Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

METHOD

Evaluation of Left Truncation and Censoring When Changing the Use of the *International Classification of Diseases Eighth Revision* Codes to *Tenth Revision* Codes in the Danish National Patient Registry

Søren Korsgaard, Frederikke Schønfeldt Troelsen, Katalin Veres, Cecilia Hvitfeldt Fuglsang, Henrik Toft Sørensen

Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Aarhus, Denmark

Correspondence: Søren Korsgaard, Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Olof Palmes Allé 43–45, Aarhus N, 8200, Denmark, Tel +45 87167212, Email soren.korsgaard@clin.au.dk

Purpose: In the Danish National Patient Registry (DNPR), covering all Danish hospitals and widely used in research, diseases have been recorded using *International Classification of Diseases* (ICD) codes, transitioning from the *Eighth* to the *Tenth revision* in 1994. Uncertainty exists regarding whether including ICD-8 codes alongside ICD-10 is needed for complete disease identification. We assessed the extent of left-truncation and left-censoring in the DNPR arising from omitting ICD-8 codes.

Patients and Methods: We sampled 500,000 Danes ≥40 years of age in 1995, 2010, and 2018. From the DNPR, we identified cardiovascular, endocrine, gastrointestinal, neurological, pulmonary, rheumatic, and urogenital diseases as well as fractures. We obtained the number of people with a disease recorded with ICD-8 codes only (*ie*, the ICD-8 record would be left-truncated by not using ICD-8 codes), ICD-8 *plus* ICD-10 codes (*ie*, the ICD-8 record would be left-censored by not using ICD-8 codes), and ICD-10 codes only. For each ICD group, we calculated the proportion of people with the disease relative to the total sample (*ie*, 500,000 people) and the total number of people with the disease across all ICD groups.

Results: Overall, the left-truncation issue decreased over the years. Relative to all people with a disease, the left-truncated proportion was for example 59% in 1995 and <2% in 2018 for diabetes mellitus; 93% in 1995, and 54% in 2018 for appendicitis. The left-truncation issue increased with age group for most diseases. The proportion of disease records left-censored by not using ICD-8 codes was generally low but highest for chronic diseases.

Conclusion: The left-truncation issue diminished over sample years, particularly for chronic diseases, yet remained rather high for selected surgical diseases. The left-truncation issue increased with age group for most diseases. Left-censoring was overall a minor issue that primarily concerned chronic diseases.

Keywords: epidemiology, methodology, bias, left-truncation, left-censoring

Introduction

The Nordic countries have a long-standing tradition of conducting health research by using data from registries that routinely collect individual-level data for administrative purposes.[1] These registries are useful to efficiently conduct longitudinal studies enabling the examination of risk factors, prognostic factors, and outcomes of diseases, as well as the assessment of therapy utilization, effectiveness, and safety.[2] The national patient registries form a cornerstone of health research across all Nordic countries.[3] The Finnish Hospital Discharge Register contains complete nationwide data since 1969 and has included outpatient hospital care since 1998.[4] The Swedish National Patient Register contains complete nationwide data since 1987 and has included outpatient care since 2001.[5] The Norwegian Patient Registry contains

Received: 12 January 2024
Accepted: 2 May 2024
Published: 18 May 2024

complete nationwide data on in- and outpatient care since 2008.[6] The Danish National Patient Registry (DNPR) contains complete inpatient data since 1978 and has included outpatient care since 1995.[7] The DNPR is the most widely used registry in Danish health research. The Nordic national patient registries record a range of information associated with hospital contacts encompassing the date of contact, date of discharge, and associated diagnoses.[3–7] The diagnoses are coded according to the World Health Organization's *International Classification of Diseases* (ICD): a classification system developed to promote comparisons over time and among countries.[8] In the DNPR, the ICD *Eighth Revision* (ICD-8) was used until the end of 1993; thereafter the *Tenth Revision* (ICD-10) was used.[7]

