Scandinavian Journal of Public Health, 1–9

ORIGINAL ARTICLE

Algorithm for forming hospital care episodes by combining attendance contacts in the Danish National Patient Register: A methodological consensus-driven study

RASMUS GREGERSEN[1,2,3], CATHRINE FOX MAULE[2†], CAMILLA BJØRN JENSEN[2], MATILDE WINTHER-JENSEN[2], ANNE HELMS ANDREASEN[2], HANNE NYGAARD[1,3], ANTON POTTEGÅRD[4], MORTEN SCHMIDT[5,6], SØREN PAASKE JOHNSEN[7], JANNE PETERSEN[2,3] & MARIE VILLUMSEN[2]

[1]Department of Emergency Medicine, Copenhagen University Hospital – Bispebjerg and Frederiksberg, Copenhagen, Denmark, [2]Department of Data, Biostatistics and Pharmacoepidemiology, Center for Clinical Research and Prevention, Copenhagen University Hospital – Bispebjerg and Frederiksberg, Copenhagen, Denmark, [3]Section of Biostatistics, Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [4]Clinical Pharmacology, Pharmacy, and Environmental Medicine, Department of Public Health, University of Southern Denmark, Odense, Denmark, [5]Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark, [6]Department of Cardiology, Aarhus University Hospital, Aarhus, Denmark, [7]Danish Center for Health Services Research, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

Abstract

Background: Studying complete hospital care episodes from register data, for instance when assessing length of stay, discharges and readmissions, can cause methodological difficulties due to the lack of a contact linkage identifier. We aimed to develop an algorithm combining sequential attendance contacts in the Danish National Patient Register (DNPR) into hospital care episodes, spanning the entire duration and all contacts from hospital arrival to departure. *Methods:* The algorithm was developed under the consensus of experts from research institutions across Denmark. It reads in second and third version DNPR data, deletes contacts without attendance, duplicates elective outpatient contacts corresponding to attendance dates and modifies contact types (e.g. repeated acute contacts), among others. Thereafter, sequential contacts within 4 h are marked as the same hospital care episode, consisting of one or more DNPR contact. We tested the algorithm in a data set of adults living in Denmark during 2013–2021 and compared different hourly cut-offs. *Results:* For the demonstration, we included 120.2 m contacts from 5.7 m persons, combined into 105.9 m hospital care episodes. Of the hospital care episodes, 6.4% were acute inpatients, 8.3% were acute outpatients, 2.0% were elective inpatients and 83.3% were elective outpatients. Using 4 h as our recommendation, 3-h, 5-h and 6-h cut-offs for contact combining revealed only minor differences in the number of hospital care episodes (<0.4%), whereas 12-h (<1.7%) and 24-h cut-offs (<43.1%) had a larger impact. **Conclusions: The algorithm automates data reading, modification and linkage of sequential attendance contacts. The algorithm can be initiated as a SAS macro and is available from an online repository.**

Keywords: *Danish National Patient Register, registries, epidemiology, algorithms, SAS macro*

Introduction

Register-based research is on the rise internationally [1, 2] and is being increasingly recognised for providing needed real-world evidence [3, 4]. Epidemiological studies based on register data can provide important knowledge on disease occurrence, patient characteristics, treatments, prognosis, side effects, risk factors and exposures to name some [5]. Register-based research is especially useful for studying conditions that are difficult to assess in clinical studies, such as rare conditions or unethical exposures and studying

†Deceased, died 11 November 2023

Correspondence: Rasmus Gregersen, Department of Emergency Medicine, Copenhagen University Hospital – Bispebjerg and Frederiksberg, Ebba Lunds Vej 40A, Building 67, 2nd floor, 2400 Copenhagen NV, Denmark. E-mail: rasmus.gregersen@regionh.dk

Date received 26 March 2024; reviewed 26 June 2024; accepted 2 September 2024

DOI: 10.1177/14034948241284711
journals.sagepub.com/home/sjp
Sage

unselected populations not restricted by the inclusion criteria of clinical trials [5–7]. As a consequence thereof, register-based studies also provide real-world evidence on the actual usage, compliance, effect, side effects and more of treatments [3, 4, 8]. Register-based research is typically cheap to conduct, requires a short timeframe from conceptualisation to finalisation, minimises the risk of selection and recall bias and yields large study populations [7]. However, projects based on registry data also introduce methodological difficulties, uncertainty due to data collection for non-research purposes and a risk of misclassification [7]. With the increasing scientific weight and recognition attributed to real-world studies, a focus is to secure transparency, correct definitions and populations, data validity and robust methodology [3, 4]. Further, the increasing complexity and amount of data calls for applicable solutions to known problems lowering the barriers to entry and educating future health researchers [9, 10].

