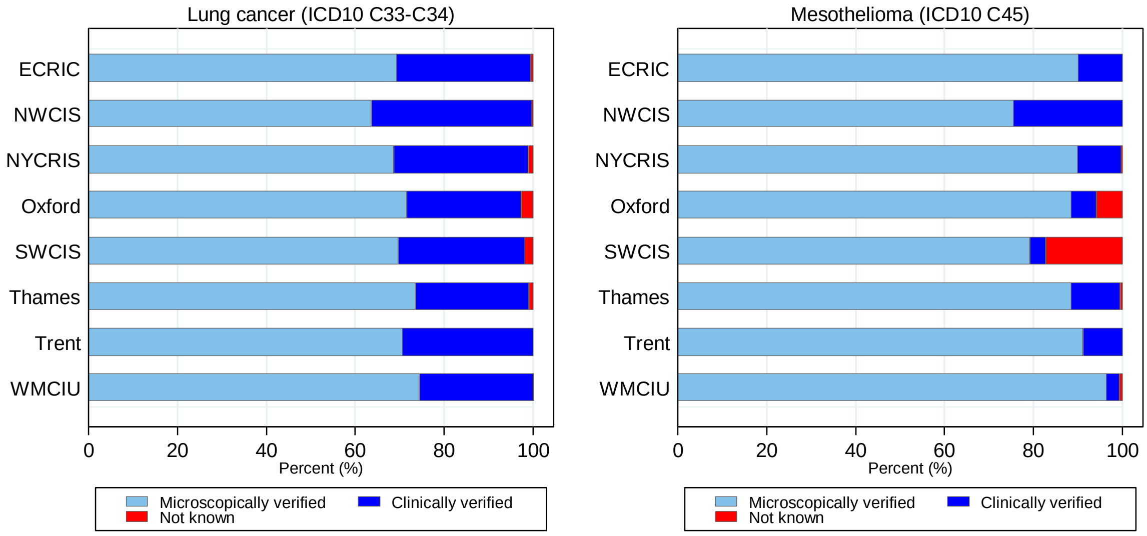
The following graphs show the proportion of death certificate only registrations for lung cancer and mesothelioma.



The proportions of death certificate only registrations were very low and did not vary much between cancer registries. In lung cancer, the overall proportion of death certificate only registrations was slightly higher (4%) than in mesothelioma (2%).

3.3 Basis of diagnosis

The following graphs show the proportion of registrations where the basis of diagnosis was microscopically verified, clinically verified or not known (specific tumour markers, not known or missing) in 2008. This analysis excludes death certificate only registrations.

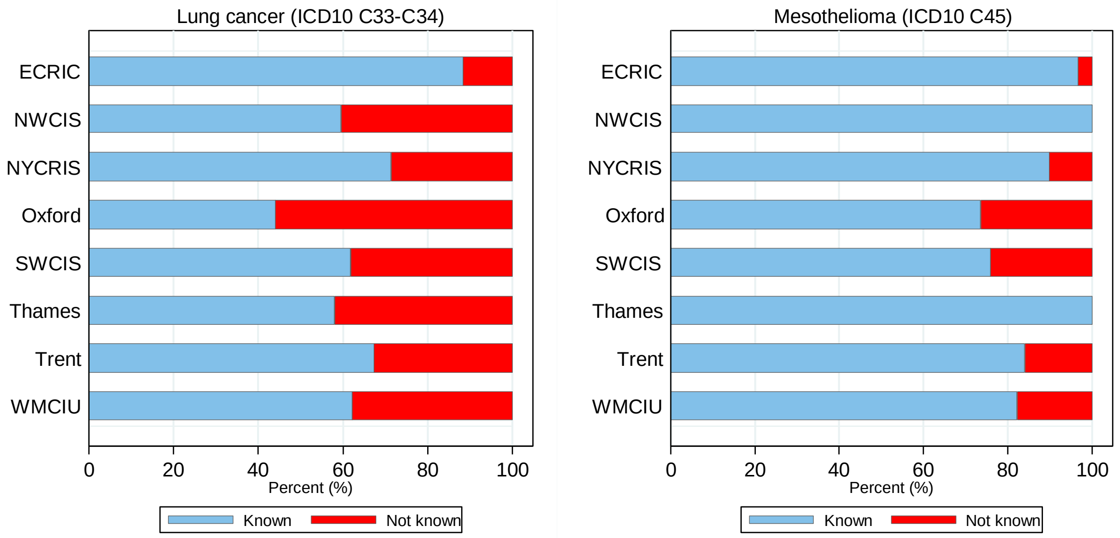


More than 70% of lung cancers and over 85% of mesotheliomas were microscopically verified. Where lung cancers were not microscopically verified, the overall majority was clinically verified and only 1% of cases were neither microscopically nor clinically verified. The microscopic verification rate was higher in mesothelioma, but only 10% of cases were clinically verified.

The higher verification rate of mesothelioma compared to lung cancer is probably related to the need for microscopic verification to arrive at its diagnosis.

3.4 *Anatomical site*

The following graphs show the proportion of registrations with known and not known anatomical site. This analysis excludes death certificate only registrations.

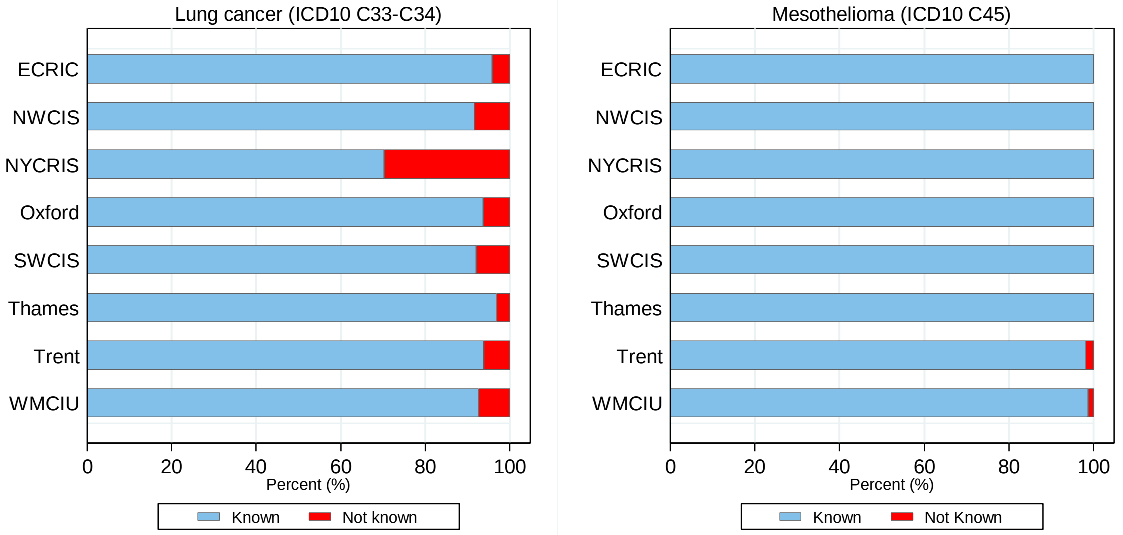


The specification of anatomical site is lower in lung cancer (65%) than in mesothelioma (90%). There is some variation by cancer registry, ranging from 44% to 88% anatomical site specification in lung cancer and from 74% to fully complete in mesothelioma.

