

A Validated Algorithm for Register-Based Identification of Patients with Relapse of Clinical Stage I Testicular Cancer

Thomas Wagner^{1,2}, Jakob Lauritsen¹, Mikkel Bandak¹, Linda Aagaard Rasmussen³, Johannes Bakker⁴, Hanna Birkbak Hovaldt⁴, Heidi Larsson⁴, Ib Jarle Christensen¹, Birgitte Grønkær Toft², Mads Agerbæk⁵, Lars Dysager⁶, Michael Kreiberg¹, Josephine Julie Rosenvilde¹, Birte Engvad⁷, Daniel M Berney⁸, Gedskede Daugaard¹

¹Department of Oncology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; ²Department of Pathology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; ³Research Unit for General Practice, Aarhus, Denmark; ⁴The Danish Clinical Quality Program – National Clinical Registries (RKKP), Aarhus, Odense and Copenhagen, Denmark; ⁵Department of Oncology, Aarhus University Hospital, Aarhus, Denmark; ⁶Department of Oncology, Odense University Hospital, Odense, Denmark; ⁷Department of Pathology, Odense University Hospital, Odense, Denmark; ⁸Centre for Cancer Biomarkers and Biotherapeutics, Barts Cancer Institute, Queen Mary University of London, London, UK

Correspondence: Thomas Wagner, Department of Oncology, Copenhagen University Hospital, Rigshospitalet, Blegdamsvej 9, Copenhagen, 2100, Denmark, Tel +45 35459682, Email thomas.wagner.nielsen@regionh.dk

Purpose: The Danish Testicular Cancer (DaTeCa) database aims to monitor and improve quality of care for testicular cancer patients. Relapse data registered in the DaTeCa database rely on manual registration. Currently, some safeguarding against missing registrations is attempted by a non-validated register-based algorithm. However, this algorithm is inaccurate and entails time-consuming medical record reviews. We aimed (1) to validate relapse data as registered in the DaTeCa database, and (2) to develop and validate an improved register-based algorithm identifying patients diagnosed with relapse of clinical stage I testicular cancer.

Patients and Methods: Patients registered in the DaTeCa database with clinical stage I testicular cancer from 2013 to 2018 were included. Medical record information on relapse data served as a gold standard. A pre-specified algorithm to identify relapse was tested and optimized on a random sample of 250 patients. Indicators of relapse were obtained from pathology codes in the Danish National Pathology Register and from diagnosis and procedure codes in the Danish National Patient Register. We applied the final algorithm to the remaining study population to validate its performance.

Results: Of the 1377 included patients, 284 patients relapsed according to the gold standard during a median follow-up time of 5.9 years. The completeness of relapse data registered in the DaTeCa database was 97.2% (95% confidence interval (CI): 95.2–99.1). The algorithm achieved a sensitivity of 99.6% (95% CI: 98.7–100), a specificity of 98.9% (95% CI: 98.2–99.6), and a positive predictive value of 95.9% (95% CI: 93.4–98.4) in the validation cohort (n = 1127, 233 relapses).

Conclusion: The registration of relapse data in the DaTeCa database is accurate, confirming the database as a reliable source for ongoing clinical quality assessments. Applying the provided algorithm to the DaTeCa database will optimize the accuracy of relapse data further, decrease time-consuming medical record review and contribute to important future clinical research.

Keywords: testicular cancer, relapse, algorithm, validation, registries

Introduction

Nationwide clinical databases provide a valuable resource for monitoring quality of care and conducting population-based research.^{1–3} However, the utility of clinical databases depends on the accuracy and completeness of the data it contains. Consequently, validation is critical to ensure reliability and accuracy.^{4,5} The Danish Testicular Cancer (DaTeCa) database is a nationwide clinical quality database containing prospectively collected data on all testicular cancer (TC) patients aged ≥ 15 years diagnosed in Denmark since 2013.⁶ Patient-level data are systematically recorded

by a combination of registration of clinical parameters by the treating oncologist and automated linkage to Danish nationwide health registries.^{6,7} Data capture from these registries decreases the registration workload and increases the completeness of the data included. In total, 99.9% of all incident TC cases diagnosed from 2013 to 2020 identified in the Danish National Patient Register⁸ and the Danish National Pathology Register⁹ are registered in the database.⁷ However, the nationwide health registries do not systematically capture information on relapse, and relapse data in the DaTeCa database rely on manual registration. Currently, some safeguarding against missing information on relapse data is attempted by crosschecking the Danish National Patient Register for chemotherapy and radiotherapy procedure codes. However, this algorithm is not validated and limited by a substantial amount of false-positive and false-negative cases upon medical record review.⁷ With 300 incident TC cases annually in Denmark,⁷ medical record review is increasingly time-consuming, impractical at a population level, and might lead to missing and erroneous registrations of relapse data in the DaTeCa database. As approximately 25% of patients with clinical stage I TC relapse after orchiectomy,⁷ information about relapse is a key clinical outcome in assessing quality of care. Consequently, efforts to improve data collection, data validity and data completeness are of high priority.^{6,7} A cancer relapse is often followed by diagnostic procedures and treatment which are registered in the Danish health registries.³ Therefore, an algorithm that uses routinely collected data in these registries has the potential to identify cancer relapses as shown for other cancer types.^{10–15} In this study, we aimed: (1) to validate the relapse data of patients with stage I TC as registered in the DaTeCa database, and (2) to develop and validate an improved register-based algorithm identifying patients diagnosed with relapse of stage I TC.

