ORIGINAL RESEARCH

Importance of Reporting Complete Procedures of Identifying Patients from the Danish National Patient Registry: The Case of Neonatal Jaundice and Epilepsy

Yuelian Sun [1–4], Julie Werenberg Dreier[3,5], Chunsen Wu[6,7], Vera Ehrenstein [4], Jakob Christensen[2–4]

[1]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark; [2]Department of Neurology, Aarhus University Hospital, Aarhus, Denmark; [3]National Centre for Register-Based Research, Department of Economics and Business Economics, Aarhus University, Aarhus, Denmark; [4]Department of Clinical Epidemiology, Department of Clinical Medicine, Aarhus University Hospital, Aarhus, Denmark; [5]Centre for Integrated Register-based Research (CIRRAU), Aarhus University, Aarhus, Denmark; [6]Department of Clinical Research, University of Southern Denmark, Odense, Denmark; [7]Department of Gynaecology and Obstetrics, Odense University Hospital, Odense, Denmark

Correspondence: Yuelian Sun, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark, Email ys@clin.au.dk

Objective: The Danish National Patient Registry (DNPR) is a valuable resource for medical and epidemiological Research. However, not all research articles fully described procedures they used to identify events. In this study, we compared two approaches in identifying persons with a disease diagnosis using neonatal jaundice and epilepsy as examples.

Methods: A cohort of singletons born alive between the 1st January 1997 and the 30th November 2016 in Denmark was used for this purpose. Diagnostic information for a hospital contact in the registry included a primary diagnosis, secondary diagnoses, referral diagnoses, and additional information to a diagnosis (associated diagnoses), if any. Approach 1 identified patients of interest by considering all diagnostic information with exclusion of referral diagnoses only. Approach 2 identified patients of interest by additionally excluding diagnoses from a hospital contact that were coded with Z00 – Z99 of ICD-10 (for health service on examination and reproduction, etc.) as the main reason of the hospital contact. We presented the proportion of people with a diagnosis of neonatal jaundice and epilepsy by the two approaches and explored the potential explanations for the difference.

Results: For the example of neonatal jaundice, the study population included N=1,186,683 persons. The proportion of children with a diagnosis of neonatal jaundice was 5.5% (n=66,736) by approach 1 and 3.9% (n=45,928) by approach 2. For the example of epilepsy, the study population included N=1,183,273 persons. The proportion of children with a diagnosis of epilepsy were 1.2% (n=14,604) by approach 1 and 0.9% (n=10,441) by approach 2.

Discussion: This study demonstrated that the two approaches identified different proportion of persons with a diagnosis of neonatal jaundice and epilepsy. We advocated researchers report complete procedures of identifying patients for making research findings reproducible and comparable.

Keywords: the Danish National Patient Registry, health administrative data, misclassification of diagnosis

Background

The Danish National Patient Registry (DNPR) is a valuable resource for medical and epidemiological research for example in identifying persons with a disease of interest as an exposure, an outcome, or a confounder.[1,2] To identify persons with a disease diagnosis of interest, researchers usually go through several procedures to include or exclude observations in the administrative data and/or diagnostic data of DNPR. However, not all researchers reports the complete procedures they have used leaving the findings unreproducible and incomparable although the importance of reporting of research has been emphasized in STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) and RECORD (REporting of studies Conducted using Observational Routinely-collected Data)

445

statement.[3,4] The STROBE statement requires "Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable" and the RECORD requires

> A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.[3,4]

In practice, authors of a paper might have described sources and International Statistical Classification of Diseases (ICD) codes for identifying patients of interest. For example two papers involving authors of this paper (YS and JWD) reported the data source and ICD codes when they identified patients with a diagnosis of epilepsy.[5,6] However, more details on how specific diagnosis types were handled were not reported, which turned out to be different (personal communication and was demonstrated later in this paper).[5,6] In this paper we compared two approaches in identifying persons with a disease diagnosis using neonatal jaundice and epilepsy as examples in a defined population since we raised up the topic when we conducted a study related to the two conditions.

