ORIGINAL RESEARCH

Comparing causes of death of Hodgkin lymphoma and breast cancer patients between medical records and cause-of-death statistics

Simone de Vries[1]
Michael Schaapveld[1]
Jan WPF Kardaun[2,3]
Kim H de Bruin[2]
Augustinus DG Krol[4]
Pieternella J Lugtenburg[5]
Judy N Jacobse[1]
Berthe MP Aleman[6]
Flora E van Leeuwen[1]

[1]Department of Epidemiology, The Netherlands Cancer Institute, Amsterdam, the Netherlands; [2]Department of Health and Care, Statistics Netherlands, The Hague, the Netherlands; [3]Department of Public Health, Academic Medical Center, Amsterdam, the Netherlands; [4]Department of Radiotherapy, Leiden University Medical Center, Leiden, the Netherlands; [5]Department of Hematology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands; [6]Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, the Netherlands

Objective: Obtaining accurate data about causes of death may be difficult in patients with a complicated disease history, including cancer survivors. This study compared causes of death derived from medical records (COD_{MR}) with causes of death derived from death certificates (COD_{DC}) as processed by Statistics Netherlands of patients primarily treated for Hodgkin lymphoma (HL) or breast cancer (BC).

Methods: Two hospital-based cohorts comprising 1,215 HL patients who died in the period 1980–2013 and 714 BC patients who died in the period 2000–2013 were linked with cause-of-death statistics files. The level of agreement was assessed for common underlying causes of death using Cohen's kappa, and original death certificates were reviewed when COD_{DC} and COD_{MR} showed discrepancies. We examined the influence of using COD_{DC} or COD_{MR} on standardized mortality ratio (SMR) estimates.

Results: Agreement for the most common causes of death, including selected malignant neoplasms and circulatory and respiratory diseases, was 81% for HL patients and 97% for BC patients. HL was more often reported as COD_{DC} (COD_{DC}=33.1% vs. COD_{MR}=23.2%), whereas circulatory disease (COD_{DC}=15.6% vs. COD_{MR}=20.9%) or other diseases potentially related to HL treatment were more often reported as COD_{MR}. Compared to SMRs based on COD_{DC}, SMRs based on COD_{MR} complemented with COD_{DC} were lower for HL and higher for circulatory disease.

Conclusion: Overall, we observed high levels of agreement between COD_{MR} and COD_{DC} for common causes of death in HL and BC patients. Observed discrepancies between COD_{MR} and COD_{DC} frequently occurred in the presence of late effects of treatment for HL.

Keywords: cause of death, Hodgkin lymphoma, breast cancer, mortality statistics

Introduction

In epidemiology, differences in mortality risks between patient cohorts and the general population are usually quantified using standardized mortality ratios (SMRs). SMR estimation requires the availability of age-, sex-, and period-specific mortality rates for the general population. Often, these mortality rates are available from nationwide registries. It is therefore important that mortality data from official death certificates reflect similar underlying cause-of-death patterns as cause-of-death information from medical records. Inconsistencies in processing of causes of death may have implications for the interpretation of SMRs in epidemiologic studies. Only few studies compared cause-of-death information from medical records or hospital discharge records with death certificates.[1–5]

In recent years, it has become challenging to obtain cause-of-death information in cancer survivor cohorts, especially in case of mortality long after initial cancer

Correspondence: Flora E van Leeuwen
Department of Epidemiology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, the Netherlands
Tel +31 20 512 2483
Fax +31 20 512 2322
Email f.v.leeuwen@nki.nl

treatment. Patients are more often cured and discharged, may have changed addresses, and may have died in another hospital, hospice or at home. Furthermore, due to the increasing administrative burden on general practitioners (GPs), response rates of clinical follow-up questionnaires have dropped considerably. To complete information on causes of death, linkage with a national cause-of-death registry can be considered. However, accurate registration of causes of death may be more difficult in patients with a long interval between disease diagnosis and date of death or a complicated disease history. For example, Hodgkin lymphoma (HL) patients are generally young at the time of treatment and have an excellent prognosis. Yet, they have an increased risk to die from various late adverse events related to treatment, including among others, second malignancies, cardiovascular diseases, and infections.[6,7] These late complications may affect cause-of-death coding in these patients. In contrast, breast cancer (BC) patients also have a good prognosis but a lower risk of adverse effects of treatment,[8] which may lead to less inconsistency in cause-of-death coding. For the interpretation of epidemiological studies on cause-specific mortality, it is important to gain insight into the effects of using cause-of-death information from different sources in different populations of cancer survivors.

In this study, we therefore compared causes of death recorded in the medical files with cause-of-death information from death certificates as processed by the Dutch cause-of-death statistics (Statistics Netherlands) in patients treated for HL or BC. We also examined to what extent differences in classification of causes of death influence SMRs.

