



ORIGINAL ARTICLE

Enhancing the accuracy of register-based metrics: Comparing methods for handling overlapping psychiatric register entries in Finnish healthcare registers

Kimmo Suokas^{1,2}  | Mai Gutvilig¹ | Sonja Lumme³  | Sami Pirkola^{2,4} | Christian Hakulinen^{1,3}

¹Department of Psychology, Faculty of Medicine, University of Helsinki, Helsinki, Finland

²Faculty of Social Sciences, Tampere University, Tampere, Finland

³Finnish Institute for Health and Welfare, Helsinki, Finland

⁴Department of Psychiatry, The Pirkanmaa Wellbeing Services County, Tampere, Finland

Correspondence

Kimmo Suokas.

Email: kimmo.suokas@helsinki.fi

Funding information

Academy of Finland, Grant/Award Number: 354237; European Research Council, Grant/Award Number: 101040247; Jalmari ja Rauha Ahokkaan Säätiö; Finnish Psychiatric Association

Abstract

Objectives: Healthcare registers are invaluable resources for research. Partly overlapping register entries and preliminary diagnoses may introduce bias. We compare various methods to address this issue and provide fully reproducible open-source R scripts.

Methods: We used all Finnish healthcare registers 1969–2020, including inpatient, outpatient and primary care. Four distinct models were formulated based on previous reports to identify actual admissions, discharges, and discharge diagnoses. We calculated the annual number of treatment events and patients, and the median length of hospital stay (LOS). We compared these metrics to non-processed data. Additionally, we analyzed the lifetime number of individuals with registered mental disorders.

Results: Overall, 2,130,468 individuals had a registered medical contact related to mental disorders. After processing, the annual number of inpatient episodes decreased by 5.85%–10.87% and LOS increased by up to 3 days (27.27%) in years 2011–2020. The number of individuals with lifetime diagnoses reduced by more than 1 percent point (pp) in two categories: schizophrenia spectrum (3.69–3.81pp) and organic mental disorders (1.2–1.27pp).

Conclusions: The methods employed in pre-processing register data significantly impact the number of inpatient episodes and LOS. Regarding lifetime incidence of mental disorders, schizophrenia spectrum disorders require a particular focus on data pre-processing.

KEYWORDS

epidemiology, healthcare register, psychiatry

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). International Journal of Methods in Psychiatric Research published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Healthcare registers provide a valuable resource for medical research due to their ability to encompass large study populations with extensive and continuous follow-up (Cheng, 2015; Laugesen et al., 2021; Lichtenberg et al., 1999; Maret-Ouda et al., 2017). In the Nordic countries, register data can be linked by a personal identification number, which enhances the reliability of the secondary use of registries and enables versatile utilization across different research settings. Furthermore, in Finland and in Sweden, for example, healthcare registers are generally considered to have good data quality for research purposes (Laugesen et al., 2021; Ludvigsson et al., 2011; Maret-Ouda et al., 2017; Sund, 2012). However, maintaining the quality of the registers is an ongoing process, marked by various technical nuances. Register-based analyses rely on several methodological assumptions that may compromise reproducibility of the results and challenge international comparisons of healthcare systems (Katschnig et al., 2019).

Registers with continuous and mainly automatized data collection, such as the Finnish healthcare registers, contain temporally overlapping entries and preliminary information. Therefore, a single inpatient episode, for example, may be recorded into multiple register entries (Pirkola & Sohlman, 2005). These multiple entries arise when there are intra-hospital transfers or shifts between distinct medical specialties within the same facility potentially resulting in temporal overlap. Moreover, entries are generated for outpatient and emergency department events that transpire at the outset or during the hospitalization and may contain possibly preliminary diagnoses recorded at the time of these events. As a result, combining multiple register entries of different treatment modalities and recognizing potential preliminary diagnoses becomes necessary in order to accurately identify the most reliable estimates of actual discharges, discharge diagnoses, and independent outpatient events that are not part of inpatient care.