Materials and Methods

Setting

Danish Health Care System and Health Registries

All permanent residents in Denmark have equal access to fully tax-funded healthcare. Utilization of these services is registered through a unique personal registration number assigned at birth or immigration, which allows individual-level linkage of data recorded in nationwide health registries and databases.^{3,16}

Clinical Stage I TC and Standard-of-Care Treatment in Denmark

Following orchiectomy, patients with TC are clinically staged with the serum tumor markers (α -fetoprotein (AFP), β -human chorionogonadotropin (β -hCG) and lactate dehydrogenase (LDH)), and a computed tomography (CT) of the thorax, abdomen and pelvis.¹⁷ Clinical stage I disease is defined as normalized postoperative serum tumor markers and no evidence of metastatic disease on the CT scan.¹⁸ In Denmark, all patients with stage I TC are offered a 5-year surveillance program at one of three university hospitals.^{6,17} A biopsy from the contralateral testis is offered to all patients at the time of orchiectomy, and in case of germ cell neoplasia in situ (GCNIS), patients are treated with radiotherapy to the testicle. Patients with relapse receive standard treatment with combination chemotherapy (bleomycin, etoposide, and cisplatin (BEP)). In cases of relapsing seminoma with limited retroperitoneal lymph node metastases, patients are offered radiotherapy.¹⁷

Data Sources

We extracted data from four Danish nationwide registries: 1) The prospective DaTeCa database⁶ with clinical data on all TC patients diagnosed in Denmark, including clinical stage at primary diagnosis and information on relapse; 2) The Danish Civil Registration System¹⁶ with daily updated vital status and information on migration; 3) The Danish National Patient Register⁸ with information on all somatic hospital contacts, including diagnosis codes according to the International Classification of Diseases, 10th version (ICD-10) system, and procedure codes according to the Health Care Classification System (Danish, Sundhedsvæsenets Klassifikation System [SKS]); and 4) The Danish National Pathology Register⁹ with information on all pathology specimens analysed in Denmark, including topography codes and morphology codes according to the Danish Systematized Nomenclature of Medicine (SNOMED) system.

Study Population

We included patients registered in the DaTeCa database with stage I TC diagnosed between 1 January 2013 and 31 December 2018. A detailed description of the study cohort has been reported elsewhere.¹⁹ In brief, we confirmed primary diagnosis by pathology review of the orchiectomy specimens. Clinical stage I disease at presentation was confirmed by medical record review. Patients fulfilling any of the following criteria were excluded: registration in the Register of Human Tissue Utilisation,²⁰ other primary histology than germ cell TC (sex-cord stromal tumors, spermatocytic tumors, prepubertal type teratomas, GCNIS only, reactive changes), ovotestis, synchronous tumors, prior TC, extragonadal germ cell tumors, orchiectomy abroad, loss to follow-up within 30 days of orchiectomy, or clinical stage IS, II or III at primary diagnosis. The study population was randomly divided into a development cohort and a validation cohort.

Gold Standard

We considered the information on relapse status and date of relapse in the medical record as the gold standard. The medical records from included patients were reviewed by TW, GD, JL or MB. In case of doubt of clinical stage at presentation or relapse status, a joint decision was made. Relapse was defined according to standard definitions: a confirmed serum tumor marker relapse (AFP, hCG) and/or radiological signs of relapse and/or histopathologically verified relapse which led to initiation of subsequent treatment (chemotherapy or radiotherapy or/and surgery).^{18,21} The time point for relapse was defined as date of biopsy or surgery in case of histopathologically proven relapse; if a relapse was defined by elevated tumor markers and/or radiologic imaging without histological verification, the date when a relapse was stated in the medical record by the treating oncologist was used.

Algorithm

An expert panel including pathologists and oncologists with expertise in TC treatment as well as data management specialists affiliated to the DaTeCa database determined pre-specified algorithm indicators of relapse, defined as types of health-care events likely to indicate relapse based on the standard-of-care treatment of stage I TC in Denmark. The indicators were identified through diagnosis codes, procedure codes and pathology diagnosis codes (Table 1). To prevent inclusion of events related to the primary clinical staging diagnostic work-up, the algorithm only searched for indicators of relapse 30 days after the date of orchiectomy and onwards. If a patient had more than one indicator of relapse, the date of the first occurring indicator was registered as the date of relapse. We applied the pre-specified algorithm to the development cohort. To optimize the algorithm, different versions with varying indicators were tested as specified in Table 1, and the performance of the algorithm was evaluated in each step. Lastly, the optimized and final algorithm was evaluated in the validation cohort.

