



# Development and Validation of an Algorithm to Achieve Complete Registration of Danish Patients with Carcinoma In Situ of the Breast, 2008–2023

Martin Kjær Simonsen <sup>1-3,\*</sup>, Trine Lønbo Grantzau <sup>4,\*</sup>, Maj-Britt Jensen <sup>5</sup>, Deirdre Cronin-Fenton <sup>6</sup>, Birgitte Vrou Offersen <sup>2,3,7,8</sup>, Trine Tramm <sup>1,3</sup>

<sup>1</sup>Department of Pathology, Aarhus University Hospital, Aarhus, Denmark; <sup>2</sup>Department of Experimental Clinical Oncology, Aarhus University Hospital, Aarhus, Denmark; <sup>3</sup>Department of Clinical Medicine, Faculty of Health, Aarhus University, Aarhus, Denmark; <sup>4</sup>Department of Pathology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark; <sup>5</sup>Danish Breast Cancer Group, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark; <sup>6</sup>Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark; <sup>7</sup>Department of Oncology, Aarhus University Hospital, Aarhus, Denmark; <sup>8</sup>Danish Center for Particle Therapy, Aarhus University Hospital, Aarhus, Denmark

\*These authors contributed equally to this work

Correspondence: Martin Kjær Simonsen, Department of Pathology, Aarhus University Hospital, Palle Juul Jensens Boulevard 99, Aarhus, 8200 Aarhus N, Denmark, Tel +45 26 80 94 91, Email [mks@oncology.au.dk](mailto:mks@oncology.au.dk)

**Purpose:** Registration of carcinoma in situ (CIS) lesions in the national Danish Breast Cancer Group (DBCG) database is incomplete. We aimed to develop and validate a computer-based algorithm designed to identify patients with pure CIS of the breast, defined as CIS without concurrent breast cancer (BC), and with no history of BC or ipsilateral CIS. A secondary aim was to assess the completeness of CIS registration in the DBCG database.

**Patients and Methods:** We developed an algorithm to identify patients diagnosed with pure CIS between 2008 and 2017 using Systematized Nomenclature of MEDicine codes from the Danish Pathology Registry (DPR). We manually reviewed pathology records of a subcohort from Aarhus University Hospital (AUH) to serve as reference standard. To test reproducibility, the algorithm was reapplied to DPR data spanning 2008 to 2023.

**Results:** Between 2008 and 2017, 13,720 patients diagnosed with CIS of any kind were identified in the DPR. Of these, 1,581 patients were diagnosed at AUH, and 572 of those were manually confirmed as cases of pure CIS. The algorithm correctly classified 560 of these cases, corresponding to a sensitivity of 97.9% (95% CI: 96.3–99.3%). The positive predictive value of the algorithm for identifying pure CIS was 100.0% (95% CI: 99.3–100.0%). On a national level, the algorithm identified 4,302 patients with pure CIS between 2008 and 2017. Of those, 1,002 (23%) were not registered in the DBCG database. In the extended study population, 7,206 patients with pure CIS were identified, and 4,266 (99.2%) cases from the original study population were reclassified as pure CIS. Of 7,206 total patients, 1,827 (25%) were not registered in the DBCG database.

**Conclusion:** Demonstrating high accuracy and reproducibility, the algorithm represents an optimal method for future identification of patients with pure CIS who are not registered in the DBCG database.

**Keywords:** epidemiology, carcinoma in situ, Systematized Nomenclature of MEDicine codes, pathology databases, algorithm development

## Introduction

Carcinoma in situ (CIS) of the breast refers to the presence of malignant epithelial cells confined to the ducts or lobules, without invasion beyond the basement membrane.<sup>1</sup> Several types of CIS of the breast exist, with ductal carcinoma in situ (DCIS) constituting more than 85% of CIS cases.<sup>2</sup> Lobular carcinoma in situ (LCIS), pleomorphic lobular carcinoma in situ (PLCIS), and other subtypes are less common. DCIS is considered a non-obligate precursor lesion to breast cancer (BC), as, in untreated cases, 25–60% of patients with DCIS may develop cancer in the same breast within 9–24 years of follow-up.<sup>3–7</sup>

Few patients with DCIS present with symptoms, and more than 80% of lesions are detected through routine mammographic screening.<sup>8</sup> Consequently, the incidence of DCIS has risen four- to fivefold over recent decades in countries where national breast screening programs have been introduced.<sup>9–12</sup> DCIS now constitutes 15–25% of all screen-detected neoplastic lesions of the breast.<sup>13,14</sup> This trend is also observed in Denmark, where nationwide mammography screening was implemented in 2008 and is offered to all women aged 50–69, every other year, with 80–86% of eligible women participating.<sup>15</sup> In Denmark, DCIS constituted 13–14% of screen-detected lesions during the roll-out of population-based screening.<sup>16</sup>

As no current analysis can accurately identify DCIS lesions without the potential for progression to BC,<sup>17,18</sup> treatment for DCIS is multimodal and extensive, involving surgery, radiotherapy (RT), and, in some countries, endocrine treatment (ET).<sup>8,14,19</sup> Both adjuvant RT and ET reduce the risk of progression to BC or local recurrence of DCIS, but have no significant impact on overall survival.<sup>20–22</sup> Moreover, the increased incidence and treatment of DCIS have not reduced breast cancer incidence in countries with long-established national mammography screening, and the BC-specific mortality after DCIS remains low (1–3%).<sup>23–26</sup> This has raised concerns about potential overdiagnosis and overtreatment of some DCIS patients, and several ongoing non-inferiority trials investigate the impact of omitting surgery<sup>27–29</sup> or RT<sup>30</sup> for low-risk DCIS lesions. However, evaluating the effect of local treatment is challenging, as it can be difficult to distinguish whether an invasive lesion developing after DCIS treatment reflects progression due to insufficient therapy or represents a new primary malignancy.<sup>31</sup>

The Danish Breast Cancer Group (DBCG) HYPO trial investigated the efficacy of moderately hypofractionated RT in patients with DCIS or BC,<sup>32</sup> and the use of sentinel node biopsy for treatment of DCIS in Denmark has also been investigated.<sup>33</sup> Apart from those studies, the clinical outcome and prognosis of Danish patients with CIS have not been described since the introduction of nationwide BC screening in 2008.<sup>34</sup> Given the evolution of diagnostic methods, treatment modalities, and guidelines since then, there is a need for current and comprehensive data on Danish patients with CIS.