Methods

We used data from a hospital-based cohort of HL patients (n=3,019), treated in the Antoni van Leeuwenhoek hospital, Erasmus MC Cancer Institute or Leiden University Medical Center, between 1965 and 2000, before the age of 51 years, and data from a hospital-based cohort of female BC patients (n=5,600) treated in the Antoni van Leeuwenhoek hospital for stages I–III invasive BC between 2000 and 2009, before the age of 61 years. Patient selection and data collection have been described in detail previously.[6,9–11,14] For the comparison of causes of death, we considered all 1,215 HL patients who died in the period 1980–2013 and all 714 BC patients who died in the period 2000–2013.

Causes of death for both HL and BC patients were obtained from hospital medical records by trained abstractors. If abstractors had difficulties in assigning a cause of death, the patient was discussed within the project team (MS,

BMPA, FEvL, SdV, JNJ) until agreement was reached. If the cause of death could not be retrieved from the hospital charts, a questionnaire was sent to the patient's GP regarding the patient's medical history and direct and underlying causes of death. Response rates from the GPs were 94% in 2004[11,12] and 83% in 2012 for the HL cohort[13] and 71% in 2016 for the BC cohort.[14] Exact dates of death were confirmed through linkage with the Central Office of Genealogy in order to allow deterministic linkage of all cohort members with Statistics Netherlands. This study was exempt from the approval of institutional review board according to Dutch law because existing data from medical files were used, which was confirmed by the institutional review board of the Netherlands Cancer Institute (PTC17.0788). Questionnaires obtained from the GPs were anonymized for the researchers, as well as the data retrieved from the hospital medical records. All analyses were performed at Statistics Netherlands: no cause-of-death registry data were exchanged. Original death certificates at Statistics Netherlands were reviewed by SdV only under a confidentiality agreement.

In the Netherlands, the underlying cause of death for each decedent is registered by Statistics Netherlands. Attending physicians or medical examiners are obliged to fill out a death certificate, which is sent to Statistics Netherlands through the municipality where the death occurred. At Statistics Netherlands, the death certificates are coded and an underlying cause is selected according to the tenth revision of the ICD (ICD-10), with a maximum of four coded diseases per decedent. The ICD-10 defines the underlying cause of death as 1) the disease or injury that initiated the train of morbid events leading directly to death or 2) the circumstances of the accident or violence that produced the fatal injury, with Volume 2 of ICD-10 providing several instructions for the application of this definition.[15] Detailed information on methods of registration of causes of death in the Netherlands has been published previously.[16,17]

The hospital-based cohorts of HL and BC patients were linked with the causes of death at Statistics Netherlands to enable comparison of the causes of death derived from the medical records (COD_{MR}) with causes of death derived from the death certificates (COD_{DC}). Individual-level linkage was based on sex, date of birth, date of death, and filing municipality (first linkage step). If linkage failed due to missing information on filing municipality in the cohort data, linkage was based on the remaining three variables (second linkage step). Additionally, we performed a third linkage procedure where we varied parts of the linkage key (eg, day of death or day of birth). By chance, it is possible that two individuals

have an identical linkage key (0.05% of total on the first linkage step and 1.3% on the second linkage step in the period 1980–2013). However, we checked for duplicate matches in both cohorts on the first two linkage keys and found none. Linkage was successful for 1,140 deceased patients (93.8%) in the HL cohort and 690 deceased patients (96.6%) in the BC cohort. The difference in successful linkages can be understood when realizing that deaths within the HL cohort partly occurred in earlier decades.

Statistical analyses

We compared the COD_{MR} and the COD_{DC} for common causes of death according to the ICD-10, including malignant neoplasms and circulatory disease, for all patients who were successfully linked with Statistics Netherlands. The causes of death of patients for whom the cause of death in our cohort was missing (unable to retrieve from the medical records and no response from the GP), or unspecified (ICD-10 codes R96–R99) were categorized as unknown or unavailable, including for 373 deceased HL patients (32.2%) and 113 deceased BC patients (16.4%). We calculated the conditional agreement between the COD_{MR} and the COD_{DC} in patients for whom the cause of death was available from both sources, thus excluding patients with unknown or unavailable cause of death. Conditional agreement for medical records was defined as the number of deceased patients for whom COD_{MR} was in agreement with COD_{DC} divided by the total number of deaths from a specific cause according to COD_{MR} and vice versa for the conditional agreement for death certificates. Overall agreement between COD_{MR} and COD_{DC} was calculated by dividing the number of patients for whom the causes of death from COD_{MR} and COD_{DC} agreed by the total number of deceased patients. To adjust for chance agreement, the Cohen's kappa statistic was calculated. Kappa values of >0.75 were considered to represent excellent agreement, values of 0.40–0.75 represented moderate agreement, and values of <0.40 represented poor agreement.[18]

The original death certificates were reviewed when COD_{MR} differed from COD_{DC}. Discrepancies were examined when HL was coded as COD_{MR} but not as COD_{DC}, or vice versa, in the HL cohort. Similarly, discrepancies were further examined when BC was coded as COD_{MR} and but not as COD_{DC}, or vice versa, in the BC cohort. Additionally, discrepancies between COD_{MR} and COD_{DC} were investigated when in the HL cohort, both COD_{MR} and COD_{DC} concerned a common malignancy as cause of death (including lung cancer, gastrointestinal cancer, non-HL and leukemia) or when COD_{MR} was recorded as a common malignancy and COD_{DC}

specified circulatory disease, or vice versa. For examination of such discrepancies, only original death certificates filed between 1980–1989 and 2000–2013 were accessible.