Statistical Analyses

The median follow-up time with interquartile range (IQR) was estimated from date of orchiectomy to the date of one of the following events: relapse, emigration, death, metachronous cancer, lost to follow-up or end of study (31 December 2021), whichever came first. To allow for data entry delays in the registries and in the DaTeCa database, data were updated 1 September 2022. Validity measures with 95% confidence intervals (CIs) were computed based on contingency tables. The completeness of the registered relapse data in the DaTeCa database was defined equivalent to sensitivity, ie, the proportion of patients with correctly registered relapse according to the gold standard.³ The positive predictive value of the registered relapse data in the DaTeCa database was estimated as the proportion of patients registered in the DaTeCa database with relapse who truly had relapse according to the gold standard. The negative predictive value was estimated as the proportion of patients registered in the DaTeCa database without relapse who truly were without relapse according to the gold standard. For algorithm development, a weighted random sample of 250 patients was used, obtained by a weighted random selection method.²² This relatively small proportion of the study cohort for development was chosen a priori for the following reasons: a) the relapse rate of stage I TC is high (20–30%),^{7,23,24} b) the application of a simple algorithm with few indicators, c) only minor refinements of the pre-specified algorithm were anticipated, and d) validation in a large study population would increase the generalizability of the algorithm. For each version of the algorithm in the development phase, we estimated the sensitivity, specificity, and positive and negative predictive values with 95% CIs from the concordant and discordant frequencies between the relapses identified by the algorithm and the gold standard. The remaining study population served as the validation cohort, and similar analyses were calculated for

Table 1 Pre-Specified and Refined Indicators of Testicular Cancer Relapse, and Different Versions of the Algorithm Based on the Various Indicators

Indicators	Source	Codes
Pre-specified		
1	Danish National Pathology Registry	SNOMED combination codes 30 or more days after the orchiectomy: any T -code except T78*** (testis) and T79*** (para-testicular structures) in combination with the Æ -code ÆF4630 (metastasis with primary tumor in the testis) and/or any of the following M -codes: M9061* (seminoma), M9065* (seminoma with syncytiotrophoblasts), M9070* (embryonal carcinoma), M907x* (embryonal carcinoma with syncytiotrophoblasts), M9071* (yolk sac tumor), M908B* (teratoma), M9084* (teratoma with somatic type malignancy), M90800 (mature teratoma), M90803 (immature teratoma), M90801 (solid teratoma), M9100* (choriocarcinoma) or M908A* (mixed germ cell tumor) with the numbers 3,4,6,7 or 9 in the fifth digit, which indicates “tumor behavior” 3: “malignant, primary”, 4: “direct spread to surrounding tissue”, 6: “malignant, metastasis”, 7: “malignant, relapse/recurrence” and 9: “malignant”, “unknown if primary or metastatic”
2	Danish National Patient Register	Procedure codes for radiotherapy (BWG*) or cytostatic therapy (BWHA***) 30 or more days after the orchiectomy
3	Danish National Patient Register	Diagnosis codes of testicular cancer relapse (ICD-10: C629X) or metastases (ICD-10: C629M) 30 or more days after the orchiectomy
4	Danish National Patient Register	Diagnosis codes of distant metastasis (ICD-10: C76*-C79*) 30 or more days after the orchiectomy
Refined		
2a	Danish National Patient Register	Procedure codes for radiotherapy (BWG*) 60 or more days after the orchiectomy, or cytostatic therapy (BWHA***) 30 or more days after the orchiectomy.
2b	Danish National Patient Register	Procedure codes for radiotherapy (BWG*) 60 or more days after the orchiectomy, or cytostatic therapy (BWHA***) 30 or more days after the orchiectomy – procedure codes combined with a simultaneously registered testicular cancer diagnosis code (ICD-10: C62*)
Different versions of the algorithm		Indicators
Algorithm 1 (pre-specified)		1,2,3 and 4
Algorithm 2		1,2 and 3
Algorithm 3		1 and 2
Algorithm 4		1 and 2a
Algorithm 5		1 and 2b

Abbreviations: ICD-10, International Classification of Disease, tenth revision; SNOMED, Systematized Nomenclature of Medicine; T, topography; M, morphology; Æ, etiology; F, function.

the final algorithm in the validation cohort. Further, to evaluate the independent information contributed by each of the indicators of the final algorithm in the validation cohort, we estimated the performance metrics separately for the pathology codes and for the procedure codes. To assess the accuracy of the relapse data estimated by the final algorithm, we estimated the proportion (and 95% CI) of relapse dates identified by the algorithm on the same date as the gold standard relapse date, and within an interval of 7,

14, 30, 60, and 90 days. The time from date of diagnosis and relapse date, obtained by the algorithm and the gold standard, were plotted against each other to spot differences in relapse dates as points diverging from line $x=y$. Analyses were performed in SAS 9.4 (SAS Institute, Cary NC).

Results

A total of 1486 patients with stage I TC were registered in the DaTeCa database. After exclusion, the final study population included 1377 patients (Figure 1). Patient characteristics are summarized in Table 2. In total, 284 patients (20.6%) relapsed according to the gold standard during a median follow-up time of 5.9 years (IQR: 4.6–7.4).

