

**Lung Cancer and Mesothelioma
Site Specific Clinical Reference Group
Data Quality Report 2009**

Sharma P Riaz
Karen M Linklater
Henrik Møller
Margreet Luchtenborg



Contents

1. Introduction.....	1
2. Methods.....	2
Data quality.....	2
Death certificate only registrations	2
Basis of diagnosis.....	2
Anatomical site.....	3
Morphology	3
Linked HES records.....	3
Ethnicity.....	3
Stage variables.....	3
3. Results.....	4
3.1.1 Quality of the lung cancer dataset, England, 2009	4
3.1.2 Quality of the mesothelioma dataset, England, 2009	5
3.2 Death certificate only.....	6
3.3 Basis of diagnosis.....	7
3.4 Anatomical site.....	8
3.5 Morphology	9
3.7 Ethnicity.....	11
3.8 Pathological stage	12
3.9 Clinical stage	14
3.10 Integrated stage.....	16
4. Key findings	18
5. Conclusions.....	19
Appendix 1: List of ICD10 4 digit codes	20
Appendix 2: List of unspecified morphology codes	21

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.

This report explores the data quality and completeness of the lung cancer and mesothelioma dataset as derived from the NCDR. It reports on data on patients diagnosed in 2009 while also exploring the trends in data quality over the 11-year period from 1999 to 2009.

2. Methods

Data were extracted from the NCDR on all cases of lung cancer (ICD-10 C33-C34) and mesothelioma (ICD10-C45) diagnosed in 1999-2009.

There were 351,701 malignant neoplasms of the trachea, bronchus and lung and 21,044 mesothelioma registrations during the 1999 to 2009 period.

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: Number and proportion of lung cancers and mesothelioma by Cancer registries in England, 1999-2009 (including DCO's).

Cancer registry codes	Cancer registry name	Lung cancer		Mesothelioma	
ECRIC	Eastern Cancer Registration Information Centre	34,364	9.8	2,557	12.2
NWCIS	North West Cancer Intelligence Service	56,851	16.2	2,785	13.2
NYCRIS	Northern & Yorkshire Cancer Registry and Information Service	60,391	17.2	3,375	16.0
Oxford	Oxford Cancer Intelligence Unit	14,602	4.2	963	4.6
SWCIS	South West Cancer Intelligence Service	43,942	12.5	3,541	16.8
Thames	Thames Cancer Registry	67,656	19.2	4,426	21.0
Trent	Trent Cancer Registry	37,670	10.7	1,762	8.4
WMCIU	West Midlands Cancer Intelligence Unit	36,225	10.3	1,635	7.8

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 and 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 retrieved. 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. Although these cases are valuable for incidence calculations, they need to be excluded from analyses of survival.

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 (not known and missing).

Anatomical site

A full list on codes for anatomical site is presented in Appendix 1. Unknown anatomical site group included tumours with an ICD10 four digit code of Cxx.8 (overlapping lesion of [specific] cancer) and Cxx.9 ([specific] cancer, unspecified). Large proportions of patients with an unspecified anatomical site will limit the ability to analyse these cancers by specific subgroups.

Morphology

Morphology was classified as known (valid morphology codes) and not known (see Appendix 2). Large proportions of tumours with an unknown morphology code will limit our ability to analyse these cancers by specific morphology subgroups.

Linked HES records

Some cancer registrations cannot be linked to an inpatient or day-case HES record and therefore no treatment information can be included in the NCDR dataset. This situation can occur as a result of unsuccessful matching of patient information, or because the subset of HES data received by the cancer registries only includes patients with a diagnosis of cancer and their treatment may not have been coded as related to a diagnosis of cancer in HES, or the patient has had no inpatient hospital activity. This is important to consider in 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 influences the treatment that patients can be offered. 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 recorded 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).

3. Results

3.1.1 Quality of the lung cancer dataset, England, 2009

Total number of registrations	England (n=33,281)		ECRIC (n=3,324)		NWCIS (n=5,447)		NYCRIS (n=5,837)		Oxford (n=1,474)		SWCIS (n=4,098)		Thames (n=6,094)		Trent (n=3,621)		WMCIU (n=3,386)	
	Number	(%)	Number	(%)	Number	(%)	Number	(%)	Number	(%)	Number	(%)	Number	(%)	Number	(%)	Number	(%)
Death certificate only																		
Death certificate only	863	2.6	3.00	0.1	300	5.5	95	1.6	43	2.9	94	2.3	206	3.4	71	2.0	51	1.5
Non-DCO registrations	32,418	97.4	3,321	99.9	5,147	94.5	5,742	98.4	1,431	97.1	4,004	97.7	5,888	96.6	3,550	98.0	3,335	98.5
Basis of diagnosis (excluding DCO registrations)																		
Microscopically verified	22,601	69.7	2,307	69.5	3,400	66.1	3,954	68.9	995	69.5	2,712	67.7	4,289	72.8	2,458	69.2	2,486	74.5
Clinically verified	9,729	30.0	985	29.7	1,724	33.5	1,788	31.1	436	30.5	1,292	32.3	1,563	26.6	1,092	30.8	849	25.5
Anatomical site (excluding DCO registrations)																		
Known anatomical site	21,751	67.1	2,887	86.9	3,330	64.7	4,305	75.0	681	47.6	2,388	59.6	3,512	59.7	2,501	70.5	2,147	64.4
Morphology (excluding DCO registrations)																		
Known	22,312	68.8	2,297	69.2	3,401	66.1	3,963	69.0	969	67.7	3,259	81.4	3,476	59.0	2,462	69.4	2,485	74.5
Linked record in Hospital Episode Statistics (excluding DCO registrations)																		
Linked	29,296	90.4	2,952	88.9	4,743	92.2	5,191	90.4	1,295	90.5	3,595	89.8	5,258	89.3	3,267	92.0	2,995	89.8
Ethnicity (excluding DCO registrations)																		
Known	28,384	87.6	2,835	85.4	4,610	89.6	5,056	88.1	1,254	87.6	3,432	85.7	5,095	86.5	3,212	90.5	2,890	86.7
Valid known stage (excluding DCO registrations)																		
Pathological																		
T	1,537	4.8	0	0.0	30	0.6	28	0.5	66	4.6	340	8.5	556	9.5	0	0.0	517	15.5
N	1,309	4.0	0	0.0	26	0.5	28	0.5	54	3.8	373	9.3	482	8.2	0	0.0	346	10.4
M	1,047	3.2	0	0.0	95	1.9	0	0.0	21	1.5	535	13.4	52	0.9	0	0.0	344	10.3
TNM	998	3.1	0	0.0	103	2.0	0	0.0	14	1.0	530	13.3	0	0.0	0	0.0	351	10.5
Clinical																		
T	4,192	13.0	0	0.0	10	0.2	0	0.0	8	0.6	829	20.7	1,756	29.9	0	0.0	1,589	47.7
N	4,110	12.7	0	0.0	10	0.2	0	0.0	9	0.6	875	21.9	1,727	29.4	0	0.0	1,489	44.7
M	4,764	14.7	0	0.0	21	0.4	0	0.0	12	0.8	1,554	38.9	1,382	23.5	0	0.0	1,795	53.8
TNM	3,019	9.3	0	0.0	20	0.4	4	0.1	7	0.5	1,186	29.7	0	0.0	0	0.0	1,802	54.1
Integrated																		
T	4,984	15.4	2,573	77.5	571	11.1	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	1,840	55.2
N	2,234	6.9	0	0.0	571	11.1	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	1,663	49.9
M	2,548	7.9	0	0.0	585	11.4	25	0.4	0	0.0	0	0.0	0	0.0	0	0.0	1,938	58.1
TNM	5,625	17.4	2,814	84.8	438	8.5	431	7.5	0	0.0	0	0.0	0	0.0	0	0.0	1,942	58.3