Little evidence exists on whether ICD-8 codes should be included in disease-identification algorithms for complete disease identification.[9] The decision may depend on the study period; the specific variable of interest (*eg*, exposure, covariables, or outcome); data availability; and the sensitivity and specificity of the ICD-8 codes recorded in the registry. Nonetheless, not using ICD-8 codes requires consideration of issues that may ensue – particularly left-censoring and left-truncation. Left-censoring and left-truncation, in principle, are limitations for any registry that does not contain information on individuals starting from their birth dates.[2,9–11] Figure 1 illustrates the definition of the two terms as employed in this study. Using the DNPR as an example, hospital contacts before 1977 will not be identifiable because the contacts predated the initiation of registration in the DNPR. Thus, if a person suffered from appendicitis before 1977 and never had a recurrent event, this person will never be identified as having had appendicitis. The appendicitis event will thus be left-truncated. Left-censoring occurs when the first record identifiable in the registry is not the actual first-time event. For instance, a person might suffer from an incident myocardial infarction in 1976 and a recurrent myocardial infarction in 1996. Using the DNPR, the myocardial infarction event in 1976 is unidentifiable, resulting in the myocardial infarction event in 1996 being incorrectly classified as the incident event. The myocardial infarction event in 1976 is therefore left-censored in the DNPR. By not using ICD-8 codes, similar issues with left-truncation and left-censoring may ensue, but with 1994 instead of 1977 as the cut-off year. Given the increasing utilization of the DNPR for longitudinal observational studies, it is important to examine these key limitations and the potential biases that may follow. Additionally, left-censoring and left-truncation are not exclusive to the DNPR. Rather they are prominent in many

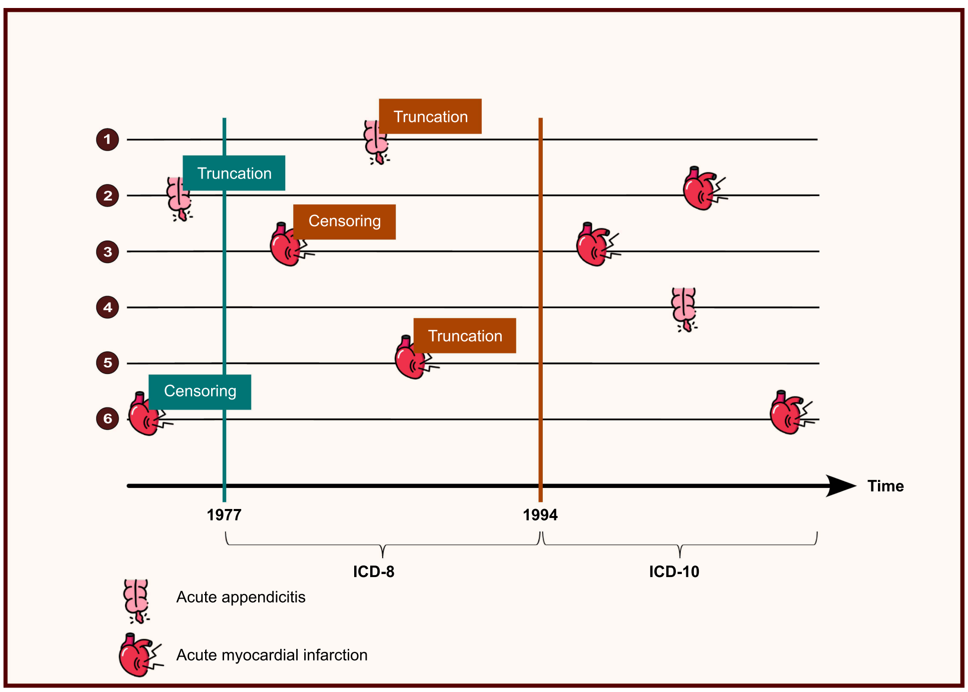

Figure 1 Illustration of left-truncation and left-censoring, using the Danish National Patient Registry as an example. Calendar time is depicted on the x-axis and divided into periods between 1977 and 1994, when ICD-8 codes were used, and ICD-10 codes thereafter. On the y-axis, six different individuals' hospital contacts with appendicitis and/or acute myocardial infarction are depicted. Events with a green label are truncated or censored in the registry because they occurred before registry establishment. Events with an Orange label would be truncated or censored if ICD-8 codes were not included in the disease-identification algorithm. Created with icons made by Freepik and Smashicons from www.flaticon.com.

international registries in which data collection occurs for relatively short periods (*eg*, the Medicare databases typically first begin the data collection at age 65[12]). To best address the issues of left-truncation and left-censoring, registries are needed that contain decades of non-age-dependent data.