Methods
Study Population
We identified 1,186,683 singletons born alive between 1 January 1997 and 30 November 2016 from the Danish Medical Birth Registry[7] who had a link to their mother in the Danish Civil Registration System.[8] In the epilepsy cohort, we excluded children who died or emigrated during the neonatal period (the first 28 days after birth) (n=3410) resulting in a study population of 1,183,273 singletons. The Civil Registration System is an administrative register that contains individual-level information on all persons residing in Denmark.[8] A unique identification number assigned to all persons in the system allows for individual-level record linkage of other registers in Denmark, the DNPR for example in this study. All data were analyzed at Statistics Denmark using encrypted identification numbers with no contact with the individuals. By Danish law, analyses of anonymous data do not require ethical review board approval. However, the study was approved by the Danish Data Protection Agency.

Introduction of the Danish National Patient Registry (DNPR)
The DNPR has collected information on discharge diagnoses of inpatient stays from 1977 and, since 1995, information on all outpatient visits and emergency room has also been registered.[1] Diagnostic information is based on the Danish version of the International Statistical Classification of Diseases (ICD), the 8th version (ICD-8) from 1977 to 1993 and the 10th version (ICD-10) from 1994 onwards. The unit of the DNPR is a hospital contact. The DNPR is organized in tables including tables for administrative data, diagnostic data, treatment, surgery, etc. The administrative information can be linked to the rest of data including diagnostic data using the unique administrative record number of hospital contacts.

The administrative data contains information on the unique administrative record number for a hospital contact, patient's civil registration number, sex, age, hospital of the health service, start date, end date, and the primary diagnosis of the hospital contact, so called action diagnosis in Danish, in one observation for the hospital contact.

The diagnostic data contains one or more observations for a hospital contact including the observations for a primary diagnosis, one or more secondary diagnoses, referral diagnoses, and additional information (diagnosis) to some primary, secondary, or referral diagnoses. The primary diagnosis indicates the condition that has led to hospitalization, need for care, outpatient contact or injury contact, and is the main reason for the examination and treatment program implemented in the hospital contact.[9] If a patient have multiple diagnoses that may be competing as the primary diagnosis, the most serious diagnosis with influence on the longer-term prognosis, is considered as the primary diagnosis. The hospital contact's other diagnoses are subsequently registered as secondary diagnoses. A referral diagnosis was the diagnosis given by the referring physician or hospital/department as a reason for referral. The reason for hospital contact with health service could also be the conditions described in Chapter 21 of ICD 10, which include health services for examination and investigation, health services in circumstances related to reproduction, rehabilitation, screening, and prophylactic measures etc coded as Z00 – Z99. For example, Z03 codes indicate "Medical observation and evaluation for suspected diseases and conditions", which are labeled as "obs. pro" codes in the documents of Danish health service, and Z38 codes indicate "Liveborn infants according to place of

birth". In the DNPR, additional codes are used to specify some primary, secondary, or referral diagnoses. For example if the primary diagnosis was "Z03 Medical observation and evaluation for suspected diseases and conditions", the additional code specifies the primary diagnosis indicating what the suspected diseases could be. We termed a diagnosis from additional codes as an associated diagnosis. Z codes could be registered in a primary, secondary, or referral diagnosis but they are often used in connection with the registration of a referral diagnosis. Z codes are only used as a primary diagnosis on contacts when no actual diagnosis of disease is given as stated in the document.[9]

Figure 1 illustrates the data structure including the variables in the administrative and diagnostic tables in the DNPR. The variable "c_adiag" in the administrative table contains the ICD codes for the primary diagnosis of a hospital contact. The variable "c_diag" in the diagnostic table contains the ICD codes for all related diagnoses such as the primary, the secondary, and the referral diagnoses. The variable "c_diagtype" contains information about the type of the diagnosis indicated in the variable "c_diag": including A (Primary diagnosis), B (Secondary diagnosis), G (Basic diagnosis when the basic diagnosis is different from the primary diagnosis), H (Referral diagnosis), M (Temporary diagnosis for unfinished outpatient contacts), C (Complication), and "+". The "c_diagtype" of "+" indicates that the diagnosis in the variable "c_diag" has additional information that is stated in the variable "c_tildiag". As we stated before that the diagnosis from the variable "c_tildiag" was termed as "an associated diagnosis" in this study.