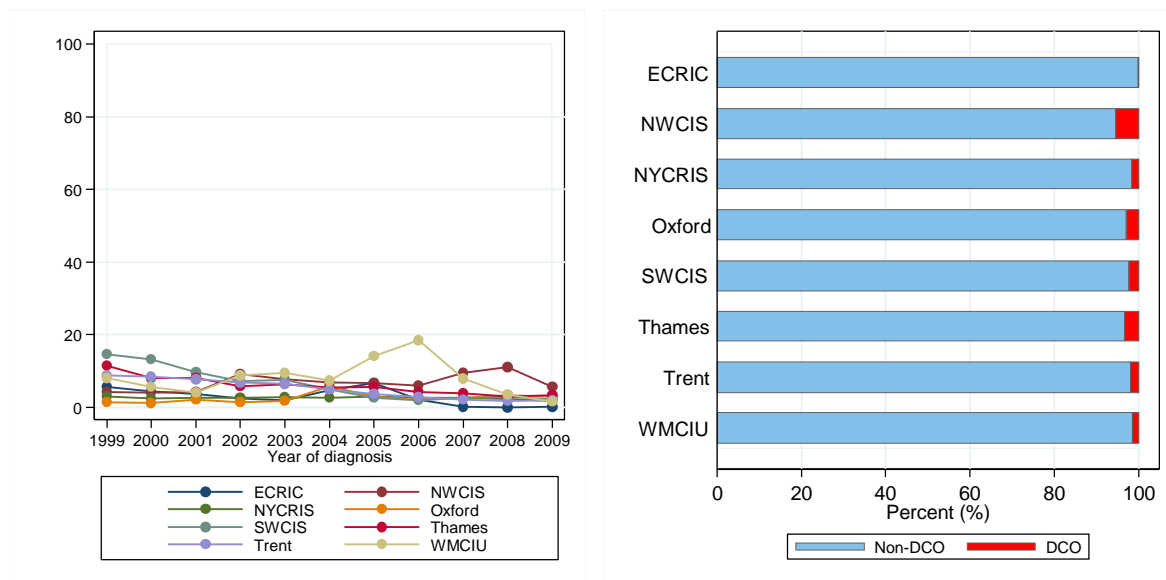
3.1.2 Quality of the mesothelioma dataset, England, 2009

Total number of registrations		England (n=33,281)	ECRIC (n=3,324)	NWCIS (n=5,447)	NYCRIS (n=5,837)	Oxford (n=1,474)	SWCIS (n=4,098)	Thames (n=6,094)	Trent (n=3,621)	WMCIU (n=3,386)
		Number	(%)	Number	(%)	Number	(%)	Number	(%)	Number
Death certificate only										
Death certificate only		41	1.9	0	0.0	14	5.0	7	1.9	1
Non-DCO registrations		2,165	98.1	268	100.0	266	95.0	362	98.1	206
Basis of diagnosis (excluding DCO registrations)										
Microscopically verified		1,832	84.6	235	87.7	223	83.8	311	85.9	261
Clinically verified		327	15.1	32	11.9	41	15.4	51	14.1	82
Anatomical site (excluding DCO registrations)										
Known anatomical site		1,947	89.9	260	97.0	259	97.4	319	88.1	260
Morphology (excluding DCO registrations)										
Known		2,163	99.9	268	100.0	266	100.0	362	100.0	343
Linked record in Hospital Episode Statistics (excluding DCO registrations)										
Linked		1,959	90.5	238	88.8	244	91.7	324	89.5	312
Ethnicity (excluding DCO registrations)										
Known		1,922	88.8	232	86.6	241	90.6	317	87.6	306
Valid known stage (excluding DCO registrations)										
Pathological										
T		8	0.4	0	0.0	8	3.0	0	0.0	0
N		5	0.2	0	0.0	3	1.1	0	0.0	2
M		16	0.7	0	0.0	8	3.0	0	0.0	8
TNM		20	0.9	0	0.0	12	4.5	0	0.0	8
Clinical										
T		49	2.3	0	0.0	0	0.0	0	0.0	4
N		49	2.3	0	0.0	0	0.0	0	0.0	2
M		61	2.8	0	0.0	2	0.8	0	0.0	18
TNM		21	1.0	0	0.0	2	0.8	0	0.0	16
Integrated										
T		57	2.6	27	10.1	5	1.9	0	0.0	0
N		31	1.4	0	0.0	5	1.9	0	0.0	0
M		30	1.4	0	0.0	5	1.9	0	0.0	0
TNM		7	0.3	0	0.0	4	1.5	1	0.3	0

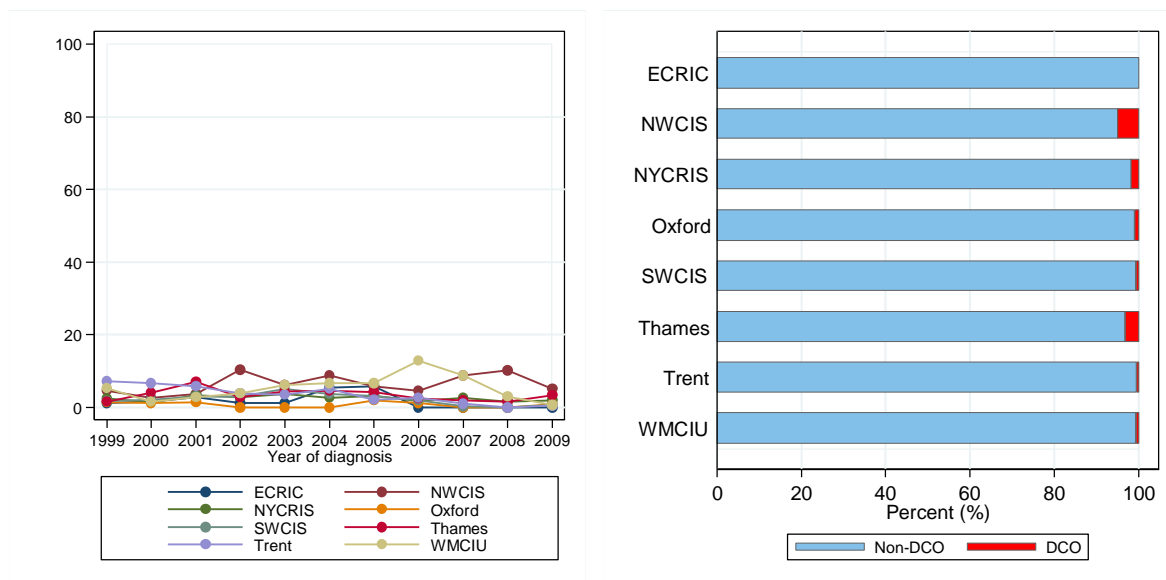
3.2 Death certificate only

The following graphs show the proportion of death certificate only registrations for lung cancer and mesothelioma by cancer registry as trends over the 11-year period (1999-2009) and in the most recent year (2009).

Lung cancer (ICD10 C33-C34)



Mesothelioma (ICD10 C45)

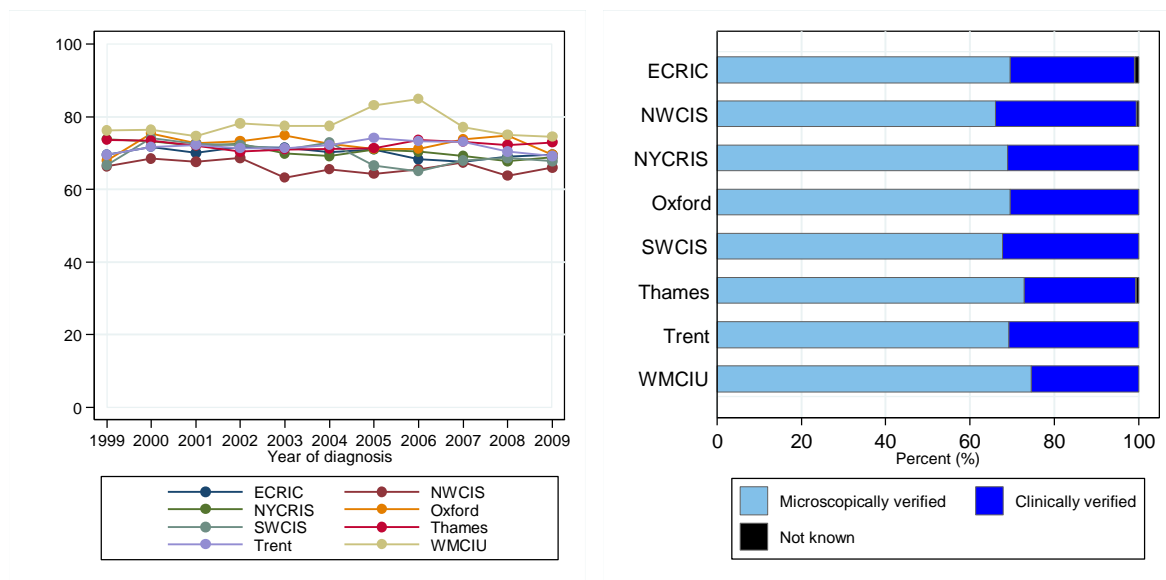


Overall, the proportions of DCO registrations were very low. The proportion of cancers with death certificate only registrations gradually decreased between 1999 and 2009. In general, in 2009, the proportion of DCO registration was higher in lung cancer (3%) than in mesothelioma (2%).