Therefore, we used the DNPR to examine the issues of left-truncation and left-censoring introduced when ICD-8 codes are not used to identify diseases. We examined several diseases from different overall disease categories. With this overview, we seek to provide a framework to assist researchers in the design of their studies, and in the interpretation of their current results and findings provided in the literature.

Materials and Methods

Setting and Data Sources

We conducted this study in Denmark, which had 5.8 million inhabitants in 2018.[13] The Danish Healthcare System provides tax-supported health care, which ensures unfettered access to private practitioners and hospitals as well as partial reimbursement for several prescription-required medications.[14] A unique ten-digit Central Personal Register number (CPR number), assigned to all Danish residents at birth or upon immigration, enables unambiguous and individual-level linkage of data across various Danish registries.[15]

We identified our study cohort through the Danish Civil Registration System.[15] This registry, established in 1968, issues the CPR number and records the vital and migration status of all Danish residents with daily electronic updates.[15]

Information on diseases diagnosed at hospitals was obtained from the DNPR. This registry, established in 1977 and providing complete national coverage since 1978, contains information on every non-psychiatric inpatient admission and, since 1995, information on outpatient clinic contacts, emergency department contacts, and psychiatric inpatient admissions.[7] The DNPR holds information on the dates of contact and discharge, the primary diagnosis (*ie*, the main reason for hospitalization), multiple optional secondary diagnoses (*eg*, comorbidities or complications), surgeries, and procedures.[7] Diagnoses were recorded according to ICD-8 until 1994 and thereafter according to ICD-10. Since 1996, surgical procedures have been recorded with the exact date and time, according to the Danish version of the Nordic Medico-Statistical Committee Classification of Surgical Procedures (NOMESCO).[7]

Study Populations

People eligible for the study populations were required to be alive, ≥40 years of age, and residing in Denmark as of January 1 of the sample years 1995, 2010, or 2018. At random dates during the respective sample years, 500,000 eligible people were sampled from the Danish Civil Registration System.

From the DNPR, we obtained information on individual diseases in cardiovascular, endocrine, gastrointestinal, neurological, pulmonary, rheumatic, and urogenital overall disease groups and fractures as well. The individual diseases and their ICD-8 and ICD-10 codes are provided in Table S1. We selected individual diseases from these overall disease groups to create a comprehensive framework that spans acute and chronic diseases, to ensure relevance for a wide range of research questions. Left-truncation and left-censoring ensuing from omitting ICD-8 codes were analyzed both for overall disease groups and for individual diseases. We used all available primary and secondary diagnosis records from inpatient, outpatient, and emergency department contacts. If a person had several hospital contacts associated with the same disease, only the most recent contact was included. The restriction to the most recent diagnosis was applied to ICD-8 and ICD-10 codes separately, thus ensuring that a patient could have both an ICD-8 and an ICD-10 code for the same disease. Likewise, people with a history of several individual diseases belonging to the same overall disease category were only counted once in the analysis of overall disease groups.

Statistical Analysis

Per individual disease and overall disease group, we obtained the number of people in three mutually exclusive ICD groups: those with only ICD-8 codes, those with both ICD-8 *plus* ICD-10 codes, and those with only ICD-10 codes. We then calculated the proportion and 95% confidence intervals of each ICD group with respect to the total sample and to the total number of people with the disease (the latter is thus the number of persons identified using all ICD groups). The formulas used are presented in Figure S1. The proportion of people with a disease identified solely by ICD-8 codes thus reflected those whose disease would have been left-truncated when not using ICD-8 codes. The proportion of people with a disease recorded with ICD-8 *plus* ICD-10 codes reflected those for whom the hospital contact recorded with the ICD-8 code would have been left-censored.

We plotted the proportions of the three ICD groups (*ie*, ICD-8 only, ICD-8 *plus* ICD-10, and ICD-10 only) with respect to the total sample and to the total number of people with the disease as stacked bar plots by sample year.

Furthermore, we stratified the analyses by sex and the following age groups: 40–54 years, 55–69 years, 70–84 years, and ≥85 years.

Data management, statistical analyses, and visualization were performed in SAS version 9.4 (SAS Institute, North Carolina, US) and RStudio version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria).