Finally, we assessed whether and to what extent discrepancies in causes of death influenced the results of person-year analyses of cause-specific mortality. We compared the observed mortality in our HL cohort and mortality in the general population, taking into account the person-years of observation of patients in our cohort by age, sex, and calendar period. Mortality data for the general population from Statistics Netherlands for the period 1980–2013 were used as reference rates. In this analysis, time at risk began at start of treatment and ended at date of death, date of emigration, or December 31, 2013, whichever came first. SMRs were calculated as the ratio of the observed and expected number of deaths in our HL cohort, and the confidence limits were calculated using exact Poisson probabilities of observed numbers. The absolute excess risk (AER) was calculated as the observed minus the expected number of deaths divided by person-years at risk, expressed per 10,000 person-years. First, we performed the analysis using COD_{MR} supplemented by COD_{DC} if COD_{MR} was missing. Subsequently, we performed the analysis using only COD_{DC}. Lastly, we performed the analysis in the same way as the first analysis, and we additionally supplemented COD_{MR} with COD_{DC} if the data between the two sources showed discrepancy.

Results

In total, we included 1,140 deceased HL patients and 690 deceased BC patients for whom linkage with Statistics Netherlands could be performed. Table 1 shows the descriptive characteristics of both patient groups. The majority of deceased HL patients were male (61.8%). Median age at primary cancer diagnosis was 30.4 years (interquartile range [IQR]=23.7–38.6) for HL patients and 50.3 years (IQR=42.7–55.6) for BC patients. For HL patients, the median interval between HL treatment and death was 17.4 (IQR=8.0–27.0), and for BC patients, this interval was 4.3 years (IQR=2.4–6.8). Cause of death was available from both the medical records and death certificates for 767 HL patients and 577 BC patients.

The level of agreement between medical records and death certificates for common causes of death in HL patients (including HL; breast, lung, and gastrointestinal cancers; non-HL and leukemia; other malignant neoplasms; circulatory disease; and respiratory disease) is shown in Table 2. Using either COD_{MR} or COD_{DC}, the majority of patients died from a malignant neoplasm (66.5% and 75.9%,

Table 1 Descriptive characteristics of deceased HL and BC patients

Characteristics	HL, n (%)	BC, n (%)
Total cohort	1,140 (100)	690 (100)
Females in cohort	435 (38.2)	690 (100)
Calendar period of cancer diagnosis		
1965–1979	598 (52.5)	
1980–1989	380 (33.3)	
1990–2000	162 (14.2)	
2000–2009		690 (100)
Age at diagnosis of cancer, median (IQR)	30.4 (23.7–38.6)	50.3 (42.7–55.6)
Age at death, median (IQR)	50.3 (38.8–59.7)	54.9 (47.4–61.0)
Time interval between cancer diagnosis and death, median (IQR)	17.4 (8.0–27.0)	4.3 (2.4–6.8)
Calendar period of death		
1980–1989	288 (25.3)	
1990–1999	336 (29.5)	
≥2000	516 (45.3)	690 (100)

Abbreviations: HL, Hodgkin lymphoma; BC, breast cancer; IQR, interquartile range.

Table 2 Level of agreement between causes of death from medical records and death certificates in HL patients

Cause of death	Number of deaths according to medical records (%)	Number of deaths according to death certificates (%)	Conditional agreement in medical records[a] (%)	Conditional agreement in death certificates[b] (%)	Cohen's kappa
Malignant neoplasm	510 (66.5)	582 (75.9)	96.5	84.5	0.66
Gastrointestinal cancer	70 (9.1)	76 (9.9)	88.6	81.6	0.83
Lung cancer	88 (11.5)	93 (12.1)	84.1	79.6	0.79
BC	26 (3.4)	31 (4.0)	100	83.9	0.91
HL	178 (23.2)	254 (33.1)	89.9	63.0	0.64
Circulatory disease	160 (20.9)	120 (15.6)	63.8	85.0	0.67
Ischemic heart disease	45 (5.9)	56 (7.3)	57.8	46.4	0.48
Other heart diseases	83 (10.8)	24 (3.1)	18.1	62.5	0.24