The anatomical site of mesothelioma is more likely to be specified because of its symptomatology and importance to treatment options.

3.5 *Morphology*

The following graphs show the proportion of registrations with known or not known morphology information. This analysis excludes death certificate only registrations.

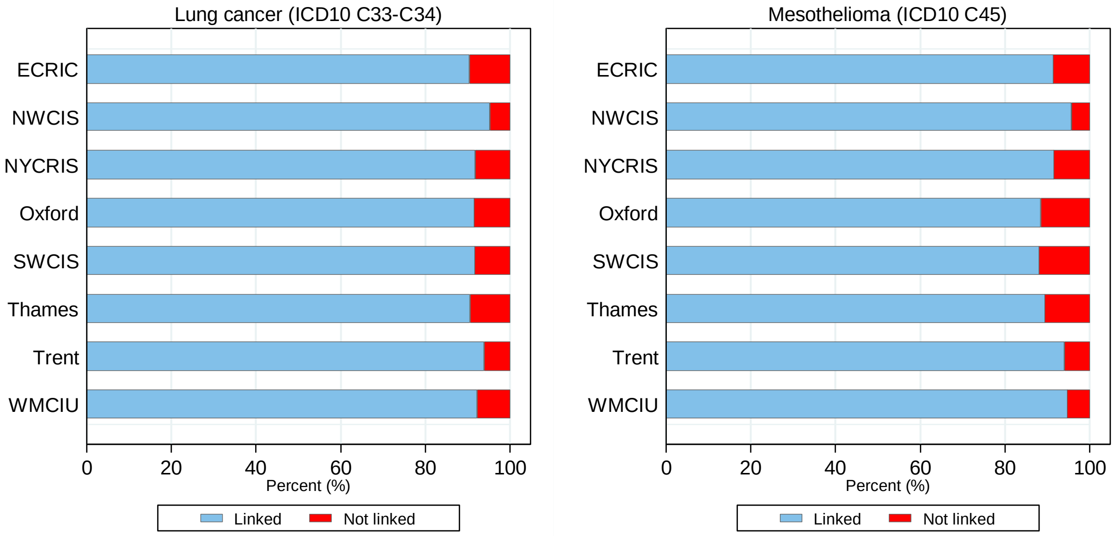


The proportion of registrations with known morphology of lung cancers was generally around 90%, with a range from 70% to 97%.

Morphology information was available for nearly all mesothelioma registrations.

3.6 *Linked HES records*

The following graphs show the proportion of registrations with and without a linked HES record. This analysis excludes death certificate only registrations.

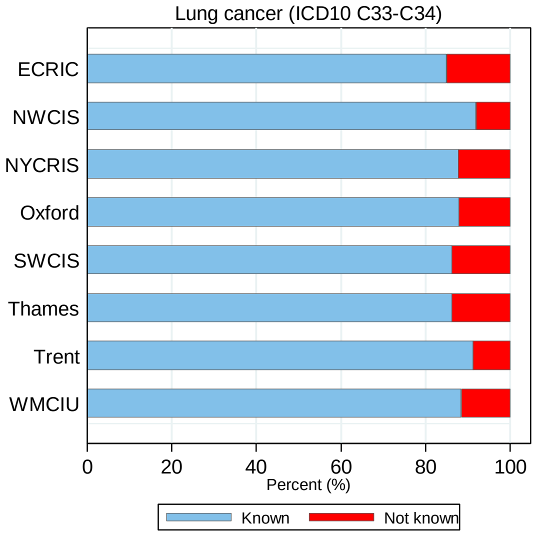


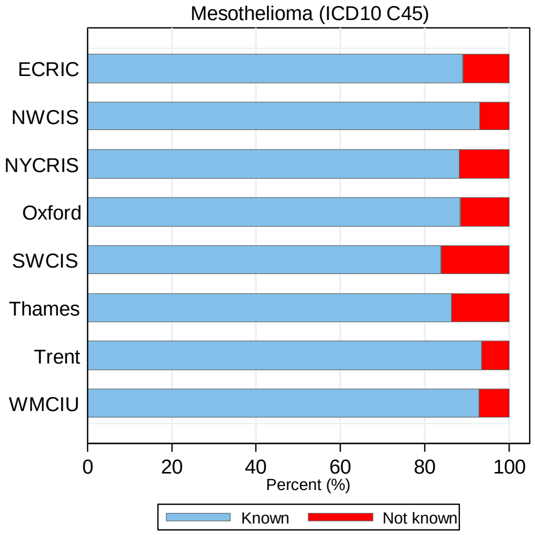
More than 90% of lung cancers and more than 88% of mesotheliomas had a linked HES record.

There was more variation between cancer registrations with a linked HES record for mesotheliomas compared with lung cancers. This is probably due to the lower number of mesothelioma than lung cancer registrations, which leads to an exaggeration of small differences.

3.7 Ethnicity

The following graphs show the proportion of registrations with known and not known ethnicity. This analysis excludes death certificate only registrations.





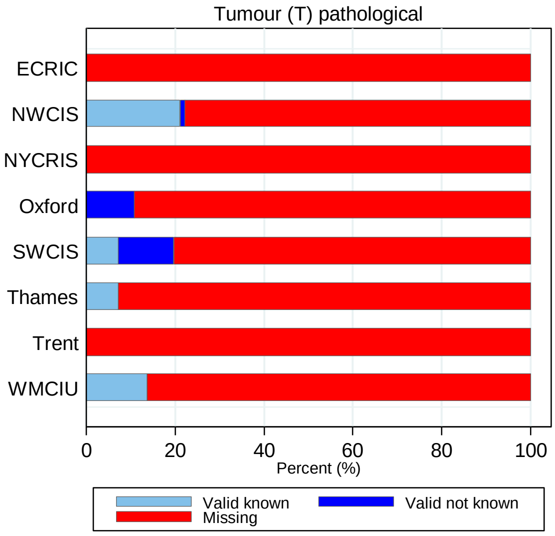
The proportion of registrations with known ethnicity is very similar at 88% of lung cancers and 89% of mesotheliomas.