Ethics, Approvals, and Registrations

According to Danish law, ethical approval and informed consent were not required for this registry-based study. The study was reported to the Danish Data Protection Agency by Aarhus University (record number 2015–57-0002, serial number 608).

Data Availability

The supporting individual-level data used for this study are not publicly available but can be obtained by application to the Danish Health Data Authority.[16]

Results

Main Analysis

We included 500,000 patients in each sample year (*ie*, 1995, 2010, and 2018). Using all ICD codes, we identified 149,385 unique diagnoses recorded before 1995, 319,720 unique diagnoses recorded before 2010, and 415,970 unique diagnoses recorded before 2018.

Figure 2 shows the proportions of persons with a diagnosis from an overall disease group recorded in the DNPR by ICD group and sample year. The proportions are shown relative to both the total sample (*ie*, 500,000 persons) and the diseased sample (*ie*, identified by all ICD groups) for each sample year. The proportion of people in the total sample with a diagnosis from an overall disease group recorded in the DNPR was higher in the more recent sample years for every disease group (*eg*, the proportion of people with a cardiovascular disease recorded was 8.4% in 1995 and 24.0% in 2018).

Left-truncation was lower in more recent sample years across all overall disease groups (*eg*, for gastrointestinal (GI) diseases, 75.4% of persons with a GI disease were affected in 1995 whereas 18.9% of persons with a GI disease were affected in 2018 (Figure 2). When examining individual diseases, the issue with left-truncation varied (Table S2 and Figures S2–S9): The proportion of people with types 1 and 2 diabetes mellitus (*ie*, a chronic medical disease most likely to be recorded in several hospital inpatient or outpatient contacts) who would not have been detected by omitting ICD-8 codes was 59.1% in 1995 and 1.6% in 2018. The proportion of people with acute appendicitis (*ie*, an acute surgical condition often without recurrence and thus usually associated with only one hospital contact) who would not have been detected by omitting ICD-8 codes was 93.0% in 1995 and 54.0% in 2018.

Left-censoring was less of an issue for all overall disease groups in the more recent sample years (*eg*, among persons with cardiovascular diseases, the proportion was 15.5% in 1995 and 2.7% in 2018) (Figure 2). For the individual diseases, the issue with left-censoring varied: The proportion of patients with type 1 and 2 diabetes mellitus with an ICD-8 *plus* ICD-10 code recorded, relative to all patients with diabetes, was 21.8% in 1995 and 6.5% in 2018 (Table S2 and Figures S2–S9). The corresponding proportions for appendicitis were 0.2% in 1995 and 0.1% in 2018 (Table S2 and Figures S2–S9).

Age-Stratified Analysis

Table S3 and Figures S10–S36 provide information on the relation between age group and the proportion of people with a diagnosis from an overall disease group (Table S3) and an individual disease (Table S3 and Figures S10–S36) by ICD