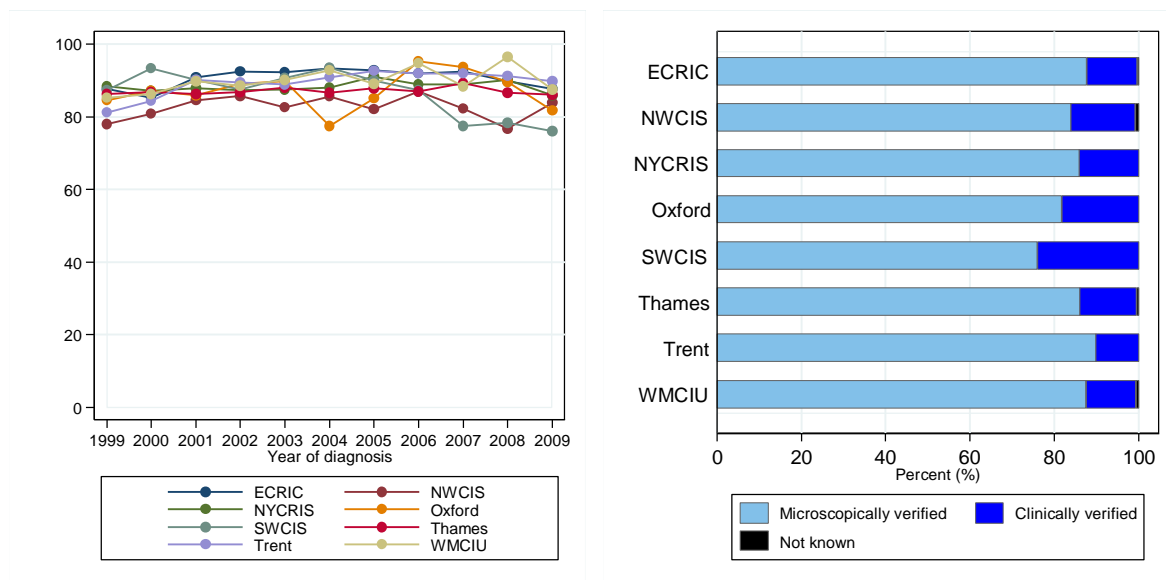
3.3 Basis of diagnosis

The following graphs show the proportion of the different bases of diagnosis of registrations for lung cancer and mesothelioma by cancer registry as trends over the 11-year period (1999-2009) and in the most recent year (2009).

Lung cancer (ICD10 C33-C34)



Mesothelioma (ICD10 C45)

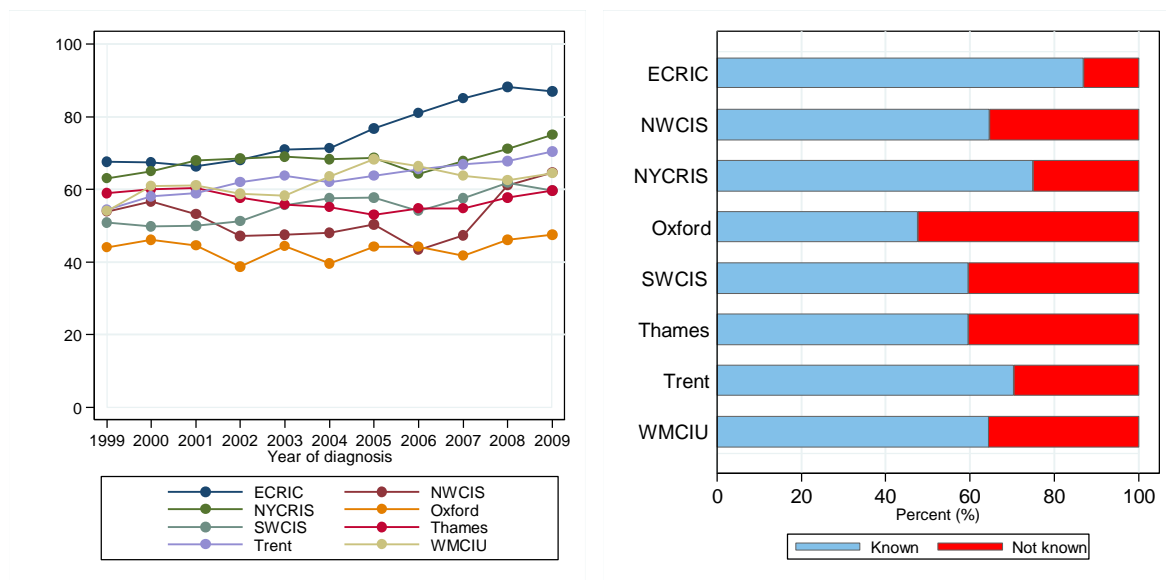


The proportion of tumours with microscopically verified information was relatively stable between 1999 and 2009 for the eight cancer registries. In 2009, over 69% of lung cancers and over 76% of mesotheliomas were microscopically verified. More than 25% of lung cancer and 10% mesothelioma were clinically verified. The microscopic verification rate was higher in mesothelioma. The higher verification rate of mesothelioma compared to lung cancer is probably related to the need for microscopic verification to arrive at its diagnosis.

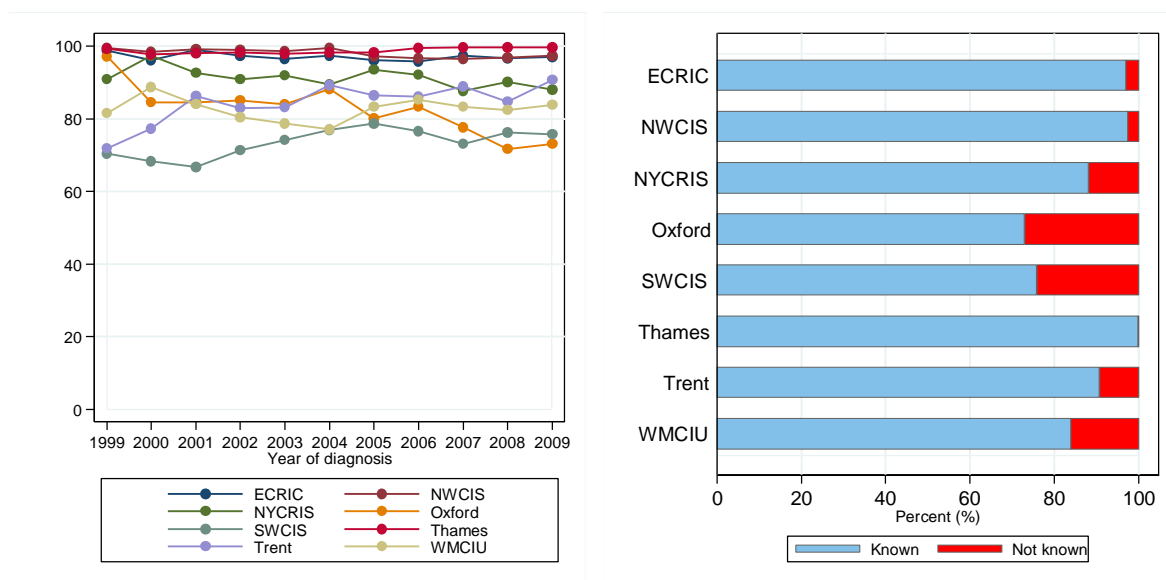
3.4 Anatomical site

The following graphs show the proportion of the registrations with anatomical site for lung cancer and mesothelioma by cancer registry as trends over the 11-year period (1999-2009) and in the most recent year (2009).