Therefore, we aimed to develop and subsequently evaluate the accuracy of a computer-based algorithm designed to identify all Danish patients diagnosed in the national Danish Pathology Registry (DPR) with pure CIS, defined as no concurrent or prior BC and no prior ipsilateral CIS, after the introduction of population-based mammography screening. We also aimed to assess the completeness of CIS registration within the DBCG database and explore the potential for supplementing these records through automated data extraction from the DPR.

## Materials and Methods

### Data Sources

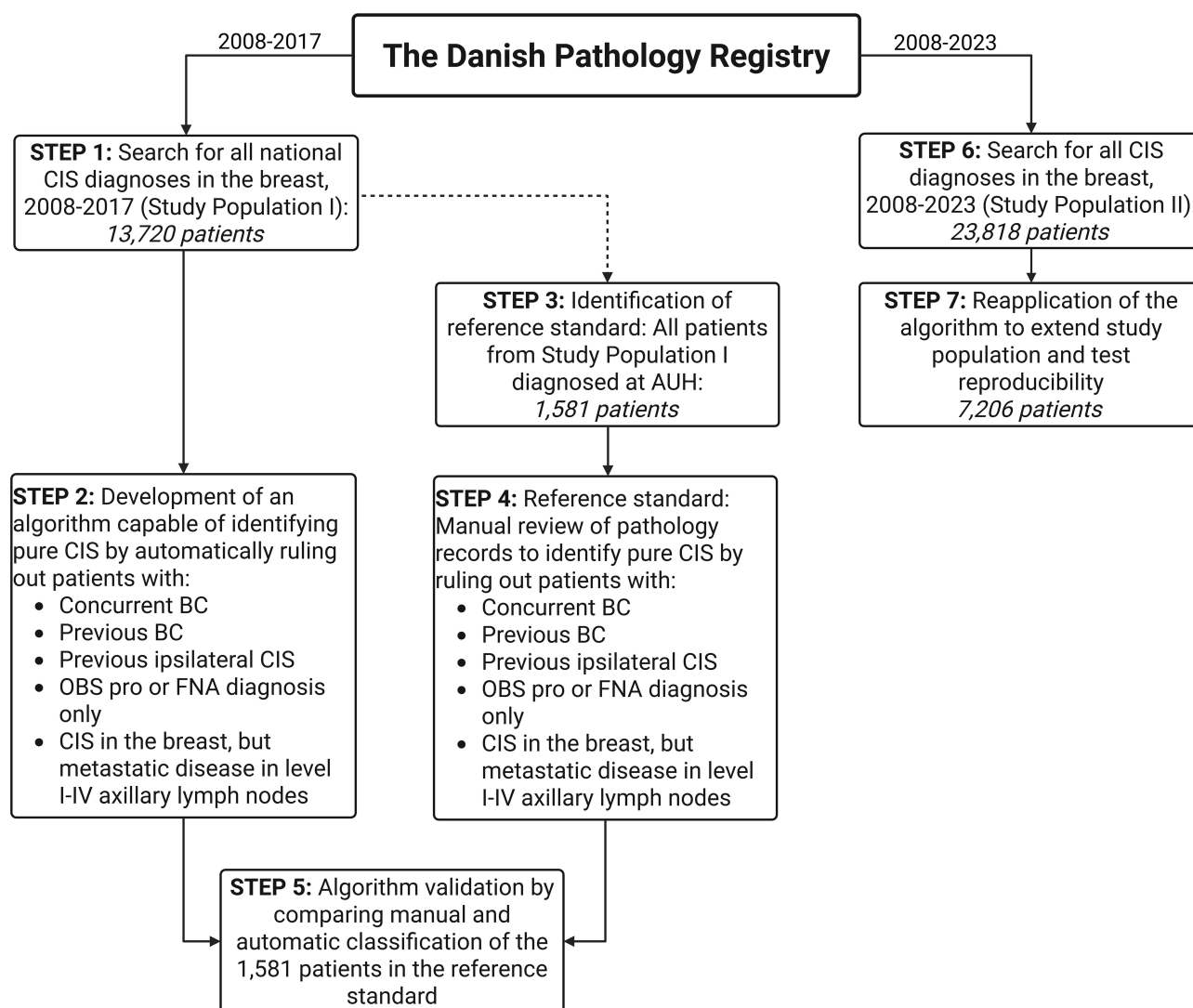
The DPR contains nationwide data for all tissue and cytology samples collected since 1997. Registration is mandatory, and the registry has been considered complete since 2000.<sup>35,36</sup> The DPR is updated daily, and includes numerous patient- and specimen-specific variables. Registration includes a sample-specific unique ID number, specimen type, gross and microscopic descriptions, and the investigating pathology department. All specimens are coded based on the Systematized Nomenclature of MEDicine (SNOMED) clinical terms classification system, and must as a minimum include topography (T-codes) and morphology codes (M-codes). Additional codes, including procedures or histopathological techniques (P-codes), function (F-codes), and etiology (AE-codes), are often added. Officially updated Danish SNOMED codes and the dates of code changes or modifications are available on the DPR Website (<http://www.patobank.dk/>).

The DBCG was established in 1977 to ensure uniform national guidelines for diagnosis and treatment of BC in Denmark.<sup>35</sup> All departments involved in BC care contribute patho-anatomical, surgical, and oncological patient data of clinical and prognostic importance. The database for invasive lesions is financially supported and includes query systems that ensure high coverage and data completeness. As a result, the completeness of BC lesions registered in the DBCG database has gradually improved to more than 95% in recent years.<sup>36</sup> While the DBCG database includes cases of in situ lesions, the absence of comparable financial support for data completeness results in a higher degree of missing or incomplete information. For this study, DBCG registrations were considered complete if they contained all descriptive variables in the DBCG reporting system that were selected for their relevance to diagnostic accuracy and clinical management. These include patho-anatomical, surgical, and oncological variables and are summarized in [Supplementary Table 1](#). In Denmark, all residents are assigned

a unique Central Person Register number, allowing individual linkage of data across multiple registries. This linkage enabled assessment of the completeness of the DBCG database and automatic supplementation of existing registrations with data extracted from SNOMED codes.