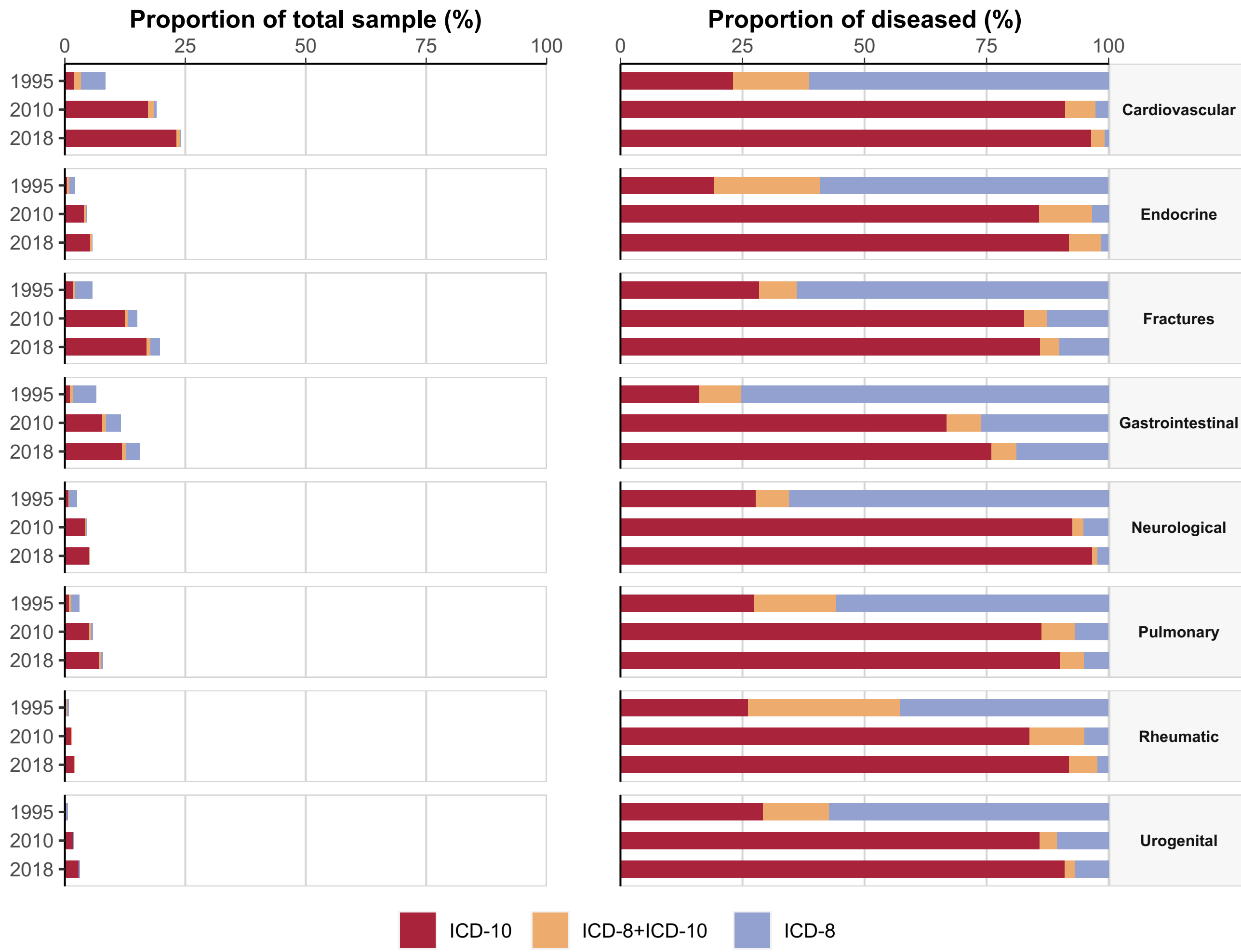

Figure 2 Proportion of people identified with ICD-10 codes only, ICD-8 plus ICD-10 codes (left-censoring), or ICD-8 codes only (left-truncation), grouped by overall disease group and sample year. The left panel shows the proportion with respect to the total sample. The right panel shows the proportion with respect to people with the disease.

group. When considering the individual diseases, the left-truncated proportion increased with higher age group for most individual diseases; however, for some diseases (*eg*, appendicitis and fractures), the proportion decreased with higher age group. Moreover, the relation between the left-truncated proportion and age group somewhat differed by sample year as the proportions were smaller and generally more stable across age groups in the sample years 2010 and 2018 than in 1995. When examining left-censoring, no systematic similar pattern in differences across age groups was found for the overall disease groups nor the individual diseases – that is, the left-censoring issue both increased, decreased, and were similar (Table S3 and Figures S10–S36).

Sex-Stratified Analysis

The sex-stratified analyses with the proportion of people identified as having a diagnosis from an overall disease group and an individual disease relative to the total sample are presented in Table S4. The issue with left-truncation and -censoring by not using IC8-codes was similar between females and males and across both overall disease groups and individual diseases (Table S4 and Figures S37–S44).

Discussion

Key Results

The proportion of patients with a disease who would not have been detected by omitting ICD-8 codes in identifying diagnoses in the DNPR (*ie*, left-truncation) varied by sample year, disease, and age group. Overall, the left-truncated proportion decreased by sample year, particularly for chronic diseases (*eg*, diabetes mellitus) but remained fairly high for acute and surgical diseases (*eg*, appendicitis and fractures). The left-truncated proportion by not using ICD-8 codes generally increased with increasing age groups for most individual diseases, particularly in the sample year 1995, whereas the proportion was generally smaller and more stable in the sample years 2010 and 2018. The results were similar between females and males.