A few research projects, such as the CEPHOS-LINK project (Katschnig & Straßmayr, 2017) or the REDD project (Kajantie et al., 2006), have addressed the importance of the procedures for identifying inpatient episodes from the partly overlapping healthcare register data. The CEPHOS-LINK project, for example, indicated that as much as 25% of the register entries associated with psychiatric inpatient care in Finland are related to transfers that take place during an ongoing hospitalization (Katschnig & Straßmayr, 2017). Despite the prevalence of this issue, a standardized consensus on best practices for handling inpatient episodes within the Finnish registers has yet to be established, leading to variations in criteria applied by different psychiatry-related research projects. For example,:

- the CEPHOS-LINK project required that a hospital stay should start and end on distinct calendar days, essentially requiring an inpatient episode to span overnight, whereas others do not have this criterion (Katschnig & Straßmayr, 2017). Similarly,
- the REDD project introduced a condition according to which a new treatment period could only commence after a full calendar day

spent outside the hospital (Kajantie et al., 2006). Any entries within the register prior to this transition were amalgamated. This approach aimed to create a clearer distinction between inter-hospital transfers and subsequent rehospitalizations.

To the best of our knowledge, these approaches have not been systematically compared, and the methods employed for pre-processing healthcare registers have generally not been publicly disclosed. This study aims to quantify differences in pre-processing strategies by comparing the number of individuals treated, the number of inpatient episodes, and the average length of stay estimates using different criteria for identifying treatment events and discharges. Our analysis incorporates inpatient, secondary outpatient, and primary care data. We hypothesized that different rules for identifying inpatient episodes lead to changes in annual and lifetime metrics of inpatient and outpatient psychiatric care with differences across diagnostic categories. In addition to general evaluation of these methods, we provide fully reproducible, open-source R scripts along with example synthetic data, in order to enable others to evaluate and benefit from this effort in the context of Finnish registers.

2 | METHOD

For this study, all Finnish healthcare register data until the end of 2020 was utilized. Individuals with a history of mental health-related contact with psychiatric inpatient care were reliably recognized since 1975, secondary outpatient care was included since 1998 and primary care since 2011.

The Research Ethics Committee of the Finnish Institute for Health and Welfare approved the study protocol (decision #10/2016S751). Informed consent is not required for register-based studies in Finland.

2.1 | Description of the registers

The Finnish Care Register for Health Care, formerly known as the Finnish Hospital Discharge Register (FHDR) prior to 1994, provides continuous nationwide inpatient data with coverage dating back to 1969, making it the first register among Nordic countries (Maret-Ouda et al., 2017; Sund, 2012). The format of the data is slightly different across the years. Initially, the FHDR consisted of separate sub-registers catering to diverse hospital types, including general, tuberculosis, psychiatric, and others. These sub-registers were subsequently consolidated into a unified system (Sund, 2012). For the years 1969–1974, the classification of medical specialties in some sub-registers may lack clarity.

Data on secondary outpatient care in the public sector have been collected since 1998, with consistent comparability across time and service providers achieved from 2006 onward (THL, 2022). The Register of Primary Health Care visits (RPHC) encompasses

publicly organized outpatient primary health care since 2011. Its inclusion has significantly enhanced the comprehensiveness of the registers, potentially influencing results in register-based psychiatric epidemiological research, especially due to the common utilization of primary care for mental health treatment in Finland (Suokas et al., 2022, 2023).

The International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) has been used in Finland since 1996. Prior to that, the Finnish version of the ICD-9 was used from 1987 to 1995, and ICD-8 from 1969 to 1986. In some primary care facilities, the International Classification of Primary Care, Second Edition (ICPC-2), is used instead of ICD-10.

2.2 | Procedures

2.2.1 | Sample definitions and initial data transformations

To ensure data quality and consistency, the following initial quality controls and data transformations were implemented: entries without a personal identification number were excluded. Entries where the admission or discharge date was missing, discharge was before admission, or entries out of the time range of the dataset were excluded.