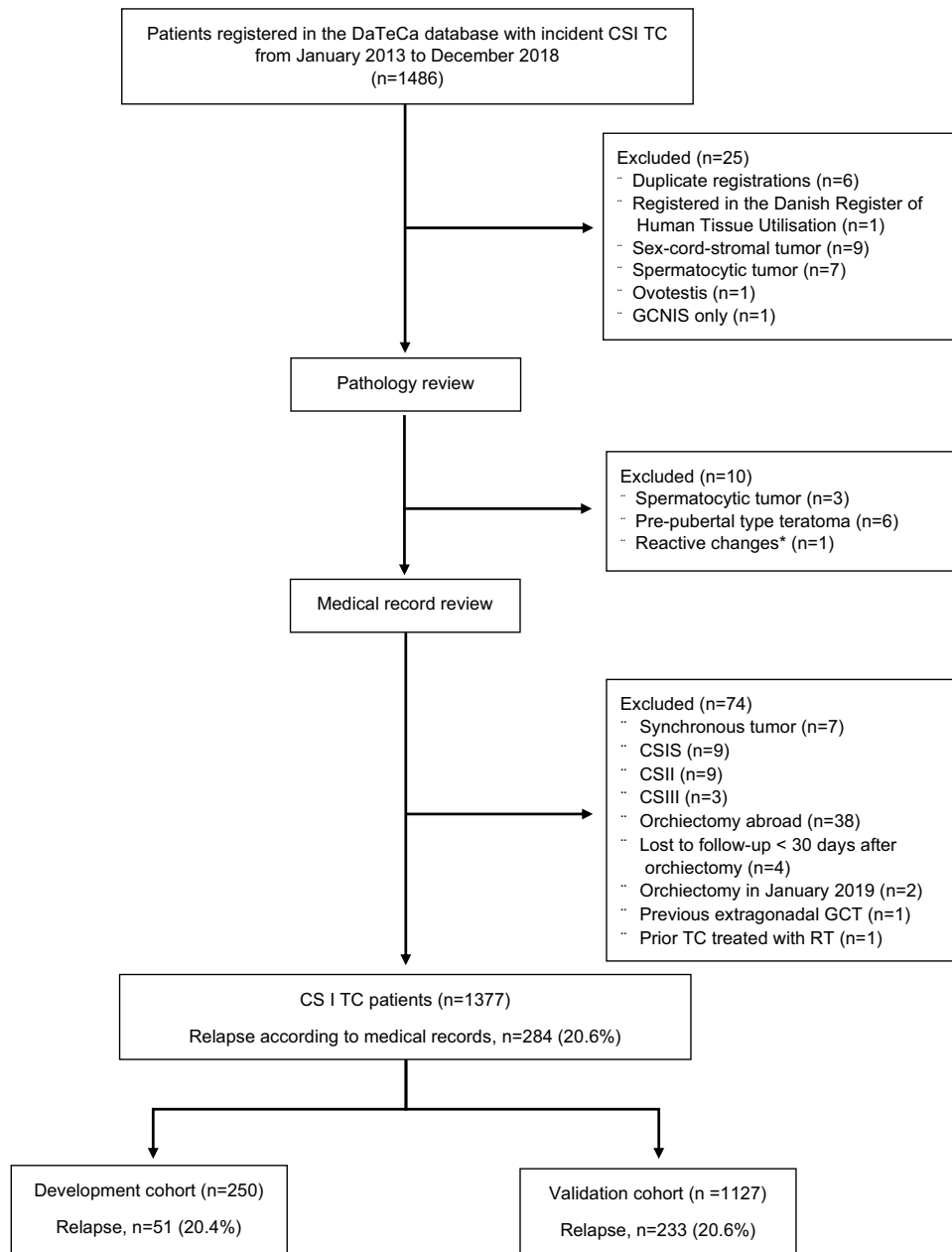


Figure 1 Flowchart of the study population.

Note: *Reactive metaplasia of epididymis.

Abbreviations: CSI, clinical stage I; CSIS, clinical stage IS; CSII, clinical stage II; CSIII, clinical stage III; DaTeCa database, Danish Testicular Cancer database; GCNIS, germ cell neoplasia in situ; GCT, germ cell tumor; RT, radiotherapy; TC, testicular cancer.

Table 2 Characteristics of the Study Cohort, and Stratified by Data Cohort

	Total Cohort	Development Cohort	Validation Cohort
Patients, n (%)	1377 (100)	250 (18)	1127 (82)
Median age at diagnosis, years (IQR)	37 (29–46)	38 (30–46)	37 (29–47)
Tumor subtype, n (%)			
Seminoma	924 (67)	173 (69)	751 (67)
Non-seminoma	453 (33)	77 (31)	376 (33)
Median follow-up, years (IQR)	5.9 (4.6–7.4)	5.9 (4.8–7.7)	5.9 (4.5–7.4)
Relapse, n (%)	284 (20.6)	51 (20.4)	233 (20.6)

Abbreviation: IQR, interquartile range.

Relapse Data in the DaTeCa Database

The positive predictive value and negative predictive value for relapse in the DaTeCa database were 99.6% (95% CI: 98.9–100) and 99.3% (95% CI: 98.8–99.8), respectively (Table 3). The completeness of relapse data registered in the DaTeCa database was 97.2% (95% CI: 94.5–98.8).