Diagnosis of Epilepsy and Neonatal Jaundice

In the analyses of jaundice, children were followed from the day of birth until the admission date with neonatal jaundice, death, emigration, the 58th day of life (neonatal period 28 days plus 30 days) or the administrative end of follow-up (31st December 2016), whichever came first. The ICD-10 codes used to identify neonatal jaundice were P57-P59.

In the analyses of epilepsy, children were followed from the 29th day after birth (ie after the neonatal period) to the admission date of diagnosis with epilepsy, death, emigration date, or the administrative end of follow-up (31 December 2016). The ICD-10 codes used to identify epilepsy were G40-G41.

Statistical Analysis and Two Approaches of Identifying Patients

We categorized each person into one of three groups according to their diagnoses related to the disease of interest in the follow up (group 1: persons with a primary or secondary diagnosis of interest and a referral or an associated diagnosis, group 2: persons with a primary or secondary diagnosis of interest only, and group 3: persons with a referral or an associated diagnosis only). We identified a primary or secondary diagnosis of interest if the variable "c_diag" was coded with P57-P59 for neonatal jaundice or G40-G41 for epilepsy, and the variable "c_diagtype" was coded with "A, B, G, and C" while the observations coded with G and C were few. Approach 1 identified patients of interest by considering all diagnostic information with exclusion of referral diagnoses only. Approach 2 identified patients of interest by additionally excluding diagnoses from a hospital contact that was coded with Z00 – Z99 of ICD-10 (codes used for circumstances related to examination, investigation, reproduction, rehabilitation, screening, and prophylactic measures, etc.) as the main reason of the hospital contact. Approach 1[5,10] and approach 2[6,11] has been used in several papers conducted by authors of this study. In Figure 1 we used mock data for nine persons to illustrate the categorization of diagnoses (a primary or secondary diagnosis, a referral diagnosis, and an associated diagnosis), groups of persons according to their diagnoses (group 1, 2, and 3), detailed procedures of identifying patients of interest by approach 1 and 2, and persons identified by approach 1 and 2.

We reported number and proportion of persons with at least one hospital contact related to neonatal jaundice or epilepsy in the DNPR and distribution of these persons in the three groups. We then reported number and proportion of persons with a diagnosis of neonatal jaundice or epilepsy identified by the two approaches and compared the two approaches in identifying persons in group 1 and 2 and excluding persons in group 3. Thirdly, we presented the number of hospital contacts among persons in the three groups to show the different characteristics of persons in the three groups. For group 1 with both a primary or secondary diagnosis and a referral or an associated diagnosis, we also presented the difference on dates of hospital contacts between the primary or secondary diagnosis and the referral or associated diagnosis. Finally, we further explored the main diagnoses of hospital contacts related to neonatal jaundice or epilepsy and diagnostic codes in the diagnostic table for a hospital contact related to neonatal jaundice or epilepsy with a Z code as the primary diagnosis for the hospital contact.

1st section		2nd section	3rd section			4th section				5th section		
Person[n] identified by		Group of persons	Data from the administrative table[a]			Data from the diagnostic table[b]				Categorization of the diagnosis		
Approach 1[l]	Approach 2[m]		pnr[c]	recnum[c]	c_adiag[c]	recnum[c]	c_diag[c]	c_diagtype[c]	c_tildiag[c]	a primary /secondary diagnosis[f]	a referral diagnosis[g]	an associated diagnosis[h]

Left portion (1st–3rd sections):

Approach 1[l]	Approach 2[m]	Group	pnr	recnum	c_adiag
x	x	1[i]	1	1	DG40[d]
x	x	1	2	2	DG40
			2	3	DG40
			2	4	DXXX
x	x	1	3	5	DZ03
			3	6	DG40
x	x	2[j]	4	7	DG40
x	x	2	5	8	DG40
			5	9	DXXX
x	.	2	6	10	DZ03
.	.	3[k]	7	11	DZ03
x	.	3	8	12	DZ03
x	x	3	9	13	DXXX
...	...		...	...	