The Danish national registries have long been an acknowledged data source for epidemiological studies, but comparable registries with some variations or limitations exist in Finland, Iceland, Sweden, Norway, the United Kingdom and The Netherlands, among other countries [5, 7, 11, 12]. The Danish National Patient Register (DNPR) holds information on hospital contacts, diagnoses, procedures and surgeries for the entire Danish population and all visitors to Danish hospitals [5, 13], linkable to other registers through the unique identifier [14]. However, contacts across different departments are registered separately and are not formally linked as complete hospital care episodes. Thus, problems arise when research projects need to consider complete hospital care episodes spanning the entire duration of a stay, for instance, when assessing length of stay, readmissions and discharge diagnoses for patients who are transferred between departments. In 2019, DNPR changed from the second (DNPR2) to the third (DNPR3) version, with a priority to introduce contact identifiers linking contacts during longer diagnostic or treatment courses and many visits at different sections under the responsibility of one department. However, these contact linkages do not cover patients who are transferred between different department responsibilities, for instance, during acute admissions, complications during planned admissions or a need to stay at the intensive care unit.

There is no consensus on combining contacts when identifying hospital care episodes in epidemiological studies. Previous studies have connected contacts within 2 h [15], 3 h [16], 4 h [17], 5 h [18], 12 h [19] or 24 h [20] of each other, or if initiated on the same or the following day [21, 22]. However, the approach was not always specified [23, 24]. Using all different subtypes of DNPR2 contacts could also be important, as a psychiatric or outpatient contact could connect two somatic contacts that would otherwise be regarded as two separate hospital care episodes. However, the types of individual contacts and data sources included when forming hospital care episodes are often not described [16, 23, 25]. A recent study sought to validate an algorithm classifying hospitalisations in the North Denmark region as either inpatient, outpatient or emergency room, and testing different hourly cut-offs for combining contacts [26]. However, the algorithm is not made publicly available, not tested on national data and not designed to incorporate all DNPR2 data types or bridge the differences between DNPR2 and DNPR3 data. There remains a lack of a universal approach, limiting comparability between studies and introducing a risk of errors.

The aim of this paper was to develop an algorithm capable of reading in and modifying the relevant raw data and combining sequential DNPR contacts into hospital care episodes. We present current difficulties and provide reasoning for our proposed solutions. Finally, we compared the effect of different hourly cut-offs and tested the practical application of the algorithm in various data sets.

Methods

This study was a methodological consensus-driven algorithm development study. The algorithm was developed in collaboration with researchers across four renowned epidemiological institutions with vast and diverse experience with register-based studies. The final result was a SAS macro named %DNPR_contact_combine. The SAS macro includes recommended features and is available from the online repository Figshare (collection: https://figshare.com/s/2122819d59fa2e728dcf) and Supplemental file 1. We are currently working on a translation for the R software, which will be made available from the repository when tested. The developed algorithm can be applied to data from 2005 and onwards, including the new data structure introduced with DNPR3 from February/March 2019.

The algorithm was initially drafted by RG based on previous experience [27]. The code was then adjusted and optimised with CFM and MV, before discussion with the remaining authors and changes as appropriate. In total, the macro has been tested by six experienced coders and used on different projects at Statistics Denmark. The purpose of the algorithm is to identify whether individual DNPR contacts in which a patient is physically present at the hospital are part of a shared hospital care episode. When determining if contacts are part of the same hospital

care episode, we only considered the time duration between the end of a contact and the initiation of the next contact.

A detailed description and walk-through of the developed algorithm can be found in Supplemental file 2. In short, the algorithm consists of six steps visualised in Figure 1. Depending on the specified start and end year, the algorithm scans the specified folders for the necessary DNPR2 and/or DNPR3 data files and checks for the needed variables. The contacts are restricted to public hospitals and adjusted only to represent the time when a patient is physically present at the hospital. Misregistration and missing information are corrected as appropriate. Then sequential contacts by the same person are linked as part of the same hospital care episode if initiated within 4 h from the previous contact. Information about the hospital care episode is added, classifying them as either elective or acute and outpatient or inpatient, and whether the hospital care episode contains somatic contacts, psychiatric contacts or both. General information on healthcare in Denmark and the development and contact structure of DNPR can be found elsewhere [5, 13], the information most important for this paper is described in Supplemental file 3. The DNPR nomenclature and definitions are listed in Table I.

To showcase the process of the algorithm, we used an established database at Statistics Denmark (project 707838) [27]. The database holds information on all adults (⩾18 years of age) living in Denmark at some point during 2013–2021 and all persons with a temporary central person register number at hospital contacts in the same period. For this population, we have information on all DNPR2 contacts, including psychiatric contacts, unfinished contacts and matching outpatient attendance dates, along with DNPR3 data. The needed files and variables of raw data, instructions on how to prepare the data and run the SAS macro and an overview of variables added to the final dataset are listed in Supplemental file 4. An introduction to the SAS macro language, variables and functions can be found elsewhere [28]. To compare the effect of different hourly cut-offs, we invoked the macro with 2-h, 3-h, 4-h, 5-h, 6-h, 8-h, 12-h and 24-h cut-offs but otherwise comparable settings, and evaluated the total number of hospital care episodes and mean number of contacts per hospital care episode in relative change (%) compared with a 4-h cut-off.

Ethics

The project was approved by the Data Protection Agency (P-2019-616). Ethical approval was not necessary for this register-based study according to Danish law.