Algorithm Development

The results from the five different versions of the algorithm in the development cohort are reported in Table 4. The sensitivity ranged from 96.1% to 98% and the positive predictive value from 73.5% to 94.2%. Contingency tables for

Table 3 Concordance of Relapse Identified by the Gold Standard and Registered in the DaTeCa Database (n = 1377)

		Relapse by Gold Standard		
		Yes	No	Total
Relapse data in the DaTeCa database	Yes	276	1	277
	No	8	1092	1100
	Total	284	1093	1377

Notes: Positive predictive value: 99.6% (95% CI: 98.9–100). Negative predictive value: 99.3% (95% CI: 98.8–99.8). Completeness: 97.2% (95% CI: 95.2–99.1).

Abbreviations: CI, confidence interval; DaTeCa database, Danish Testicular Cancer database.

Table 4 Performance of Different Versions of the Algorithm to Identify Patients with Relapse from Stage I Testicular Cancer in the Development Cohort (n = 250)

Algorithm	Indicators	Sensitivity % (95% CI)	Specificity % (95% CI)	Positive Predictive Value % (95% CI)	Negative Predictive Value % (95% CI)
1	1,2,3,4	98.0 (94.2–100)	90.1 (87.0–94.9)	73.5 (63.0–84.0)	99.5 (98.4–100)
2	1,2,3	98.0 (94.2–100)	92.0 (88.2–95.7)	75.8 (65.4–86.1)	99.5 (98.4–100)
3	1,2	98.0 (94.2–100)	92.0 (88.2–95.7)	75.8 (65.4–86.1)	99.5 (98.4–100)
4	1,2a	98.0 (94.2–100)	95.5 (92.6–98.4)	84.8 (75.6–93.9)	99.5 (98.5–100)
5	1,2b	96.1 (90.1–100)	98.5 (96.8–100)	94.2 (87.9–100)	99.0 (97.6–100)

Notes: Indicator 1: pathology codes: SNOMED combinations, 30 days. Indicator 2: procedure codes: radiotherapy and cytostatic therapy, 30 days. Indicator 3: diagnosis codes: testicular cancer relapse and metastasis codes, 30 days. Indicator 4: diagnosis codes: distant metastasis codes. Indicator 2a: procedure codes: radiotherapy 60 days, cytostatic therapy 30 days. Indicator 2b: procedure codes in combination with testicular cancer diagnosis code: radiotherapy 60 days, cytostatic therapy 30 days.

Abbreviation: CI, confidence interval.

each algorithm evaluation are provided in [Tables S1–S5](#). As shown in [Table 4](#), the refined and final algorithm (algorithm 5) outperformed the pre-specified algorithm (algorithm 1) improving the specificity markedly with minimal change in sensitivity. A case-by-case review of the false positives in algorithm 1 revealed that most cases were due to (a) radiotherapy for contralateral GCNIS, or (b) radiotherapy and/or chemotherapy for non-TC malignancies or non-malignant diseases. To reduce false positives from (a), the period where procedure codes for radiotherapy would be considered as treatment of relapse was extended from 30 to 60 days after orchiectomy (algorithm 4). To reduce false positives from (b), the procedure codes were restricted to registrations combined with a TC diagnosis code (algorithm 5).

Algorithm Validation

Applying the final algorithm (algorithm 5) to the validation cohort, the algorithm identified 232 of the 233 relapses and additional 10 false positives ([Table 5](#)), corresponding to a sensitivity of 99.6% (95% CI: 98.7–100) and a specificity of 98.9% (95% CI: 98.2–100) ([Table 6](#)). With a relapse prevalence of 20.6% according to the gold standard, the positive predictive value was 95.9% (95% CI: 93.4–98.4), and the negative predictive value was 99.9% (95% CI: 99.7–100). The stratified analyses showed that the sensitivity was highest for the indicators of procedure codes, whereas the specificity for the indicators of pathology codes was 100% ([Table 6](#)). The relapse date estimated by the algorithm was the same date as the gold standard relapse date in 44.4% of the concordant relapse cases, and within 30 days in 95.7% of the cases ([Table 7](#)). In six cases (2%), the estimated date of the algorithm was >90 days from the relapse date according to the gold standard ([Figure 2](#)). This was due to delayed treatment of contralateral GCNIS beyond 60 days from orchiectomy; thus, the radiotherapy treatment dates for GCNIS were captured by the algorithm as the relapse dates rather than the following genuine relapse dates.

Discussion

In this large nationwide, population-based cohort study we: 1) confirmed high completeness and accuracy of the registered relapse data in the DaTeCa database for patients with stage I TC; and 2) developed and validated an algorithm

Table 5 Concordance of Relapse Identified by the Gold Standard and the Final Algorithm (Algorithm 5) in the Validation Cohort (n = 1127)

		Relapse by Gold Standard		
		Yes	No	Total
Relapse by algorithm	Yes	232	10	242
	No	1	884	885
	Total	233	894	1127

Table 6 Performance of the Final Algorithm (Algorithm 5) in the Validation Cohort (n = 1127) to Identify Patients with Relapse from Stage I Testicular Cancer

Algorithm Performance	All Indicators Combined	Indicators Stratified on Codes Related to Pathology and Procedure	
	% (95% CI)	Pathology Codes % (95% CI)	Procedure Codes % (95% CI)
Sensitivity	99.6 (98.7–100)	57.9 (51.6–64.2)	97.0 (94.8–99.2)
Specificity	98.9 (98.2–99.6)	100 (100–100)	98.9 (98.2–99.6)
Positive predictive value	95.9 (93.4–98.4)	100 (100–100)	95.8 (93.2–98.3)
Negative predictive value	99.9 (99.7–100)	90.1 (88.3–92.0)	99.2 (98.6–99.8)

Abbreviation: CI, confidence interval.