Right portion (4th–5th sections):

recnum[c]	c_diag[c]	c_diagtype[c]	c_tildiag[c]	a primary/secondary diagnosis[f]	a referral diagnosis[g]	an associated diagnosis[h]
1	DG40[d]	A[e]	.	x		
1	DZ03	H[e]	.			
1	DZ03	"+"[e]	DG40			x
2	DG40	A	.	x		
2	DG40	H	.		x	
3	DG40	A	.	x		
4	Dxxx	A	.			
4	DZ03	H	.			
4	DG40	B[e]	.	x		
5	DZ03	A				
5	DZ03	"+"	DG40			x
6	DG40	A	.	x		
7	DG40	A	.	x		
8	DG40	A	.	x		
9	DXXX	A	.			
9	DG40	B	.	x		
10	DZ03	A	.			
10	DG40	B	.	x		
11	DZ03	A	.			
11	DG40	H	.		x	
12	DZ03	A	.			
12	DZ03	*+*	DG40			x
13	DXXX	A	.			
13	DZ03	*+*	DG40			x
...						

Figure 1 Illustration of the data in the administrative table [a](the 3rd section) and diagnostic table [b](the 4th section) for 9 persons (each had a unique person's identification number "pnr" and was marked with a color) including main variables [c], diagnostic ICD codes [d], and diagnostic types [e] in the Danish National Patient Registry using diagnosis of epilepsy as an example, and process in identifying patients with a diagnosis of epilepsy including categorization the diagnoses (a primary or secondary diagnosis [f], a referral diagnosis [g], and an associated diagnosis [h]) (the 5th section), groups of persons according to their diagnoses [i, j, k] (the 2nd section), and persons identified by approach 1[l] and 2[m] (the 1st section) [n].
Notes: [a]The administrative table contains one observation for one hospital contact with a unique 'recnum'. [b]The diagnostic table contains one or more observations related to one hospital contact. [c]Label of variables. pnr: Person's identification. recnum: Administrative record number of hospital contacts. c_adiag: Primary diagnosis of the hospital contact. c_diag: All diagnostic information including primary, secondary, referral diagnosis, and additional diagnostic information to the primary, secondary, referral diagnosis . c_diagtype: Type of diagnosis. c_tildiag: Additional diagnostic codes to the primary, secondary, referral diagnosis, if any. [d]DG40: G40 is the ICD 10 code for epilepsy, which is added a capital D in the data; DZ03: an example of Z-codes in the ICD 10 ranging from DZ00-DZ99. Dxxx: other codes besides codes for epilepsy and Z-codes. [e]Value of the variable 'c_diagtype' (type of diagnosis). A: Main / primary diagnosis. B: Auxiliary / secondary diagnosis. C: Complication. G: Basic diagnosis when the basic diagnosis is different from the main diagnosis. H: Referral diagnosis. +: indicates that the diagnosis in the variable 'c_diag' has an associated diagnosis and that the associated diagnosis is stated in the variable 'c_tildiag'. [f]We identified a primary or secondary diagnosis when the diagnostic observation with the variable 'c_diag' was coded with the codes of interest (epilepsy or jaundice), and the variable 'c_diagtype' was coded with 'A, B, G, and C'. However, the number of observations with the value 'G and C' were few. [g]A referral diagnosis which was the diagnosis given by the referring physician or hospital / department as a reason for referral. [h]An associated diagnosis which was originated from the additional codes to a primary, secondary, or referral diagnosis in the DNPR. [i]Group 1: persons who had a primary or secondary diagnosis and a referral or an associated diagnosis as well (person 1, 2, 3), [j]Group 2: persons who had a primary or secondary diagnosis only (person 4, 5, 6). [k]Group 3: persons with a referral or an associated diagnosis only (7,8, 9). [l]Approach 1: Procedures of identifying persons with a disease diagnosis, epilepsy as an example. 1. Link the administrative data with the diagnoses data. 2. Identify all records if the variable 'c_diag' or 'c_tildiag' contained the code of interest for the diagnosis, DG40-DG41 for epilepsy. 3. Exclude records for the referral diagnosis if the variable 'c_diagtype' was coded with 'H'. [m]Approach 2: Procedures of identifying persons with a disease diagnosis, epilepsy as an example. 1. Exclude hospital contacts if the variable 'c_adiag' was coded with Z00 – Z99. 2. Link the administrative data with the diagnoses data. 3. Identify all records if the variable 'c_diag' or 'c_tildiag' contained the code of interest for the diagnosis, DG40-DG41 for epilepsy. 4. Exclude records for the referral diagnosis if the variable 'c_diagtype' was coded with 'H'.. 5. Exclude records if the variable 'c_diag' was coded with Z00 – Z99. [n]Approach 1 identified person 1, 2, 3, 4, 5, 6, 8, and 9 with a diagnosis of epilepsy excepting person 7. Approach 2 identified person 1, 2, 3, 4, 5, 9 with a diagnosis of epilepsy excepting for person 6, 7, 8.