Results

To demonstrate the algorithm, we invoked the SAS macro with the code shown in Supplemental files 1 and 2. The results generated by the optional data report are provided in Supplemental file 5. The run completed in 119 min; the same run without the data report completed in 85 min. In years 2013–2019 we included 76.8 m somatic, 8.3 m psychiatric and 224,023 unfinished DNPR2 contacts (prior to attendance date duplication) and in years 2019–2021 42.6 m DNPR3 contacts (Supplemental file 4). Our final dataset consisted of 120.2 m contacts (recounting duplicated, elective DNPR2 contacts) from 5.7 m individuals, which were linked into 105.9 m hospital care episodes. Of the final hospital care episodes, 6.8 m (6.4%) were considered acute inpatient, 8.8 m (8.3%) acute outpatient, 2.1 m (2.0%) elective inpatient and 88.2 m (83.3%) elective outpatient. In addition, 97.2% involved only somatic contacts, 2.8% involved only psychiatric contacts and 0.05% involved both somatic and psychiatric contacts.

Different hourly cut-offs

As the DNPR3 raw data include precise information on the start time and end time of all contacts, it can be used to recommend an optimal number of hours for contact combining [26]. Therefore, we invoked the SAS macro on the DNPR3 data with a 1-h contact combination window, thereby combining contacts that start and end very closely to each other but leaving contacts initiated more than 1 h after uncombined. These are regarded as new, separate hospital care episodes. Afterwards, we evaluated if a new hospital care episode was initiated within the next 36 h and, if so, how much time had passed, as depicted in Figure 2. It showed that a large proportion of subsequent hospital care episodes were initiated within 4 h (marked with red), after which new hospital care episodes declined less steeply or plateaued. Many histograms have new bulges of hospital care episodes after 8–30 h, especially pronounced for subsequent elective outpatient hospital care episodes, which were not likely to be part of the same hospital care episode. Therefore, we recommend 4 h as a standard time cut-off for combining hospital care episodes, but 3–6-h windows also seem commendable.

Following this, we investigated the effect of 2-h, 3-h, 5-h, 6-h, 8-h, 12-h and 24-h cut-offs compared with a 4-h cut-off, as our suggested reference. An increase in hourly cut-offs caused a decrease in acute