Starting from 1996, patients in inpatient care on the last day of the year were reported in the register. In these entries, the discharge day should be left blank; however, this was not consistently observed and was accounted for in the scripts.

The coding of psychiatric care varies across years. Uniform specialty coding in all register entries commenced in 1987, with further coding changes introduced in 1994. Prior to 1987, register entries from both mental hospitals and general hospitals with psychiatry as a specialty were considered psychiatric care.

Prior to 1998, all entries pertained to inpatient care. Subsequently, distinguishing entries related to inpatient services became necessary. The coding of treatment type underwent substantial changes starting in 2019. In 2019, both the old and new coding systems were in use, and some entries even featured a mixture of both, which needs to be accounted for. For outpatient and emergency department events, consecutive start and end dates were allowed because the event may have taken place around midnight. However, if the stay in a non-inpatient facility spanned over more than two calendar days, it was considered inpatient care. This might be the case in emergency departments on some occasions. Day hospital care was defined as outpatient care, following the conventions used in Finland's official statistics (THL, 2021).

The RPHC contains data on assessment of the need for care, scheduling of the appointments, consultations between professionals and beyond. For the purpose of this study, in-person and virtual real-time contacts were included.

Mental health-related diagnoses under ICD-8, ICD-9, and ICPC-2 were converted to corresponding ICD-10 sub-chapter categories whenever possible. Conversion tables provided by the classification developers were utilized when possible (World Health Organization, 1994; WONCA International Classification Committee (WICC), 2005). ICPC-2 concepts that did not have exact counterparts in ICD-10 were grouped separately and were not included in the ICD-10 sub-chapter categories. See Supporting Information S1 for information on the register variables.

2.2.2 | Identification of inpatient episodes and outpatient events during the inpatient episode

Using the criteria outlined in the introduction section, we have derived four distinct models for identifying inpatient episodes from the healthcare registers (Table 1). Inpatient episodes were identified as follows:

TABLE 1 Possible models for identifying inpatient episodes.

Model	Description
1	A new hospitalization may begin the day after a previous hospitalization ended, with no specific minimum length required for a hospitalization. This represents the most liberal approach
2	A new hospitalization may begin the day after a previous hospitalization ended. Valid hospitalizations are those that extend over a minimum of two consecutive days, incorporating at least one overnight stay. If both admission and discharge take place on the same day, the visit is classified as an outpatient visit. This model was used in the CEPHOS-LINK project
3	A new hospitalization is allowed after a full day has been spent outside the hospital after the previous hospitalization ended. There is no specific minimum duration required for a hospitalization. This model was used in the REDD project
4	A new hospitalization is allowed after a full day has been spent outside the hospital after the previous hospitalization ended. Valid hospitalizations are those that extend over a minimum of two consecutive days, incorporating at least one overnight stay. If both admission and discharge take place on the same day, the visit is classified as an outpatient visit. This represents the most conservative model

1. We identified overlapping register entries related to inpatient care, with transfers recognized in two different ways: First, discharges and new admissions on the same day were considered transfers during an inpatient episode (Models 1 and 2, Table 1), and second, discharges and new admissions without a full calendar day outside the hospital were considered transfers (Models 3 and 4, Table 1). After that,
2. Overnight stay was examined and for Models 2 and 4, inpatient episodes starting and ending on the same day were reclassified as outpatient events.

After identifying inpatient episodes with each model, possible secondary outpatient and primary care events during the inpatient episode were identified. These register entries were not considered independent treatment events but were considered a part of the inpatient episode. Hence, the model selected for identifying inpatient episodes affected the number of outpatient and primary care events in the fully processed data.

If a single hospital stay involved treatment in multiple wards of various medical specialties, the overall inpatient episode was defined as the total time spent in the hospital. However, the psychiatric inpatient episode was considered to start from the initial psychiatric admission and end at the last psychiatric discharge.