Table 7 Accuracy of Relapse Date as Estimated from the Final Algorithm (Algorithm 5) Compared with the Gold Standard in the Full Study Cohort (n = 1377)

Algorithm Estimated Date	% (95% CI)
Same date	44.4 (38.6–50.5)
Within 7 days	70.1 (64.4–75.4)
Within 14 days	85.8 (81.1–89.6)
Within 30 days	95.7 (92.7–97.8)
Within 60 days	97.5 (94.9–99.0)
Within 90 days	97.9 (95.4–99.2)

Abbreviation: CI, confidence interval.

using routinely registered administrative data which identified patients with relapse of stage I TC with a sensitivity of 99.6% and a positive predictive value of 95.9%.

Implications

Our study implies that despite close time-consuming medical record follow-up and crosschecking with Danish National Patient Register for procedure codes, a stage I TC relapse may be missing for 3% of the patients in the DaTeCa database. This will underestimate the risk of relapse, and it might impact the estimates in studies on prognostic factors for relapse. Therefore, to optimize the completeness of the relapse data in the DaTeCa database, we prioritized an algorithm with high sensitivity. However, of almost equal importance by aiming to reduce the increasingly time-consuming medical record review, we refined the algorithm to optimize the positive predictive value. With an achieved sensitivity of 99.6% combined with a positive predictive value of 95.9%, applying the algorithm to the prospective DaTeCa database will

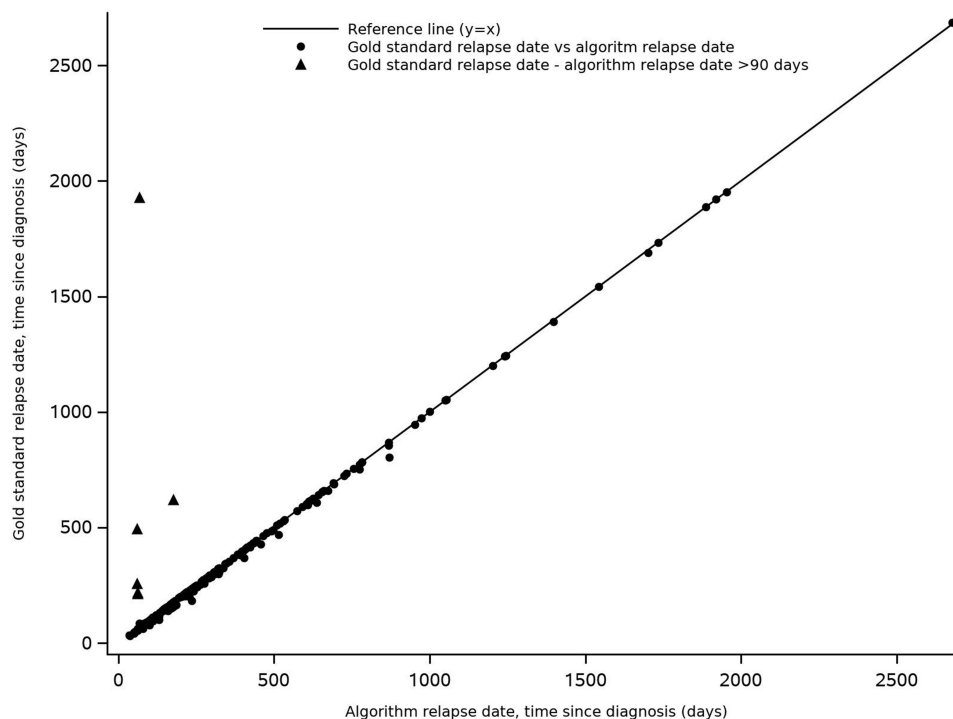


Figure 2 Concordance of relapse date between the final algorithm (algorithm 5) and by the gold standard.

increase the completeness of relapse data and reduce the resources allocated to medical record review markedly. The algorithm identified relapse date accurately; the estimated relapse date was within 30 days of that obtained by gold standard in 95.7% of the cases. Valid information on date of relapse is important to evaluate the surveillance program, time to relapse and timing of radiologic examinations.⁶ This makes the algorithm a valuable resource for future research.