Results
Results for Neonatal Jaundice

Among the study population (N=1,186,683), we identified 65,938 (5.6%) persons with at least one hospital contact related to neonatal jaundice in the DNPR including 4964 (7.5%) persons with a primary or secondary diagnosis and a referral or an associated diagnosis as well (group 1), 53,527 (81.2%) persons with a primary or secondary diagnosis

only (group 2), and 7447 (11.3%) persons with only a referral or associated diagnosis (group 3) (Table 1). The detailed information about the process is presented in S-Figure 1a. We identified 65,736 (5.5%) and 45,928 (3.9%) persons with a diagnosis of neonatal jaundice by approach 1 and 2, respectively (Table 1). Approach 1 identified all persons in group 1 and 2, and excluded only 2.7% of persons in group 3 while approach 2 identified less persons in group 1 (94.0%) and group 2 (77.0%) and excluded most persons (99.2%) in group 3 (Table 1).

Persons in group 1 had more hospital contacts related to neonatal jaundice than group 2, even if we did not take the hospital contacts due to a referral diagnosis or an associated diagnosis into account in group 1 (S_Table 1a). About 39.6% (27.7%+11.9%) of persons in group 1 had two or more hospital contacts, while 25.2% (20.8%+4.4%) of persons in group 2 did. Among persons in group 3, only 9.1% (8.5%+0.6%) had two or more hospital contacts (S_Table 1a). Among group 1 who had both a primary or secondary diagnosis and a referral or an associated diagnosis, 684 (13.8%) persons had different hospital contact dates between the primary or secondary diagnosis and the referral or associated diagnosis of neonatal jaundice while 164 (3.3%) had a difference more than one week (S_Table 2a).

Among the 65,938 persons with at least one hospital contact related to neonatal jaundice we identified a total of 87,624 related hospital contacts in the follow up period. The primary diagnosis (the main reason) for these hospital contacts were neonatal jaundice (38.0%), health service for circumstance related to examination, investigation, reproduction, etc. that was coded with Z (29.8%), and others disorders (32.2%) (S_Table 3a). Among the service coded with Z, 63.7% were service related to birth, which was coded with Z38 (S_Table 3a). Among 26,100 hospital contact related to neonatal jaundice with a Z code as the main diagnosis, there were 26,611 diagnostic records related to neonatal jaundice in the diagnostic table including 65.9% of them had neonatal jaundice as the secondary diagnosis, 33.1% of them had neonatal jaundice as an associated diagnosis, and 1% of them had neonatal jaundice as a referral diagnosis (S_Table 4a).