2.2.3 | Identification of discharge diagnoses and outpatient diagnoses

We recorded discharge diagnoses on the last day of the psychiatric inpatient episode. If the patient was transferred to a ward of another specialty before the final discharge, we also recorded discharge diagnoses on the final day of the overall inpatient episode. Additionally, we documented all diagnoses made during the inpatient episode. Emergency department, outpatient and primary care events during the course of inpatient episodes were also identified. Psychiatric outpatient diagnoses registered on the day of or the days following discharge from a psychiatric ward before the final discharge were also included in the final discharge diagnoses.

Inpatient, outpatient, and primary care diagnoses established before the end of the inpatient episode were categorized as preliminary and were not considered discharge diagnoses. If an inpatient episode included psychiatric care, mental disorder diagnoses from specialties other than psychiatry were excluded.

In the Finnish healthcare registers, one main diagnosis and several additional diagnoses can be and often are included in a single register entry. We collected a list of diagnoses that included all of the diagnoses at the date of discharge from the psychiatric ward and possible additional diagnoses issued by psychiatrists in the case where the overall inpatient episode ended after discharge from psychiatric wards. All diagnoses meeting these criteria were included in the list of each patient's final discharge diagnoses; the number of possible discharge diagnoses was not limited.

The earliest age at which a person might possibly develop a specific disorder were set in a similar way as previously (Pedersen et al., 2014).

2.3 | Analysis

After identifying inpatient episodes and outpatient and primary care events associated with inpatient episodes, we compared the number of episodes and treated individuals calculated with the four models to corresponding numbers from the non-processed data (with initial data transformations only).

First, we conducted separate analyses for the years 2011–2020 to assess the impact of episode identification on annual descriptive statistics, including the number of psychiatric inpatient episodes, median length of stay of psychiatric inpatient episodes, and the number of treated individuals within specific diagnostic categories.

Second, we calculated the total number of individuals within specific diagnostic categories between 1975 and 2020.

Numbers and differences in percentages are presented throughout the study. We used R, version 4.2.2 for all data processing. The scripts have been made publicly available and contain supplementary description of each step of the process summarized above (https://github.com/kmmsks/hilmo_identify_episodes/).

3 | RESULTS

Between the years 1975 and 2020, 2,130,468 individuals had a valid registered medical contact related to mental disorders. Between years 2011 and 2020, less than 0.4% of observations were excluded due to missing personal identification number, missing admission date or discharge date, discharge recorded before admission, or entries out of the time range of the data set.

3.1 | Annual number of psychiatric inpatient episodes, outpatient events and number of patients

After the identification of inpatient episodes, accounting for the overlapping days, and recognizing outpatient events during inpatient episodes, the most liberal model (Model 1) resulted in 5.85%–7.63% less inpatient episodes than non-processed data. Applying the most conservative model (Model 4) for detecting inpatient episodes reduced the numbers of inpatient episodes by additional 2.31–3.63pp (Table 2).

Outpatient events were recognized after the identification of inpatient episodes. Model 1 resulted in 1.76%–2.84% less outpatient events compared to non-processed data. There was practically no difference between the models (less than 0.1pp) (Table 2).

The annual number of individuals with inpatient treatments in the year 2020 is shown in Table 3. If an overnight stay was required for a visit to count as inpatient treatment, the number of individuals

TABLE 2 Annual number of register entries related to psychiatric care and reduction in the number of inpatient episodes and outpatient events after identification of inpatient episodes with different models.

