Original Article

Agreement between the Danish Cancer Registry and the Danish Lung Cancer Registry

Jane Christensen[1], Anne Mette Tranberg Kejs[2], Lise Kristine Højsgaard Schmidt[3], Jes Søgaard[2], Margit Caroline Rasted[3], Ole Andersen[1] & Erik Jakobsen[4]

1) Danish Cancer Society Research Center, Danish Cancer Society, 2) Documentation and Quality, Danish Cancer Society, 3) Data Quality and Documentation, Danish Health Data Authority, 4) Department of Thoracic Surgery, Odense University Hospital, Denmark

ABSTRACT

Introduction: The Danish Cancer Registry (DCR) and the Danish Lung Cancer Registry (DLCR) are nation-wide registries recording Danish patients with lung cancer (LC). The aim of this study was to assess data agreement and possible consequences hereof on estimation of survival between patients in the two registries.

Methods: Descriptive statistics were used for comparison of registered patients in 2013-2014 in the DCR and the DLCR. Furthermore, the one-year relative survival (1y-RS) and Cox proportional mortality hazard rates (MRR) were calculated.

Results: In 2013-2014, a total of 9,111 Danish residents were identified with LC in the DCR and 9,316 were found in the DLCR. Merging the two registries showed an agreement of 87%, whereas 6% were included only in the DCR and 8% only in the DLCR. Including patients only registered in one registry, but who seemed to meet the inclusion criteria of both registries, would increase the agreement to 95%. No differences were seen for 1y-RS. However, MRR for patients in the DLCR was significantly lower than for patients in the DCR: 0.94 (95% confidence interval: 0.91-0.98).

Conclusions: Surprisingly, the DCR registered fewer patients in 2013-2014 than the DLCR, even though they employ the same primary data source. The agreement between the DCR and the DLCR was 87%; this may be increased to 95% if patients who seemed to meet the inclusion criteria of the other register were also included. The discrepancies found were mainly due to different definitions of dates of diagnosis, registrations probably missed by the algorithms and possible registration errors. Discrepancies resulted in a significant difference in MRR, but not in 1y-RS.

Funding: none.
Trial registration: not relevant.

Lung cancer (LC) is the second most incident cancer in Denmark [1]. Furthermore, the incidence and mortality of LC in Denmark are higher than in the other Nordic countries [2]. To remedy this situation, reliable and up-to-date statistics on incidence and mortality are required. It is therefore important that the Danish registries are complete and validated.

The Danish Cancer Registry (DCR), established in 1943, is one of the world's oldest cancer registries. It is population-based and contains information about all incident cancers and pre-stages. The Danish Lung Cancer Registry (DLCR) is population-based and registers the outcome of LC patients diagnosed and/or treated in Danish departments specialising in pulmonary diseases.

After shifting from manual registration to automated capture in 2004, the DCR was compared to the DLCR and the Danish Breast Cancer Cooperative Group (DBCG) [3]. For patients diagnosed in 2006, the agreement between the DCR and the DLCR was 86%; and for the DCR and the DBCG, the corresponding percentage was 92%. Recently, the DCR was compared to the Danish Colorectal Cancer Group Database for patients registered in 2014-2015, and the agreement was 86% [4].

The Danish registries are widely used by epidemiologists, clinicians and administrators for

estimation of incidence and mortality and for planning of prophylaxis, capacity for diagnostic work-up and treatment. Possible differences between registrations in the DCR and the DLCR may thus have important consequences.

The aim of this study was to compare the agreement of LC patients registered in the DCR and the DLCR and to evaluate any influence on estimates of survival.

METHODS

Data sources

Since 1968, all permanent residents in Denmark have been assigned a unique personal identification number (CPR). The CPR number is used in all national registries supporting inter-registry linkage. The Danish Civil Registration System includes information on gender, date of birth, place of residence, emigration, immigration, disappearance and vital status [5].