Results for Epilepsy

Among the study population (N=1,183,273), we identified 19,486 (1.7%) persons with at least one hospital contact related to epilepsy in the DNPR, including 6925 (35.5%) persons with a primary or secondary diagnosis and a referral or an associated diagnosis as well (group 1), 4011 (20.6%) persons with a primary or secondary diagnosis only (group 2), and 8550 (43.9%) persons with only a referral or associated diagnosis (group 3) (Table 2). The detailed information about the process is presented in S-Figure 1b. We identified 14,604 (1.2%) and 10,441 (0.9%) persons with a diagnosis of epilepsy by approach 1 and 2, respectively (Table 2). Approach 1 identified all persons in group 1 and 2 and excluded

Table 1 Summary of Persons with a Hospital Contact Related to Neonatal Jaundice and Summary of Approach 1 and Approach 2 in Identifying Persons with a Diagnosis of Neonatal Jaundice Among the Study Population of Singletons Born Between 01 Jan 1997 and 30 Nov 1996 in Denmark (N=1,186,683)

	All		**Approach 1**		**Approach 2**	
	Number	**%**	**Number**	**%**	**Number**	**%**
Persons with at least one hospital contact related to neonatal jaundice Persons with a diagnosis of neonatal jaundice according to the approach	65,938	5.6	65,736	5.5	45,928	3.9
			Identifying persons in group 1, 2, and 3 by the approach (numbers in the brackets are persons excluded by the approach)			
Categorization of persons according to their diagnostic observations	**Number**	**Col %**	**Number**	**Row %**	**Number**	**Row %**
Group 1: persons with a primary or secondary diagnosis and a referral or an associated diagnosis of neonatal jaundice as well	4964	7.5	4964	100.0	4667	94.0
Group2: persons with a primary or secondary diagnosis of neonatal jaundice only	53,527	81.2	53,527	100.0	41,201	77.0
Group 3: persons with a referral or an associated diagnosis of neonatal jaundice only	7447	11.3	7245 (202)	97.3 (2.7)	60 (7387)	0.8 (99.2)

Table 2 Summary of Persons with a Hospital Contact Related to Epilepsy and Summary of Approach 1 and Approach 2 in Identifying Persons with a Diagnosis of Epilepsy Among the Study Population of Singletons Born Between 01 Jan 1997 and 30 Nov 1996 in Denmark (N=1,183,273[a])

	All		Approach 1		Approach 2	
	Number	**%**	**Number**	**%**	**Number**	**%**
Persons with at least one hospital contact related to epilepsy	19,486	1.7				
Persons with a diagnosis of epilepsy according to the approach			14,604	1.2	10,441	0.9
			Identifying persons in group 1, 2, and 3 by the approach (numbers in the brackets are persons excluded by the approach)			
Categorization of persons according to their diagnostic observations	**Number**	**Col %**	**Number**	**Row %**	**Number**	**Row %**
Group 1: persons with a primary or secondary diagnosis and a referral or an associated diagnosis of epilepsy as well	6925	35.5	6925	100.0	6722	97.1
Group2: persons with a primary or secondary diagnosis of epilepsy only	4011	20.6	4011	100.0	3693	92.1
Group 3: persons with a referral or an associated diagnosis of epilepsy only	8550	43.9	3668 (4882)	42.9 (57.1)	26 (8524)	0.3 (99.7)

Notes: [a]Children who died or emigrated during the neonatal period (the first 28 days after birth) were excluded (n=3410).

57.1% persons in group 3. Approach 2 identified slightly less persons in group 1 (97.1%) and group 2 (92.1%), and excluded 99.7% persons in group 3 (Table 2).

Persons in group 1 had more hospital contacts related to epilepsy than group 2, even we did not take the hospital contacts due to a referral diagnosis or an associated diagnosis into account in group 1 (S_Table 1b). About 69.2% of persons in group 1 had two or more hospital contacts, while 48.1% of persons in group 2 did. Among persons in group 3, only 20.3% had two or more hospital contacts. Furthermore, the percentages of persons who had five or more hospital contacts were 36.9%, 10.9%, and 0.3% for the three groups, respectively (S_Table 1b). Among group 1 who had both a primary or secondary diagnosis and a referral or an associated diagnosis, 1389 (20.1%) persons had different hospital contact dates between the primary or secondary diagnosis and the referral or associated diagnosis of epilepsy while 534 (7.7%) had a difference more than half year (S_Table 2b). Among 19,486 persons with at least one hospital contact related to epilepsy, we identified 73,929 hospital contacts related to epilepsy among the study population in the follow up period. The primary diagnosis for these hospital contacts were epilepsy (58.4%), health service circumstance related to examination, investigation, reproduction, etc., that was coded with Z (20.3%), and others (21.3%) (S_Table 3b). Among the service coded with Z, 84.9% were service related to medical observation and evaluation for ruling out suspected disease (epilepsy), which was coded with Z03 (S_Table 3b). Among 15,012 hospital contacts related to epilepsy with a Z code as the primary diagnosis, there were 16,568 diagnostic records related to epilepsy in the diagnostic table including 14.3% of them with epilepsy as the secondary diagnosis, 39.6% of them with epilepsy as an associated diagnosis, and 46.1% of them with epilepsy as a referral diagnosis (S_Table 4b).

