

Lung cancer data in the National Cancer Data Repository, Hospital Episode Statistics and National Lung Cancer Audit datasets

A comparison of case ascertainment and recording
of lung cancer histology and treatment

Sharma P. Riaz
Karen M. Linklater
Marie Horton
Michael D. Peake
Henrik Møller
Margreet Lüchtenborg



Introduction

Because of the poor survival of lung cancer patients in England compared to other countries in Europe with similar health care expenditure, there has been an increase in epidemiological research into the reasons behind this. In order to arrive at meaningful conclusions, it is crucial that any research in this area is based on complete and high quality data.

In England, data on all cancer diagnoses are collected by the eight regional cancer registries (Eastern Cancer Registration and Information Centre, North West Cancer Intelligence Service, Northern and Yorkshire Cancer Registry and Information Service, Oxford Cancer Intelligence Unit, South West Cancer Intelligence Service, Thames Cancer Registry, Trent Cancer Registry and West Midlands Cancer Intelligence Unit). Data from the cancer registries are merged to form the core of the National Cancer Data Repository (NCDR). Completeness of lung cancer case ascertainment between 2001 and 2007 was estimated to be over 99%. [1] The NCDR is supplemented with data on in-patient admissions and treatments from the Hospital Episode Statistics (HES) data.

The National Lung Cancer Audit (NLCA) was established in 2004 and has resulted in a database that holds detailed clinical information on individuals diagnosed with lung cancer in England and Wales. Data entry is non-mandatory, but its case ascertainment has been shown to have increased over the years and was reported to be approximately 85% of the total incidence in 2008. [2]

The purpose of this report was to assess the case ascertainment and comparability of the lung cancer data on histology, year and type of surgery, chemotherapy and radiotherapy between the NCDR, HES and NLCA datasets.

Methods

Patients

We extracted data on 32,841 English lung cancer patients (ICD-10 C33-C34) diagnosed in 2008 from the NCDR dataset. From HES, 29,012 patients with a diagnosis of lung cancer in 2008 were identified. Data on 23,199 patients diagnosed with lung cancer (ICD-10 C33-C34) in England were extracted from the NLCA dataset.

We excluded 1,244 lung cancer patients identified from death certificate only (DCO) registrations in the NCDR and 221 patients who were diagnosed with carcinoid tumours (ICD-10 M8240). From the NLCA dataset, we excluded 97 patients with carcinoid tumours and 56 patients who had a primary diagnosis of mesothelioma (ICD-10 C45) in the NLCA dataset, but were classified as lung cancer diagnoses (ICD-10 C34) in the NCDR dataset.

Information on surgery, chemotherapy and radiotherapy of the patients identified in the NCDR was supplemented with information on these treatment from patient-linked HES records, thus giving rise to a single dataset (NCDR & HES) for comparison to NLCA with regard to histology and treatments.

We used the NHS number to link the NCDR, HES and NLCA datasets. Patients who were not identified in all three datasets were excluded, leaving 18,949 patients eligible for analysis of data comparability between the combined NCDR & HES and NLCA datasets.

Histology

Histology in the combined NCDR & HES dataset was classified using the ICD3/ICD4 digital codes (Appendix, Table A). In the NLCA dataset histology was classified using the ICD-O morphology

(Appendix, Table B). Lung cancer histology was categorised into non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) in both datasets.

Treatments

Treatment procedures in HES are defined according to codes from the Office of Population, Censuses and Surveys Classification of Surgical Operations and Procedures (4th revision) (OPCS-4). [3] Surgery information for patients in the combined NCDR & HES dataset was included from one month before to six months after the date of diagnosis. If patients had more than one recorded surgery type, the most extensive procedure was used in the analysis. We classified the surgical procedures into pneumonectomies and partial procedures (Appendix, Table C). Patients were defined as having received chemotherapy or radiotherapy if a relevant OPCS4 code was found from one month before to six months after the date of diagnosis.