Limitations

Some limitations should be acknowledged when interpreting our results. Because we collected random samples, the proportions might potentially have differed if we had collected other samples. Nonetheless, the 500,000 patients ≥40 years of age randomly sampled from the Danish population in each sample year constituted with respect to the total population 21% in 1995, 18% in 2010, and 17% in 2018. Accordingly, the samples should be representative.

Because we relied on ICD codes to detect diseases, a potential risk of false-positive misclassification of persons affected with a given disease exists, depending on the validity of the diagnostic and coding process in the DNPR. A thorough description of the DNPR as a data source including a contemporary summary of validation studies for different diseases has been published previously.[7] Moreover, depending on the completeness of disease recording in the DNPR, a risk of false-negative misclassification of people with no disease might have existed, particularly for diseases often managed solely in general practice (*eg*, hypertension, dyslipidemia, diabetes, and chronic obstructive pulmonary disease). In addition, the likelihood of having a hospital contact, and consequently a disease recorded in the DNPR, was higher in 2018 than in 1995. Moreover, changes in diagnostic criteria and tests might have influenced the detected number of diseased persons across different sample years without reflecting genuine changes in disease occurrence (*eg*, the adaptation of cardiac troponins in the universal definition of acute myocardial infarction[17]). Likewise, the shifting trend of treating some patients with certain common diseases (*eg*, diabetes mellitus type 2 and chronic obstructive pulmonary disease) exclusively in general practice may impact observed trends in prevalence proportions when relying solely on ICD codes for disease detection. Furthermore, older patients who had only one hospital contact with a disease before 1977, followed by no subsequent contacts, would not have been identified as prevalent cases on the basis of the DNPR, because such events would have been left-truncated. Altogether, comparing proportions across sample years and making inferences regarding trends in prevalence proportions might not be appropriate. However, rather than presenting the true prevalence proportions in the Danish population, our study was aimed at examining the proportion of people identified as having a disease as a function of the use of different ICD code revisions, and thus addressing issues with left-truncation and left-censoring when ICD-8 codes are not used.

Comparison with Findings from Other Studies

To the best of our knowledge, no prior studies have evaluated the left-truncation issue arising from disregarding previous ICD versions when identifying diseases in health registries. However, studies aimed at answering a different question, have examined the effects of the length of the look-back period (*ie*, searching the patient's hospital history in a registry for previous events), which is used when defining incident events, *eg*, in instances of gynecological diseases,[18] cancers,[19] and cardiovascular diseases.[20–24] Consulting such studies may be valuable when seeking to establish cohorts consisting of incident cases only because those studies were aimed specifically at addressing the appropriate length of the look-back period for a particular disease. The present study extends these studies by examining the effects of left-truncation and left-censoring arising from not using ICD-8 codes and notably included several disease groups and individual diseases.

Interpretation

Our study was aimed at providing a better understanding of issues related to left-truncation and left-censoring arising from not using ICD-8 codes. Our findings suggest that in deciding whether to use ICD-8 codes, the following three factors are particularly important: 1) the time from the end of ICD-8 use to the start of the study period; 2) the type of individual disease; and 3) the age of the study participants.

First, the proportion of people with a disease who remained undetected by not using ICD-8 codes (*ie*, left-truncated) was, as expected, higher in 1995 than in 2018 across all individual diseases, sexes, and age groups. Therefore, not including ICD-8 codes should be performed with caution, particularly when the study period commences close to the transition from ICD-8 to ICD-10 coding.

Second, even when the study period is more recent, the disease of interest must be considered. Our results suggested that only 1.6% of people with a history of diabetes mellitus would have been left-truncated in the sample year 2018 by not using ICD-8 codes. Although left-censoring may still be an issue, only 6.5% of patients with diabetes mellitus had both an ICD-8 *plus* an ICD-10 code in the sample year 2018. In contrast, 54.0% of people with a history of appendicitis (an acute, often non-recurring disease) were left-truncated in the sample year 2018 by not using ICD-8 codes. As expected, left-censoring was a minor issue in appendicitis, because the proportion of people with both an ICD-8 and an ICD-10 code was negligible in all sample years. Whether left-truncation or left-censoring is the greater concern may depend on the variable of interest. For instance, when identifying covariables, not identifying a comorbidity for a small proportion of patients might be relatively less troublesome as long as a similar approach is applied to both exposed and unexposed patients. However, when the variable of interest is the exposure, maximizing the certainty of having all patients properly adjudicated as exposed or unexposed may be desirable. Even if the undetected proportion is considered acceptable, not only the exposure status but also when the exposure began may be of interest because the date of exposure is often used to define the start of follow-up in cohort studies or in stratification according to the temporal relation between exposure and outcome in case–control studies.[25,26]