Notes: Cause of death according to the ICD-10: malignant neoplasms (C00–C97), gastrointestinal cancer (C15–C26), lung cancer (C33–C34), BC (C50), HL (C81), circulatory disease (I00–I99), ischemic heart disease (I20–I25), and other heart diseases (I30–I51). [a]Conditional agreement in medical records is defined as number of deaths in agreement with death certificates divided by the total number of deaths from a specific cause of death. [b]Conditional agreement in death certificates is defined as number of deaths in agreement with medical records divided by the total number of deaths from a specific cause of death.
Abbreviations: HL, Hodgkin lymphoma; BC, breast cancer.

respectively). According to COD_{MR}, 23.2% of all patients died from HL compared to 33.1% according to COD_{DC}. When the COD_{DC} stated HL, only in 63% of deaths, COD_{DC} and COD_{MR} agreed. The conditional agreement for HL as COD_{MR} was 89.9%, while the corresponding Cohen's kappa value of 0.64 indicated moderate agreement. On the other hand, for breast, lung, and gastrointestinal cancers, excellent agreement was observed between COD_{MR} and COD_{DC} in HL patients (corresponding Cohen's kappa: 0.91, 0.79, and 0.83, respectively). HL patients more often died from cardiovascular disease according to COD_{MR} than according to COD_{DC} (20.9% vs. 15.6%, respectively). For ischemic heart disease, the conditional agreement was 57.8% for COD_{MR} and 46.4% for COD_{DC}, with a corresponding kappa value of 0.48. The conditional agreement for other heart diseases was low for COD_{MR} (18.1%), and the corresponding kappa

value of 0.24 reflected poor level of agreement between COD_{MR} and COD_{DC}. The overall agreement between COD_{MR} and COD_{DC} for common causes of death in HL patients was 81.4%, excluding other, unknown, and unavailable causes of death (Table 3). Including all other causes of death, the overall agreement was 70.8%. Among patients for whom the COD_{DC} was HL, 29 out of 365 patients died from non-HL or leukemia according to COD_{MR}. Similarly, 10 out of 180 patients who died of HL according to COD_{MR}, died from non-HL or leukemia according to COD_{DC}. Additionally, a substantial part of HL patients who died of HL according to COD_{DC} died from circulatory disease according to COD_{MR} (22 out of 164 patients).

In addition in BC patients, malignant neoplasm was the most important cause of death (Table 4). When the COD_{MR} was BC, in 95.4% of deaths, the COD_{DC} was also BC,

Table 3 Causes of death according to medical records and death certificates in HL patients

Death certificates / Medical records	HL	BC	Lung cancer	Gastrointestinal cancer	NHL and leukemia	Other malignant neoplasms	Circulatory disease	Respiratory disease	Other causes of death	Unknown and unavailable[a]	Total medical records
HL	160	1	1	3	10	0	0	0	3	2	180
BC	0	26	0	0	0	0	0	0	0	—	27
Lung cancer	4	1	74	0	1	6	1	1	0	0	88
Gastrointestinal cancer	2	0	0	62	0	3	2	1	0	0	70
NHL and leukemia	29	0	0	0	40	0	1	0	3	—	74
Other malignant neoplasms	4	—	9	6	2	47	—	1	4	—	76
Circulatory disease	22	2	7	4	5	3	102	7	8	4	164
Respiratory disease	12	0	—	0	3	0	4	4	0	0	24
Other causes of death	21	0	—	—	4	4	9	5	28	4	77
Unknown and unavailable[a]	111	6	28	30	20	29	80	12	39	5	360
Total number in death certificates	365	37	121	106	85	92	200	31	85	18	1,140

Notes: Cause of death according to the ICD-10: HL (C81), BC (C50), lung cancer (C33–C34), gastrointestinal cancer (C15–C26), NHL and leukemia (C82–C96), circulatory disease (I00–I99), and respiratory disease (J00–J99). [a]Unknown and unavailable causes of death include ICD-10 codes R96–R99 and cases where no information on cause of death in medical records or no response from the GP could be obtained.
Abbreviations: HL, Hodgkin lymphoma; BC, breast cancer; NHL, non-Hodgkin lymphoma; GP, general practitioner.

resulting in a conditional agreement in COD_{DC} of 95.0%. The corresponding Cohen's kappa value of 0.75 indicated high agreement between COD_{MR} and COD_{DC}. Kappa values also indicated excellent agreement between COD_{MR} and COD_{DC} for lung cancer (0.78) and gastrointestinal cancer (0.80) in BC patients. For BC patients, a higher overall agreement between COD_{MR} and COD_{DC} for common causes of death was observed compared to HL patients (Table S1). Here, the overall agreement for breast, lung, and gastrointestinal cancers; other malignant neoplasms; and circulatory disease was 96.6%. Including other causes of death, the overall agreement was 91.7%.