The Danish National Patient Register (NPR) is used for allocation of resources to the clinical departments and records information on diagnoses and treatment in all somatic Danish hospitals [6].

The Danish Pathology Register (DPR) collects information from all Danish departments of pathology. Data are coded according to the Danish SNOMED classification and include information on topography and morphology [7].

Data in the NPR and the DPR were used in this study to confirm the registrations in the DCR and the DLCR. Since 1943, the DCR has recorded all incident cancers in Denmark. The primary data sources are the NPR and death certificates, whereas the DPR is, to some extent, a supplementary source. The DCR holds information on tumour characteristics including diagnoses according to the International Classification of Diseases-7 and -10, morphology, topography, laterality, stage, grade and date of diagnosis [8]. From 2004 to 2008, the DCR changed from paper-based notifications to using an automated algorithm. The DCR captures 80-90% of information of all incident cancers automatically in the NPR. The final 10-20% is assessed manually and regards missing information on morphology or multiple primaries. The algorithm used is built on recommendations from the European Association of Cancer Registries and the European Network of Cancer Registries with some exceptions on multiple primaries and certain benign and precancerous lesions [9]. The DCR includes all cases of LC with the International Statistical Classification of Diseases and Related Health Problems, tenth Revision (ICD-10) diagnoses C33-C34 if also registered as new incident cases. The DCR defines the date of diagnosis as the date on which a Danish cancer pathway is initiated and in which LC is later registered for the first time.

Since 2000, the DLCR [10] has monitored and evaluated the quality of treatment of all registered Danish LC patients. Since 2003, cases have been identified by an algorithm from

the NPR which is the primary source of data. Patients are included if registered for the first time with a C33-C34 diagnosis subsequently verified by the notifying clinician. About 75% of the information registered is available in central Danish registries, and the clinicians are requested to provide about 25% from medical records and to verify data from the registries. Information included in the DLCR from the NPR regards surgical and oncological treatment, stage, and Charlson's Comorbidity Index. In the DLCR, the date of diagnosis is defined as the first registered date of the contact with the clinical department initiating the patient's trajectory, subsequently leading to a registered LC diagnosis. Death certificates are not used by the DLCR.

Permanent Danish residents registered with LC in the DCR and the DLCR in the 2013-2014 period were identified by their CPR number and included in this study. If a patient had been registered both in the DCR and the DLCR, the registrations were considered identical if the interval between the registered dates of diagnoses was ≤ 120 days.

Statistical methods

The merger of the DCR and the DLCR denotes the sum of patients in both registries, and the intersection denotes patients found in both registries. Patients registered in the DCR with identical dates of diagnosis and death are named 'possible case of death certificate only' ('possible DCOs').

One-year relative survival (1y-RS) was calculated using the method proposed by the International Cancer Survival [11]. Relative survival is defined as the observed survival divided by the expected survival. Observed survival was estimated by the actuarial method and expected survival by the Ederer II method [12]. Relative survival may be interpreted as survival if the only cause of death was LC.

Mortality rate ratio (MRR) was calculated by Cox proportional hazards models using time since diagnosis as the underlying timescale and gender and age at the time of diagnosis as strata. The underlying assumption of proportional hazards was evaluated and was not violated. The MRR may be interpreted as mortality of all causes.

Kaplan-Meier curves were used for graphical presentation of survival.

SAS statistical software (release 9.4) was used for all statistical analyses.

Trial registration: not relevant.

RESULTS

Patient cohort

In the 2013-2014 period, a total of 9,111 patients were identified in the DCR and 9,316 patients in the DLCR. When merging the DCR and the DCLR, 9,872 individual patients were found of

whom 8,555 were in the intersection yielding an agreement of 87% (**Table 1**).