Year	Non-processed data, ^a N	Processed data, reduction %			
		Model 1 ^b	Model 2 ^c	Model 3 ^d	Model 4 ^e
Inpatient episodes					
2011	40,969	6.28	7.92	7.03	8.59
2012	39,721	5.85	7.52	6.56	8.20
2013	37,912	6.55	8.52	7.32	9.23
2014	36,996	6.77	9.13	7.64	9.91
2015	37,017	7.16	10.00	8.03	10.79
2016	35,191	6.35	8.49	7.19	9.25
2017	36,294	6.16	8.25	7.17	9.15
2018	36,912	6.50	8.60	7.40	9.39
2019	37,852	7.63	9.93	8.70	10.87
2020	36,207	7.03	9.44	8.17	10.42
Outpatient events					
2011	1,770,009	2.88	2.84	2.89	2.85
2012	1,794,413	2.77	2.74	2.78	2.75
2013	1,811,596	2.31	2.27	2.32	2.28
2014	1,856,891	2.29	2.24	2.29	2.25
2015	2,000,159	2.58	2.52	2.58	2.53
2016	2,056,493	2.07	2.04	2.08	2.05
2017	2,239,948	2.07	2.03	2.08	2.05
2018	2,184,623	2.14	2.10	2.15	2.11
2019	2,188,546	2.06	2.02	2.07	2.03
2020	2,281,021	1.80	1.76	1.81	1.77

^aValid non-processed data after initial quality checks.

^bHospitalizations can start the day after a previous one, with no minimum length requirement.

^cValid hospitalizations require at least two consecutive days with one overnight stay. Same-day admission and discharge count as outpatient visits.

^dA new hospitalization can start after a full day outside the hospital, with no minimum duration requirement.

^eValid hospitalizations require at least two consecutive days with one overnight stay. Same-day admission and discharge count as outpatient visits.

with inpatient episodes decreased by 1.46%–1.46%. Diagnosis-specific values changed by less than 3% in all ICD-10 sub-chapter categories. It is worth noticing that the number of individuals increased with pre-processing when psychiatric outpatient events during an inpatient stay were included.

3.2 | Length of stay of psychiatric inpatient care

The median length of stay increased by 1–3 days in different years with different models compared to the non-processed data (Table 4). In 2013–2015, the difference was the greatest, 3 days in models 2 and 4, resulting in a 27.27% change in the median length of stay. In 2020, the median length of stay increased from 8 to 9 or 10 days

(25%). The distribution of the number of episodes by length of stay is shown in Figure 1.

3.3 | Overall number of individuals covered by diagnosis

Between the years 1975 and 2020, 2,130,468 individuals had medical contacts related to mental disorders (Table 5). After the identification of inpatient episodes, the number of individuals with a lifetime diagnosis was reduced by more than 1 pp only in two of the ICD-10 sub-chapter categories: schizophrenia spectrum disorders (F20–F29) by 3.69–3.81pp and organic, including symptomatic, mental disorders (F00–F09) by 1.2–1.27pp. More specifically, organic

TABLE 3 Annual number of individuals with psychiatric inpatient episodes by discharge diagnosis in non-processed data and after identification of inpatient episodes with different models in 2020.

Diagnosis	Earliest possible age of onset, y	Non-processed data, ^a N	Processed data, change %			
			Model 1 ^b	Model 2 ^c	Model 3 ^d	Model 4 ^e
Any mental disorder		24,322	0.00	-1.46	0.00	-1.46
Organic mental disorders (F0)	5	968	-1.24	-2.69	-1.14	-2.58
Dementias (F00-03)	35	534	-1.12	-2.43	-1.31	-2.62
Other (F04-09)	5	520	-3.08	-4.42	-2.69	-4.04
Substance use disorders (F1)	5	4569	0.26	-1.97	0.22	-1.99
Schizophrenia spectrum (F2)	5	9366	0.14	-0.73	0.09	-0.78
Schizophrenia (F20)	5	3909	0.00	-0.64	0.00	-0.64
Other (F21-29)	5	5858	0.10	-0.90	0.02	-0.99
Mood disorders (F3)	5	10,098	0.50	-1.06	0.40	-1.14
Mania and bipolar disorder (F30-31)	5	2392	0.25	-0.79	0.25	-0.79
Depressive disorders (F32-F33)	5	7490	0.64	-1.01	0.49	-1.13
Other (F34-F39)	5	635	2.20	0.00	1.73	-0.47
Neurotic, stress-related and somatoform disorders (F4)	5	5346	1.05	-0.95	0.95	-1.05
Anxiety disorders (F40-41)	5	3721	1.56	-0.51	1.40	-0.67
Obsessive - compulsive disorder (F42)	5	598	0.17	-1.34	0.00	-1.51
Behavioral syndromes (F5)	5	1088	0.64	-0.74	0.55	-0.83
Personality disorders (F6)	5	2323	0.39	-0.99	0.26	-1.12
Mental retardation (F7)	1	405	-0.25	-2.22	-0.25	-2.22
Disorders of psychological development (F8)	1	1210	0.33	-0.66	0.08	-0.91
Behavioral and emotional disorders (F90-98)	1	2603	0.81	-0.58	0.73	-0.65
ADHD (F90)	1	1262	0.87	-0.63	0.87	-0.63
Other (F91-98)	1	1726	0.52	-0.75	0.41	-0.87