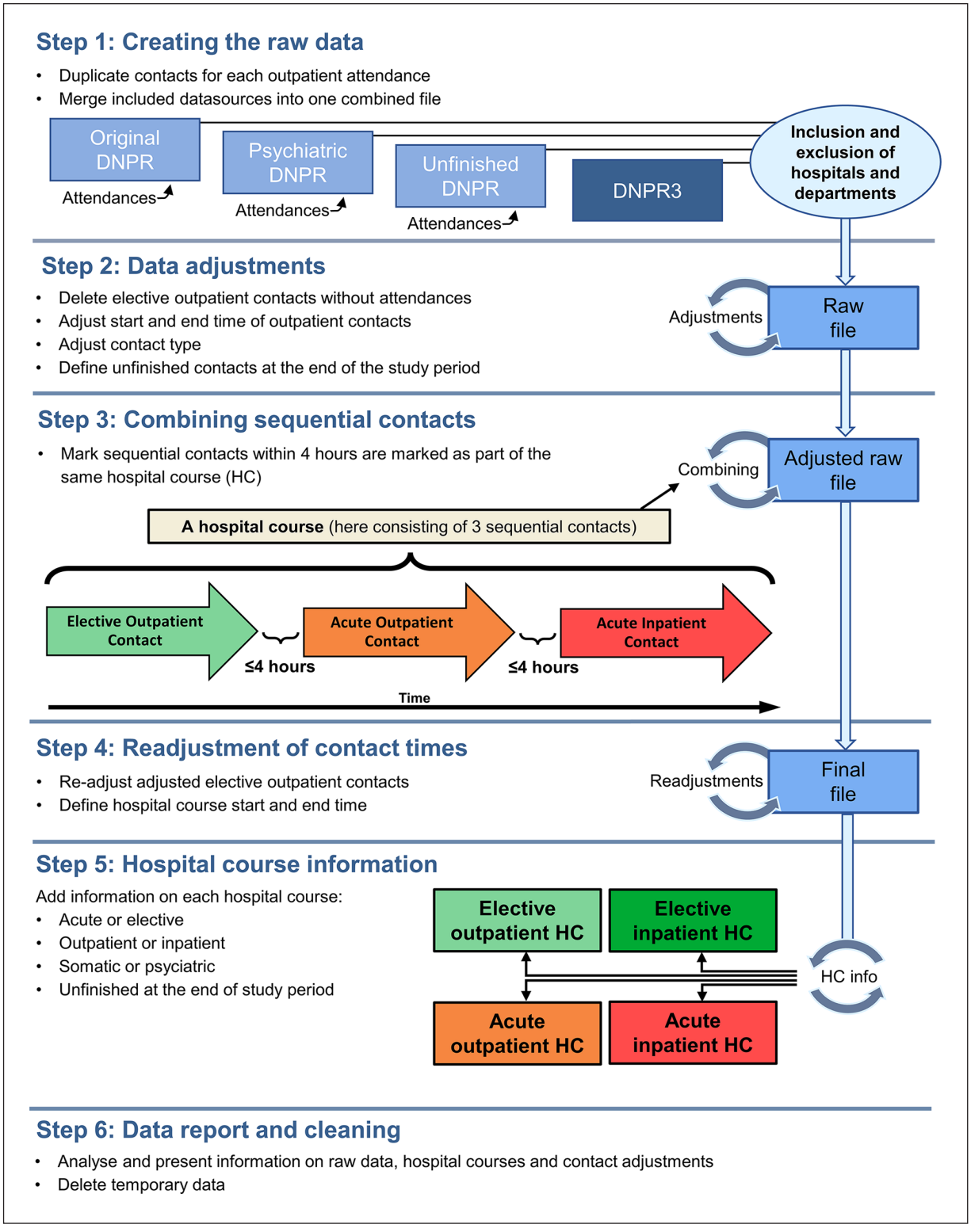

Figure 1. Overview of the algorithm process in which contacts from different Danish National Patient Register (DNPR) sources are compiled in a shared file and adjusted before sequential contacts are combined into hospital care episodes (HCEs).

inpatient, acute outpatient and elective outpatient hospital care episodes and an increase in elective inpatient hospital care episodes (Table II). When compared with a 4-h cut-off, we only saw minor relative changes compared with 3-h (+0.1% to +0.2%), 5-h (−0.1 to −0.2%) and 6-h cut-offs (−0.1% to −0.4%). However, the corresponding absolute difference of acute inpatient hospital care

Table I. Definitions and denominations from the Danish National Patient Register (DNPR).

Term/group (synonyms)	Definition
Contact	A record in DNPR representing either (i) a physical stay (of any duration) or (ii) a diagnostic or treatment course. A record in DNPR3 representing a patient-hospital contact, either physical or digital.
Inpatient contact (hospitalisation contact)	A contact in which the patient is considered hospitalised, meaning intended kept for longer or overnight stay, and assigned a hospital bed capacity (Danish: Normeret sengeplads). With DNPR3, contacts are no longer registered as inpatient contacts or outpatient contacts.
Emergency department contact	Acute outpatient contacts at emergency departments, discontinued from 2014 and onwards.
Outpatient contact (visit contact)	A contact in which the patient is not considered hospitalised.
Acute contact (unplanned contact)	Contacts that were not planned.
Elective contact (planned contact)	Contacts that were planned.
Attendance dates	The dates during an outpatient contact in which the patient was physically present at the hospital. Mandatory to register for elective outpatient contacts, not for acute outpatient contacts.
Unfinished contacts (ongoing contacts)	Contacts which are still ongoing when the data are delivered.
Psychiatric contacts	Contacts taking place at a psychiatric department.
Somatic contacts	Contacts that are not taking place at a psychiatric department.
Hospital care episode	Linked individual contacts representing the complete hospital care episode from a patient arriving at the hospital until the patient is discharged. Can consist of a single contact if there was no transfer.
Hospital course types	
• Acute inpatient hospital care episode	At least one acute contact and one inpatient contact.
• Acute outpatient hospital care episode	At least one acute contact but no inpatient contact.
• Elective inpatient hospital care episode	No acute contact but at least one inpatient contact.
Elective outpatient hospital care episode.	No acute contact and no inpatient contact.

episodes differed by +10,310 (3 h) to −12,146 (6 h) and acute outpatient hospital care episodes differed by +20,106 (3 h) to −32,645 (6 h), which might affect outcome measures such as readmission or incidence notably. The mean number of contacts per hospital care episode increased for all hospital care episode types with increasing hourly cut-offs (Table II). A 24-h cut-off was markedly different from the others, probably as many patients with outpatient visits the following day were linked.

Discussion

We outlined the difficulties in combining DNPR contacts into hospital care episodes, developed an algorithm for handling the issues and provided an SAS macro executing the algorithm. The algorithm modifies the raw data as needed: duplicating DNPR2 outpatient contacts by attendance dates, deleting contacts without attendances or outside public hospitals, modifying DNPR2 misregistration and missing data and combining the different DNPR2 and DNPR3 data sources, among others. Subsequently, the algorithm marks sequential contacts by the same patient within a specified duration as part of a combined hospital care episode, reflecting the entire duration from the patient entering the hospital to the patient leaving the hospital, including inter and intra-hospital transfers. Hospital care episodes thereby reflect single or more DNPR contacts considered as part of the same stay. Based on DNPR3 data we recommend a 4-h cut-off for combining contacts. The algorithm is compatible with data from 2005 and onwards and incorporates both DNPR2 somatic, psychiatric and unfinished contacts, as well as DNPR3 data. Six programmers confirmed the practical application of the algorithm.