Lung cancer (ICD10 C33-C34)



Mesothelioma (ICD10 C45)

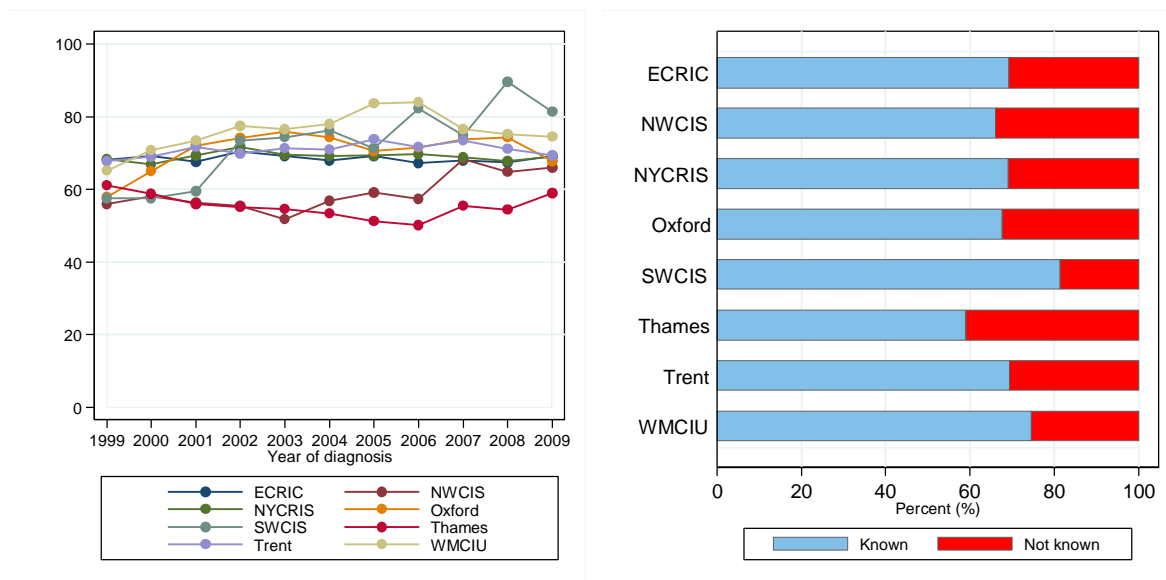


Overall, the proportion of lung cancer registrations with a known anatomical site increased between 1999 and 2009 for all cancer registries. The trends of mesothelioma registrations with a known anatomical site varied across the different cancer registries. In the most recent year the specification of anatomical site was lower in lung cancer (67%) than in mesothelioma (90%). The anatomical site of mesothelioma is more likely to be specified because of its symptomatology and importance to treatment options.

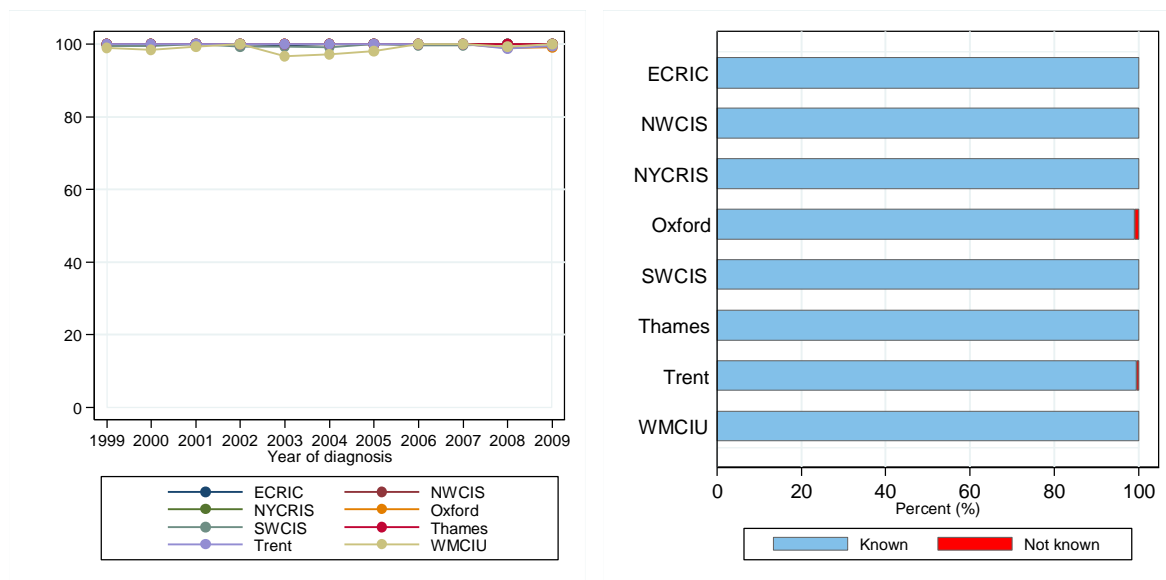
3.5 Morphology

The following graphs show the proportion of registrations with known morphology for lung cancer and mesothelioma by cancer registry as trends over the 11-year period (1999-2009) and in the most recent year (2009).

Lung cancer (ICD10 C33-C34)



Mesothelioma (ICD10 C45)

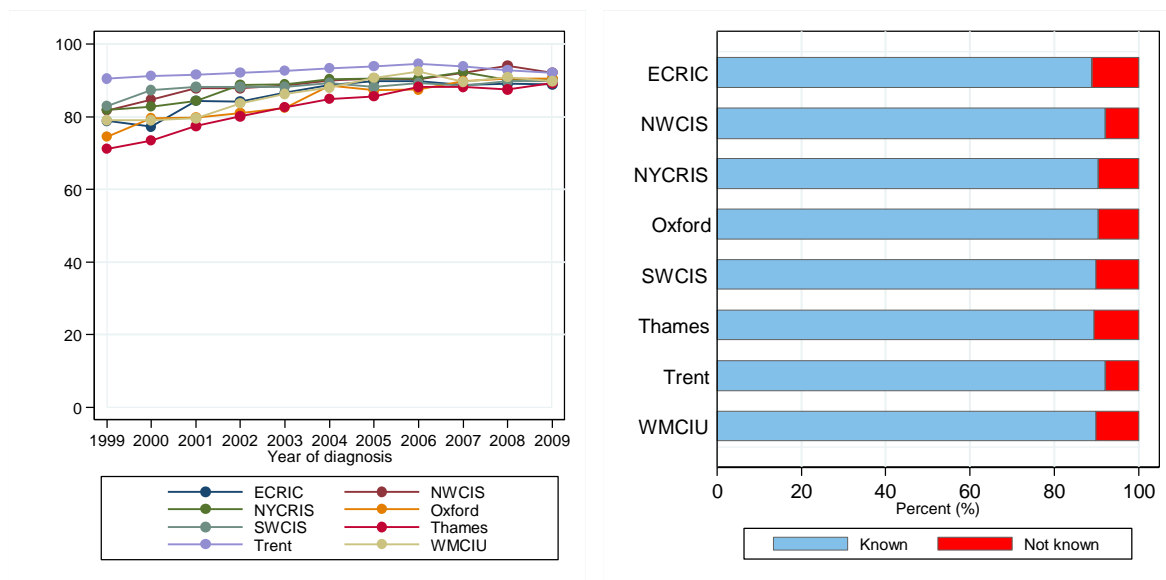


On average, the proportion of registrations with known morphology of lung cancers increased from around 63% in 1999 to 69% in 2009. Morphology information was available for nearly all mesothelioma registrations.

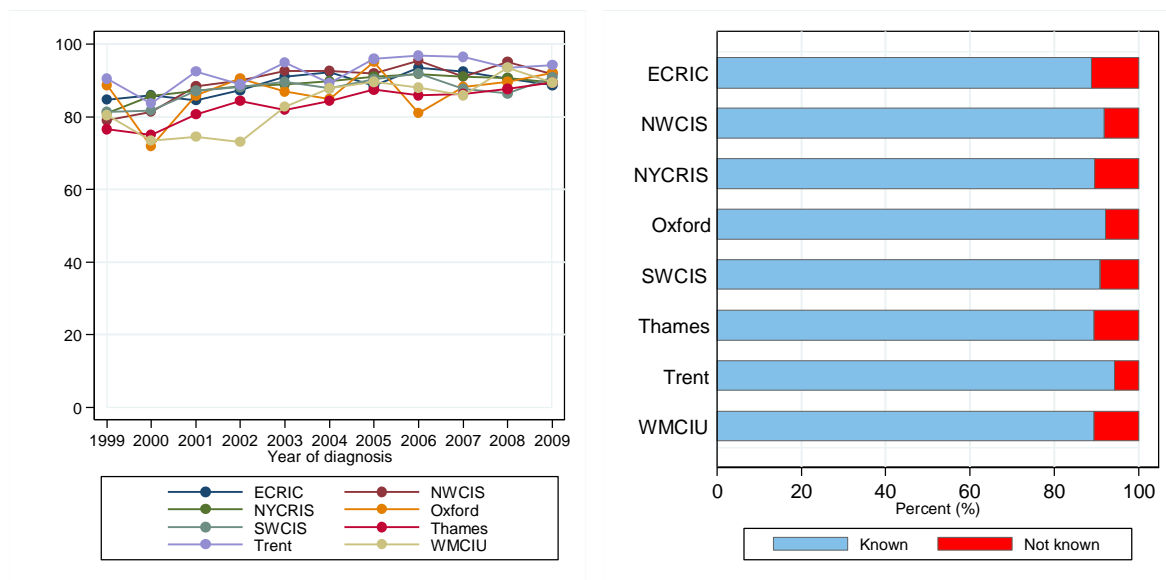
3.6 Linked HES records

The following graphs show the proportion of registrations with a linked HES record for lung cancer and mesothelioma by cancer registry as trends over the 11-year period (1999-2009) and in the most recent year (2009).