TABEL 1 / Patients registered with lung cancer in the Danish Cancer Registry (DCR) and the Danish Lung Cancer Registry (DLCR) in the 2013-2014 period before and after merging the registries with registrations according to the match criterion of no more than 120 days between the registered dates of diagnosis.

	n (%)
Total DCR	9,111
Total DCLR	9,316
Intersection[a] of DCR and DLCR	8,555 (87)
Only found in DCR	556 (6)
Only found in DLCR	761 (8)
Total in merger[b]	9,872 (100)

a) Patients found in both registries.
b) Total number of individual patients in the DCR and the DLCR.

Among 556 patients only found in DCR (**Table 2**), 171 patients were registered in the DLCR, but outside the study period; 98 were possible DCOs. The inclusion criteria of the DLCR were met for 240 patients who were registered in NPR both with LC and with another type of cancer; among these patients 88 were also registered in the DCR with LC in or outside the study period, whereas 188 were not registered in the DPR – i.e. with unknown morphology. Of the remaining 47 patients, five were only found in the DPR and not in the NPR, and 42 were not registered either in the DPR or in the NPR with LC. Among the 98 possible DCOs, 43 were found in the DPR, and all were found in the NPR with a LC diagnosis.

TABEL 2 / Patients registered with lung cancer in the 2013-2014 period and only found in either the Danish Cancer Registry or the Danish Lung Cancer Registry after merging of the two registries.

	n
Patients only found in the DCR	
Registered in the DLCR with a lung cancer diagnosis before or after the study period	171
Possible DCOs with same registered date of diagnosis and death	
Hereof:	98
Registered in the NPR with a lung cancer diagnosis	98
Also registered in the DPR	43
Registered in the NPR with a lung cancer diagnosis, meeting the inclusion criteria of the DLCR	
Hereof:	240
Registered both with a lung cancer diagnosis and with another type of cancer than lung cancer in the study period	240
Not registered in the DPR with a lung cancer diagnosis in the study period	188
Registered in the DPR with a lung cancer diagnosis in the study period	52
Registered in the DPR with a lung cancer diagnosis outside the study period	36
Not registered in the NPR with a lung cancer diagnosis	
Hereof:	47
Registered in the DPR with a lung cancer diagnosis	5
Not registered in the DPR with a lung cancer diagnosis	42
Total	556
Patients only found in the DLCR	
Registered in the DCR with a lung cancer diagnosis before or after the study period	40
Registered in the NPR with a lung cancer diagnosis, meeting the inclusion criteria of the DCR	236
Hereof: registered in the DPR with a lung cancer diagnosis	211
Not registered in the NPR with a lung cancer diagnosis, but registered in the DPR	171
Not registered neither in the NPR nor in the DPR, and not indicated as a "clinical diagnosis" in the DLCR, and not meeting the inclusion criteria of the DCR	314
Total	761

DCO = possible case of death certificate only, registered with identical date of diagnosis and death; DCR = Danish Cancer Registry; DLCR = Danish Lung Cancer Registry; DPR = Danish Pathology Register; NPR = Danish National Patient Register.

Of the 761 patients with LC only found in the DLCR, 40 were also registered in the DCR, but outside the study period (Table 2). In the NPR, 236 patients were registered with LC, and 171 were only registered in the DPR with LC, but not in the DCR at all – either with LC or with

another cancer. Among the 236 patients registered in NPR with LC, 211 had a registration of LC in the DPR. The inclusion criteria of the DCR were fulfilled for these 407 (236 + 171) patients. Furthermore, 314 patients registered in the DLCR were not found in either the NPR or the DPR, and they were not in the DCR at all – either with LC or with another cancer type or premalignancy and are therefore unexplained.

If all patients only found in the DCR and/or in the DLCR who seemed to meet the inclusion criteria of both registries were included in the intersection, the agreement between the registries - disregarding the study period - would increase to 95%.

Relative survival

The 1y-RS for all patients in the DCR did not differ from the 1y-RS for all patients in the DLCR: 50% (95% confidence interval (CI): 48-52) versus 52% (95% CI: 50-54) (**Table 3**). No difference was found in 1y-RS when using the different dates of diagnosis of the two registries as entry dates for patients in the intersection of the DCR and the DLCR. Patients found only in the DCR had a significantly lower 1y-RS than all patients in the DCR. Likewise, patients found only in the DLCR had a significantly lower 1y-RS than all patients in the DLCR.