## Source Population

The source population consisted of all women in Denmark aged  $\geq 18$  years who were diagnosed with any type of CIS of the breast. These included patients diagnosed with DCIS, PLCIS, florid-subtype LCIS, and classical LCIS. The initial study period encompassed all patients diagnosed between January 1, 2008, and December 31, 2017, referred to as *Study Population I*. To test the reproducibility of the algorithm, we expanded the study period to include all patients diagnosed between January 1, 2008, and December 31, 2023, referred to as *Study Population II* (Figure 1). Therefore, there is an expected overlap between the patients in Study Population I and those in the early period of Study Population II, allowing assessment of reproducibility. Data on CIS and BC lesions were obtained from the DPR.



**Figure 1** Flowchart of algorithm development and validation for identifying cases of pure CIS.

**Abbreviations:** CIS, carcinoma in situ; AUH, Aarhus University Hospital; BC, breast cancer; OBS Pro, observation for provisional diagnosis; FNA, fine-needle aspiration.

## Algorithm for Identification of Pure CIS

From the DPR, patients diagnosed with CIS of the breast for Study Populations I and II were identified using the following criteria: All DPR registered CIS cases were identified based on SNOMED code combinations, including the T-codes for the breast or papilla/areola (ie, T04XXX), combined with the M-codes M85XX2 or M85403\* for CIS lesions of the breast (\*code applicable for Paget's disease of the nipple until December 31, 2010) (See [Supplementary Material 1](#) for all included SNOMED codes). This yielded a total of 13,720 and 23,818 patients diagnosed with any CIS in Study Populations I and II, respectively.

To restrict the study population to include pure CIS of the breast only, information on all previous invasive or CIS lesions from January 1, 1970, to December 31, 2023, for all patients was extracted from the DPR using the following criteria: 1) Any T-code in combination with morphology codes M8 and M9 with the number 3 (carcinoma), 6 (carcinoma metastasis), 7 (carcinoma recurrence), and 9 (carcinoma, uncertain whether primary or metastatic) in the fifth position (ie, M8XXX3); and 2) T-codes T04XXX for the breast or papilla/areola in combination with the morphology codes M85XX2 or M85403 for CIS lesions of the breast. Patients with a history of BC or ipsilateral CIS of the breast diagnosed before January 1, 2008, were identified and excluded. Patients with an earlier diagnosis of BC in the DBCG database were also excluded. To avoid inclusion of concurrent BC or patients diagnosed with CIS on biopsy, but upgraded to invasive disease on the surgical specimen, patients diagnosed with invasive disease 90 days before or after the primary CIS diagnosis were disregarded. The following cases were also excluded: 1) CIS diagnosis exclusively based on fine-needle aspiration (FNA), 2) CIS diagnosis suspected but not definitive (observation for provisional diagnosis (OBS pro)), 3) cases of microinvasive DCIS, and 4) cases of pure CIS of the breast but with regional node metastasis (see [Supplementary Table 2](#) for full code text) ([Figure 1](#)). Additionally, automatic extraction of histopathological data using SNOMED codes (ie, lesion size, resection margins, and tumor grade) was incorporated into the algorithm.

## Reference Standard

To validate the algorithm, we assembled a sub-cohort consisting of all patients diagnosed with CIS at Aarhus University Hospital (AUH) from Study Population I. The pathology records of each patient were manually reviewed and classified according to the predefined inclusion and exclusion criteria to ensure correct case assignment. This reference standard served as the gold standard for assessing the accuracy of the algorithm, comprising 1,581 patients ([Figure 1](#)).

## Statistical Analysis

The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the algorithm for identifying pure CIS cases were computed by comparing the patient classification by the algorithm to the reference standard. The Clopper-Pearson exact method was used to calculate binomial 95% confidence intervals (CI).

## Ethics and Approvals

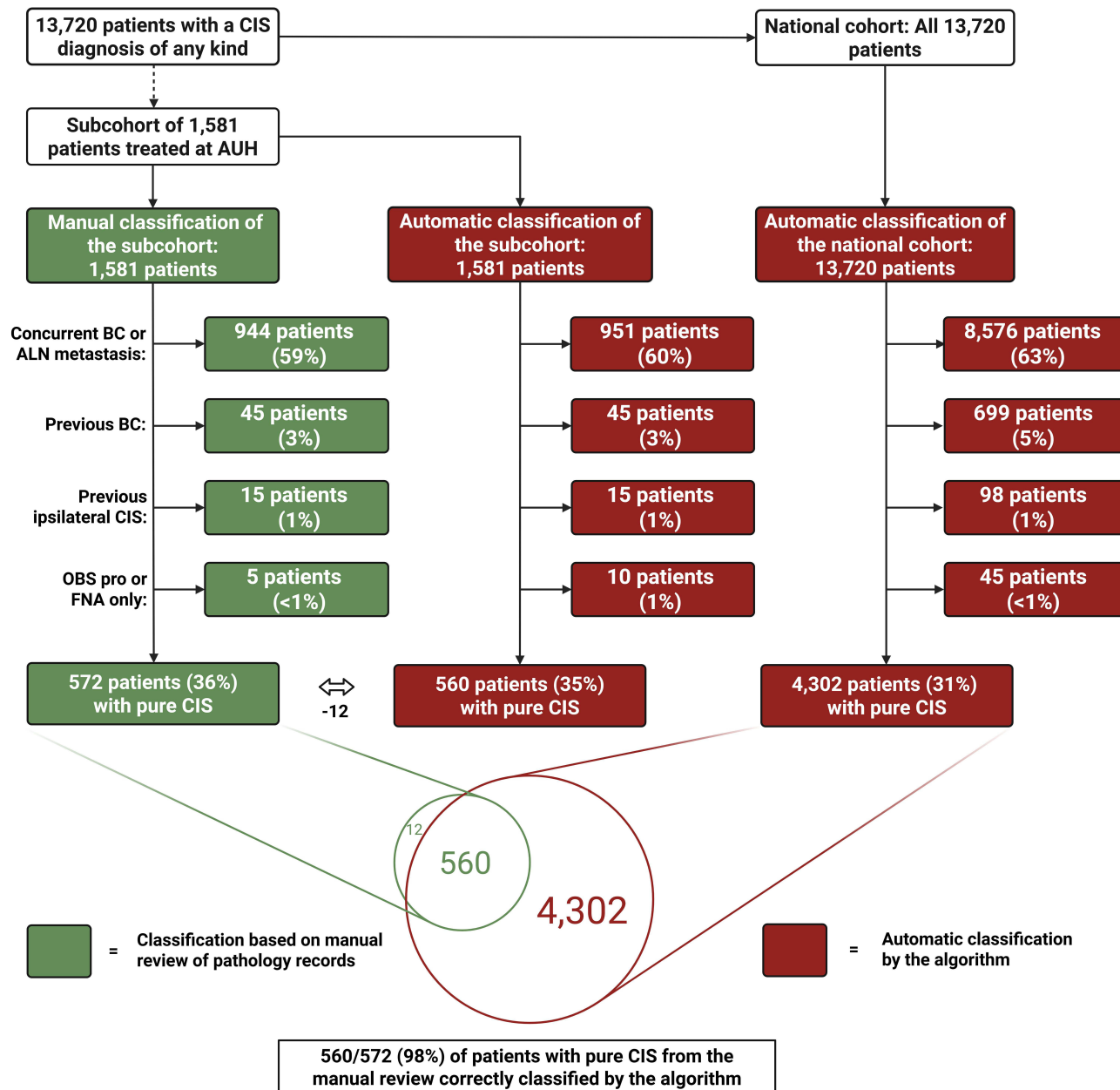
This study was initiated and supported by the DBCG. This study was approved by the Regional Committee on Health Research Ethics for the Central Denmark Region (journal no. 1–10-72-11-20) and registered in the Internal Research Project Registry of the Central Denmark Region (journal no. 1–16-02-456-19). The data has been accessed and processed under the Danish Act on Processing of Personal Data and Health Act (“Databeskyttelsesforordningen og Databeskyttelsesloven”).