For 70 out of 109 deceased HL patients for whom COD_{MR} and COD_{DC} differed, death certificates were filed in the years that were available for review (Table 5). When COD_{MR} was attributed to HL, seven discrepancies were due to the lack of specification by the physician, eg, NHL or "lymphoma" was stated on the death certificate rather than HL, and three discrepancies reflected coding errors by the registration clerk. When COD_{DC} was recorded as HL, 16 discrepancies were caused by the fact that the physician filing the death certificate related the cause of death to late adverse events of treatment for HL, which was, however, diagnosed >10 years prior to death in patients cured of HL, eg, "lung fibrosis after radiotherapy for the treatment of HL" was stated on the death certificate.

For 41 out of 42 deceased BC patients for whom we observed discrepancies between COD_{MR} and COD_{DC}, death certificates were available for review (Table S2). Of the 21 deaths attributed to BC in the COD_{MR} but not in the COD_{DC}, the discrepancy was once due to a coding error and the COD_{DC} was once attributed to an external cause of death without BC being mentioned on the death certificate. No explanation could be found for the remaining 19 discrepancies. In none of the 20 discrepancies where COD_{DC} was attributed to BC while the COD_{MR} was recorded as another disease, an explanation could be found.

Table 6 shows the results of the person-year analysis of cause-specific mortality for common causes of death in HL patients, using COD_{MR} complemented with COD_{DC} when COD_{MR} was missing and, alternatively, using COD_{DC} only. Compared to SMRs based on COD_{DC} only, SMRs based on COD_{MR} complemented with COD_{DC} were higher for circulatory disease (SMR=5.9; 95% CI=5.2–6.6 vs. SMR=4.6; 95% CI=4.0–5.2, $P<0.05$). Conversely, SMRs were lower for HL (SMR=1,030; 95% CI=917–1,152 vs. SMR=1,249; 95% CI=1,124–1,384, $P<0.05$). These differences are reflected in the AERs as well, especially for HL as cause of death.

Table 4 Level of agreement between causes of death from medical records and death certificates in BC patients

Cause of death	Number of deaths according to medical records (%)	Number of deaths according to death certificates (%)	Conditional agreement in medical records[a] (%)	Conditional agreement in death certificates[b] (%)	Cohen's kappa
Malignant neoplasm	544 (94.3)	554 (96.0)	99.1	96.6	0.63
Gastrointestinal cancer	12 (2.1)	18 (3.1)	100	66.7	0.80
Lung cancer	22 (3.8)	29 (5.0)	90.9	69.0	0.78
BC	478 (82.8)	476 (82.5)	95.4	95.0	0.75
Circulatory disease	11 (1.9)	4 (0.7)	36.4	100	0.53

Notes: Cause of death according to the ICD-10: malignant neoplasms (C00–C97), gastrointestinal cancer (C15–C26), lung cancer (C33–C34), BC (C50), and circulatory disease (I00–I99). [a]Conditional agreement in medical records is defined as number of deaths in agreement with death certificates divided by the total number of deaths from a specific cause of death. [b]Conditional agreement in death certificates is defined as number of deaths in agreement with medical records divided by the total number of deaths from a specific cause of death.
Abbreviation: BC, breast cancer.

Table 5 Causes of discrepancies between causes of death from medical records and causes of death from death certificates in HL patients

n	%	Comments
HL as cause of death from medical records, other cause of death based on death certificates		
10	**66**	**Coding or filing error**
7	47	Filing error on death certificate (physician stated "NHL" or "lymphoma")
3	20	Coding error by registration clerk
2	**13**	**External cause of death**
1	7	Cause of death from medical records mentioned on death certificate[a]
1	7	Cause of death from death certificates registered in medical records[b]
3	**20**	**No explanation[c]**
2	13	Malignant neoplasm stated as cause of death on death certificate, not registered in medical records
1	7	Malignant neoplasm stated as cause of death on death certificate, registered in medical records
3	20	Cause of death from medical records not stated on death certificate
15	**100**	
HL as cause of death from death certificates, other causes of death based on medical records		
16	**29**	**Physician relates cause of death to HL diagnosis >10 years prior to death[d]**
7	34	Cause of death from medical records stated on death certificate, but not as underlying cause of death[e]
4	7	Cause of death from medical records related to diagnosis stated on death certificate
5	9	Cause of death from medical records not stated on death certificate
39	**71**	**No explanation**
15	27	Cause of death from medical records stated on death certificate, but not as underlying cause of death
7	13	Cause of death from medical records related to diagnosis stated on death certificate[f]
17	31	Cause of death from medical records not stated on death certificate
55	**100**	

Notes: [a]The death certificate stated that the patient committed suicide because there were no further options for the treatment of HL. [b]The death certificate stated that the patient had euthanasia for BC. [c]Numbers do not add up to 100%; multiple comments possible when no explanation was found. [d]The physician stated on the death certificate: "chemotherapy or radiotherapy for the treatment of HL". [e]The death certificate stated for example "cardiomyopathy after chemotherapy and radiotherapy for the treatment of HL". [f]The death certificate stated for example "heart failure", whereas the medical records stated "cardiomyopathy".
Abbreviations: HL, Hodgkin lymphoma; NHL, non-Hodgkin lymphoma; BC, breast cancer.