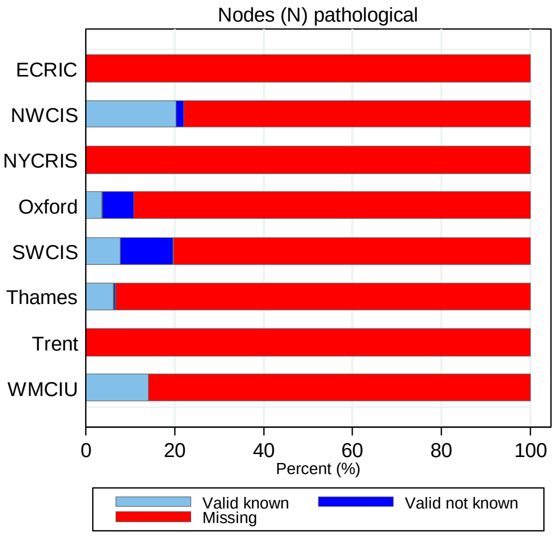
The variation in proportions of registrations with known ethnicity between the cancer registries is mainly due to the completeness of record linkage to HES. Therefore, the variation in known ethnicity between the registries is similar to the variation in proportions of registrations with a linked HES record.

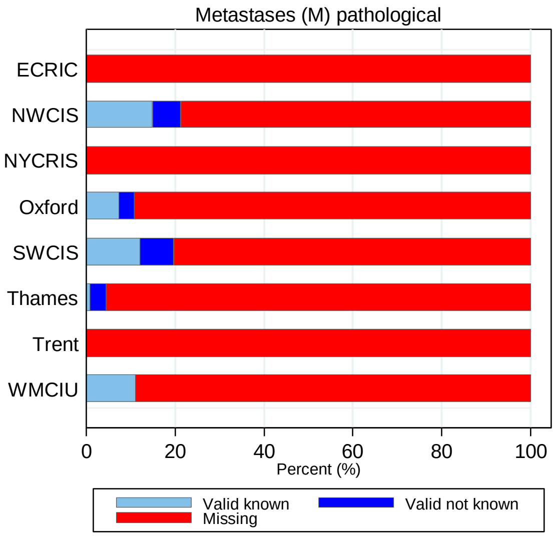
3.8 Pathological stage

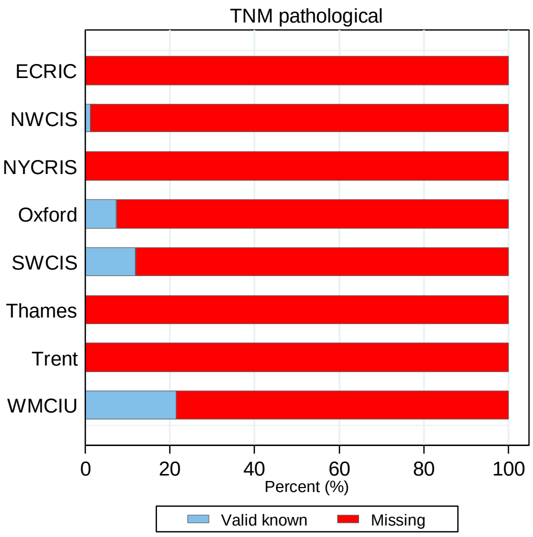
The following graphs show the proportion of registrations with pathological T, N, M and TNM stage information in 2008. This analysis excludes death certificate only registrations.

Lung cancer (ICD10 C34)

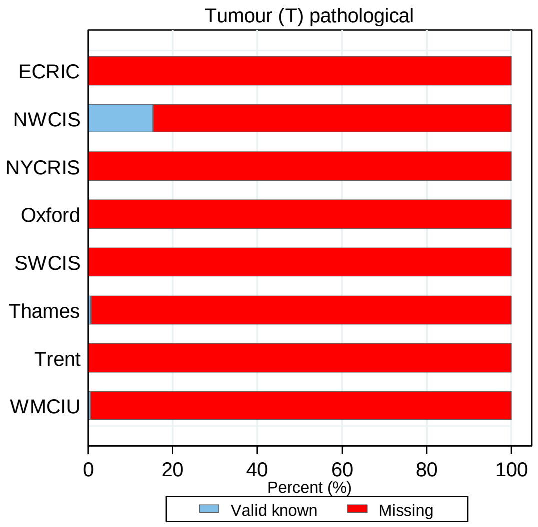


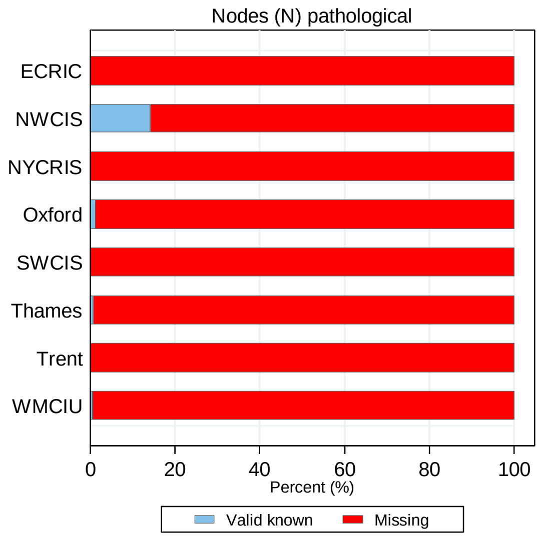


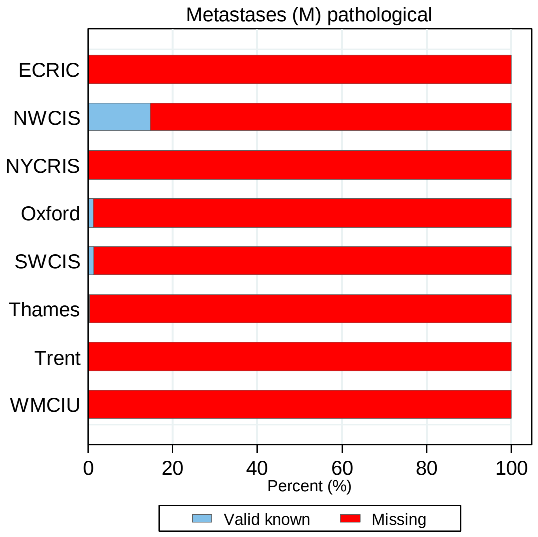


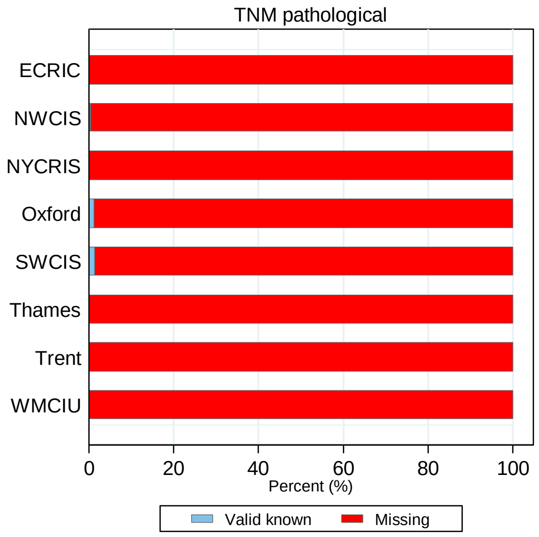


Mesothelioma (ICD10 C45)