In the NLCA dataset, surgical procedures with curative intent were recorded. We classified the surgical procedures in NLCA into pneumonectomies, partial procedures and mesothelioma-related procedures (Appendix, Table D). Chemotherapy treatment was recorded as chemotherapy alone, neo-adjuvant chemotherapy before surgery, part of chemotherapy/radiotherapy treatment plan, adjuvant chemotherapy post-surgery and induction chemotherapy to downstage before surgery. Radiotherapy was recorded as curative (radical) radiotherapy, curative (Continuous Hyperfractionated Accelerated Radiotherapy (CHART)/ CHART Weekend-less (CHARTWELL)), part of a chemotherapy/radiotherapy treatments plan, adjuvant radiotherapy following surgical treatment, palliative radiotherapy. Binary variables were created to indicate whether a patient received chemotherapy or radiotherapy.

Data analysis

A Venn diagram was constructed, where each circle represents one of the different datasets (NLCA, NCDR and HES) and the overlap of any of the three circles represents patients identified in more than one dataset.

We performed cross tabulations to assess the agreement on histology, year and type of surgery, chemotherapy and radiotherapy between the linked NCDR, HES and NLCA dataset.

Results

Figure 1 shows the Venn diagram of lung cancer registrations in the NCDR, HES and NLCA datasets. In the NCDR dataset, 31,376 lung cancer patients were identified, 92.2% of whom were also identified in HES and 63.7% of whom were also identified in NLCA. In NLCA, 23,046 lung cancer patients were identified, 86.8% of whom were also identified in the NCDR. Of 34,458 lung cancer patients identified in any of the three datasets, 91.1% were identified in the NCDR and 66.9% in NLCA. In HES, 34 patients were identified with a lung cancer diagnosis, who were not identified in the NCDR or NLCA datasets.

Further investigation of the 2,989 patients identified in the NLCA dataset but not in NCDR, revealed that most of these patients had a record in the NCDR but were excluded from the 2008 dataset because they were identified as a DCO registration (146), had an earlier lung cancer diagnosis (541), a morphology or site that was excluded (197), a record of a lung metastasis (130), cancer of a different site (920) or any other valid cancer registration exclusion reason (175). This meant that out of these NLCA-only patients 880 could not be identified in the 2008 NCDR, of which 472 were identified in the 2009 NCDR.

Table 1 show that 96.9% of patients who were recorded with a diagnosis of NSCLC in the combined NCDR & HES dataset were also recorded as such in NLCA. The percentage of patients recorded as being diagnosed with SCLC in both datasets was lower at 88.8%. Overall, agreement of recorded histology between both datasets was 95.9%. Of patients who were recorded with a diagnosis of NSCLC in the combined NCDR & HES dataset, 3.1% were recorded as having a diagnosis of SCLC in NLCA. Conversely, 11.2% of patients were recorded as having SCLC in the combined NCDR & HES dataset and as NSCLC in NLCA.

Table 2 shows that 98.1% of patients had no recorded surgery date in either dataset. Of 1,869 patients with a surgical procedure recorded in the combined NCDR & HES dataset in 2008 85.3% also had a record of surgery in 2008, 14.5% did not have a recorded surgery year, and four patients were recorded as having undergone surgery in 2009 in NLCA. Of 145 patients with a record of surgery in 2009 in the combined NCDR & HES dataset 80.0% also had a surgery record in 2009 in NLCA, 18.6% had no recorded surgery year and two patients were recorded as having undergone surgery in 2008 in NLCA. One patient was recorded in NLCA as having undergone surgery in 2007 and was not recorded with a surgery date in the combined NCDR & HES dataset. Overall, concordance of recorded surgery year between both datasets was 96.7%.