Lung cancer (ICD10 C33-C34)



Mesothelioma (ICD10 C45)

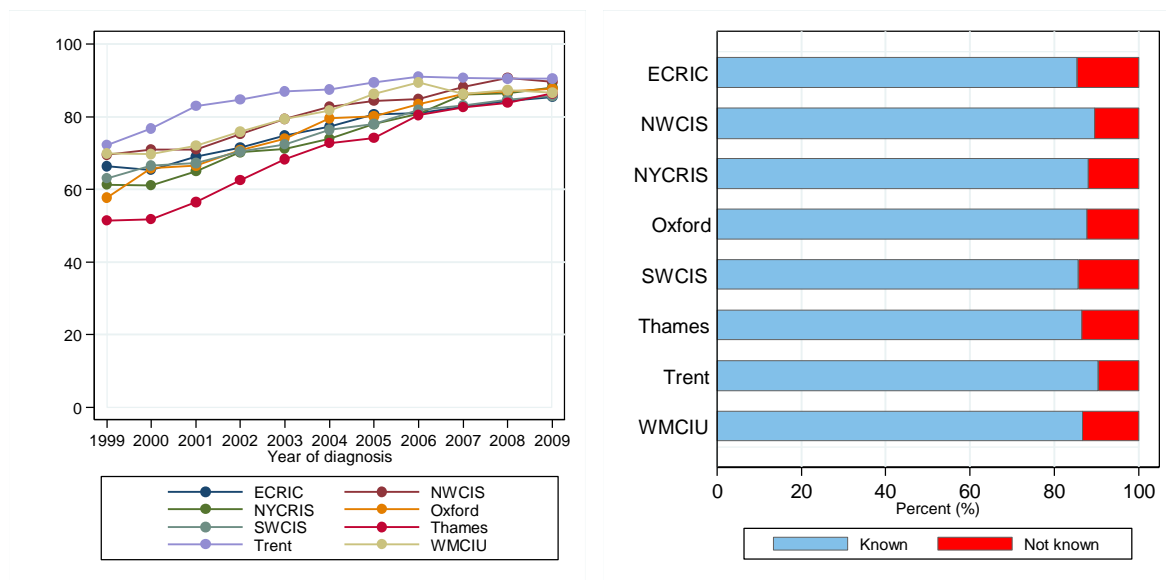


Overall, the proportion of patients with linked HES record information increased from 1999 to 2009 across all cancer registries. In the most recent year, around 90% of cancers had a linked HES record. There was more variation between cancer registrations with a linked HES record for mesotheliomas compared 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.

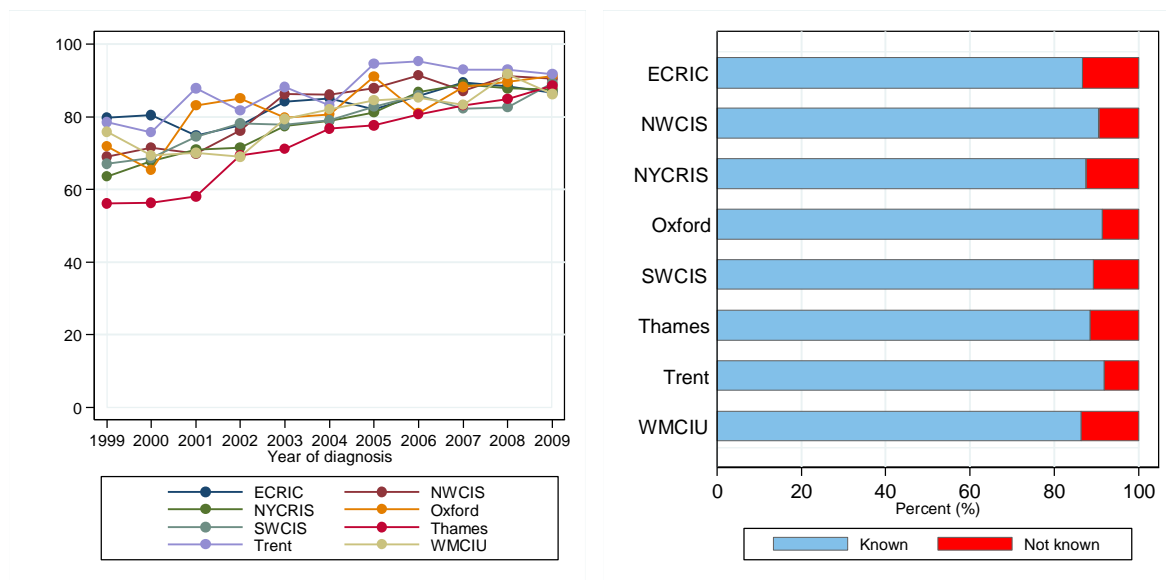
3.7 Ethnicity

The following graphs show the proportion of registrations with known ethnicity for lung cancer and mesothelioma by cancer registry as trends over the 11-year period (1999-2009) and in the most recent year (2009).

Lung cancer (ICD10 C33-C34)



Mesothelioma (ICD10 C45)

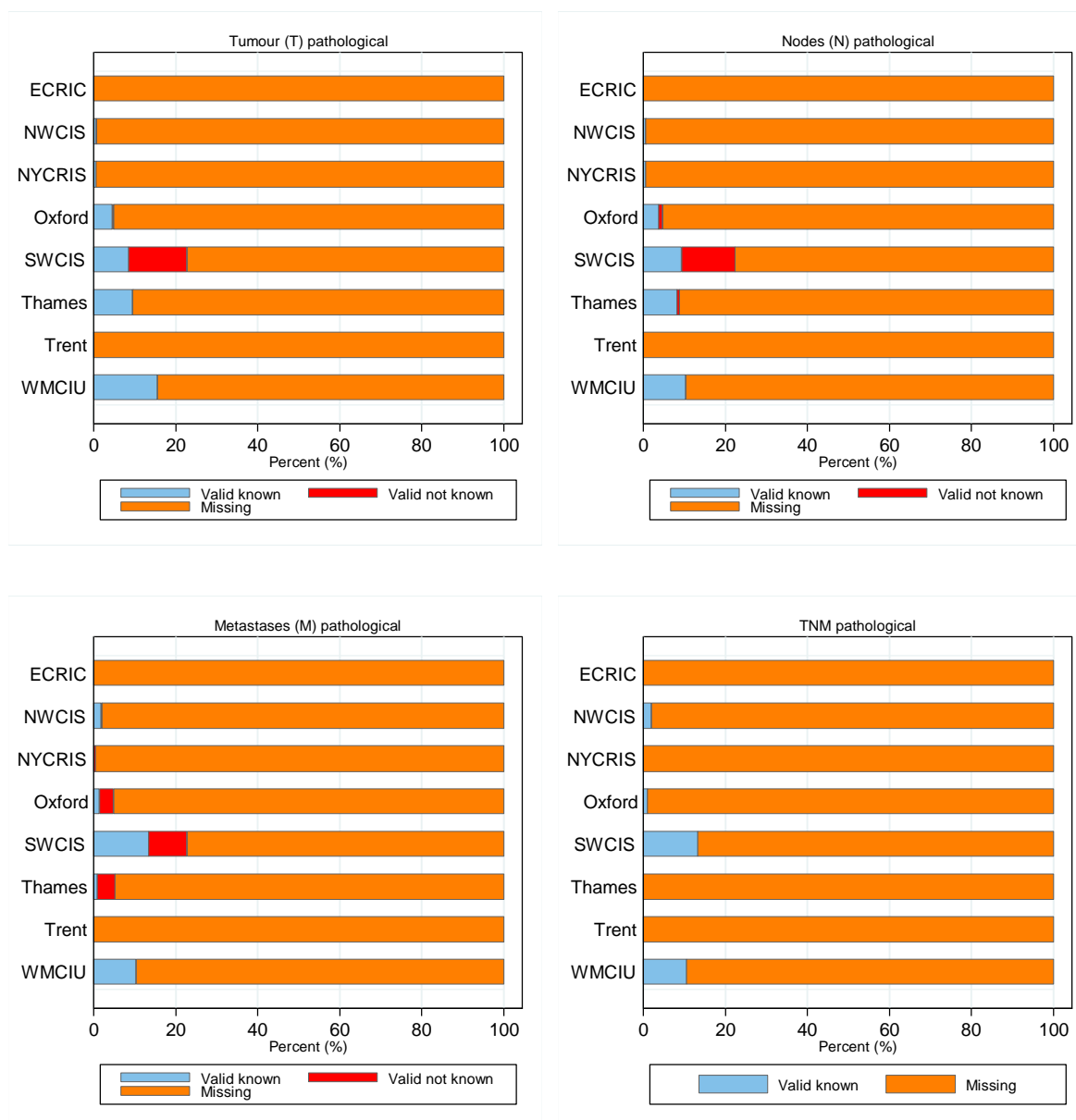


Across the cancer registries, there was an increase in the proportion of patients with known ethnicity information between 1999 and 2009. In 2009, the proportion of registrations with known ethnicity was very similar at 88% of lung cancers and 89% of mesotheliomas. The variation in proportions of registrations with known ethnicity between the cancer registries was 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