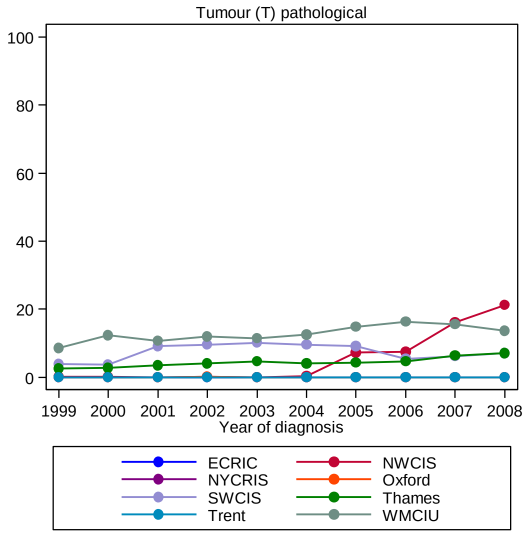


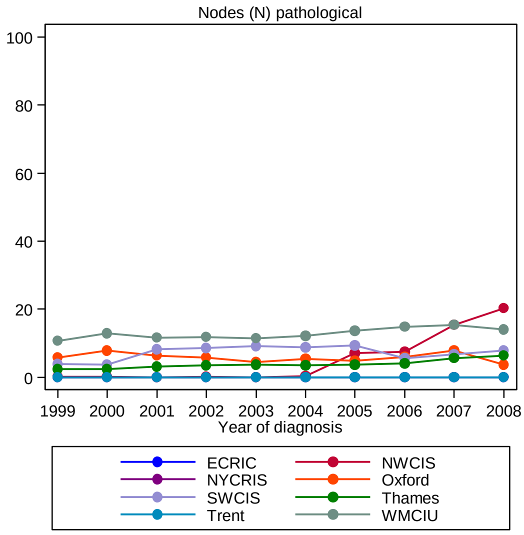


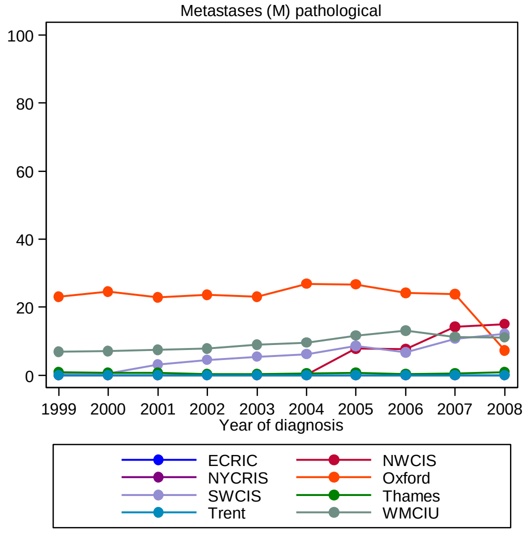


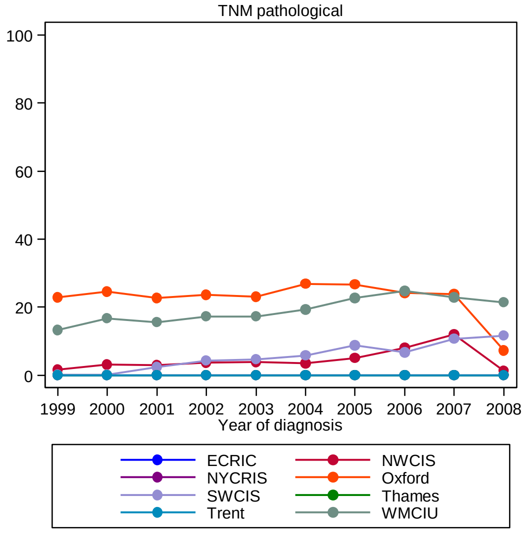
Overall, there were very low proportions of pathological T, N, M, and TNM stage recorded in both lung cancer and mesothelioma. Pathological T, N, and M stage information was missing for more than 91%, and pathological TNM stage for 96% of all lung cancer registrations, and almost all mesothelioma registrations.

The following graphs show the proportion of lung cancer registrations with pathological T, N, M and TNM stage information between 1999 and 2008. This analysis excludes death certificate only registrations.









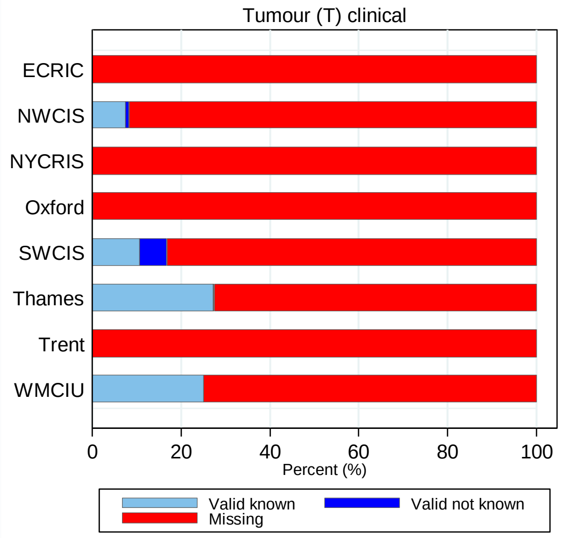
The availability of the separate pathological T, N, M and TNM stage information has remained constantly low throughout the ten-year period 1999 to 2008.

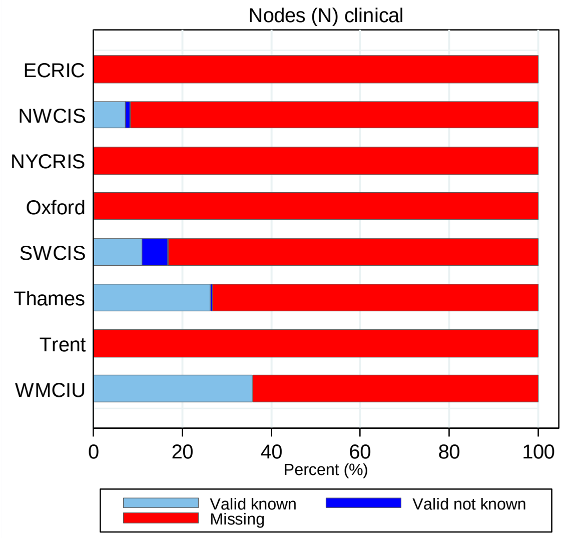
Trends were not run for mesothelioma due to the small number of registrations with recorded stage.