## Results

Study Population I included 13,720 patients who were diagnosed with CIS of the breast of any kind. The algorithm classified 8,576 (63%) as having concurrent BC, 699 (5%) as having previous BC, 98 (1%) as having previous CIS, and 45 (<1%) as having an OBS pro or FNA diagnosis only, leaving 4,302 (31%) classified as incident cases of pure CIS. The distribution of the manually reviewed reference standard was similar ([Figure 2](#)). Of the 572 manually confirmed cases of pure CIS in the reference standard, 560 were correctly classified by the algorithm (sensitivity 97.9%, 95% CI: 96.4–98.9) ([Figure 2](#)). Five of the 12 misclassified patients were due to an initial suspicion of CIS that was later confirmed, but the

# The Danish Pathology Registry

Search for all CIS diagnoses in the breast, 2008-2017



**Figure 2** Patient classification by the algorithm compared to the reference standard.

**Abbreviations:** CIS, carcinoma in situ; AUH, Aarhus University Hospital; BC, breast cancer; ALN, axillary lymph node; OBS pro, observation for provisional diagnosis; FNA, fine-needle aspiration.

SNOMED code for an OBS pro diagnosis (ÆYYYY00) remained in the DPR data. The algorithm classified the remaining seven patients as having concurrent BC, either because an initial diagnosis of invasive carcinoma was later revised to carcinoma in situ (three patients) or because of an erroneous assignment of the general SNOMED code for carcinoma (M80103) to samples without histological evidence of carcinoma. The positive predictive value (PPV) of the algorithm for identifying pure CIS from any CIS diagnosis was 100.0% (95% CI: 99.4–100.0%), and the negative predictive value (NPV) was 98.8% (95% CI: 98.0–99.3%) (Table 1).

Of the 4,302 patients identified as cases of pure CIS by the algorithm in Study Population I, 1,002 (23%) were not registered in the DBCG database. Of the remaining patients, 2,709 (63%) were registered in the database with varying

**Table 1** Validation Metrics for the Algorithm

Reference Standard	Algorithm		
Pure CIS	Pure CIS 560	Excluded 12	In total 572
Excluded	0	1,009	1,009
In total	560	1,021	1,581
<b>Results</b>	<b>% (95% CI)</b>		
Sensitivity (TP/(TP+FN))	97.9% (96.4–98.9%)		
Specificity (TN/(TN+FP))	100.0% (99.6–100.0%)		
PPV (TP/(TP+FP))	100.0% (99.4–100.0%)		
NPV (TN/(TN+FP))	98.8% (98.0–99.3%)		

**Abbreviations:** CIS, carcinoma in situ; TP, true positive; FN, false negative; TN, true negative; FP, false positive; PPV, positive predictive value; NPV, negative predictive value.

degrees of missing information, leaving only 591 (14%) with a complete DBCG registration (Figure 3). In Study Population II, 23,818 patients with any CIS diagnosis were identified from the DPR. The algorithm classified 7,206 of those as having pure CIS. Of the 4,302 pure CIS cases in Study Population I, 4,266 (99.2%) were reclassified as such by the algorithm in Study Population II. The discrepancy of 36 patients was mainly due to the inclusion of the SNOMED code applicable for Paget’s disease of the nipple until December 31, 2010 (M85403) in the DPR data extraction criteria for Study Population II, and not Study Population I. As this code often co-occurred with T-codes representing biopsies from the papilla/areola complex, which lacked information on laterality, the corresponding DPR entries could not be assigned to a specific side (ie either the left or right breast). Consequently, these patients were excluded by the algorithm. However, the presence of M85403 also enabled the identification of 40 additional patients with pure CIS (diagnosed in 2008–2017) in Study Population II, who were not captured in Study Population I.

Of the 7,206 patients in Study Population II with pure CIS, 1,827 (25%) were not registered in the DBCG database. The temporal and geographical variations in DBCG registrations are shown in Figure 3. The inter-departmental variation in registrations was low, whereas the frequency of completely registered patients has been increasing in recent years.

Table 2 presents the data completeness for Study Population I across the two data sources and the reference standard. Most variables in the DBCG database had completion rates between 80% and 90%, with higher rates observed for pathological variables related to DCIS such as lesion size, surgical margins, and Van Nuys classification. Several variables could be modestly improved, typically by 1–3 percentage points, by adding data extracted from SNOMED codes. In the reference standard, data completeness was high, with the completion rate of most variables exceeding 90%.