Correct contact combination is essential in many observational study designs, as it is the only way to determine at which department the patient initially arrived, which departments the patient had been transferred to, where the patient was discharged from, how long the patient stayed, and if contacts were part of the same hospital care episode or could be considered a readmission. Therefore, incorrect contact combining may cause errors when studying length of stay, in-hospital mortality, transfers to the intensive care unit, outcomes after discharge, incidence of diseases and readmissions, among others. Only considering some types of available contacts might also introduce errors. In our demonstration, 0.05% (n=56,771) of hospital care episodes contained contacts from both psychiatric and somatic departments. However, when only considering acute inpatient hospital care episodes, 0.5% included both somatic and psychiatric contacts. This highlights that hospital care episodes could be misclassified as too short or as two separate instances if all data sources had not been used. For the DNPR2 structure, elective outpatient contacts do not reflect the physical attendance but rather the entire outpatient course. Therefore, to identify physical attendance correctly across an outpatient course, it is necessary to consider the matching visit dates and, in some instances, define attendance times. All these data manipulations are necessary for reducing the risk of errors, and are

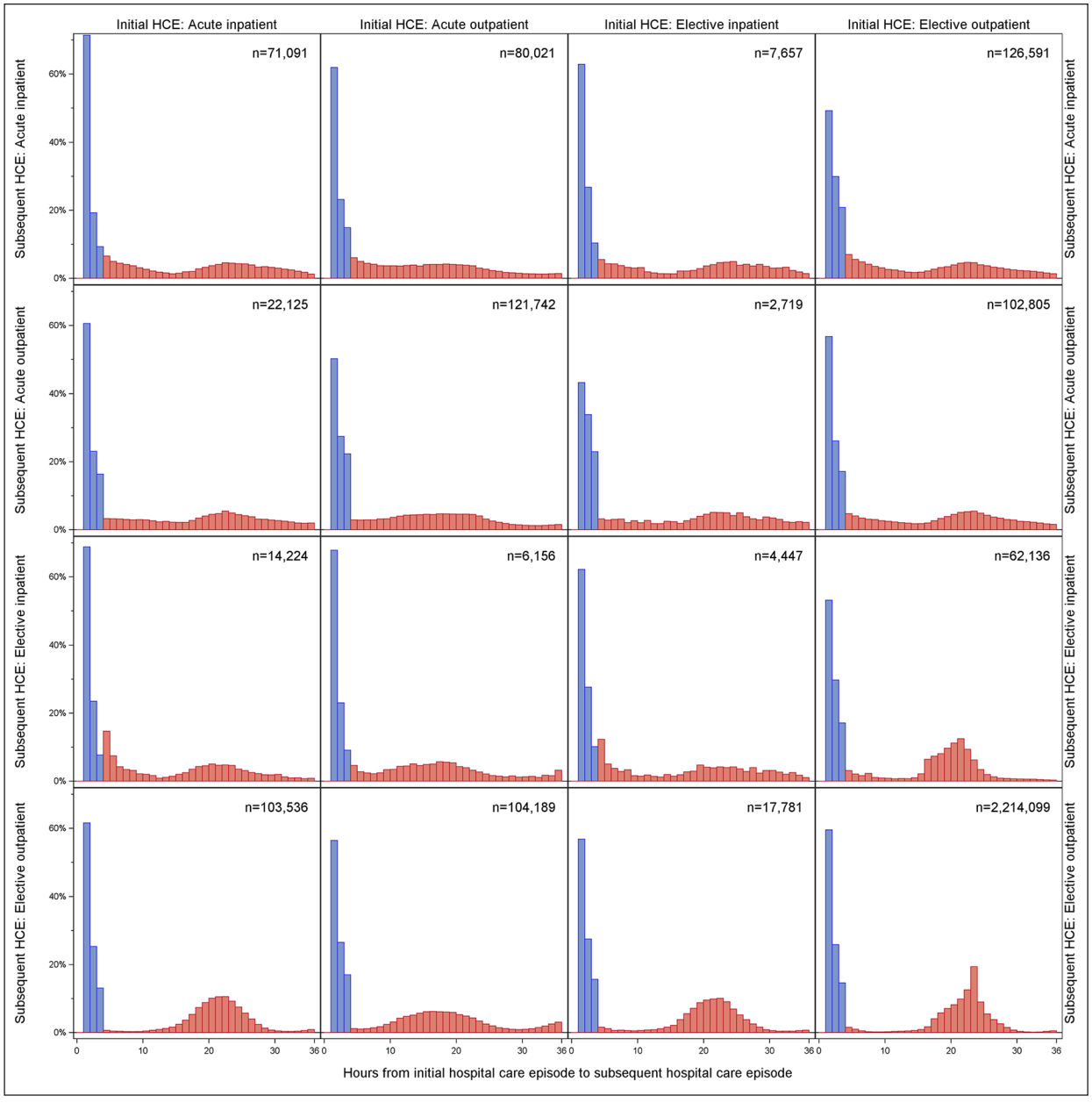

Figure 2. Histograms of hours from ended hospital care episode in the Danish National Patient Register third version (formed by linking contacts <1 h from each other) to a subsequent hospital care episode within 36 h, stratified by type of initial and subsequent hospital care episode. The graph shows a slower declining or plateauing proportion in most histograms after 4 h, marked with blue. HCE: hospital care episode.

included in the presented algorithm as a suggestion for a standardised approach for combining DNPR contacts to hospital care episodes. Further, this allows for better comparisons between studies and lowers the barriers to entry for new researchers.