Note: Negative change refers to reduction, positive to increase in the number of individuals. In the processed data, psychiatric outpatient diagnoses at or after the day of discharge from psychiatric ward are included and are the source of diagnoses causing percentages over 100.

^aValid non-processed data after initial quality checks. All registered diagnoses considered.

^bHospitalizations can start the day after a previous one, with no minimum length requirement.

^cValid hospitalizations require at least two consecutive days with one overnight stay. Same-day admission and discharge count as outpatient visits.

^dA new hospitalization can start after a full day outside the hospital, with no minimum duration requirement.

^eValid hospitalizations require at least two consecutive days with one overnight stay. Same-day admission and discharge count as outpatient visits.

mental disorders other than dementias (ICD-10: F04-09) as well as unspecified nonorganic psychosis (F29) and acute and transient psychotic disorders (F23) were noteworthy sources of variation in these sub-chapter categories (Table 5).

4 | DISCUSSION

This study demonstrated that the pre-processing of partly overlapping register entries is a complex process that has important implications for the scientific results and administrative metrics of psychiatric care derived from healthcare registers. The annual

count of inpatient episodes showed a reduction ranging from 6% to 11%, while the median length of stay increased by up to 27% following the identification of inpatient episodes. However, the overall number of individuals with registered mental disorders in psychiatric services or primary care remained relatively stable, with two notable exceptions: the exclusion of preliminary diagnoses proved to be of clear significance when assessing psychotic disorders or organic mental disorders. To enhance the precision of register-based analyses, the management of overlapping register entries should be carried out and reported systematically. We have introduced a reproducible open-source method for this process.

TABLE 4 Annual median length of stay of psychiatric inpatient episodes in non-processed data and after identification of inpatient episodes with different models.

Year	Non-processed data ^a	Processed data			
		Model 1 ^b	Model 2 ^c	Model 3 ^d	Model 4 ^e
2011	12	14	14	14	14
2012	12	14	14	14	14
2013	11	13	14	14	14
2014	11	13	14	13	14
2015	11	13	14	13	14
2016	11	13	13	13	13
2017	11	12	12	12	13
2018	10	11	12	12	12
2019	9	11	11	11	11
2020	8	9	9	9	10

^aValid non-processed data after initial quality checks.

^bHospitalizations can start the day after a previous one, with no minimum length requirement.

^cValid hospitalizations require at least two consecutive days with one overnight stay. Same-day admission and discharge count as outpatient visits.

^dA new hospitalization can start after a full day outside the hospital, with no minimum duration requirement.

^eValid hospitalizations require at least two consecutive days with one overnight stay. Same-day admission and discharge count as outpatient visits.