The SMRs in the person-year analysis, where we complemented COD_{MR} with COD_{DC} and where we supplemented COD_{MR} with COD_{DC} when these causes of death showed discrepancy, approximated the results of the analysis where we used COD_{DC} only.

Discussion

Overall, we observed high agreement when comparing causes of death from medical records and death certificates. Among HL patients, an overall agreement of 81% was observed for common causes of death. However, COD_{MR} more often stated cardiovascular disease, whereas COD_{DC} was more often recorded as HL. Differences between COD_{MR} and COD_{DC} were much smaller for BC patients. The majority of BC patients died from BC itself, and a high overall agreement (97%) was observed between COD_{MR} and COD_{DC}. We compared SMRs based on different sources of data in our HL cohort and observed a lower SMR for HL and a higher SMR

Table 6 Person-year analysis on cause-specific mortality using different cause-of-death data in HL patients

Cause of death	Observed causes of death from medical records, complemented with death certificates[a]				Observed causes of death from death certificates only[b]				Observed causes of death from medical records, discrepancies supplemented with death certificates[c]			
	O	E	SMR (95% CI)	AER	O	E	SMR (95% CI)	AER	O	E	SMR (95% CI)	AER
Malignant neoplasms	743	69.8	10.6 (10.0–11.6)	129.3	804	69.8	11.5 (10.7–12.3)	141.0	828	69.8	11.9 (11.1–12.7)	145.6
Gastrointestinal cancer	105	18.4	5.7 (4.7–6.9)	16.6	108	18.4	5.9 (4.8–7.0)	17.2	111	18.4	6.0 (5.0–7.2)	17.8
Lung cancer	120	18.4	6.5 (5.4–7.8)	19.5	121	18.4	6.6 (5.5–7.9)	19.7	125	18.4	6.8 (5.7–8.1)	20.5
BC	34	7.4	4.6 (3.2–6.4)	5.1	37	7.4	5.0 (3.5–6.9)	5.7	38	7.4	5.1 (3.6–7.0)	5.9
HL	301	0.3	1,029.7 (916.6–1,152.8)	57.8	365	0.3	1,248.6 (1,123.8–1,383.5)[d]	70.1	368	0.3	1,258.9 (1,133.5–1,394.3)	70.6
NHL	50	2.0	24.6 (18.2–32.4)	9.2	60	2.0	29.5 (22.5–38.0)	11.1	67	2.0	32.9 (25.5–41.8)	12.5
Leukemia	29	1.9	15.0 (10.0–21.5)	5.2	23	1.9	11.9 (7.5–17.8)	4.0	24	1.9	12.4 (8.0–18.5)	4.2
Infectious and parasitic diseases	32	2.6	12.4 (8.4–17.4)	5.6	16	2.6	6.2 (3.5–10.0)	2.6	18	2.6	6.9 (4.1–11.0)	3.0
Circulatory disease	257	43.8	5.9 (5.2–6.6)	41.0	200	43.8	4.6 (4.0–5.2)[d]	30.0	213	43.8	4.9 (4.2–5.6)	32.5
Respiratory disease	37	8.0	4.6 (3.2–6.4)	5.6	31	8.0	3.9 (2.6–5.5)	4.4	32	8.0	4.0 (2.7–5.6)	4.6

Notes: [a]Cause of death was based on information from the medical records and supplemented with information from death certificates when the cause of death was unknown or unavailable from the medical records. [b]Cause of death was based on information from the death certificates only. [c]Cause of death was based on information from the medical records and supplemented with information from death certificates when the cause of death was unknown or unavailable from the medical records or when the cause of death was discrepant between the medical records and death certificates. [d]Statistical comparison between SMRs from the analysis using observed causes of death from medical records complemented with death certificates and using observed causes of death from death certificates only showed a statistically significant difference at a *P*-value of <0.05. Here, we assumed that the covariance between the different SMRs was zero. Since the two samples were not independent, covariance will be positive, resulting in a smaller *P*-value. Comparisons remain inconclusive when we observe a nonstatistically significant difference between SMRs at a covariance of zero.
Abbreviations: HL, Hodgkin lymphoma; NHL, non-Hodgkin lymphoma; O, observed; E, expected; SMR, standardized mortality ratio; AER, absolute excess risk.