With the introduction of DNPR3, many of the previously needed data modifications are no longer needed. As the register is contact based, there is no need to merge with attendance dates nor any need to correct or define contact durations because these are registered for each contact. We only used information from the contact file, reflecting which department holds the responsibility during the contact, as with the DNPR2 structure. With DNPR3, further information on the actual department, division, or clinic location(s) is available through Sundhedsvæsenets Organisationsregister (SOR, the register of healthcare organisation) codes [26]. For instance, this could reveal if a patient has been in an operating room during the contact or was transferred between

Table II. Difference in number of hospital care episodes and number of contacts per hospital care episode when running the %DNPR_contact_combine macro with different hourly cut-offs, compared with a 4-h cut-off.

	Hospital care episode type	2 h (difference)	3 h (difference)	4 h (reference)	5 h (difference)	6 h (difference)	8 h (difference)	12 h (difference)	24 h (difference)
Total number of hospital care episodes	Acute inpatient hospital care episode	+28,983 (+0.4%)	+10,310 (+0.2%)	6,786,581	–7054 (–0.1%)	–12,146 (–0.2%)	–18,766 (–0.3%)	–13,302 (–0.2%)	46,622 (0.7%)
	Acute outpatient hospital care episode	+48,469 (+0.6%)	+20,106 (+0.2%)	8,802,300	–17,029 (–0.2%)	–32,645–0.4%)	–64,157 (–0.7%)	–149,572 (–1.7%)	–494,011 (–5.6%)
	Elective inpatient hospital care episode	+14,655 (+0.7%)	+4408 (+0.2%)	2,123,817	–2734 (–0.1%)	–4724 (–0.2%)	–7485 (–0.4%)	–4209 (–0.2%)	915,368 (43.1%)
	Elective outpatient hospital care episode	+263,548 (+0.3%)	+103,966 (+0.1%)	88,211,414	–73,856 (–0.08%)	–131,705 (–0.1%)	–209,422 (–0.2%)	–372,067 (–0.4%)	–7,124,915 (–8.1%)
Number of contacts per hospital care episode	Acute inpatient hospital care episode	–0.03 (–1.5%)	–0.01 (–0.6%)	1.91	+0.01 (+0.4%)	+0.02 (+0.8%)	+0.03 (+1.4%)	+0.05 (+2.6%)	+0.21 (+10.8%)
	Acute outpatient hospital care episode	–0.01 (–0.6%)	<–0.01 (–0.3%)	1.12	<+0.01 (+0.2%)	+0.01 (+0.5%)	+0.01 (+0.8%)	+0.02 (+1.4%)	+0.04 (+3.1%)
	Elective inpatient hospital care episode	–0.01 (–0.6%)	<–0.01 (–0.2%)	1.14	<+0.01 (+0.2%)	<+0.01 (+0.3%)	+0.01 (+0.5%)	+0.02 (+1.6%)	+0.64 (+56.3%)
	Elective outpatient hospital care episode	<–0.01 (–0.2%)	<–0.01 (–0.06%)	1.05	<+0.01 (+0.035%)	<+0.01 (+0.057%)	<+0.01 (+0.073%)	<+0.01 (+0.1%)	+0.04 (+4.0%)

Differences displayed by actual difference (relative change in percentage).

two different wards under the same overall department's responsibility. DNPR3 lacks information on whether the contact was considered outpatient or inpatient, but this information is also somewhat arbitrary and was possibly registered differently across different institutions in DNPR2. Rather than considering each contact, we defined the entire hospital care episode as inpatient if lasting more than 12 h and spanning at least two dates [26]. However, in a clinical setting, patients can be considered as acute outpatients for up to 48–72 h staying in the emergency departments [29].

Different hourly cut-offs for contact combining showed some effect on the final number of hospital care episodes – increasing the duration caused a decrease in both the number of acute inpatient and outpatient hospital care episodes, whereas the mean number of contacts per hospital care episode increased. Ideally, determining the most optimal hourly cut-off would be based on clinical validation evaluating the correct or incorrect combination of contacts. However, this was not possible within the scope of this paper. Contact combining should be long enough to include even long interhospital transfers while not being unnecessarily long, and including contacts which should not have been a separate hospital care episode (e.g. readmissions). Based on our visualisation and this clinical reasoning, we suggest 4 h as a standard. This agrees with a recent study showing that most subsequent DNPR3 contacts within 36 h occur within the first 4–6 h. However, there is also a concentration of subsequent outpatient contacts after approximately 8–24 h [26]. Another study compared 6-h and 12-h cut-offs with 3-h cut-offs resulting in ~120,000 and ~219,000 fewer acute contacts, respectively, over a 12-year period in DNPR2 [16]. As an alternative to duration, a contact combination of sequential contacts could also be conducted based on arrival and discharge information (for instance, one contact ending with 'referred to another department' and the next contact, within some timeframe, initiated with 'referred from other department'). However, when a contact is just registered as terminated rather than transferred, this will also erroneously be reflected in a missing combination of contacts that should have been combined [5]. The increase in elective inpatient hospital care episodes with increasing hourly cut-off is probably caused by our DNPR3 inpatient definition, in which contacts lasting more than 12 h and spanning over at least two dates are considered inpatient.