Concordance of recorded type of surgical procedure is shown in table 3 and was found to be 70.1% overall. Of 1,782 patients recorded with partial surgery in the combined NCDR & HES dataset, 76.1% also had a record of a partial surgery, 23.3% did not have a recorded surgery, and nine patients were recorded as having undergone pneumonectomy in NLCA. Of 232 patients with a record of pneumonectomy in the combined NCDR & HES dataset, 65.9% had a pneumonectomy, 21.6% had no recorded surgery and 11.2% had partial surgery in NLCA

Tables 4 and 5 show the concordance of treatment with chemotherapy and radiotherapy between the combined NCDR & HES dataset and the NLCA datasets. The concordance was very similar for both at 80.7% for chemotherapy and 80.2% of radiotherapy. However, 47.9% of patients that had a record of having received chemotherapy in the combined NCDR & HES dataset were not recorded as such in NLCA and 7.1% of patients who had a record of receiving chemotherapy in NLCA were not recorded as such in the combined NCDR & HES dataset. The corresponding figures for radiotherapy were 58.6% and 17.6%.

Key findings and conclusion

Of all patients identified in the NCDR, HES and NLCA datasets combined, 91.1% has a record in NCDR and 66.9% has a record in NLCA. It is striking that NLCA identifies 2,989 patients who are not recorded in the NCDR, whereas only 63.7% of patients identified in the NCDR are also identified in the NLCA.

The majority of patients identified by NLCA but not in the NCDR have a registered lung cancer diagnosis in NCDR in a different year, or with a different morphology or site, or are suspected to have a lung metastasis instead of a primary lung cancer. Further enquiry into patients with a cancer diagnosis of a different site or suspicion of lung metastasis is on-going.

Of patients identified in the combined NCDR & HES and NLCA dataset, overall concordance of histological type was high. Of recorded treatments, the concordance of information on year and type of lung cancer surgery between the combined NCDR & HES dataset and NLCA is high.

Concordance of the recording of chemotherapy and radiotherapy is relatively poor and probably reflects the underreporting of these treatments in the combined NCDR & HES dataset and NLCA. This could be due to the different ways in which the data are collected. The cancer registries collect data on treatment, but the completeness in the NCDR is poor. Supplementation through record linkage to HES improves the data completeness to some extent, however, since linkage is only to HES in-patient and day case records, data relating to chemotherapy and radiotherapy received by patients during out-patient appointments will be missed. Although treatment data are collected in NLCA, information regarding treatment is generally collected in the MDT meetings and therefore it may not cover all chemotherapy and radiotherapy given to patients.

Identification of all lung cancer patients in NCDR is high, and although NLCA has lower case ascertainment, it has been shown to improve. This report shows that the data quality and completeness of treatments given to lung cancer patients, in particular with respect to chemo- and radiotherapy, gives reason to exercise caution when interpreting any analysis incorporating these data.

References

1. Moller H, Richards S, Hanchett N, Riaz SP, Lüchtenborg M, Holmberg L, et al. Completeness of case ascertainment and survival time error in English cancer registries: impact on 1-year survival estimates. *Br J Cancer*. 2011; 105:170-6.
2. National lung cancer audit report 2011. Leeds 2011.
3. NHS. OPCS-4 Classification. 2012. Available at: <http://www.connectingforhealth.nhs.uk/systemsandservices/data/clinicalcoding/codingstandards/opcs4>. Accessed 7/01/12.