for circulatory disease when using COD_{MR} complemented with COD_{DC} compared to using COD_{DC} only. We only calculated SMRs for the HL cohort, since very few discrepancies were observed for common causes of death in the BC cohort, and only a small number of missing causes of death from the medical records would need to be complemented with the causes of death from the death certificate. Therefore, we do not expect to find more than minor differences in the calculated SMRs in the BC cohort. We assessed discrepancies at an individual level by reviewing original death certificates when differences were observed between COD_{MR} and COD_{DC}. However, no explanations could be found for most of these discrepancies. In our study, linkage errors are minimal and we do not expect that these errors explain the observed discrepancies. Part of the discrepancies between COD_{MR} and COD_{DC} in HL patients could be explained by the fact that the physician attributed the cause of death on the death certificate to a potential late adverse effect of treatment for HL. However, as the patient was cured of HL according to the medical record, recording of HL as a cause of death does not seem to be correct. In addition, for an individual patient, causality of an exposure-disease sequence can never be established. Here, we must keep in mind that COD_{MR} and COD_{DC} are collected for different purposes. In research

focusing on the prognosis of HL or long-term adverse events of HL treatment, it is undesirable to code HL as the underlying cause of death when a patient is cured of HL and dies from a potential late effect of HL treatment. For example, when a physician states "cardiomyopathy as a consequence of chemotherapy for the treatment of HL", the COD_{DC} will be coded as HL, whereas in late effects research, the COD_{MR} would be attributed to cardiomyopathy.

When interpreting the results of this study, it is important to keep in mind that COD_{MR} and COD_{DC} have been assessed in different ways. To obtain causes of death from HL and BC patients, we reviewed the medical files. If the cause of death could not be retrieved, a questionnaire was sent to the patient's GP. Thus, where Statistics Netherlands has to rely on the physician who fills in the death certificate, in our research setting, we were able to utilize more information from medical records to obtain COD_{MR}. Acknowledging these differences between COD_{MR} and COD_{DC}, the question arises what level of agreement can be reached when comparing COD_{MR} and COD_{DC}. Statistics Netherlands previously assessed the reliability of coding an underlying cause of death from death certificates.[16] The reliability of the cause-of-death statistics, assessed by coding the same death certificate by different registration clerks, ranged from high (>90%) for major causes

of death such as cancers and acute myocardial infarction to low (<70%) for chronic diseases. In the current study, we observed an overall agreement for common causes of death of 71% in HL patients and 92% in BC patients. Relating these findings to the reliability study of Statistics Netherlands, the level of agreement in the current study is comparable with known variability in determining cause of death.

Among HL patients, our most important finding was that COD_{MR} was more often attributed to cardiovascular disease, whereas COD_{DC} was more often attributed to HL. This indicates that HL is more often recorded in the cause-of-death statistics files than we would expect from a clinical perspective, leading to an overestimation of mortality from HL. Our comparison between COD_{MR} and COD_{DC} showed that discrepancies were more frequent within the HL cohort than those in the BC cohort. BC patients most often died from BC itself within a short period of time, whereas a substantial number of HL patients died years after the initial HL diagnosis. Although there are differences between HL and BC patients in the calendar period of cancer diagnosis, follow-up time, and age at treatment, discrepancies in cause of death appear to be more frequent when deaths occurred after a relatively long period of follow-up, when the likelihood of deaths due to late adverse events related to treatment is higher.