Comparison with Other Studies

Our algorithm performed slightly better than algorithms based on same data sources to identify relapse of malignant melanoma,¹² colorectal,¹⁴ breast,^{13,15} bladder,¹¹ and endometrial cancer.¹⁰ In these studies, the sensitivity ranged from 85% to 97.3%, and the specificity ranged from 90% to 97.2%. There may be several reasons for this slightly better performance. All Danish patients with stage I TC regardless of prognostic factors are followed on a surveillance program; no one is treated with adjuvant treatment.^{6,7,17} This facilitates a clear distinction of de novo metastatic versus stage I disease with subsequent relapse. Additionally, TC patients are young with a low frequency of other malignancies. Finally, TC is a histopathologically distinct type of cancer (germ cell cancer) with disease-specific TC SNOMED codes in the Danish National Pathology Register. Using SNOMED codes as an indicator of relapse increased the performance of the algorithm markedly, identifying the few patients who were treated with surgery alone and not captured by procedure codes for chemotherapy or radiotherapy.

Strengths and Limitations

This study has several strengths. The algorithm is based on continuously updated Danish national registries with near complete high-quality records ensuring highly valid data.^{1,3,5} The study includes an unselected, large nationwide, population-based cohort established in a tax-funded health-care system with free and equal access for all residents in Denmark which limits the risk of selection bias, and provides generalizability of the results to the entire population of patients with stage I TC in Denmark. Finally, the gold standard data originated from a thorough medical record review of the entire cohort minimizing misclassification of relapse status and stage I disease at primary diagnosis. Some limitations should be considered when applying the algorithm. The study cohort included correctly classified stage I patients based on medical record review. As shown in [Figure 1](#), using the data extract from the DaTeCa database without medical record review would have led to inclusion of 21 patients who turned out to have de novo metastatic disease at presentation (clinical stage IS, II and III). If these patients were treated with chemotherapy 30 days or radiotherapy 60 days from orchiectomy and onwards, the algorithm would have captured this as indicators of relapse. This potential misclassification infers a risk of overestimated relapse rates; further, it might overestimate the overall treatment burden as well as underestimate the overall prognosis of stage I disease.^{25,26} This underscores that accurate reporting to the DaTeCa database of the clinical stage at primary diagnosis is important for reliable relapse data generated by the algorithm. Surprisingly, the performance of the algorithm in the validation cohort marginally improved compared to the development cohort where the algorithm was optimized. However, this might be a chance finding in relation to the random division of the study population, considering the few false-positive and false-negative cases overall.

Conclusion

The completeness and accuracy of the relapse data in the DaTeCa database is high, confirming the database as a reliable source for future scientific studies and ongoing clinical quality assessments. We developed and validated a register-based algorithm to identify patients with relapse from stage I TC. Applying this algorithm to the DaTeCa database will optimize the accuracy of the relapse data further and decrease time-consuming medical record review. Optimized registrations on relapse data facilitates populations-based research on several important issues including prognostic factors for relapse, treatment burden and rational surveillance programs for patients with stage I TC.^{6,18}

Data Sharing Statement

The data are stored at The Danish Clinical Quality Program – National Clinical Registries (RKKP). The data is not publicly available due to the Danish data protection legislation as the data contains information that could compromise the privacy of the research participants.

Ethics

The study is approved by the Danish Data Protection Agency, approval no. VD-2018-433, the Regional Ethics Committee, approval no. SJ-690, and the Danish Patient Safety Authority, approval no. 31-1521-341.

Funding

The study is supported by the Danish Clinical Quality Program – National Clinical Registries, the Danish Cancer Society, the Danish Cancer Research Foundation, and Preben and Anna Simonsen's Foundation. The funding sources had no influence on study conceptualization, data collection, data analysis or manuscript preparation.

Disclosure

Daniel M Berney and Gedske Daugaard are co-senior authors for this study. Daniel M Berney is supported by Orchid. The authors report no other conflicts of interest in this work.