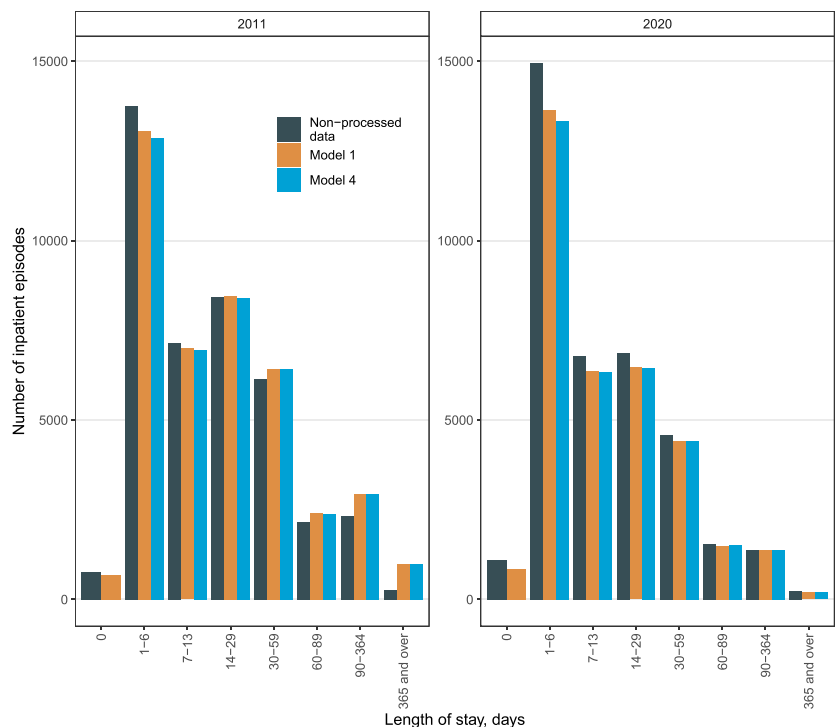


FIGURE 1 Length of stay of psychiatric inpatient episodes in the years 2011 and 2020. Categorization is based on the official reports. Non-processed data refers to valid non-processed data after initial quality checks. In Model 1, hospitalizations can start the day after a previous one, with no minimum length requirement. In Model 4, valid hospitalizations require at least two consecutive days with one overnight stay; same-day admission and discharge count as outpatient visits.

The methods employed in this study identify days in inpatient care that are covered in multiple register entries with high certainty. Therefore, using any of these models improves accuracy of any metrics derived from the registers. On the other hand, selecting the most conservative model instead of the most liberal, yielded only minor further adjustments in the estimates. While these

models are likely to overlook a small proportion of patients with same-day re-hospitalizations, this has been regarded as a lesser concern compared to erroneously categorizing transfers as readmissions (Katschnig et al., 2019). Whether there is a need for a minimum length requirement for a hospitalization is a subjective judgment, as the registers do not provide a definitive answer, given

TABLE 5 Overall number of individuals with registered mental disorders by diagnosis in non-processed data and after identification of inpatient episodes with different models, years 1975–2020 and 1996–2020.