TABEL 3 / Survival estimates for patients registered with lung cancer in Danish Cancer Registry and Danish Lung Cancer Registry in 2013-2014.

	1-yr RS (95% CI)	Cox proportional hazard models		
		MRR1 (95% CI)	MRR2 (95% CI)	MRR3 (95% CI)
DCR	50 (48-52)	1.00 (ref.)		
DLCR	52 (50-54)	0.94 (0.91-0.98)		
Intersection of the DCR and the DLCRᵃ	52 (50-54)		1.00 (ref.)	
Intersection of the DCR and the DLCRᵇ	53 (51-54)			1.00 (ref.)
Only found in the DCR	22 (19-27)		2.15 (1.95-2.38)	2.21 (2.00-2.45)
Only found in the DLCR	39 (36-43)		1.09 (0.99-1.20)	1.11 (1.01-1.23)

CI = confidence interval; DCR = Danish Lung Cancer Registry; DCR = Danish Cancer Registry; MRR = mortality rate ratio; RS = relative survival.
a) DCR date of diagnosis used as date of entry.
b) DLCR date of diagnosis used as date of entry.

Cox proportional hazard models

The MRR was significantly lower for all patients in the DLCR than for all patients in the DCR, 0.94 (95% CI: 0.91-0.98) (Table 3). When comparing patients only found in one registry with patients in the intersection of the DCR and the DLCR, patients only found in the DCR had a

significantly higher mortality.

Kaplan-Meier curves

Survival curves for all patients in the DCR and the DLCR were similar, whereas curves for patients only found in the DCR or the DLCR had a lower survival than all patients in the DCR and all patients in the DLCR, confirming the estimations of survival and mortality (**Figure 1**).

FIGURE 1 / Kaplan-Meier curves for selected groups.
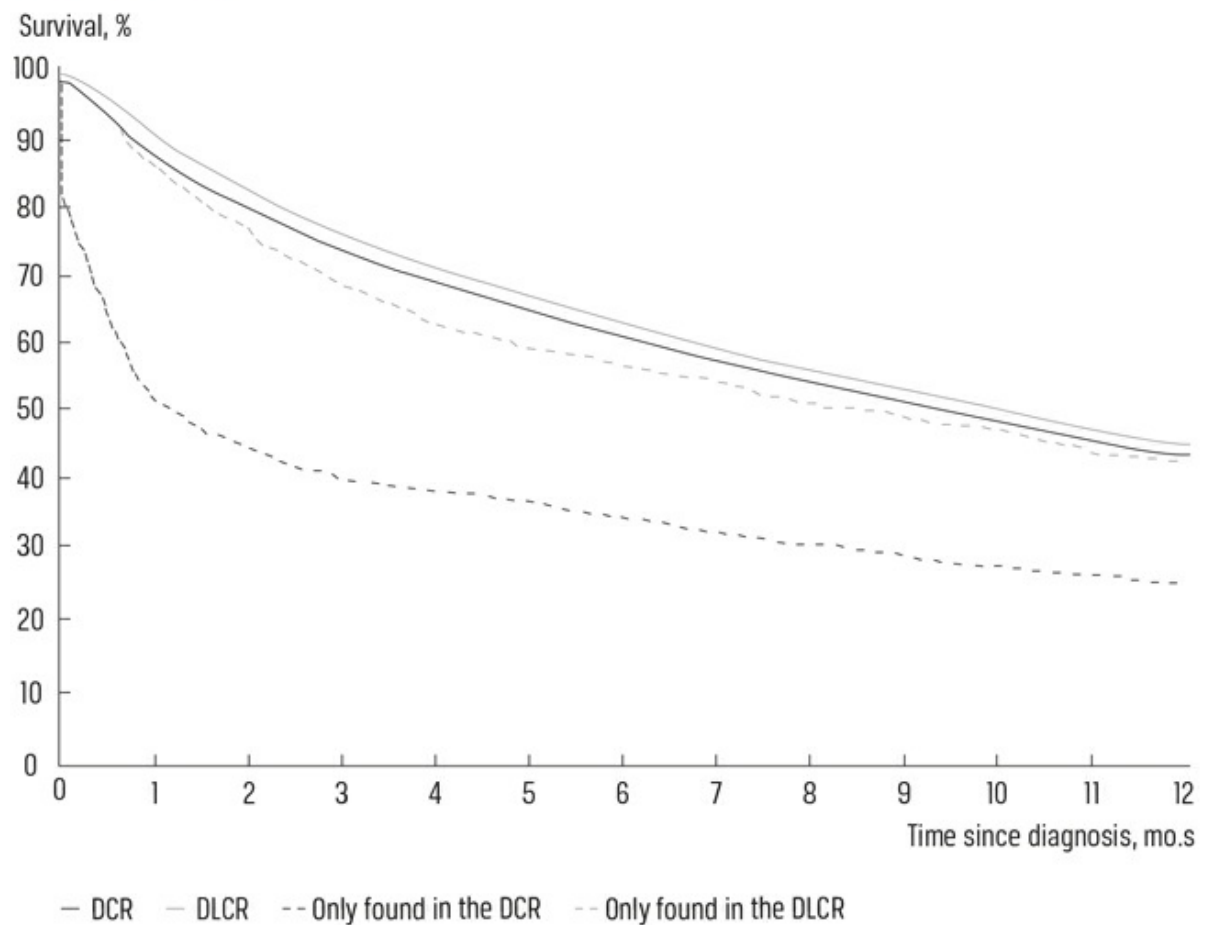
DISCUSSION