References

1. Sørensen HT, Pedersen L, Jørgensen J, Ehrenstein V. Danish clinical quality databases – an important and untapped resource for clinical research. *Clin Epidemiol.* 2016;8:425–427. doi:10.2147/CLEP.S113265
2. Green A. Danish clinical databases: an overview. *Scand J Public Health.* 2011;39(7 Suppl):68–71. doi:10.1177/1403494811402413
3. Schmidt M, Schmidt SAJ, Adelborg K, et al. The Danish health care system and epidemiological research: from health care contacts to database records. *Clin Epidemiol.* 2019;11:563–591. doi:10.2147/CLEP.S179083
4. Nørgaard M, Johnsen SP. How can the research potential of the clinical quality databases be maximized? The Danish experience. *J Intern Med.* 2016;279(2):132–140. doi:10.1111/joim.12437
5. Thygesen LC, Ersbøll AK. When the entire population is the sample: strengths and limitations in register-based epidemiology. *Eur J Epidemiol.* 2014;29(8):551–558. doi:10.1007/s10654-013-9873-0
6. Daugaard G, Kier M, Bandak M, et al. The Danish testicular cancer database. *Clin Epidemiol.* 2016;8:703–707. doi:10.2147/CLEP.S99493
7. Datecadatabase. Annual report (in Danish); 2021. Available from: https://www.sundhed.dk/content/cms/86/15686_dateca-aarsrapport-2021_offen_tlig-version.pdf. Accessed November 14, 2022.
8. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish national patient registry: a review of content, data quality, and research potential. *Clin Epidemiol.* 2015;7:449–490. doi:10.2147/CLEP.S91125
9. Erichsen R, Lash TL, Hamilton-Dutoit SJ, Bjerregaard B, Vyberg M, Pedersen L. Existing data sources for clinical epidemiology: the Danish national pathology registry and data bank. *Clin Epidemiol.* 2010;2(1):51–56. doi:10.2147/CLEP.S9908
10. Rasmussen LA, Jensen H, Virgilsen LF, et al. Identification of endometrial cancer recurrence – a validated algorithm based on nationwide Danish registries. *Acta Oncol.* 2021;60(4):452–458. doi:10.1080/0284186X.2020.1859133
11. Rasmussen LA, Jensen H, Virgilsen LF, Jensen JB, Vedsted P. A validated algorithm to identify recurrence of bladder cancer: a register-based study in Denmark. *Clin Epidemiol.* 2018;10:1755–1763. doi:10.2147/CLEP.S177305
12. Rasmussen LA, Jensen H, Virgilsen LF, Rosenkrantz L, Hölmich VP. A validated register-based algorithm to identify patients diagnosed with recurrence of malignant melanoma in Denmark. *Clin Epidemiol.* 2021;13:207–214. doi:10.2147/CLEP.S295844
13. Aagaard Rasmussen L, Jensen H, Flytkjær Virgilsen L, Jellesmark Thorsen LB, Vrou Offersen B, Vedsted P. A validated algorithm for register-based identification of patients with recurrence of breast cancer – based on Danish Breast Cancer Group (DBCG) data. *Cancer Epidemiol.* 2019;59 (January 2019):129–134. doi:10.1016/j.canep.2019.01.016
14. Lash TL, Riis AH, Ostenfeld EB, Erichsen R, Vyberg M, Thorlacius-Ussing O. A validated algorithm to ascertain colorectal cancer recurrence using registry resources in Denmark. *Int J Cancer.* 2015;136(9):2210–2215. doi:10.1002/ijc.29267
15. Pedersen RN, Öztürk B, Mellemkjær L, et al. Validation of an algorithm to ascertain late breast cancer recurrence using Danish medical registries. *Clin Epidemiol.* 2020;12:1083–1093. doi:10.2147/CLEP.S269962
16. Pedersen CB. The Danish civil registration system. *Scand J Public Health.* 2011;39(7):22–25. doi:10.1177/1403494810387965
17. Datecadatabase. National guidelines (in Danish); 2022. Available from: https://www.dmcc.dk/siteassets/kliniske-retningslinjer—skabeloner-og-vejledninger/kliniske-retningslinjer-opdelt-pa-dmcc/dateca/dateca_testikel-kraft_v2.0_admgodk041121.pdf. Accessed March 30, 2023.
18. Oldenburg J, Berney DM, Bokemeyer C, et al. Testicular seminoma and non-seminoma: ESMO-EURACAN clinical practice guideline for diagnosis, treatment and follow-up ☆. *Ann Oncol.* 2022;33(4):362–375. doi:10.1016/j.annonc.2022.01.002
19. Wagner T, Toft BG, Engvad B, et al. Prognostic factors for relapse in patients with clinical stage I testicular cancer: protocol for a Danish nationwide cohort study. *BMJ Open.* 2019;9(10):1–9. doi:10.1136/bmjopen-2019-033713
20. Bjerregaard B, Larsen OB. The Danish pathology register. *Scand J Public Health.* 2011;39(7):72–74. doi:10.1177/1403494810393563
21. Honecker F, Aparicio J, Berney D, et al. ESMO consensus conference on testicular germ cell cancer: diagnosis, treatment and follow-up. *Ann Oncol.* 2018;29(8):1658–1686. doi:10.1093/annonc/mdy217

22. Ciol MA, Hoffman JM, Dudgeon BJ, Shumway-Cook A, Yorkston KM, Chan L. Understanding the use of weights in the analysis of data from multistage surveys. *Arch Phys Med Rehabil.* 2006;87(2):299–303. doi:10.1016/j.apmr.2005.09.021
23. Mortensen MS, Lauritsen J, Gundgaard MG, et al. A nationwide cohort study of stage I seminoma patients followed on a surveillance program. *Eur Urol.* 2014;66(6):1172–1178. doi:10.1016/j.eururo.2014.07.001
24. Daugaard G, Gundgaard MG, Mortensen MS, et al. Surveillance for stage I nonseminoma testicular cancer: outcomes and long-term follow-up in a population-based cohort. *J Clin Oncol.* 2014;32(34):3817–3823. doi:10.1200/JCO.2013.53.5831
25. Kier MG, Lauritsen J, Mortensen MS, et al. Prognostic factors and treatment results after bleomycin, etoposide, and cisplatin in germ cell cancer: a population-based study. *Eur Urol.* 2017;71(2):290–298. doi:10.1016/j.eururo.2016.09.015
26. Chovanec M, Lauritsen J, Bandak M, et al. Late adverse effects and quality of life in survivors of testicular germ cell tumour. *Nat Rev Urol.* 2021;18(4):227–245. doi:10.1038/s41585-021-00440-w

Clinical Epidemiology

Dovepress

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>