Discussion

This study showed that the two approaches, which researchers used to identify patients with a diagnosis of a disorder from the DNPR,[5,6,10,11] identified different number and proportion of persons with a diagnosis of neonatal jaundice and epilepsy while the complete procedures of the approaches were not reported in the published papers. The way of handling hospital contacts due to circumstances related to examination, investigation, reproduction, rehabilitation, screening, and prophylactic measures, etc, which are coded Z00-Z99 in ICD-10 was one of potential causes of the difference. Researchers who excluded any diagnosis of a hospital contact with a Z code as a primary diagnosis considered such diagnoses may have a lower degree of validity in identifying persons with the disease of interest although they did not report the procedures (personal communication).[6,11] Researchers who included diagnoses secondary to a primary Z diagnosis, might have done so either because they did not realize the influence of different

approaches,[5,10] or they accepted the validity of diagnosis secondary to a primary Z diagnosis. We, further showed that 29.8% of hospital contacts related to neonatal jaundice were coded with a Z code while 20.3% of hospital contacts related to epilepsy were coded with a Z code. The profile of Z code varied according to disease of interest. For neonatal jaundice the majority of Z codes were for birth service while for the epilepsy the majority of Z codes were examination for ruling out the suspected disease. In the documents from the Danish Health Authority it states that Z codes are only used as a primary diagnosis on contacts where there is no "real disease".[9] A study from Denmark, which included a cohort of all patients with a Z03 diagnosis from an acute medical admission unit showed that 92% of them obtained a specific diagnosis during the entire hospital stay and 8% were discharged with Z03 as the main diagnosis.[12] We found, however, quite a lot of hospital contacts with a Z code as the primary diagnosis (most of them were related to birth) were given a secondary diagnosis for neonatal jaundice. That explained why approach 2 identify less persons from group 2 than approach 1 for neonatal jaundice diagnosis. The way of handling an associated diagnosis was another potential cause of the difference. Whether an associated diagnosis of a disorder alone is reliable or not need more research to verify it. Our data showed that persons with a referral or an associated diagnosis alone had a limited number of hospital contacts related to the disease of interest compared to persons with both a primary or secondary diagnosis and a referral or an associated diagnosis as well and persons with a primary or secondary diagnosis only. Even the persons with a primary or secondary diagnosis had different profiles on the number of hospital contacts depending on if they had a referral or an associated diagnosis. It was unclear if the severity of the disorder was different between them.

In this paper we presented two approaches to demonstrate the importance of reporting complete procedures of identifying patients from the DNPR. In practice, other approaches can be used, for example, researchers might identify patients who had a primary or a secondary diagnosis of a disease of interest,[13] patients who had a primary diagnosis of disease of interest leaving alone other kinds of information.[14,15] Which approach should be used depends on type of disease and purpose of research and needs to be explored further, which was, however, beyond the scope of this study.

The Danish National Patient Registry (DNPR) is a valuable resource for research. Researchers should be well notified about the structure and contents of the DNPR, which could help them to make decision when they extract the information from the DNPR. In practice, the inclusion or exclusion criteria in identifying patients is related to balance between acceptable sensitivity and specificity given the aims of a specific study. More research is needed to explore the optimal way of identifying patients of interest in different conditions like stroke with an abrupt onset, cancer with a chronic process, febrile seizure that occurs in early life or dementia that occurs in elderly. The appropriate approach might vary for studies depending on preferred sensitivity or specificity of a diagnosis.