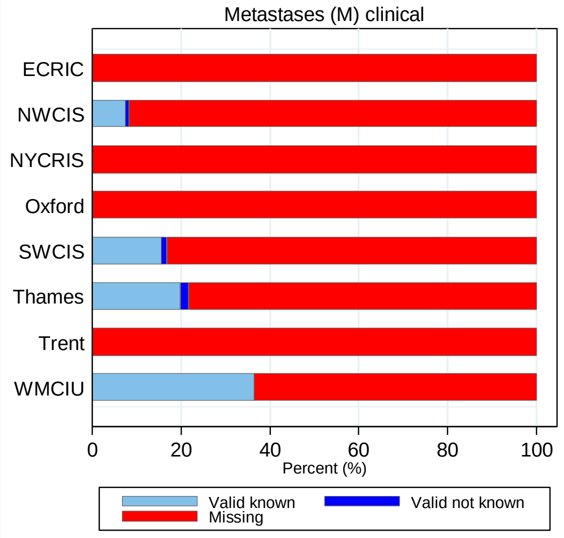
3.9 Clinical stage

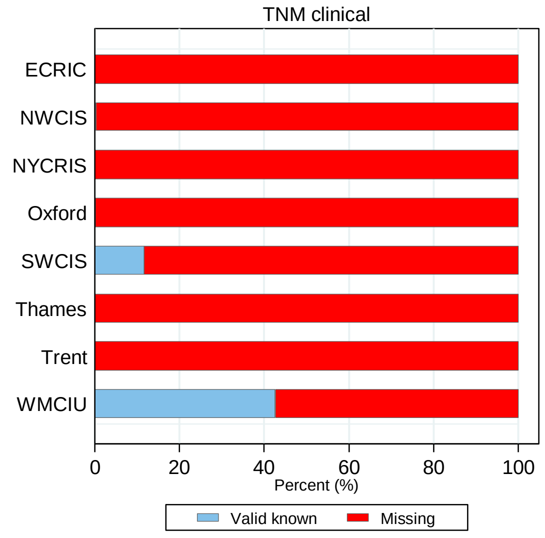
The following graphs show the proportion of registrations with clinical T, N, and M and TNM stage information in 2008. This analysis excludes death certificate only registrations.

Lung cancer (ICD10 C34)