The algorithm was developed in collaboration with a group of experienced researchers with different perspectives and backgrounds to ensure that relevant aspects have been considered and that the algorithm is broadly applicable. Besides thorough development and testing in various data sets, we did not compute quantitative measures of validation. Thus, we cannot rule out that we have missed out on information which should have been considered, incorporated, or handled differently. If major issues are identified, or future updates change the data structure, we will strive to upload updated versions to the online repository. When working with the final data, it is important to be aware of the contact identifiers. Due to our duplication of elective outpatient contacts by attendance dates, there will be duplicate values of the unique identifier recnum for DNPR2. For DNPR3 data, other identifiers are used.

The presented algorithm provides a solution applicable for the DNPR. Several studies have previously combined study populations across Nordic countries using different national patient registers [30–32]. This is possible, as the national patient registers across the Nordic countries are very similar in data structure and content [12]. It is likely that an approach similar to ours with adaptations towards local data structures is also relevant in these or other countries. For creating an adapted approach in other country settings, the most important steps are to include all types of patient contacts (inpatient, outpatient and emergency department contacts, as well as psychiatric and somatic contacts), modify the contacts to represent actual attendances including arrival and departure times, exclude contacts not taking place at public hospitals and combine sequential contacts within 4 h to hospital care episodes. A key understanding of the underlying data mechanisms, reporting and definitions are of the utmost importance throughout development.

Conclusions

This new algorithm for forming hospital care episodes by combining sequential attendance contacts in the DNPR contacts was capable of handling and processing different types of DNPR2 and DNPR3 data. The algorithm can be initiated as an SAS macro and is freely available from an online repository.

Acknowledgements

In memory of Cathrine Fox Maule: It is with great sorrow that we have lost our valued colleague Cathrine Fox Maule after a long illness. We will remember Cathrine for all her help, insights and partnership on many different projects. She was a mild, positive and warm person, always helpful, constructive, supportive and patient in all aspects of her work. She remained active in our research group until the very end. Cathrine passed away on 11 November 2023, leaving behind her husband and two kids. May she rest in peace.

Author contributions

RG conceptualised the study and developed the initial model in collaboration with CFM, supervised by JP and MV. The model was further adjusted with inputs from all authors, and tested by CBJ, MWJ and AHA. RG performed the formal analyses, created graphical illustrations and wrote the initial draft. All authors critically reviewed the manuscript.

Data availability

Due to the rules of protection of individuals' data from Statistics Denmark, it is not possible to share the data in any raw or anonymised form. Danish research institutions can obtain permission to access the data on equal terms. The algorithm has been made freely and publicly available through the online repository Figshare (collection: https://figshare.com/s/2122819d59fa2e728dcf).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work was financed by a grant from the University of Copenhagen (2023 Strategy Funds: Data+ grant), a grant from the Research Council at Bispebjerg and Frederiksberg Hospital, and a grant from the Department of Emergency Medicine.

Ethics approval and consent to participate

Ethical approval is not required for registry-based studies in Denmark. The project was approved by Statistics Denmark (project number: 707838).

Presentations

Oral presentation at the NordicEpi 2024 conference (Copenhagen, June 2024).

Software availability and requirements

Project name: %DNPR_contact combine SAS macro
Project home page: https://figshare.com/s/2122819d59fa2e728dcf

Operating system: NA
Programming language: SAS
Licence: CC BY 4.0

ORCID iD

Rasmus Gregersen https://orcid.org/0000-0001-9439-4933

Supplemental material

Supplemental material for this article is available online.

References

[1] 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.

[2] Romanini E, Schettini I, Torre M, et al. The rise of registry-based research: a bibliometric analysis. *Acta Orthop* 2021;92:628–632.

[3] Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence – what is it and what can it tell us? *N Engl J Med* 2016;375:2293–2297.

[4] Yuan H, Ali MS, Brouwer ES, et al. Real-world evidence: what it is and what it can tell us according to the International Society for Pharmacoepidemiology (ISPE) Comparative Effectiveness Research (CER) Special Interest Group (SIG). *Clin Pharmacol Ther* 2018;104:239–241.

[5] Schmidt M, Schmidt SAJ, Sandegaard JL, et al. The Danish National patient registry: a review of content, data quality, and research potential. *Clin Epidemiol* 2015;7:449–490.

[6] Rothman KJ. Six persistent research misconceptions. *J Gen Intern Med* 2014;29:1060–1064.

[7] Thygesen LC and Ersbøll AK. When the entire population is the sample: strengths and limitations in register-based epidemiology. *Eur J Epidemiol* 2014;29:551–558.