Agreement between the DCR and the DLCR in this study was 87%, which is a slight improvement compared with a previous study on data from 2006 where it was 86% [3]. The previous study was performed after the automatisation of registrations of the DCR in 2004; and in 2006, further changes were made to the automated algorithm. If all patients identified in only one registry, but probably meeting the inclusion criteria of both registries, were included in the intersection, the agreement would increase to 95%.

For patients only found in one registry, 211 (171 + 40) patients were actually found in the other registry, but before or after 120 days. The different definitions of dates of diagnosis but also any differences in registration practice in the clinical departments may mean that registrations in both registries are not quite up-to-date.

In the present study, 407 patients only found in the DLCR may have been candidates for inclusion in the DCR, and 240 patients only found in the DCR – according to registrations suffering from both LC and another type of cancer - may have been candidates for inclusion in the DLCR. The findings indicate that the capture algorithms have missed these patients. Alternatively, these patients may be under further investigation in the registries before inclusion. Such investigations may last several years. Either the DCR or the DLCR publishes the numbers of patients who are undergoing further investigation in their annual reports.

Both the DCR and the DLCR use the NPR as their primary data source, and therefore errors in the NPR could have implications for both registries. To our knowledge, no validation studies have verified diagnoses of LC in the NPR, but a study from 2015 reported a positive overall predictive value of 98% (95% CI: 89.4-99.9) for selected cancers [13]. The main discrepancies between registrations in the DCR and the DLCR are due to the different capture algorithms, different definitions of the dates of diagnosis or the use of slightly different data sources.

The finding that the number of registered patients in the DLCR was larger than the number in the DCR is unexpected, because the DLCR primarily registers patients in order to monitor the outcome of diagnostic work-up and treatment, whereas the DCR aims at registering all incident cancers. However, quite similar findings were found when the DCR was compared to the clinical database for patients with colorectal cancer which also had more registered cases of colorectal cancer than the DCR did [4]. Lastly, 42 patients in the DCR and 314 patients in the DLCR were not registered either in the NPR or in the DPR or in the other registry before or after the study period. These registrations must therefore be categorised as unexplained. For the DLCR, 314 patients corresponding to 3% of all patients in the DLCR, is a rather high number of unexplained cases.