## Discussion

With almost perfect sensitivity and PPV, the computer-based algorithm proved to be highly reliable in identifying Danish patients with pure CIS using data from the DPR. All 12 misclassified patients in the reference cohort were attributable to an erroneous use of SNOMED codes, suggesting that, under current coding practices, the algorithm performs at its theoretical maximum. The algorithm successfully reclassified most patients with pure CIS in Study Population I as pure CIS in Study Population II, demonstrating high consistency and reproducibility. The slight discrepancy was confined to cases diagnosed between 2008 and 2010 and did not occur among patients diagnosed after 2010. As one-quarter of the patients identified by the algorithm were not registered in the DBCG database, and several registered cases contained incomplete information, the completeness of the DBCG database for CIS lesions could be further improved.

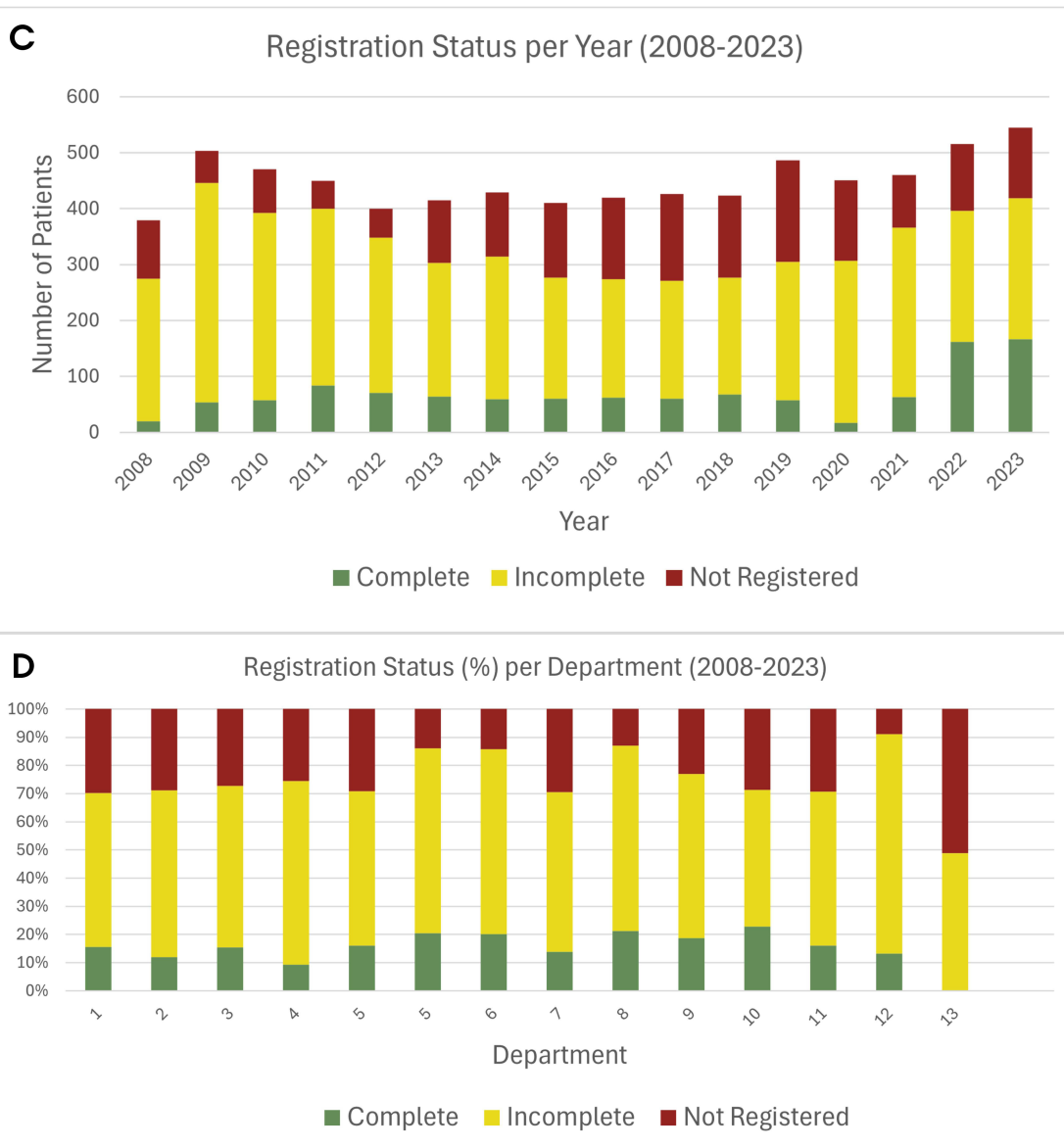
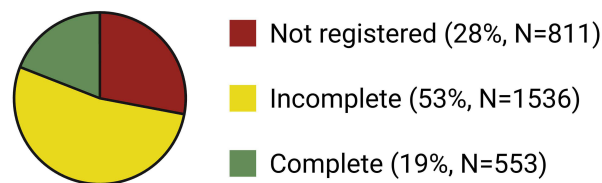
In a similar American study, two algorithms were used to differentiate between patients with pure DCIS and BC, based on registry data. The first algorithm had a sensitivity, specificity, and PPV of 79%, 89%, and 62%, respectively. The second algorithm had a sensitivity of 50%, a specificity of 97%, and a PPV of 77%.<sup>37</sup> The study was later retracted due to misclassification of some patients, but highlights the challenges in accurately identifying pure CIS using registry data.

Previous studies in Denmark have utilized DPR SNOMED codes either to validate the national usage of the codes,<sup>38</sup> or to identify recurrences in specific cancer groups.<sup>39–42</sup> High sensitivities (88–100%), PPV (73–94%), and NPV

**A: 2008-2017**



**B: 2018-2023**



**Figure 3** Temporal and geographical variance for in situ registrations in the Danish Breast Cancer Group database. **Notes:** (A) Completeness of the database, 2008–2017; (B) Completeness of the database, 2018–2023; (C) Temporal variation in completeness; (D) Geographical variation in completeness.