Tables & Figures

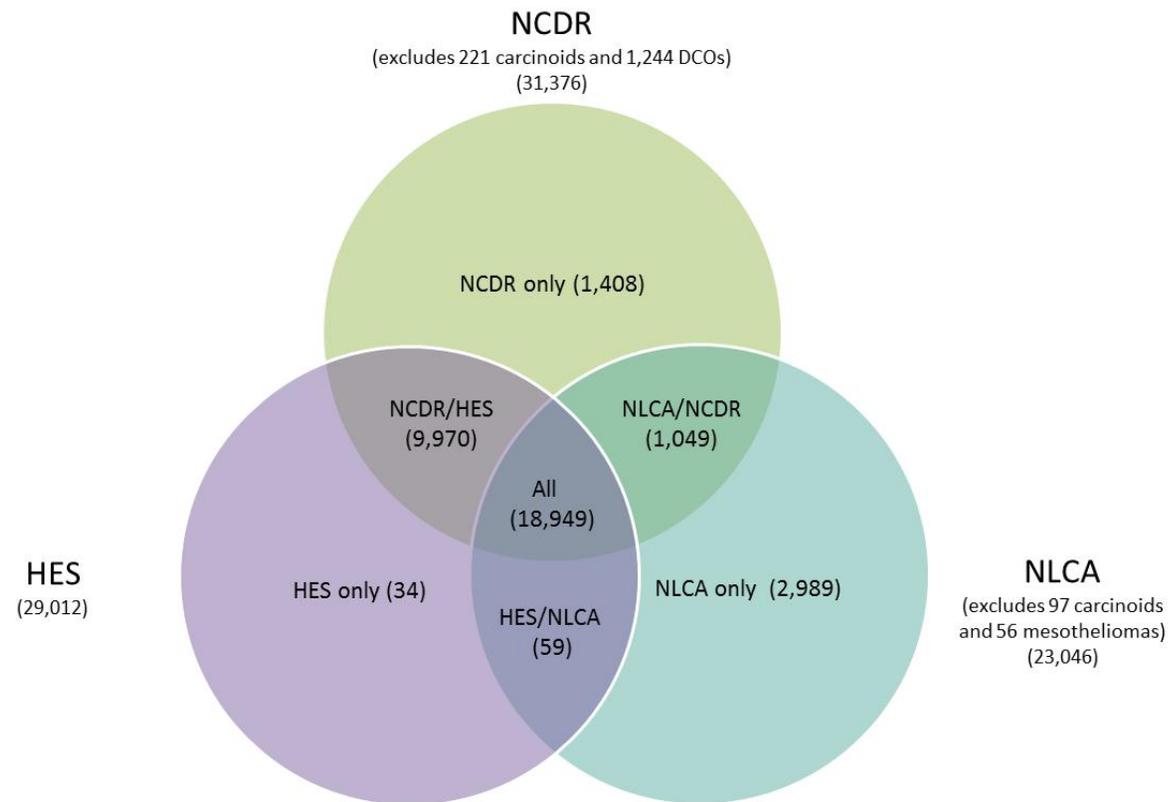


Figure 1: Venn diagram showing the number of lung cancer patients recorded in the National Cancer Data Repository (NCDR) (excluding death certificate only (DCOs) and carcinoids), Hospital Episodes Statistics (HES) and the National Lung Cancer Audit (NLCA) datasets (excluding carcinoids and mesotheliomas), 2008, England.

Table 1: Number and proportion of patients by histology in the combined NCDR & HES and the NLCA datasets

NLCA		NCDR & HES			
		NSCLC		SCLC	
		N	%	N	%
	NSCLC	16,027	96.9	511	11.2
	SCLC	271	3.1	2,140	88.8
	Total	16,298	100	2,651	100

Table 2: Number and proportion of patients by year of surgery in the combined NCDR & HES and the NLCA datasets

NLCA	Year of surgery	NCDR & HES					
		Not recorded		2008		2009	
		N	%	N	%	N	%
	Not recorded	16,618	98.1	271	14.5	27	18.6
	2007	1	0.0	0	0.0	0	0.0
	2008	301	1.8	1,594	85.3	2	1.4
	2009	15	0.1	4	0.2	116	80.0
	Total	16,935	100	1,869	100	145	100

Table 3: Number and proportion of patients by type of surgical procedure in the combined NCDR & HES and the NLCA datasets

NLCA	Surgical procedure	NCDR & HES					
		Not recorded		Partial		Pneumonectomy	
		N	%	N	%	N	%
	Not recorded	126	39.7	416	23.3	50	21.6
	Mesothelioma	44	13.9	1	0.1	3	1.3
	Partial	142	44.8	1,356	76.1	26	11.2
	Pneumonectomy	5	1.6	9	0.5	153	65.9
	Total	317	100	1,782	100	232	100.0