Increasing use of health administrative data calls for guidance on better use of the data. Accurate identification of patients of interest means less misclassification of exposure, outcomes, and confounders in research. Accuracy of diagnostic codes depends on multiple factors including database quality, the specific condition being identified, and the validity of the codes in patient groups as mentioned in a systematic review.[16] The authors of the review recommended complete and accurate reporting of algorithm of identifying patients in validation research and having a clear label of studies as using health administrative data.[16] The STROBE and RECORD statement should be further applied in preparing articles.

Conclusions

This descriptive study demonstrated two approaches that researchers might use in identifying patients with a disease identified quite different proportion of persons with the diagnosis. We advocate that researchers should report the complete procedures of identifying patients from the DNPR, and validate the approach if necessary for making research findings reproducible and comparable and for better use of the Danish health data.

Funding

Disclosure

References

1. Schmidt M, Schmidt SA, Sandegaard JL, et al. 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
2. 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
3. STrengthening the Reporting of OBservational studies in Epidemiology (STROBE). Available from: https://www.strobe-statement.org/. Accessed March 17, 2022.
4. REporting of studies Conducted using Observational Routinely-collected Data (RECORD). Available from: https://www.record-statement.org/. Accessed March 17, 2022.
5. Sun Y, Christensen J, Olsen J. Childhood epilepsy and maternal antibodies to microbial and tissue antigens during pregnancy. *Epilepsy Res*. 2013;107(1–2):61–74. doi:10.1016/j.eplepsyres.2013.08.019
6. Dreier JW, Li J, Sun Y, et al. Evaluation of long-term risk of epilepsy, psychiatric disorders, and mortality among children with recurrent febrile seizures: a National Cohort Study in Denmark. *JAMA Pediatr*. 2019;173(12):1164–1170. doi:10.1001/jamapediatrics.2019.3343
7. Knudsen LB, Olsen J. The Danish medical birth registry. *Dan Med Bull*. 1998;45(3):320–323.
8. Schmidt M, Pedersen L, Sorensen 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
9. (Authority) STDHaM. Fællesindhold for basisregistrering af sygehuspatienter 2019, Vejledningsdel (Common content for basic registration of hospital patients 2019, Guidance section). 2019-VEJL ed; 2019.
10. Mao Y, Pedersen LH, Christensen J, et al. Prenatal exposure to antidepressants and risk of epilepsy in childhood. *Pharmacoepidemiol Drug Saf*. 2016;25(11):1320–1330. doi:10.1002/pds.4072
11. Aagaard SK, Dreier JW, Sun Y, et al. Accidental deaths in young people with epilepsy and psychiatric comorbidity-A Danish nationwide cohort study. *Epilepsia*. 2020;61(3):479–488. doi:10.1111/epi.16453
12. Hansen KM, Nielsen H, Vest-Hansen B, et al. Readmission and mortality in patients discharged with a diagnosis of medical observation and evaluation (Z03*-codes) from an acute admission unit in Denmark: a prospective cohort study. *BMC Health Serv Res*. 2017;17(1):211. doi:10.1186/s12913-017-2156-9
13. Cocoros NM, Svensson E, Szépligeti SK, et al. Long-term risk of Parkinson disease following influenza and other infections. *JAMA Neurol*. 2021;78(12):1461–1470. doi:10.1001/jamaneurol.2021.3895
14. Vaeth S, Andersen H, Christensen R, et al. A search for undiagnosed Charcot-Marie-tooth disease among patients registered with unspecified polyneuropathy in the Danish National Patient Registry. *Clin Epidemiol*. 2021;13:113–120. doi:10.2147/clep.S292676
15. Rye C, Rubin KH, Moller FT, et al. Positive predictive value of diagnostic codes for inflammatory bowel disease in the Danish National Patient Registry among individuals 50+ years, using patient records as reference standard. *Clin Epidemiol*. 2021;13:335–344. doi:10.2147/clep.S298770
16. Benchimol EI, Manuel DG, To T, et al. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol*. 2011;64(8):821–829. doi:10.1016/j.jclinepi.2010.10.006