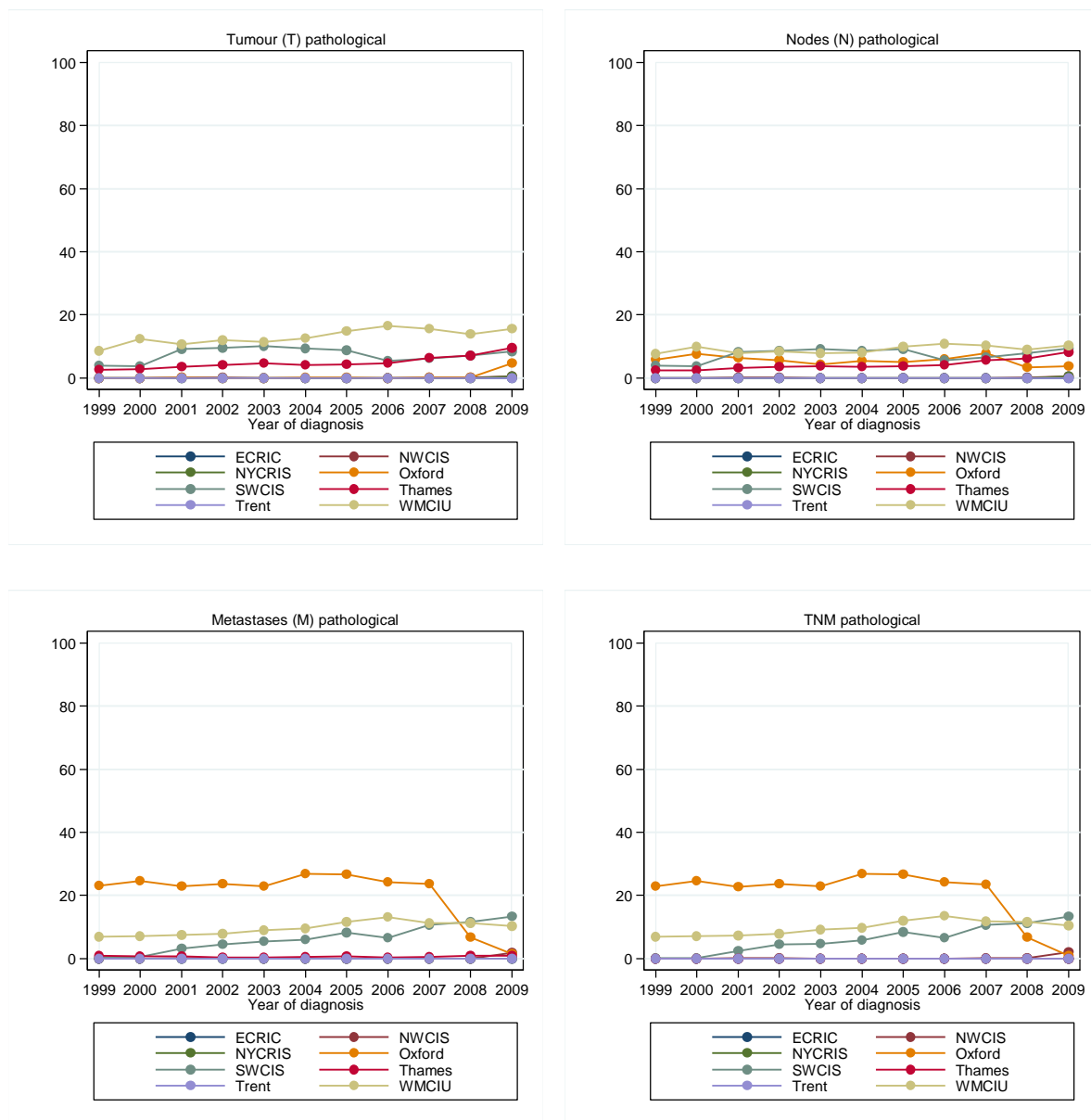
The following graphs show the proportion of registrations with pathological T, N, M and TNM stage information by cancer registry in 2009. Stage information for mesothelioma is not included due to the small number of mesothelioma registrations with a recorded stage.

Lung cancer (ICD10 C33-C34)



Overall, there were very low proportions of pathological T, N, M, and TNM stage recorded for lung cancer. Pathological T, N, and M stage information was missing for more than 96%, and pathological TNM stage for 95% of all lung cancer registrations.

The following graphs show the trends in the proportion of lung cancer registrations with pathological T, N, M and TNM stage information by cancer registry between 1999 and 2009.

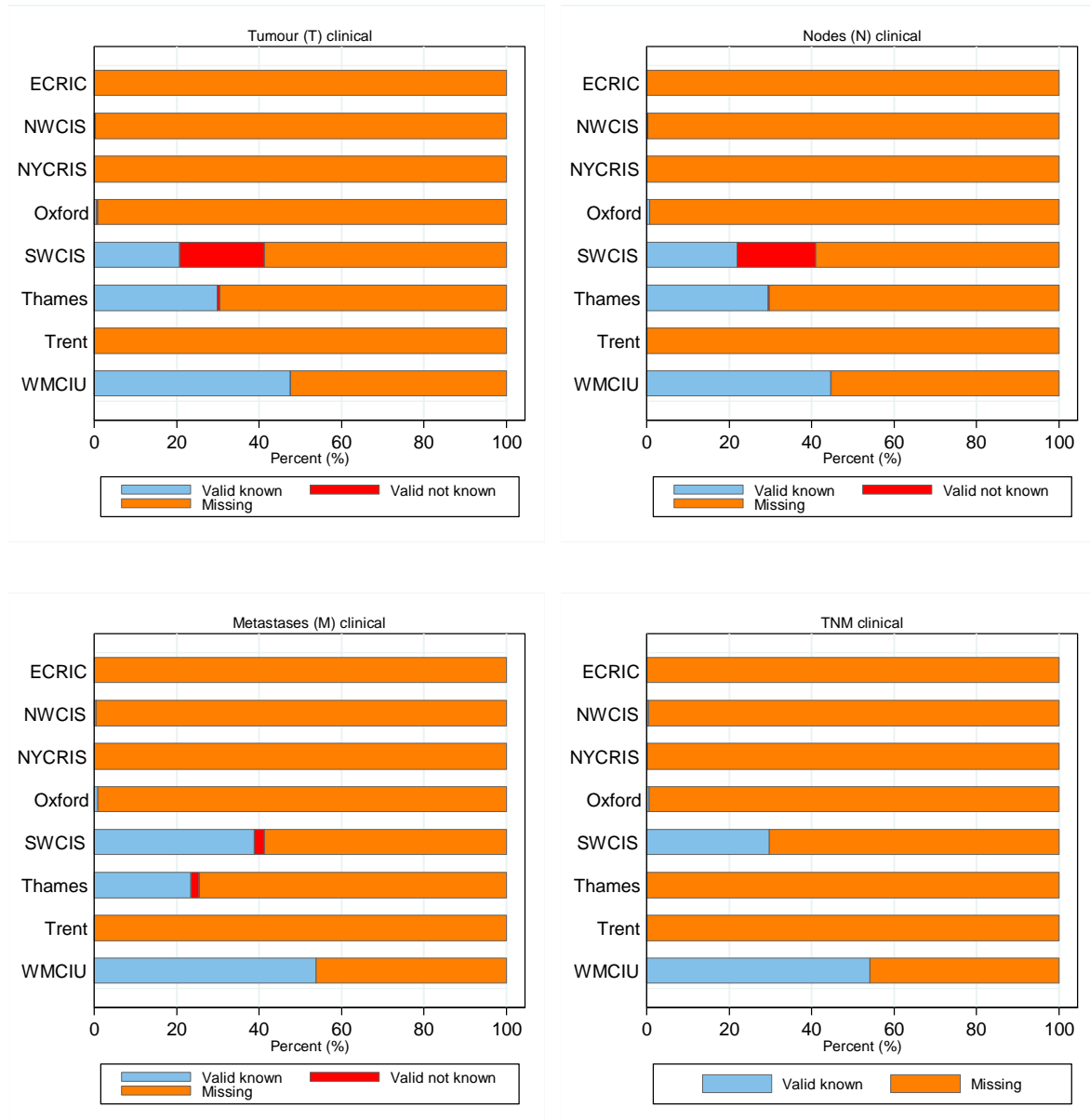


The availability of the separate pathological T, N, M as well as TNM stage information has remained constantly low throughout the eleven-year period 1999 to 2009.