Table 4: Number and proportion of patients by chemotherapy in the combined NCDR& HES and the NLCA datasets

NLCA		NCDR & HES			
		Chemotherapy		No chemotherapy	
		N	%	N	%
	Chemotherapy	2,949	52.1	944	7.1
	No chemotherapy	2,708	47.9	12,348	92.9
	Total	5,657	100	13,292	100

Table 5: Number of patients by radiotherapy in the combined NCDR & NLCA and the NLCA dataset

NLCA		NCDR & HES			
		Radiotherapy		No radiotherapy	
		N	%	N	%
	Radiotherapy	412	41.4	3,163	17.6
	No radiotherapy	584	58.6	14,790	82.4
	Total	996	100	17,953	100

Appendix

Table A: Histology codes used in NCDR (ICD3/ICD4 digital code).

Description	Code
NSCLC	
Neoplasm, malignant	8000
Tumour cells, malignant	8001
Malignant tumour, small cell type	8002
Malignant tumour, giant cell type	8003
Malignant tumour, fusiform cell type	8004
Carcinoma NOS	8010
Epithelioma, malignant	8011
Large cell carcinoma NOS	8012
Large cell neuroendocrine carcinoma	8013
Carcinoma, undifferentiated NOS	8020
Carcinoma, anaplastic type NOS	8021
Pleomorphic carcinoma	8022
Giant cell and spindle cell carcinoma	8030
Giant cell carcinoma	8031
Spindle cell carcinoma	8032
Pseudosarcomatous carcinoma	8033
Polygonal cell carcinoma	8034
Tumorlet	8040
Non-small cell carcinoma	8046
Papillary carcinoma NOS	8050
Papillary squamous cell carcinoma	8052
Squamous cell carcinoma NOS	8070
Squamous cell carcinoma, keratinising NOS	8071
Squamous cell carcinoma, large cell, non-keratinising	8072

Squamous cell carcinoma, small cell, non-keratinising	8073
Squamous cell carcinoma, spindle cell	8074
Adenoid squamous cell carcinoma	8075
Squamous cell carcinoma, microinvasive	8076
Lymphoepithelial carcinoma	8082
Basaloid squamous cell carcinoma	8083
Basaloid carcinoma	8123
Adenocarcinoma NOS	8140
Scirrhous adenocarcinoma	8141
Superficial spreading adenocarcinoma	8143
Adenocarcinoma, interstitial type	8144
Carcinoma, diffuse type	8145
Monomorphic adenoma	8146
Cholangiocarcinoma	8160
Adenoid cystic carcinoma	8200
Cribriform carcinoma	8201
Tubular adenocarcinoma	8211
Solid carcinoma NOS	8230
Carcinoid tumour NOS (except of appendix M8240/1)	8240
Carcinoid tumour, argentaffin, malignant	8241
Goblet cell carcinoid	8243
Composite carcinoid	8244
Tubular carcinoid	8245
Neuroendocrine carcinoma	8246
Atypical carcinoid tumour	8249
Bronchiolo-alveolar adenocarcinoma	8250
Alveolar adenocarcinoma	8251
Bronchio-alveolar carcinoma, non-mucinous	8252
Bronchio-alveolar carcinoma, mucinous	8253
Bronchio-alveolar carcinoma, mixed mucinous and non-mucinous	8254
Adenocarcinoma with mixed sub-types	8255
Papillary adenocarcinoma NOS	8260