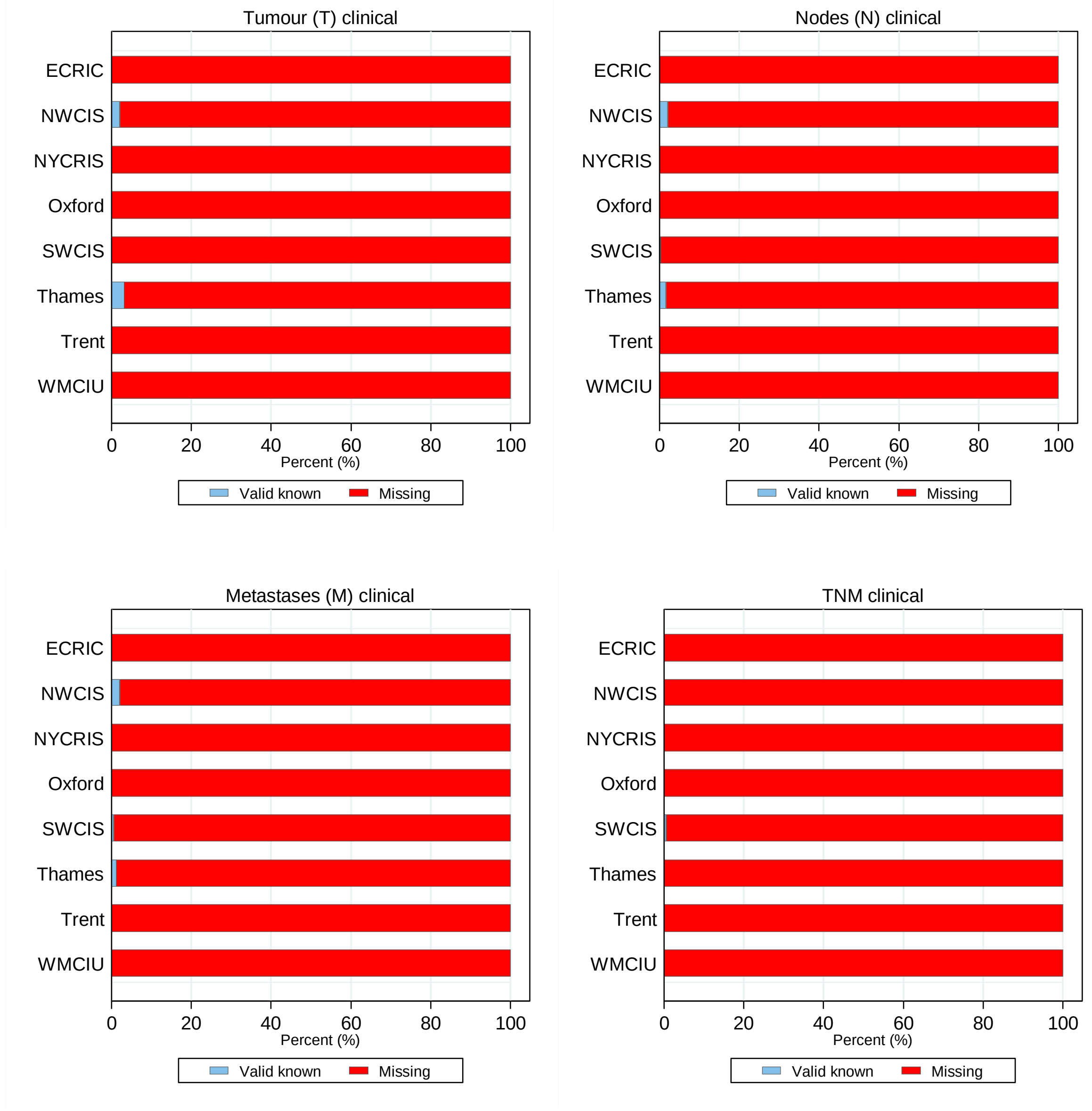






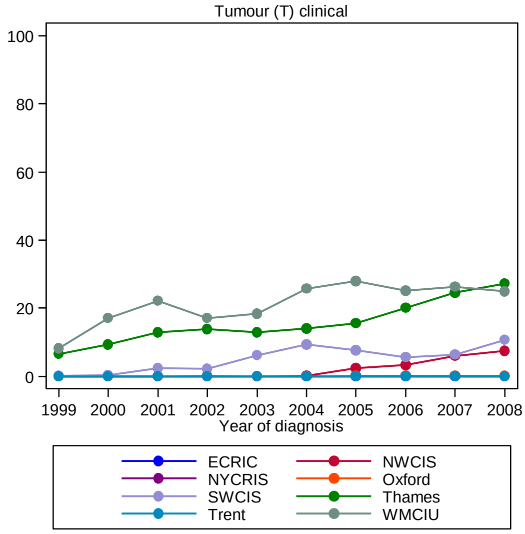


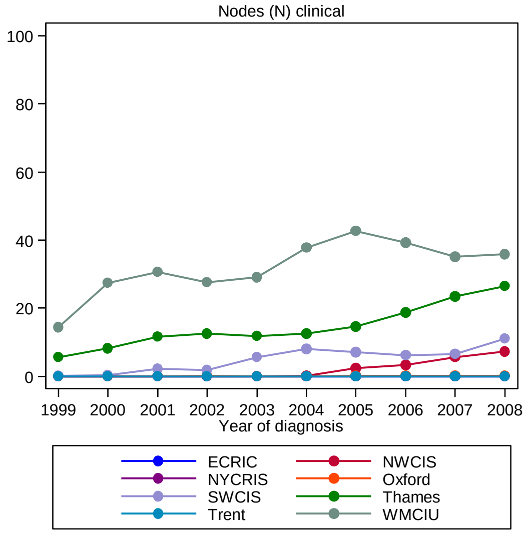
Mesothelioma (ICD10 C45)

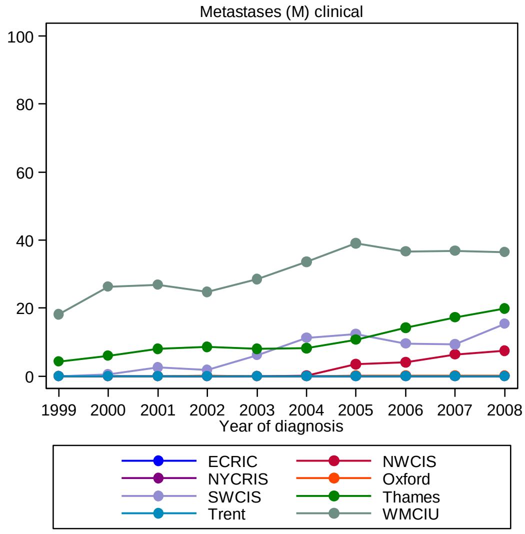


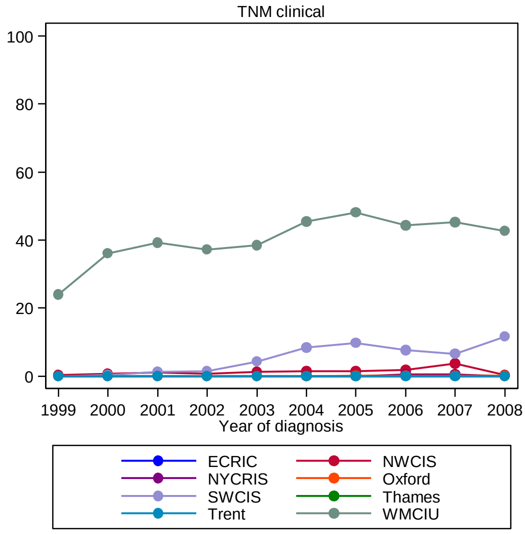
Overall, there were low proportions of clinical T, N, M, and TNM stage recorded in both lung cancer and mesothelioma. Clinical T, N, and M stage information was missing for more than 88%, and clinical TNM stage for 94% of all lung cancer registrations, and almost all mesothelioma registrations.

The following graphs show the proportion of lung cancer registrations with clinical T, N, M and TNM stage information between 1999 and 2008. This analysis excludes death certificate only registrations.







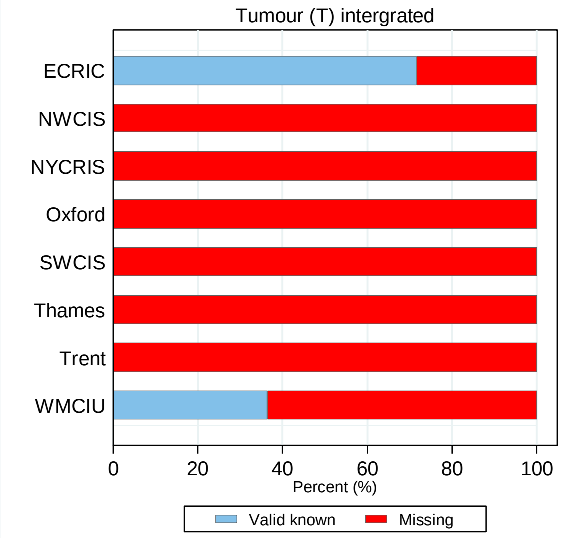


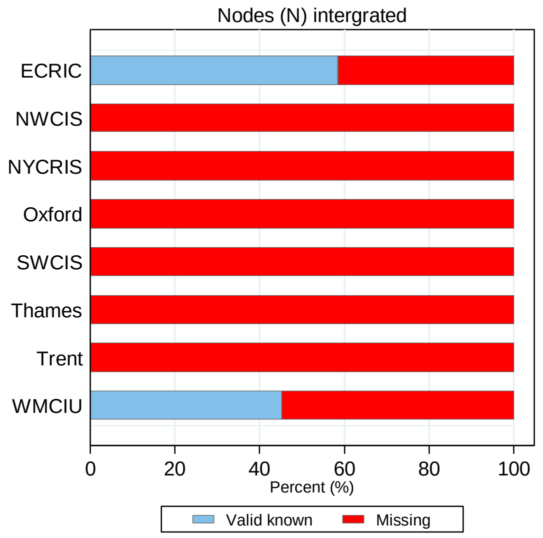
In general, the availability of clinical T, N, M and TNM stage information was higher than pathological stage information and has increased somewhat between 1999 and 2008.

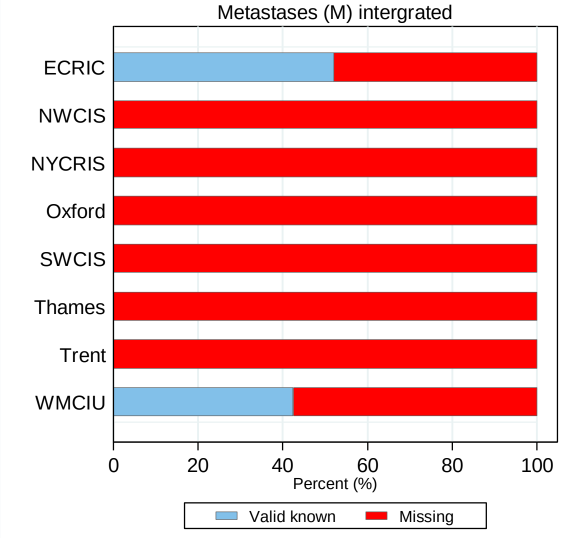
3.10 *Integrated stage*

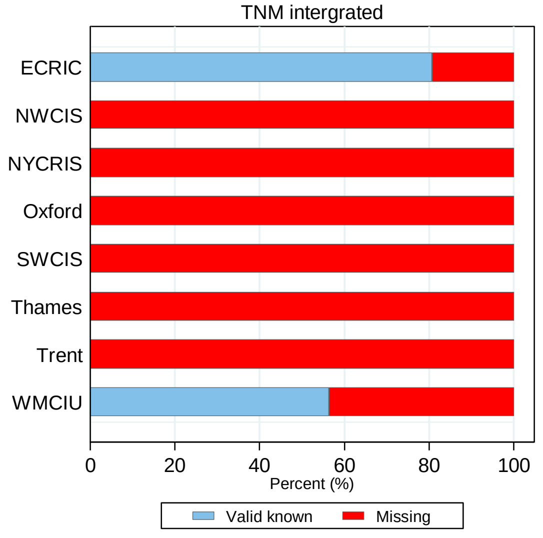
The following graphs show the proportion of registrations with integrated T, N, and M and TNM stage information in 2008. This analysis excludes death certificate only registrations.