[8] Khunti K, Almalki M, Chan JCN, et al. The role of real-world evidence in treatment decision-making, regulatory assessment, and understanding the perspectives of people with type 2 diabetes: examples with gliclazide MR. *Diabetes Ther* 2023;14:1609–1625.

[9] Lau B, Duggal P, Ehrhardt S, et al. Perspectives on the future of epidemiology: a framework for training. *Am J Epidemiol* 2020;189:634–639.

[10] Ebrahim S, Ferrie JE and Smith GD. The future of epidemiology: methods or matter? *Int J Epidemiol* 2016;45:1699–1716.

[11] Maret-Ouda J, Tao W, Wahlin K, et al. Nordic registry-based cohort studies: possibilities and pitfalls when combining Nordic registry data. *Scand J Public Health* 2017;45:14–19.

[12] Laugesen K, Ludvigsson JF, Schmidt M, et al. Nordic health registry-based research: a review of health care systems and key registries. *Clin Epidemiol* 2021;13:533–554.

[13] 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.

[14] Schmidt M, Pedersen L and Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol* 2014;29:541–549.

[15] Søvsø MB, Hermansen SB, Færk E, et al. Diagnosis and mortality of emergency department patients in the North Denmark region. *BMC Health Serv Res* 2018;18:1–9.

[16] Fløjstrup M, Bogh SB, Henriksen DP, et al. Increasing emergency hospital activity in Denmark, 2005–2016: a nationwide descriptive study. *BMJ Open* 2020;10:1–8. DOI: 10.1136/bmjopen-2019-031409.

[17] Bogh SB, Fløjstrup M, Nissen SK, et al. Nationwide study on trends in unplanned hospital attendance and deaths during the 7 weeks after the onset of the COVID-19 pandemic in Denmark. *BMJ Qual Saf* 2021;30:986–995.

[18] Pedersen MK, Nielsen GL, Uhrenfeldt L, et al. Description of OPRA: a Danish database designed for the analyses of risk factors associated with 30-day hospital readmission of people aged 65+ years. *Scand J Public Health* 2017;45:595–604.

[19] Bodilsen J, Nielsen PB, Søgaard M, et al. Hospital admission and mortality rates for non-covid diseases in Denmark during covid-19 pandemic: nationwide population based cohort study. *BMJ* 2021;373:.

[20] 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;17:1–7.

[21] Vest-Hansen B, Riis AH, Sørensen HT, et al. Acute admissions to medical departments in Denmark: diagnoses and patient characteristics. *Eur J Intern Med* 2014;25:639–645.

[22] Gregersen R, Andresen K, Burcharth J, et al. Short-term mortality, readmission, and recurrence in treatment of acute diverticulitis with abscess formation: a nationwide register-based cohort study. *Int J Colorectal Dis* 2016;31:983–990.

[23] Ridgeway G, Nørgaard M, Rasmussen TB, et al. Benchmarking Danish hospitals on mortality and readmission rates after cardiovascular admission. *Clin Epidemiol* 2019;11:67–80.

[24] Ruwald MH, Hansen ML, Lamberts M, et al. The relation between age, sex, comorbidity, and pharmacotherapy and the risk of syncope: a Danish nationwide study. *Europace* 2012;14:1506–1514.

[25] Elkjær M, Wolff DL, Primdahl J, et al. Older adults who eceive homecare are at increased risk of readmission and mortality following a short ED admission: a nationally register-based cohort study. *BMC Geriatr* 2021;21:1–9.

[26] Skjøth F, Nielsen H and Bodilsen J. Validity of algorithm for classification of in- and outpatient hospital contacts in the Danish National Patient Registry. *Clin Epidemiol* 2022;14:1561–1570.

[27] Gregersen R, Maule CF, Bak-Jensen HH, et al. Profiling Bispebjerg Acute Cohort: database formation, acute contact characteristics of a metropolitan hospital, and comparisons to urban and rural hospitals in Denmark. *Clin Epidemiol* 2022;14:409–424.

[28] UCLA: Statistical Consulting Group. *Introduction to SAS macro language*. stats.oarc.ucla.edu. 2021. https://stats.oarc.ucla.edu/sas/seminars/sas-macros-introduction/ (accessed 1 March 2022).

[29] Tanderup A, Lassen AT, Rosholm JU, et al. Disability and morbidity among older patients in the emergency department: a Danish population-based cohort study. *BMJ Open* 2018;8:1–11. DOI: 10.1136/bmjopen-2018-023803.

[30] Olén O, Erichsen R, Sachs MC, et al. Colorectal cancer in ulcerative colitis: a Scandinavian population-based cohort study. *Lancet* 2020;395:123–131.

[31] Halvorsen S, Ghanima W, Johnsen SP, et al. Effectiveness and safety of non-vitamin K antagonist oral anticoagulants and warfarin in atrial fibrillation: a Scandinavian population-based cohort study. *Eur Heart J – Qual Care Clin Outcomes* 2022;8:577–587.

[32] Stephansson O, Kieler H, Haglund B, et al. Selective serotonin reuptake inhibitors during pregnancy and risk of stillbirth and infant mortality. *JAMA* 2013;309:48–54.