Despite the differences between the registries, no significant differences were found in the estimated 1y-RS between patients registered in the DCR and the DLCR. The similar values of 1y-RS are in accordance with previous reports [14, 15]. The MRR was lower for patients registered in the DCLR than for patients in the DCR, which may reflect DCOs in the DCR. Also, 1y-RS was lower and MRR higher for patients found only in one registry than for the rest of the patients. This observation reflects that patients not registered in both registries are patients with a poor prognosis who may not have been offered complete diagnostic work-up or treatment.

To improve the agreement and completeness of the DCR and the DCLR, regular linkages

between the two registries is recommended to identify patients not captured by one of the registries. Such linkage would decrease the number of patients under further investigation. A change in the notification system so that the clinical departments needed to notify one registry only might produce a similar effect. Furthermore, the registries might consider publishing in their annual reports the number of patients undergoing further investigation. This would increase the transparency for users of the registries.

CONCLUSIONS

Agreement between the DCR and the DLCR was 87% and would have been increased to 95% if the registries had included patients found in one registry only but who seemed to meet the inclusion criteria of both. The discrepancies between the DCR and the DLCR were mainly due to registrations in the NPR that had been missed by the capture algorithms, different determinations of dates of diagnosis and the requirement for clinicians to report to both registries. Before patients are included in the DLCR, the LC diagnosis must be verified by the notifying clinician. Therefore, information in the DLCR may be different from information in the NPR records. The observed differences lead to differences in the estimated mortality rates, but not in 1y-RS, except for patients identified in one registry only. To improve the agreement and completeness of the DCR and the DCLR, regular linkages between the two registries are recommended.

LITERATURE

1. Danish Health Data Protection Agency. Nye Kræfttilfælde i Danmark 2014. Copenhagen: Danish Health Data Protection Agency, 2015.
2. Engholm G, Ferlay J, Christensen N et al. NORDCAN: cancer incidence, mortality, prevalence and survival in the Nordic countries, Version 8.1 (28.06.2018). Association of the Nordic Cancer Registries. Copenhagen: Danish Cancer Society, 2018.
3. Statens Serum Institut. Validation of the Danish Cancer Registry and selected clinical databases. Copenhagen: Statens Serum Institut,2012.
4. Christensen J, Hojsgaard Schmidt LK et al. Agreement between the Danish Cancer Registry and the Danish Colorectal Cancer Group Database. Acta Oncol 2020;59:116-23.
5. Pedersen CB. The Danish Civil Registration System. Scand J Public Health 2011;39:22-5.
6. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. Scand J Public Health 2011;39:30-

3.

7.   Bjerregaard B, Larsen OB. The Danish Pathology Register. Scand J Public Health 2011;39:72-4.

8.   Gjerstorff ML. The Danish Cancer Registry. Scand J Public Health 2011;39:42-5.

9.   Jensen OM SH, Jensen HS. Cancer registration in Denmark and the study of multiple primary cancers, 1943-80. Natl Cancer Inst Monogr 1985;68:245-51.

10. Jakobsen E, Green A, Oesterlind K et al. Nationwide quality improvement in lung cancer care: the role of the Danish Lung Cancer Group and Registry. J Thoracic Oncol. 2013;8:1238-47.

11. Engholm G, Gislum M, Bray F et al. Trends in the survival of patients diagnosed with cancer in the Nordic countries 1964-2003 followed up to the end of 2006. Material and methods. Acta Oncol 2010;49:545-60.

12. Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. Biometrics 1982;38:933-42.

13. Thygesen SK, Christiansen CF, Christensen S et al. The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. BMC Med Res Methodol 2011;11:83.

14. Esundhed.dk (1 Mar 2020).

15. DMCG.dk Benchmarking II Consortium. Uddybende rapport om canceroverlevelse i Danmark 1995-2014 (1 Mar 2020)