Third, the age of study participants is another factor that should be considered. The issue of left-truncation introduced by not using ICD-8 codes was overall more apparent in the highest age groups. This finding was expected because older patients generally had a higher likelihood of having a disease recorded with only an ICD-8 code. However, exceptions to this overall tendency were observed: For appendicitis, the issue of left-truncation by not using ICD-8 codes was greatest in the youngest age group; however, this finding likely reflected left-truncation among the highest age groups, because recording in the DNPR started in 1977. The oldest persons were likely to have had appendicitis before 1977, and this event accordingly was not likely to be found in the DNPR. In contrast, the youngest persons born after 1977 had their entire hospital history recorded in the DNPR, and diseases recorded with only ICD-8 codes would remain left-truncated by not using ICD-8 codes in the disease-identification algorithm (Figure 1).

In addition, the given study aim may affect the decision of whether to include ICD-8 codes. In descriptive studies aimed at assessing temporal trends in absolute disease occurrence, not using ICD-8 codes in the disease identification-algorithm may result in an underestimation of disease occurrence with a magnitude likely to depend on both the study period and the age of the study participants. In contrast, not using ICD-8 codes, thereby potentially using a less sensitive but more specific disease identification-algorithm, may still be sufficient for obtaining valid relative measures of associations. Before not using ICD-8 codes, the risk of introducing reverse causation must also be considered if the outcome under investigation actually occurred before the exposure, but the first event was left-censored because ICD-8 codes were not used.

When data availability impedes inclusion of ICD-8 codes in a disease-identification algorithm, researchers may use the results presented in this study to guide choices of whether this limitation may introduce left-truncation and left-censoring issues. Importantly, in addition to the considerations presented above, whether the recording of the specific ICD code is sufficiently valid to appropriately detect the disease of interest is a critical consideration.

Conclusion

The left-truncation issue differed by time from the end of ICD-8 use to the sample year, disease, and age group. Overall, the left-truncated proportion diminished over sample years, particularly for chronic diseases, yet remained rather high for certain surgical diseases. The left-truncation issue generally increased with age group for most individual diseases. The left-censoring issue was overall minor, but highest for chronic diseases. These results may assist researchers in designing studies and interpreting results.