(84–100%) were also observed during the validation of these algorithms. Although the similarities between these studies are limited, this pattern highlights the untapped potential and high quality of the DPR.

In this study, a substantial manual effort was dedicated to reviewing pathology records to create a reference standard, ensuring accurate validation of the algorithm. The high sensitivity, specificity, PPV, and NPV of the algorithm in identifying pure CIS cases minimized the risk of patient misclassification on a national level. The validated and

**Table 2** Variable Registration of in situ Lesions in the DBCG Database Compared to Manual Retrieval of Data from Pathology Records, 2008–2017: Number of Patients with Documented Information per Variable

	Data Registered in the DBCG Database (N=3,300)	Data Registered in the DBCG Database + Additional Data from SNOMED Codes (n=3,300)	Reference Standard, Manual Retrieval of Data (n=572)
Variable (Type of Data Reported in the DBCG Database)	n with a Registration/ Total n Available (%)	n with a Registration/Total n Available (%)	n with Data Description in Records/Total n Available <sup>a</sup> (%)
<b>Patho-anatomical information:</b>			
Hospital department (categorical)	2,986/3,300 (90.5%)	3,300/3,300 (100.0%)	572/572 (100.0%)
Specimen number (number)	2,968/3,300 (89.9%)	3,300/3,300 (100.0%)	572/572 (100.0%)
Laterality (left/right)	3,294/3,300 (99.9%)	3,300/3,300 (100.0%)	572/572 (100.0%)
DCIS (yes/no/unknown)	2,975/3,300 (90.1%)	3,076/3,300 (93.2%)	572/572 (100.0%)
Pure micropapillary type (yes/no/unknown)	2,908/3,300 (88.0%)	2,911/3,300 (88.2%)	536/572 (93.8%)
LCIS (yes/no/unknown)	2,892/3,300 (87.6%)	2,916/3,300 (88.3%)	571/572 (99.8%)
PLCIS (yes/no/unknown)	2,863/3,300 (86.8%)	2,867/3,300 (86.9%)	570/572 (99.6%)
PDN (yes/no/unknown)	2,870/3,300 (86.9%)	2,881/3,300 (87.3%)	569/572 (99.4%)
Microcalcifications (yes/no/unknown)	2,858/3,300 (86.6%)	2,898/3,300 (87.8%)	499/572 (87.2%)
Lesion size, DCIS (mm) <sup>b</sup>	2,716/2,891 (93.4%)	2,725/2,891 (94.2%)	502/539 (93.1%)
Lesion size, PLCIS (mm) <sup>b</sup>	24/42 (57.1%)	28/46 (61.9%)	6/6 (100.0%)
Surgical margins, DCIS (mm) <sup>b</sup>	2,787/2,891 (96.4%)	2,843/2,891 (98.6%)	514/539 (95.3%)
Surgical margins, PLCIS (mm) <sup>b</sup>	30/42 (71.1%)	34/46 (73.9%)	6/6 (100.0%)
Nuclear grade, DCIS (1/2/3) <sup>b</sup>	2,852/2,891 (98.6%)	2,852/2,891 (98.6%)	489/539 (90.7%)
Necrosis, DCIS (yes/no) <sup>b</sup>	2,714/2,891 (93.9%)	2,714/2,891 (93.9%)	497/539 (92.2%)
Van Nuys group, DCIS (1/2/3) <sup>b</sup>	2,853/2,891 (98.7%)	2,874/2,891 (99.4%)	534/539 (99.1%)
Sentinel Node procedure (yes/no)	2,983/3,300 (90.3%)	3,057/3,300 (92.6%)	572/572 (100.0%)
Sentinel nodes removed (1 2 3 >3) <sup>b</sup>	1,697/1,712 (99.1%)	1,771/1,786 (99.1%)	225/225 (100.0%)
<b>Surgical information:</b>			
Biopsy date (dd.mm.yyyy) <sup>c</sup>	2,841/3,300 (86.1%)	2,970/3,300 (90.0%)	479/572 (83.7%)
Lumpectomy date (dd.mm.yyyy) <sup>c</sup>	2,626/3,300 (79.5%)	2,773/3,300 (84.0%)	402/572 (70.3%)
Mastectomy date (dd.mm.yyyy) <sup>c</sup>	942/3,300 (28.5%)	1,024/3,300 (31.0%)	167/572 (29.2%)
Palpable tumor (yes/no)	2,896/3,300 (87.7%)	2,896/3,300 (87.7%)	N/A
Clinical PDN (yes/no)	2,855/3,300 (86.5%)	2,855/3,300 (86.5%)	N/A
Profound fascia on specimen (yes/no)	2,852/3,300 (86.4%)	2,852/3,300 (86.4%)	506/572 (88.5%)
Tumor on imaging (yes/no)	2,717/3,300 (82.3%)	2,717/3,300 (82.3%)	N/A
Tissue density on imaging (yes/no)	2,691/3,300 (81.5%)	2,691/3,300 (81.5%)	N/A

(Continued)

Table 2 (Continued).

	Data Registered in the DBCG Database (N=3,300)	Data Registered in the DBCG Database + Additional Data from SNOMED Codes (n=3,300)	Reference Standard, Manual Retrieval of Data (n=572)
Variable (Type of Data Reported in the DBCG Database)	n with a Registration/ Total n Available (%)	n with a Registration/Total n Available (%)	n with Data Description in Records/Total n Available <sup>a</sup> (%)
Microcalcifications (yes/no)	2,810/3,300 (85.1%)	2,810/3,300 (85.1%)	N/A
Quadrant (upper lateral/upper medial/lower lateral/lower medial/central)	2,772/3,300 (84.0%)	2,772/3,300 (84.0%)	N/A
<b>Oncological information:</b>			
Menopausal status (categorical)	2,386/3,300 (72.3%)	2,386/3,300 (72.3%)	N/A
Adjuvant radiotherapy (yes/no) <sup>b</sup>	1,846/2,298 (55.9%)	1,846/2,298 (55.9%)	N/A
Indication for follow-up (yes/no)	3,011/3,300 (91.2%)	3,011/3,300 (91.2%)	N/A

**Notes:** Column 1 represents the 3,300 patients registered in the DBCG database and identified as pure CIS by the algorithm. Column 2 represents the data from column 1 with the addition of data extracted automatically from SNOMED codes using the algorithm. Column 3 represents the standard reference cohort, and thus manually retrieved data from pathology records. <sup>a</sup>In the reference standard, only information documented in the pathology reports was considered. Thus, unknown values were disregarded. <sup>b</sup>Variables applicable only to relevant patient subgroups. <sup>c</sup>Some patients are represented in multiple procedure groups due to the documentation of more than one procedure.