3.9 Clinical stage

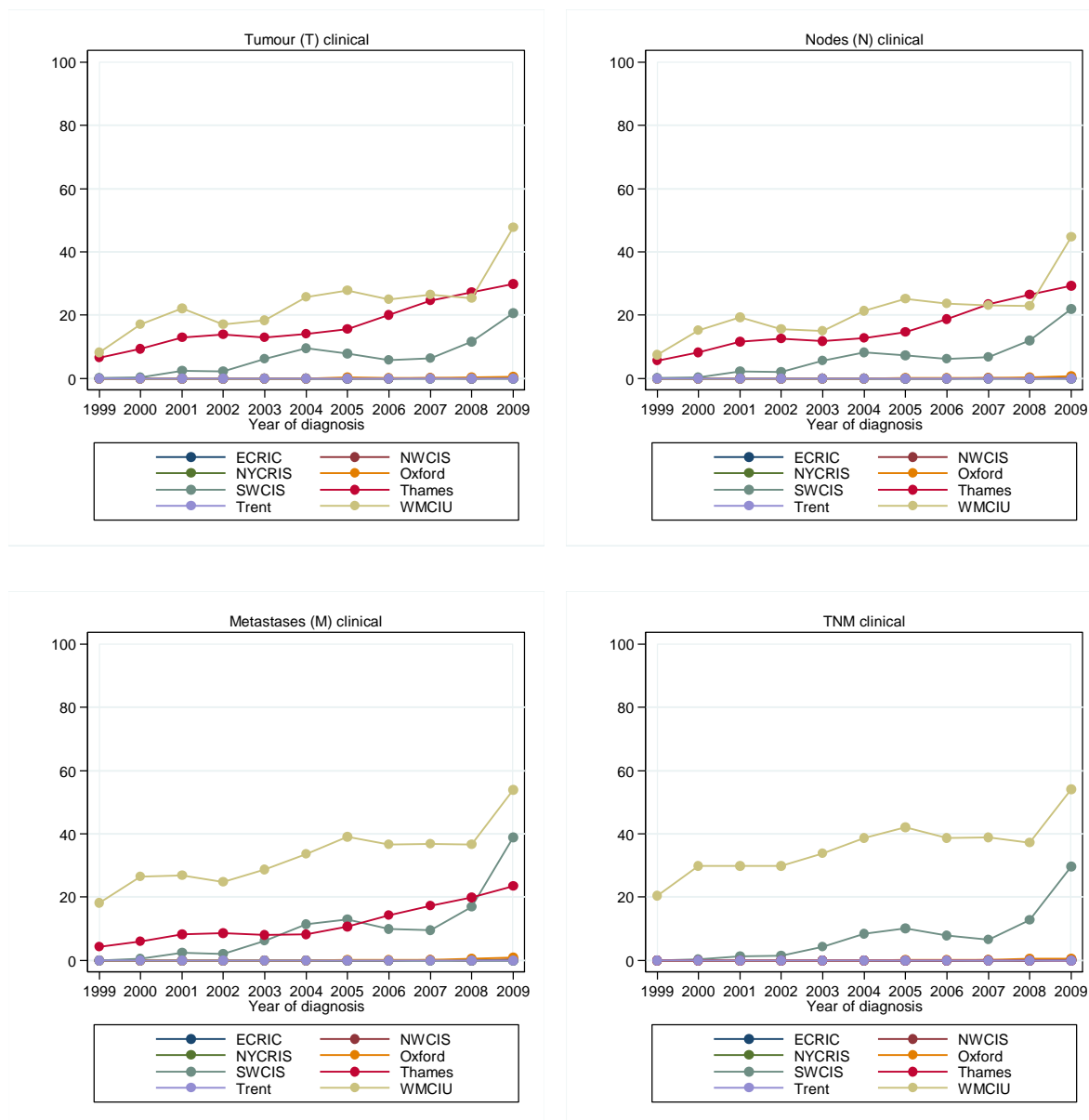
The following graphs show trends in the proportion of registrations with clinical T, N, and M and TNM stage information by cancer registry in 2009.

Lung cancer (ICD10 C33-C34)



Overall, there were low proportions of clinical T, N, M, and TNM stage recorded in the lung cancer dataset. Clinical T, N, and M stage information was missing for more than 84%, and clinical TNM stage for 91% of all lung cancer registrations.

The following graphs show the trends in the proportion of lung cancer registrations with clinical T, N, M and TNM stage information by cancer registry between 1999 and 2009.

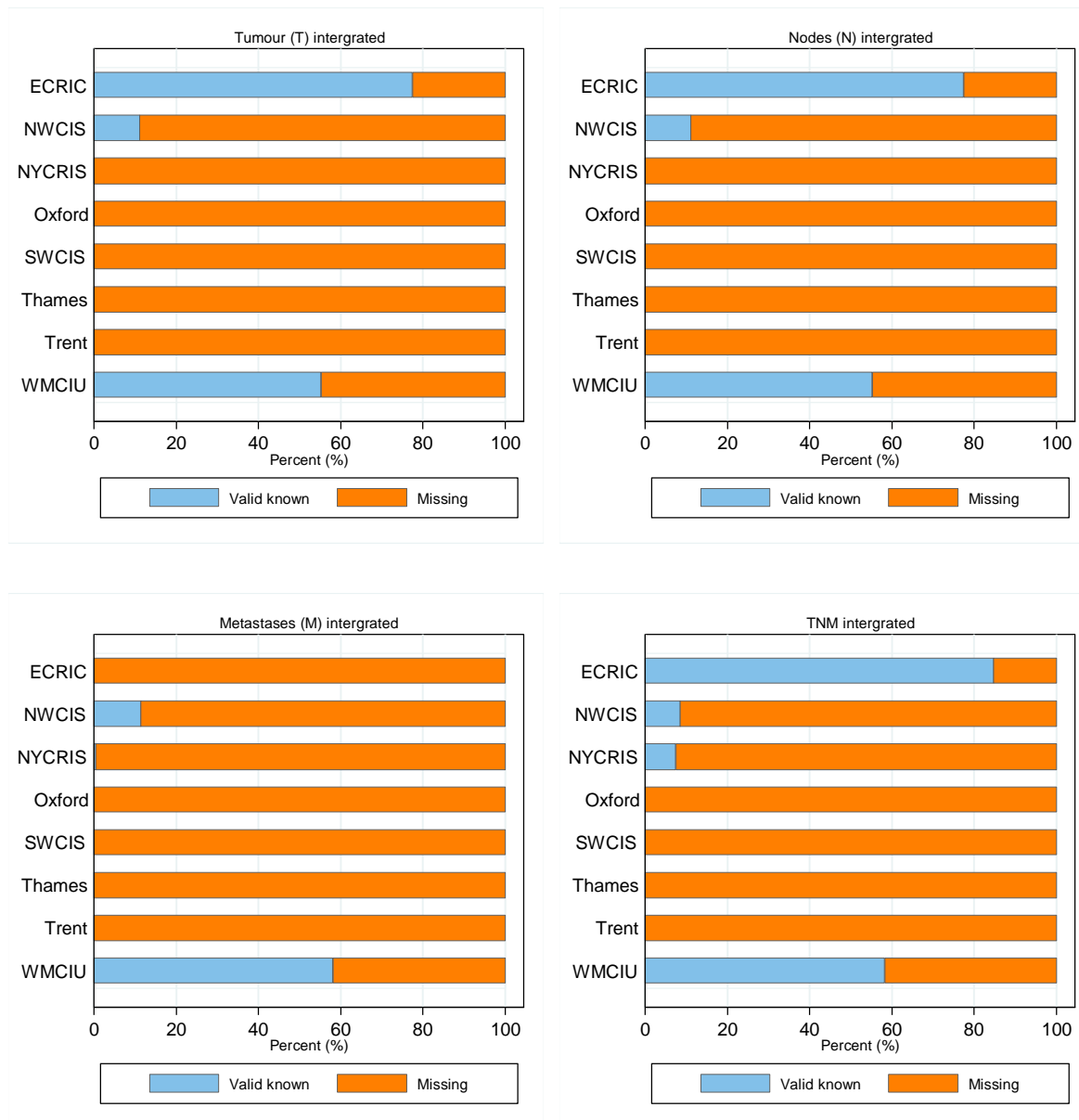


In general, the availability of clinical T, N, M and TNM stage information was higher than pathological stage information and has increased somewhat between 1999 and 2009. The proportions of cancer registrations with T, N, and M stage increased in the WMCIU, NYCRIS, and Thames Cancer Registry, and particularly in the last registration year.