Funding

This study was funded by the Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University. The Department of Clinical Epidemiology is involved in studies with funding from various companies as research grants to (and administered by) Aarhus University, but none of those studies have any relation to the submitted work.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Furu K, Wettermark B, Andersen M, Martikainen JE, Almarsdottir AB, Sorensen HT. The Nordic countries as a cohort for pharmacoepidemiological research. *Basic Clin Physiol Pharmacol*. 2010;106(2):86–94. doi:10.1111/j.1742-7843.2009.00494.x
2. Sørensen HT, Baron JA. Teaching Epidemiology: a guide for teachers in epidemiology, public health and clinical medicine. In: Olsen J, Greene N, Saracci R, Trichopoulos D, editors. *Medical Databases*. Oxford University Press; 2015.
3. Laugesen K, Ludvigsson JF, Schmidt M, et al. Nordic health registry-based research: a review of health care systems and key registries. *CLEP*. 2021;13:533–554. doi:10.2147/CLEP.S314959
4. Sund R. Quality of the Finnish Hospital discharge register: A systematic review. *Scand J Public Health*. 2012;40(6):505–515. doi:10.1177/1403494812456637
5. Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health*. 2011;11(1):450. doi:10.1186/1471-2458-11-450
6. Bakken IJ, Ariansen AMS, Knudsen GP, Johansen KI, Vollset SE. The Norwegian patient registry and the Norwegian registry for primary health care: Research potential of two nationwide health-care registries. *Scand J Public Health*. 2020;48(1):49–55. doi:10.1177/1403494819859737
7. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen 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
8. World Health O. international statistical classification of diseases and related health problems (ICD); 2004.
9. Sørensen ST, Kristensen FP, Troelsen FS, Schmidt M, Sørensen HT. Health registries as research tools: a review of methodological key issues. *Dan Med J*. 2023;70(4):A12220796.
10. Cain KC, Harlow SD, Little RJ, et al. Bias due to left truncation and left censoring in longitudinal studies of developmental and disease processes. *Am J Epidemiol*. 2011;173(9):1078–1084. doi:10.1093/aje/kwq481
11. Groenwold RHH, Dekkers OM. Look left, look right: censoring in clinical research. *Eur J Endocrinol*. 2023;189(5):E1–e3. doi:10.1093/ejendo/lvad152
12. Mues KE, Liede A, Liu J, et al. Use of the Medicare database in epidemiologic and health services research: a valuable source of real-world evidence on the older and disabled populations in the US. *Clin Epidemiol*. 2017;9:267–277. doi:10.2147/CLEP.S105613
13. Statistics D. Denmark in Figures 2018;2018.
14. Schmidt M, Schmidt SAJ, Adelborg K, et al. The Danish health care system and epidemiological research: from health care contacts to database records. *CLEP*. 2019;11:563–591.
15. Schmidt M, Pedersen L, Sørensen HT. The Danish civil registration system as a tool in epidemiology. *Eur J Epidemiol*. 2014;29(8):541–549. doi:10.1007/s10654-014-9930-3
16. The Danish Health Data A. How to Get Started; 2019
17. Schmidt M, Jacobsen JB, Lash TL, Bøtker HE, Sørensen HT. 25 year trends in first time hospitalisation for acute myocardial infarction, subsequent short and long term mortality, and the prognostic impact of sex and comorbidity: a Danish nationwide cohort study. *BMJ*. 2012;344:e356. doi:10.1136/bmj.e356
18. Kim M, Chae K-H, Chung Y-J, et al. The effect of the look-back period for estimating incidence using administrative data. *BMC Health Serv Res*. 2020;20(1):166. doi:10.1186/s12913-020-5016-y
19. Czwikla J, Jobski K, Schink T. The impact of the lookback period and definition of confirmatory events on the identification of incident cancer cases in administrative data. *BMC Med Res Methodol*. 2017;17(1):122. doi:10.1186/s12874-017-0407-4
20. Nedkoff L, Knuiman M, Hung J, Sanfilippo FM, Katzenellenbogen JM, Briffa TG. Concordance between administrative health data and medical records for diabetes status in coronary heart disease patients: a retrospective linked data study. *BMC Med Res Methodol*. 2013;13(1):121. doi:10.1186/1471-2288-13-121
21. Sulo G, Igland J, Vollset SE, et al. Effect of the lookback period's length used to identify incident acute myocardial infarction on the observed trends on incidence rates and survival: cardiovascular disease in Norway project. *Circ Cardio Qual Out*. 2015;8(4):376–382. doi:10.1161/CIRCOUTCOMES.114.001703
22. Osler M, Rostgaard K, Sørensen T IA, Madsen M. The effect of recurrent events on register-based estimates of level and trends in incidence of acute myocardial infarction. *J Clin Epidemiol*. 1999;52(7):595–600. doi:10.1016/S0895-4356(99)00032-3
23. Korsgaard S, Christiansen CF, Schmidt M, Sørensen HT. Impact of the look-back period on identifying recurrent myocardial infarctions in the Danish national patient registry. *CLEP*. 2021;13:1051–1059. doi:10.2147/CLEP.S334546
24. Worthington JM, Gattellari M, Goumas C, Jalaludin B. differentiating incident from recurrent stroke using administrative data: the impact of varying lengths of look-back periods on the risk of misclassification. *Neuroepidemiology*. 2017;48(3–4):111–118. doi:10.1159/000478016
25. Johannesdottir SA, Erichsen R, Horváth-Puhó E, Schmidt M, Sørensen HT. Coeliac disease and risk of venous thromboembolism: a nationwide population-based case-control study. *Br J Haematol*. 2012;157(4):499–501. doi:10.1111/j.1365-2141.2012.09030.x
26. Erichsen R, Christiansen CF, Mehnert F, Weiss NS, Baron JA, Sørensen HT. Colorectal cancer and risk of atrial fibrillation and flutter: a population-based case–control study. *Intern Emerg Med*. 2012;7(5):431–438. doi:10.1007/s11739-011-0701-9