**Abbreviations:** DBCG, Danish Breast Cancer Group; SNOMED, Systematized Nomenclature of Medicine; DCIS, ductal carcinoma in situ; LCIS, lobular carcinoma in situ; PLICIS, pleomorphic LCIS; PDN, Paget's disease of the nipple.

comprehensive national registry of pathological records within the DPR represents a robust, systematic, and highly effective system of data collection and management. No other pathology registries worldwide have achieved the level of completeness observed in the DPR.<sup>43</sup>

Although usage of SNOMED codes is nationally standardized, some degree of personal, institutional and temporal variation cannot be excluded, potentially leading to misclassification of patients. As the algorithm was only validated in a single-institution cohort, a potential limitation of its external validity on a national level cannot be excluded. However, the low inter-departmental variation in registrations suggests that geographical differences are minimal.

## Conclusion

The registration of Danish patients with CIS in the DBCG database is incomplete, limiting the reliability of the current registry data for clinical decision-making and academic research. The computer-based algorithm developed in this study appears to be an optimal tool for identifying CIS cases not registered in DBCG. The identified patients will form the basis of a national cohort study investigating the clinical outcomes and recurrence patterns following a diagnosis of CIS.

## Acknowledgments

This project was financially supported by the Danish Cancer Society (R231-A13893, applicant Trine Tramm), the Health Research Foundation of Central Denmark Region, and the Danish Cancer Research Foundation.

## Disclosure

Dr Maj-Britt Jensen reports personal fees from Novartis, outside the submitted work. The authors report no other conflicts of interest in this work.

## References

- Allred DC. Ductal carcinoma in situ: terminology, classification, and natural history. *J Natl Cancer Inst Monogr.* 2010;2010(41):134–138. doi:10.1093/jncimonographs/igq035
- Timbres J, Kohut K, Caneppele M, et al. DCIS and LCIS: are the risk factors for developing in situ breast cancer different? *Cancers.* 2023;15(17):4397. doi:10.3390/cancers15174397
- Sanders ME, Schuyler PA, Dupont WD, Page DL. The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. *Cancer.* 2005;103(12):2481–2484. doi:10.1002/cncr.21069
- Collins LC, Tamimi RM, Baer HJ, Connolly JL, Colditz GA, Schnitt SJ. Outcome of patients with ductal carcinoma in situ untreated after diagnostic biopsy: results from the nursespaget health study. *Cancer.* 2005;103(9):1778–1784. doi:10.1002/cncr.20979
- Page DL, Dupont WD, Rogers LW, Landenberger M. Intraductal carcinoma of the breast: follow-up after biopsy only. *Cancer.* 1982;49(4):751–758. doi:10.1002/1097-0142(19820215)49:4<751::AID-CNCR2820490426>3.0.CO;2-Y
- Betsill WL, Rosen PP, Lieberman PH, Robbins GF. Intraductal carcinoma. Long-term follow-up after treatment by biopsy alone. *JAMA.* 1978;239(18):1863–1867. doi:10.1001/jama.1978.03280450035020
- Wang J, Li B, Luo M, et al. Progression from ductal carcinoma in situ to invasive breast cancer: molecular features and clinical significance. *Signal Transduct Target Ther.* 2024;9(1):83. doi:10.1038/s41392-024-01779-3
- Kalwaniya DS, Gairola M, Gupta S, Pawan G. Ductal carcinoma in situ: a detailed review of current practices. *Cureus.* 2023;15(4):e37932. doi:10.7759/cureus.37932
- Virnig BA, Tuttle TM, Shamlivan T, Kane RL. Ductal carcinoma in situ of the breast: a systematic review of incidence, treatment, and outcomes. *J Natl Cancer Inst.* 2010;102(3):170–178. doi:10.1093/jnci/djp482
- Sorum R, Hofvind S, Skaane P, Haldorsen T. Trends in incidence of ductal carcinoma in situ: the effect of a population-based screening programme. *Breast.* 2010;19(6):499–505. doi:10.1016/j.breast.2010.05.014
- van Steenberghe LN, Voogd AC, Roukema JA, et al. Screening caused rising incidence rates of ductal carcinoma in situ of the breast. *Breast Cancer Res Treat.* 2009;115(1):181–183. doi:10.1007/s10549-008-0067-5
- Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. *N Engl J Med.* 2012;367(21):1998–2005. doi:10.1056/NEJMoa1206809
- Barrio AV, Van Zee KJ. Controversies in the treatment of ductal carcinoma in situ. *Annu Rev Med.* 2017;68(1):197–211. doi:10.1146/annurev-med-050715-104920
- Delaloge S, Khan SA, Wesseling J, Whelan T. Ductal carcinoma in situ of the breast: finding the balance between overtreatment and undertreatment. *Lancet.* 2024;403(10445):2734–2746. doi:10.1016/S0140-6736(24)00425-2
- Christiansen P, Vejborg I, Kroman N, et al. Position paper: breast cancer screening, diagnosis, and treatment in Denmark. *Acta Oncol.* 2014;53(4):433–444. doi:10.3109/0284186X.2013.874573
- Lynge E, Bak M, von Euler-Chelpin M, et al. Outcome of breast cancer screening in Denmark. *BMC Cancer.* 2017;17(1):897. doi:10.1186/s12885-017-3929-6
- Dettoni RS, Stur E, Laus AC, et al. Potential biomarkers of ductal carcinoma in situ progression. *BMC Cancer.* 2020;20(1):119. doi:10.1186/s12885-020-6608-y
- Casasent AK, Almekinders MM, Mulder C, et al. Learning to distinguish progressive and non-progressive ductal carcinoma in situ. *Nat Rev Cancer.* 2022;22(12):663–678. doi:10.1038/s41568-022-00512-y
- Martinez-Perez C, Turnbull AK, Ekatah GE, et al. Current treatment trends and the need for better predictive tools in the management of ductal carcinoma in situ of the breast. *Cancer Treat Rev.* 2017;55:163–172. doi:10.1016/j.ctrv.2017.03.009
- Early Breast Cancer Trialists' Collaborative G; Correa C, McGale P. Overview of the randomized trials of radiotherapy in ductal carcinoma in situ of the breast. *J Natl Cancer Inst Monogr.* 2010;2010(41):162–177.
- Stuart KE, Houssami N, Taylor R, Hayen A, Boyages J. Long-term outcomes of ductal carcinoma in situ of the breast: a systematic review, meta-analysis and meta-regression analysis. *BMC Cancer.* 2015;15(1):890. doi:10.1186/s12885-015-1904-7
- Kim R, Kawai A, Wakisaka M, et al. Outcomes in patients with non-invasive breast carcinoma. *Cancer Rep.* 2023;6(4):e1768. doi:10.1002/cnr2.1768
- Narod SA, Iqbal J, Giannakeas V, Sopik V, Sun P. Breast cancer mortality after a diagnosis of ductal carcinoma in situ. *JAMA Oncol.* 2015;1(7):888–896. doi:10.1001/jamaoncol.2015.2510
- Giannakeas V, Sopik V, Narod SA. Association of a diagnosis of ductal carcinoma in situ with death from breast cancer. *JAMA Network Open.* 2020;3(9):e2017124. doi:10.1001/jamanetworkopen.2020.17124
- Ernster VL, Barclay J, Kerlikowske K, Wilkie H, Ballard-Barbash R. Mortality among women with ductal carcinoma in situ of the breast in the population-based surveillance, epidemiology and end results program. *Arch Intern Med.* 2000;160(7):953–958. doi:10.1001/archinte.160.7.953
- Moller MH, Lousdal ML, Kristiansen IS, Stovring H. Effect of organized mammography screening on breast cancer mortality: a population-based cohort study in Norway. *Int J Cancer.* 2019;144(4):697–706. doi:10.1002/ijc.31832
- Youngwirth LM, Boughy JC, Hwang ES. Surgery versus monitoring and endocrine therapy for low-risk DCIS: the COMET Trial. *Bull Am Coll Surg.* 2017;102(1):62–63.
- Francis A, Fallowfield L, Rea D. The LORIS Trial: addressing overtreatment of ductal carcinoma in situ. *Clin Oncol.* 2015;27(1):6–8. doi:10.1016/j.clon.2014.09.015
- Elshof LE, Schaapveld M, Rutgers EJ, et al. The method of detection of ductal carcinoma in situ has no therapeutic implications: results of a population-based cohort study. *Breast Cancer Res.* 2017;19(1):26. doi:10.1186/s13058-017-0819-4
- Radiotherapy omission in low risk ductal in situ carcinoma breast (ROMANCE). Available from: <https://clinicaltrials.gov/study/NCT03878342>. Accessed January 19, 2025.
- Smith TE, Lee D, Turner BC, Carter D, Haffty BG. True recurrence vs. new primary ipsilateral breast tumor relapse: an analysis of clinical and pathologic differences and their implications in natural history, prognoses, and therapeutic management. *Int J Radiat Oncol Biol Phys.* 2000;48(5):1281–1289. doi:10.1016/S0360-3016(00)01378-X