3.10 Integrated stage

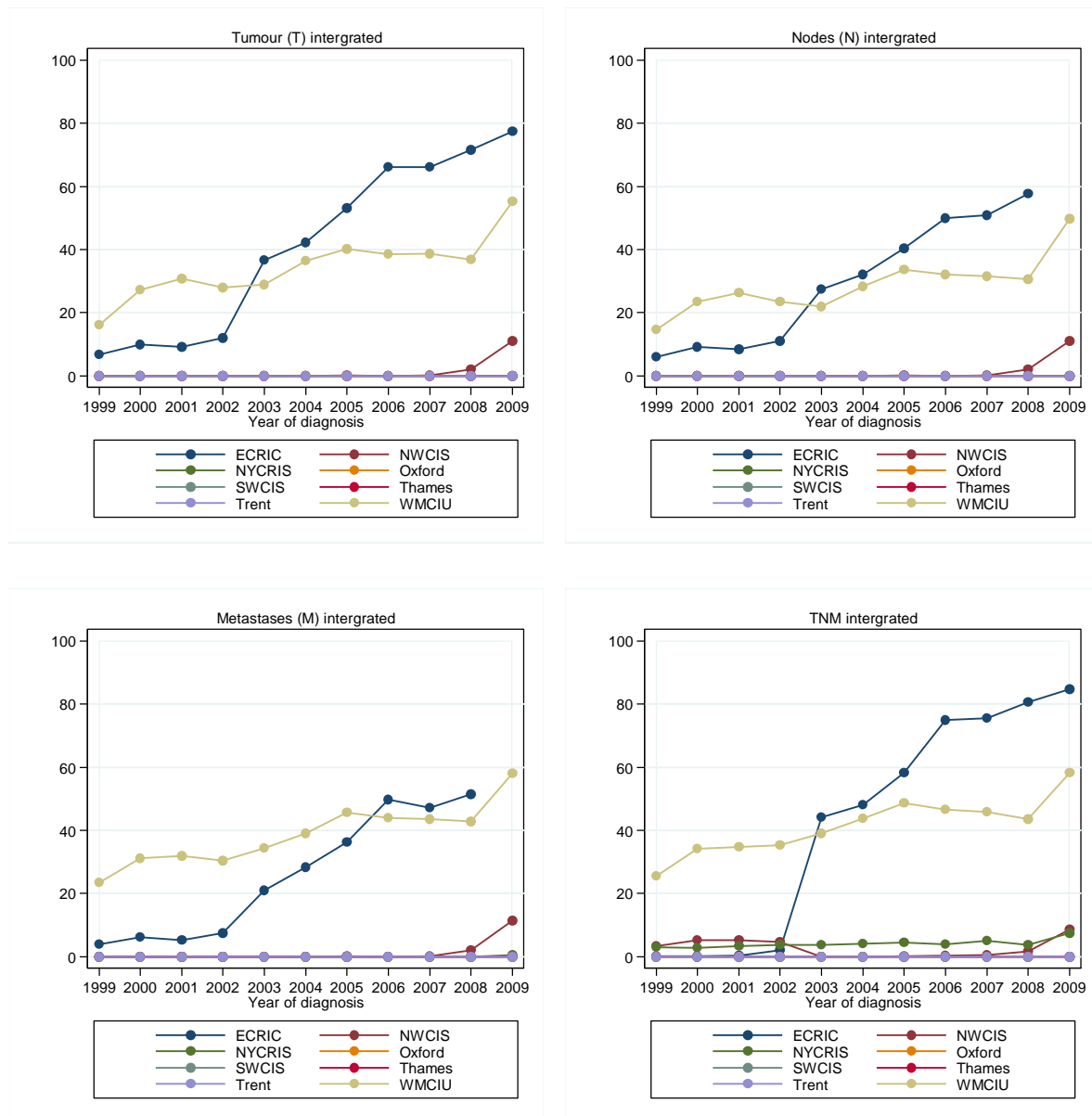
The following graphs show the proportion of registrations with integrated T, N, and M and TNM stage information by cancer registry in 2009.

Lung cancer (ICD10 C34)



Only two cancer registries (ECRIC and WMCIU) submitted their staging information using the TNM (integrated) stage field. The availability of T and TNM stage information was high in ECRIC, whereas information availability was quite high for all parameters in the data submitted by WMCIU.

The following graphs show trends in the proportion of lung cancer registrations with integrated T, N, M, and TNM stage information by cancer registry between 1999 and 2009.



The availability of the integrated stage information has increased in WMCIU between 1999 and 2009, and a rapid increase was observed in ECRIC registrations from 2002 onwards. In 2009, ECRIC had no stage information for the nodes and metastases fields.

4. Key findings

- The proportion of death certificate only registrations decreased over the 11-year period (1999-2009). Overall, proportions of DCO registrations were low in lung cancer (3%) and in mesothelioma (2%).
- Between 1999 and 2009 the information of patients with microscopically verified information was relatively stable for all eight cancer registries. In the most recent year, more than 69% of lung cancers were microscopically and 25% clinically verified, whereas over 76% of mesotheliomas were microscopically verified and 10% were clinically verified.
- The proportion of lung cancers with known anatomical site information increased over time. Overall, the specification of anatomical site is 67% in lung cancer and 90% in mesothelioma.
- Over the 11-year period morphology information increased for lung cancers. Morphology information was available for nearly all mesotheliomas.
- The proportion of cancer registration with a linked HES record increased between 1999 and 2009. In 2009, more than 90% of cancers had a linked HES record.
- The proportion of registrations with known ethnicity increased over the 11-year period. In the earliest year ethnicity information was 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 the 11-year period analysed. Very little stage information for mesotheliomas was available.

5. Conclusions

This report has investigated the data quality of the lung cancer and mesothelioma registrations held within the National Cancer Data Repository, with a focus on the most recent year and the trends between 1999 and 2009.

The proportion of death certificate only registrations of both lung cancer and mesothelioma was low and declined over the 11-year period (1999-2009). These registrations would have to be excluded from any analysis that investigates survival of these patients. It is important that work continues to further reduce the number of these registrations.

Morphological classification of lung cancer was low but increased between 1999 and 2009. A high proportion of morphology availability allows for the possibility of analysing specific lung cancer groups; hence it is important the upward trend is continued.

The proportion of lung cancer and mesothelioma registrations with a linked record in HES and the recording of ethnicity have increased over the study period.

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 further improvements will be seen with time.

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, Carina, 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 unspecified morphology codes

Lung cancer

M8000	Neoplasm, malignant
M8001	Tumour cells, malignant
M8002	Malignant tumour, small cell type
M8003	Malignant tumour, giant cell type
M8004	Malignant tumour, fusiform cell type
M8010	Carcinoma NOS
M8011	Epithelioma, malignant
M8020	Carcinoma, undifferentiated NOS
M8021	Carcinoma, anaplastic type NOS
M8022	Pleomorphic carcinoma
M8030	Giant cell and spindle cell carcinoma
M8031	Giant cell carcinoma
M8032	Spindle cell carcinoma
M8033	Pseudosarcomatous carcinoma
M8034	Polygonal cell carcinoma
M8040	Tumorlet

Missing

Mesothelioma

M8000	Neoplasm, malignant
M8001	Tumour cells, malignant

Missing

FIND OUT MORE:

[Thames Cancer Registry](#) is the lead cancer registry for lung cancer and mesothelioma.

The NCIN is a UK-wide initiative, working closely with cancer services in England, Scotland, Wales and Northern Ireland, and the NCRI, to drive improvements in standards of cancer care and clinical outcomes by improving and using the information it collects for analysis, publication and research. In England, the NCIN is part of the National Cancer Programme.