Conclusion

We observed high levels of agreement between COD_{MR} and COD_{DC} for common causes of death in HL and BC patients. However, differences were observed in the attribution of HL and circulatory disease as causes of death between COD_{MR} and COD_{DC}. Observed discrepancies between COD_{MR} and COD_{DC} frequently occurred in the presence of late effects of treatment for HL; HL was more often recorded as COD_{DC}, whereas the COD_{MR} more often concerned potential late effects of treatment. In future studies, researchers should keep in mind that differences in causes of death may arise when examining long-term mortality in cancer patients using different sources of cause-of-death data.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Johansson LA, Westerling R. Comparing Swedish hospital discharge records with death certificates: implications for mortality statistics. *Int J Epidemiol.* 2000;29(3):495–502.
2. Johansson LA, Westerling R. Comparing hospital discharge records with death certificates: can the differences be explained? *J Epidemiol Community Health.* 2002;56(4):301–308.
3. Johansson LA, Björkenstam C, Westerling R. Unexplained differences between hospital and mortality data indicated mistakes in death certification: an investigation of 1,094 deaths in Sweden during 1995. *J Clin Epidemiol.* 2009;62(11):1202–1209.
4. Klijs B, Nusselder WJ, Mackenbach JP. Nationwide individual record linkage study showed poor agreement of causes of death and hospital diagnoses at individual level but reasonable agreement at population level. *J Clin Epidemiol.* 2014;67(2):160–168.
5. Schaffar R, Rapiti E, Rachet B, Woods L. Accuracy of cause of death data routinely recorded in a population-based cancer registry: impact on cause-specific survival and validation using the Geneva Cancer Registry. *BMC Cancer.* 2013;13:609.
6. Aleman BM, van den Belt-Dusebout AW, Klokman WJ, van 't Veer MB, Bartelink H, van Leeuwen FE. Long-term cause-specific mortality of patients treated for Hodgkin's disease. *J Clin Oncol.* 2003;21(18):3431–3439.
7. Ng AK, Bernardo MP, Weller E, et al. Long-term survival and competing causes of death in patients with early-stage Hodgkin's disease treated at age 50 or younger. *J Clin Oncol.* 2002;20(8):2101–2108.
8. Hooning MJ, Aleman BM, van Rosmalen AJ, Kuenen MA, Klijn JG, van Leeuwen FE. Cause-specific mortality in long-term survivors of breast cancer: a 25-year follow-up study. *International journal of radiation oncology, biology*. *Physics*. 2006;64(4):1081–1091.
9. van Leeuwen FE, Klokman WJ, Hagenbeek A, et al. Second cancer risk following Hodgkin's disease: a 20-year follow-up study. *J Clin Oncol.* 1994;12(2):312–325.
10. van Leeuwen FE, Klokman WJ, Veer MB, et al. Long-term risk of second malignancy in survivors of Hodgkin's disease treated during adolescence or young adulthood. *J Clin Oncol.* 2000;18(3):487–497.
11. de Bruin ML, Burgers JA, Baas P, et al. Malignant mesothelioma after radiation treatment for Hodgkin lymphoma. *Blood.* 2009;113(16):3679–3681.
12. Aleman BM, van den Belt-Dusebout AW, de Bruin ML, et al. Late cardiotoxicity after treatment for Hodgkin lymphoma. *Blood.* 2007;109(5):1878–1886.
13. van Nimwegen FA, Schaapveld M, Janus CP, et al. Cardiovascular disease after Hodgkin lymphoma treatment: 40-year disease risk. *JAMA Intern Med*. 2015;175(6):1007–1017.
14. Boekel NB, Jacobse JN, Schaapveld M, et al. Cardiovascular disease incidence associated with internal mammary chain irradiation and other breast cancer treatments: a large cohort study of patients treated from 1970 through 2009. *Br J Cancer*. Epub 2018 Aug 1.
15. Organization WH. *ICD-10: International Statistical Classification of Diseases and Related Health Problems*. Geneva: WHO; 1992–1994.
16. Harteloh P, de Bruin K, Kardaun J. The reliability of cause-of-death coding in The Netherlands. *Eur J Epidemiol.* 2010;25(8):531–538.
17. van Sonsbeek JLA. *Van de schaduw des doods tot een licht ten leven. De historie van de methodiek van de doodsoorzakenstatistiek in Nederland 1865-2005* [From the shadow of death to a light of life. The history of the methodology of cause-of-death statistics in the Netherlands 1865-2005]. Voorburg/Heerlen: Centraal Bureau voor de Statistiek; 2005. Dutch.
18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.

Supplementary materials

Table S1 Causes of death according to medical records and death certificates in BC patients

Medical records	Death certificates BC	Lung cancer	Gastrointestinal cancer	Other malignant neoplasms	Circulatory disease	Other causes of death	Unknown and unavailable[a]	Total medical records
BC	456	7	3	8	0	4	5	483
Lung cancer	2	20	0	0	0	0	0	22
Gastrointestinal cancer	0	0	12	0	0	0	0	12
Other malignant neoplasms	5	1	2	25	0	1	0	32
Circulatory disease	5	0	0	0	4	2	0	11
Other causes of death	8	1	1	0	0	12	0	22
Unknown and unavailable[a]	69	8	8	4	8	11	0	108
Total number in registry data	545	37	26	35	12	30	5	690

Notes: Cause of death according to the ICD-10: BC (C50), lung cancer (C33–C34), gastrointestinal cancer (C15–C26), and circulatory disease (I00–I99). [a]Unknown and unavailable causes of death include ICD-10 codes R96–R99 and cases where no information on cause of death in medical records or no response from the GP could be obtained.
Abbreviations: BC, breast cancer; GP, general practitioner.

Table S2 Causes of discrepancies between causes of death from medical records and causes of death from death certificates in BC patients

n	%	Comments
BC as cause of death from medical records		
1	5	**Coding error**
1	5	Coding error by registration clerk ("ovarian cancer" instead of "BC")
1	5	**External cause of death**
1	5	Cause of death from medical records not mentioned on death certificate
19	21	**No explanation[a]**
9	47	Malignant neoplasm stated as cause of death on death certificate, not registered in medical records
4	21	Malignant neoplasm stated as cause of death on death certificate, registered in medical records
6	32	Cause of death from medical records stated on death certificate, but not as underlying cause of death
1	5	Cause of death from medical records related to diagnosis stated on death certificate
12	63	Cause of death from medical records not stated on death certificate
21	100	
BC as cause of death from death certificates		
20	100	**No explanation**
7	35	Cause of death from medical records stated on death certificate, but not as underlying cause of death
6	30	Cause of death from medical records related to diagnosis stated on death certificate
7	35	Cause of death from medical records not stated on death certificate
20	100	

Note: [a]Numbers do not add up to 100%; multiple comments possible when no explanation was found.
Abbreviation: BC, breast cancer.