Adenocarcinoma in tubulovillous adenoma	8263
Oxyphilic adenocarcinoma	8290
Clear cell adenocarcinoma NOS	8310
Granular cell carcinoma	8320
Mixed cell adenocarcinoma	8323
Adrenal cortical carcinoma	8370
Mucoepidermoid carcinoma	8430
Cystadenocarcinoma NOS	8440
Mucinous cystadenocarcinoma NOS	8470
Mucinous adenocarcinoma	8480
Mucin-producing adenocarcinoma	8481
Signet ring cell carcinoma	8490
Lobular carcinoma NOS	8520
Acinar cell carcinoma	8550
Adenosquamous carcinoma	8560
Epithelial-myoepithelial carcinoma	8562
Adenocarcinoma with squamous metaplasia	8570
Adenocarcinoma with spindle cell metaplasia	8572
Adenocarcinoma with neuroendocrine differentiation	8574
Metaplastic carcinoma NOS	8575
Malignant melanoma NOS	8720
Sarcoma NOS	8800
Spindle cell sarcoma	8801
Giant cell sarcoma (except of bone M9250/3)	8802
Small cell sarcoma	8803
Epithelioid sarcoma	8804
Fibrosarcoma NOS	8810
Fibromyxosarcoma	8811
Solitary fibrous tumour, malignant	8815
Fibrous histiocytoma, malignant	8830
Liposarcoma NOS	8850
Leiomyosarcoma NOS	8890

Angiomyosarcoma	8894
Rhabdomyosarcoma NOS	8900
Pleomorphic rhabdomyosarcoma	8901
Adenosarcoma	8933
Mixed tumour, malignant NOS	8940
Rhabdoid sarcoma	8963
Pulmonary blastoma	8972
Carcinosarcoma NOS	8980
Synovial sarcoma NOS	9040
Teratoma, malignant NOS	9080
Choriocarcinoma NOS	9100
Haemangiosarcoma	9120
Haemangioendothelioma, malignant	9130
Epithelioid haemangioendothelioma, malignant	9133
Lymphangiosarcoma	9170
Osteosarcoma NOS	9180
Mesenchymal chondrosarcoma	9240
Peripheral neuroectodermal tumour	9364
Primitive neuroectodermal tumour	9473
Neurofibroma	9540
Missing	
SCLC	
Small cell carcinoma NOS	8041
Oat cell carcinoma	8042
Small cell carcinoma, fusiform cell	8043
Small cell carcinoma, intermediate cell	8044
Small cell-large cell carcinoma	8045

Table B: Histology codes used in NLCA (ICD-O morphology)

Description	Code
NSCLC	
Carcinoma in situ	M 8010/2
Large cell Carcinoma NOS	M 8012/3
Large cell neuroendocrine (ICD-O-3)	M 8013/3
Large cell - undifferentiated	M 8020/3
Non-small cell carcinoma (ICD-O-3) (includes adenosquamous carcinoma)	M 8046/3
Squamous cell carcinoma NOS	M 8070/3
Adenocarcinoma NOS (Adenocarcinoma without alveolar cell features)	M 8140/3
Bronchio-alveolar cell carcinoma (Adenocarcinoma with alveolar cell features)	M 8250/3
Carcinosarcoma NOS	M 8980/3
Mixed tumour (malignant)	M 8940/3
Other	M 9999/9
Missing	
SCLC	
Small cell carcinoma	M 8041/3

Table C: Classification of OPCS-4 codes for main surgical procedures in HES

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

Table D: Classification of OPCS-4 codes for main surgical procedures in NLCA

Surgical Procedure	Code
Pneumonectomy	
Pneumonectomy	E54.1
Partial	
Carinal resection	E44.1
Bilobectomy	E54.2
Lobectomy	E54.3
Wedge resection of lesion of lung (segment)	E54.4A
Segmental resection	E54.4B
Multiple wedges resected	E54.8A
Sleeve resection	E54.8B
Lung resection with resection of chest wall (not identifying which lobe resection)	E54.8 + T01
Segmental resection	E55_4B
Open operation on lung (open and close) (Incision of lung neck)	E57.4
Other open operation on lung	E57.8
Mesothelioma	
Extrapleural pneumonectomy	01
Debulking pleurectomy	02
Pleurodesis	03