Lung cancer (ICD10 C34)

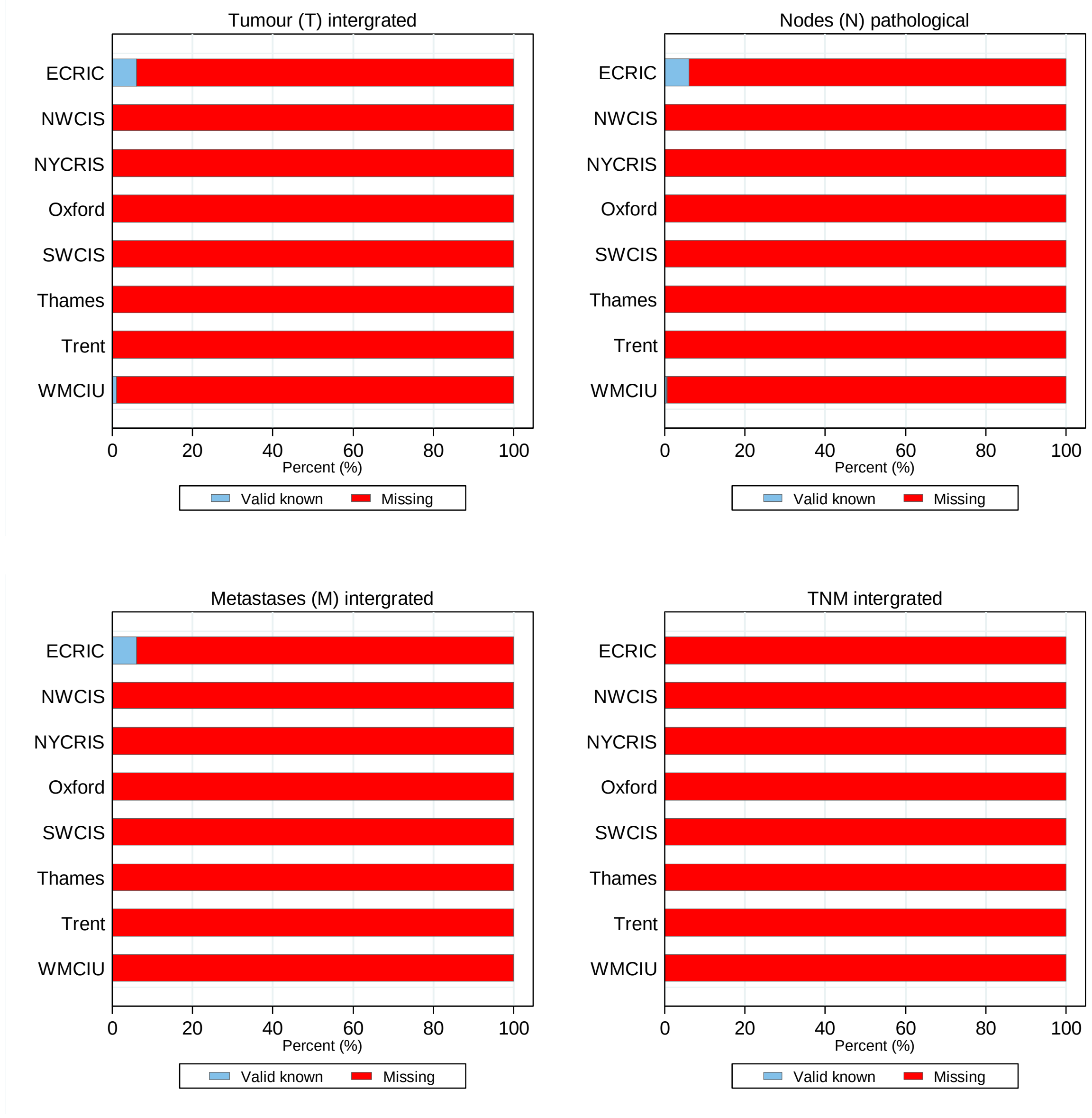






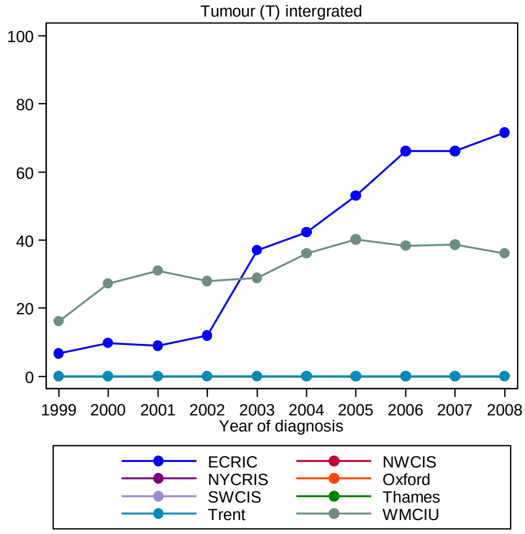


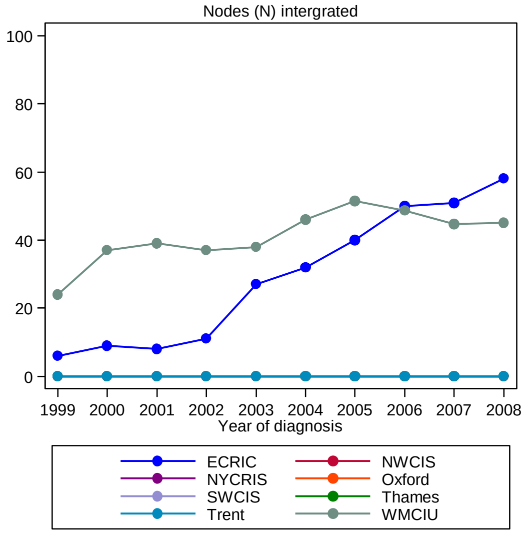
Mesothelioma (ICD10 C45)