32. Offeren BV, Alsner J, Nielsen HM, et al. Hypofractionated versus standard fractionated radiotherapy in patients with early breast cancer or ductal carcinoma in situ in a randomized Phase III trial: the DBCG HYPO Trial. *J Clin Oncol.* 2020;38(31):3615–3625. doi:10.1200/JCO.20.01363
33. Holm-Rasmussen EV, Jensen MB, Balslev E, Kroman N, Tvedskov TF. The use of sentinel lymph node biopsy in the treatment of breast ductal carcinoma in situ: a Danish population-based study. *Eur J Cancer.* 2017;87:1–9. doi:10.1016/j.ejca.2017.09.037
34. Laenkholm AV, Jensen MB, Kroman N, Rank F. Breast cancer in situ. from pre-malignant lesion of uncertain significance to well-defined non-invasive malignant lesion. the Danish breast cancer cooperative group register 1977-2007 revisited. *Acta Oncol.* 2008;47(4):765–771. doi:10.1080/02841860802043931
35. Christiansen P, Ejlersen B, Jensen MB, Mouridsen H. Danish breast cancer cooperative group. *Clin Epidemiol.* 2016;8:445–449. doi:10.2147/CLEP.S99457
36. DBCG. The Danish breast cancer group quality database for breast cancer – national annual report 2024. Available from: [https://sundk.dk/media/12rp4h4c/dbcg-a-rsrapport-2024\\_offentliggjort-version-20250627.pdf](https://sundk.dk/media/12rp4h4c/dbcg-a-rsrapport-2024_offentliggjort-version-20250627.pdf). Accessed July 13, 2025.
37. Hirth JM, Hatch SS, Lin YL, Giordano SH, Silva HC, Kuo YF. Development and validation of algorithms to differentiate ductal carcinoma in situ from invasive breast cancer within administrative claims data. *Cancer.* 2018;124(13):2815–2823. doi:10.1002/cncr.31393
38. Moller M, Bressendorff I, Borg R, et al. The validity of pathology codes for biopsy-confirmed kidney disease in the Danish national patobank. *Clin Kidney J.* 2024;17(8):sfae203. doi:10.1093/ckj/sfae203
39. 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
40. Pedersen RN, Ozturk B, Mellemkjaer 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
41. Cronin-Fenton D, Kjaersgaard A, Norgaard M, et al. Breast cancer recurrence, bone metastases, and visceral metastases in women with stage II and III breast cancer in Denmark. *Breast Cancer Res Treat.* 2018;167(2):517–528. doi:10.1007/s10549-017-4510-3
42. Aagaard Rasmussen L, Jensen H, Flytkjaer Virgilsen L, Jellesmark Thorsen LB, Vrou Offeren 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;59:129–134. doi:10.1016/j.canep.2019.01.016
43. 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;2:51–56. doi:10.2147/CLEP.S9908

## Clinical Epidemiology

### 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>

**Dovepress**  
Taylor & Francis Group