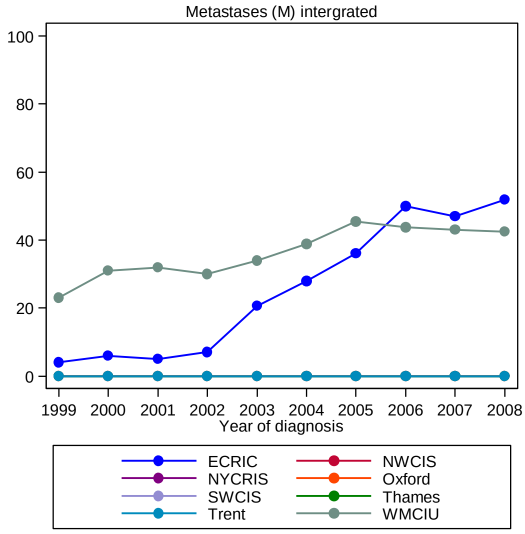


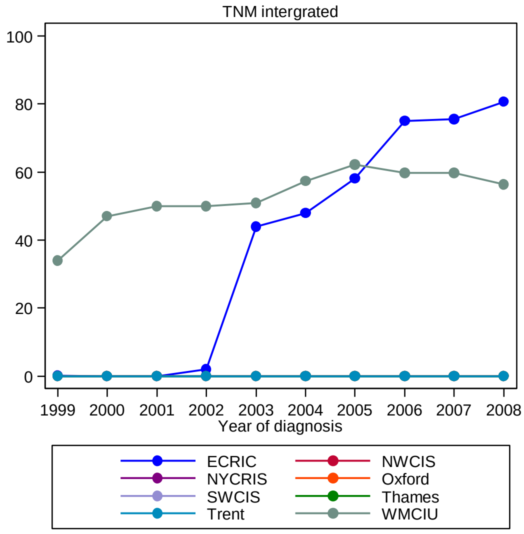
Only two cancer registries (ECRIC and WMCIU) submitted their staging information using the TNM (integrated) stage field.

The following graphs show the proportion of lung cancer registrations with integrated T, N, M, and TNM stage information between 1999 and 2008. This analysis excludes death certificate only registrations.


Tumour (T) intergrated


Nodes (N) intergrated


Metastases (M) intergrated


TNM intergrated

The availability of the integrated stage information has increased in WMCIU between 1999 and 2008, and a rapid increase was observed in ECRIC registrations from 2002 onwards.

3.11 Completeness

The completeness of the lung cancer dataset was good. In 2008, only 78 patients with lung cancer were estimated to be missed by the cancer registration process to a total of 31,598 patients that were registered, amounting to 0.2%.

The completeness of the lung cancer dataset was calculated by extracting HES records with a cancer diagnosis and relevant surgical procedure that were not matched to a cancer registration. The combination of these codes increased the certainty that these patients were true cancer cases and not just a suspicion of cancer. However, a low proportion of patients with lung cancer will have surgery. Therefore, this method may over-estimate the completeness of ascertainment of these cancers.

Since the surgery rate for mesothelioma is even lower than for lung cancer, we did not analyse the completeness for mesothelioma.

4. Key findings

- The overall proportion of death certificate only registrations was low and slightly higher in lung cancer (4%) than in mesothelioma (2%).

- More than 70% of lung cancers and over 85% of mesotheliomas were microscopically verified. Of lung cancers, 29% were clinically verified, whereas only 10% of mesotheliomas were clinically verified.

- The specification of anatomical site is lower in lung cancer (65%) than in mesothelioma (90%).

- Morphology information was available in 70% of lung cancers and in nearly all mesotheliomas.

- More than 90% of lung cancers and more than 88% of mesotheliomas had a linked HES record.

- Ethnicity information is available in 88% of lung cancers and 89% of mesotheliomas.

- In lung cancer, the availability of information from the studied stage fields (pathological, clinical and integrated T, N, M and TNM) was poor, although in some cases there was an increase in the proportion of records with a valid known stage over time. Very little stage information for mesotheliomas is available.

- Using a method that identifies inpatient records with both a lung cancer diagnosis and relevant procedure code, only 0.2% of lung cancer patients were estimated to have been missed by the cancer registration process in England in 2008.

5. Conclusions

This report has investigated the data quality of the lung cancer and mesothelioma registrations held within the National Cancer Data Repository.

The proportion of death certificate only registrations of both lung cancer and mesothelioma was low. These registrations would have to be excluded from any analysis that investigates survival of these patients, because they appear in the data repository as having zero survival time and therefore they may indicate incomplete case ascertainment and could potentially bias the survival estimates. It is important that work continues to further reduce the number of these registrations.

Morphological classification of lung cancer is low and therefore limits the possibility of analysing specific lung cancer groups.

The proportion of lung cancer and mesothelioma registrations with a linked record in HES is high. As improvements in the linkage between the two datasets continue, this proportion is probably going to increase further.

Overall, the availability of stage information was poor, and only moderate increases in availability of stage information was observed. Stage information is important and as national projects are underway to improve its availability, it is expected that improvements will be seen with time.

Although we do not have a valid method to assess the potentially missed mesothelioma cases, it is encouraging to estimate the completeness of lung cancer registrations to be well over 99%.

Appendix 1: List of ICD10 4 digit codes

C33 Malignant neoplasm of trachea

C34 Malignant neoplasm of bronchus or lung

C34.0	Malignant neoplasm: Main bronchus, Carnia, hilus of lung
C34.1	Malignant neoplasm: Upper lobe, bronchus or lung
C34.2	Malignant neoplasm: Middle lobe (or lingular lobe on left), bronchus of lung
C34.3	Malignant neoplasm: Lower lobe, bronchus or lung
C34.8	Malignant neoplasm: Overlapping lesion of bronchus and lung
C34.9	Malignant neoplasm: Bronchus or lung, unspecified

C45 Malignant neoplasm of mesothelioma

C45.0	Mesothelioma of pleura
C45.1	Mesothelioma of peritoneum
C45.2	Mesothelioma of pericardium
C45.7	Mesothelioma of other sites
C45.9	Mesothelioma, unspecified

Source: http://apps.who.int/classifications/apps/icd/icd10online/

Appendix 2: List of procedure codes used in the completeness analysis

E541	Total pneumonectomy
E391	Open excision of lesion of trachea
E398	Other specified partial excision of trachea
E399	Unspecified partial excision of trachea
E441	Excision of carina
E461	Sleeve resection of bronchus and anastomosis HFQ
E542	Bilobectomy of lung
E543	Lobectomy of lung
E544	Excision of segment of lung
E545	Partial lobectomy of lung NEC
E548	Other specified excision of lung
E549	Unspecified excision of lung
E552	Open excision of lesion of lung
E559	Unspecified open extirpation of lesion of lung
T013	Excision of lesion of chest wall
T023	Insertion of prosthesis into chest wall NEC

FIND OUT MORE:

Thames Cancer Registry is the lead cancer registry for lung cancer and mesothelioma.