Diagnosis	Earliest possible age of onset, y	Non-processed data, ^a N	Processed data, reduction %			
			Model 1 ^b	Model 2 ^c	Model 3 ^d	Model 4 ^e
Any mental disorder		2,130,468	0.01	0.01	0.01	0.01
Organic mental disorders (F0)	5	197,915	1.20	1.20	1.27	1.27
Dementias (F00-03) ^f	35	127,875	0.61	0.61	0.66	0.66
Other (F04-09) ^f	5	47,271	4.93	4.93	5.14	5.14
Substance use disorders (F1)	5	286,587	0.65	0.65	0.68	0.68
Schizophrenia spectrum (F2)	5	186,718	3.69	3.69	3.81	3.81
Schizophrenia (F20) ^f	5	78,855	1.60	1.60	1.68	1.68
Schizotypal disorder (F21) ^f	5	7079	3.84	3.84	3.97	3.97
Persistent delusional disorders (F22) ^f	5	28,820	5.06	5.06	5.22	5.22
Acute and transient psychotic disorders (F23) ^f	5	24,783	9.78	9.78	10.06	10.06
Induced delusional disorder (F24) ^f	5	223	14.35	14.35	14.35	14.35
Schizoaffective disorders (F25) ^f	5	16,123	3.11	3.11	3.21	3.21
Other nonorganic psychotic disorders (F28) ^f	5	3728	5.95	5.95	6.12	6.12
Unspecified nonorganic psychosis (F29) ^f	5	66,166	16.34	16.34	16.72	16.72
Mood disorders (F3)	5	654,874	0.66	0.66	0.68	0.68
Mania and bipolar disorder (F30–31) ^f	5	69,150	1.82	1.82	1.88	1.88
Depressive disorders (F32–F33) ^f	5	604,304	0.71	0.71	0.73	0.73
Other (F34–F39) ^b	5	78,838	1.59	1.59	1.64	1.64
Neurotic, stress-related and somatoform disorders (F4)	5	760,972	0.43	0.43	0.45	0.45
Anxiety disorders (F40-41) ^f	5	405,956	0.64	0.64	0.66	0.66
Obsessive - compulsive disorder (F42) ^f	5	29,242	0.55	0.55	0.58	0.58
Behavioral syndromes (F5)	5	359,745	0.38	0.38	0.38	0.38
Personality disorders (F6)	5	134,291	0.80	0.80	0.86	0.86
Mental retardation (F7)	1	27,259	0.55	0.55	0.59	0.59
Disorders of psychological development (F8)	1	126,565	0.69	0.69	0.76	0.76
Behavioral and emotional disorders (F90-98)	1	269,365	0.55	0.55	0.56	0.56
ADHD (F90) ^f	1	79,257	0.10	0.10	0.11	0.11
Other (F91-98) ^f	1	114,770	0.40	0.40	0.41	0.41

^aValid non-processed data after initial quality checks.

^bHospitalizations can start the day after a previous one, with no minimum length requirement.

^cValid hospitalizations require at least two consecutive days with one overnight stay. Same-day admission and discharge count as outpatient visits.

^dA new hospitalization can start after a full day outside the hospital, with no minimum duration requirement.

^eValid hospitalizations require at least two consecutive days with one overnight stay. Same-day admission and discharge count as outpatient visits.

^fSpecific diagnoses only for the period of 1996–2020, to avoid inaccuracy in conversion of ICD-8 and ICD-9 diagnoses.

their inability to discern the reasons for episodes lasting less than a day. The relevance of a minimum length criterion depends on the primary focus of the analysis; for instance, if the emphasis is on admissions, a minimum length might not be advantageous, whereas if inpatient episodes are the primary focus, it could be justified.

With the provided scripts, however, these criteria can be easily customized.

Identifying the length of an inpatient stay from registers is a complex task in many European countries (Katschnig et al., 2019). The current scripts serve one solution that can be easily customized

based on the needs at hand. Finnish official statistics report an annual median length of stay in psychiatric inpatient care that closely aligns with our non-processed statistics, with 2017 as the only exception, where there was a one-day discrepancy (THL, 2021). However, after the identification of inpatient episodes, our estimates for the length of stay tend to be one to 3 days higher. The current results highlight that length of stay is a measure that is prone to variation based on underlying assumptions on the definition of psychiatric inpatient episodes.

The role of identifying discharge diagnoses differed across diagnostic categories, with schizophrenia and organic disorders exhibiting the highest prevalence of preliminary diagnoses. This observation aligns with expectations, as the diagnoses of psychotic disorders are known to frequently evolve during follow-up (Bromet et al., 2011; Coulter et al., 2019; Köhler-Forsberg et al., 2023). A landmark study indicated that although the Finnish healthcare registers are effective in screening for possible psychotic disorders, they may not be optimal for complete case ascertainment (Perälä et al., 2007). Therefore, the application of conservative models for the exclusion of preliminary diagnoses may prove advantageous. Conversely, it is worth noting that during the 1990s, a trend toward a more restrictive definition of schizophrenia in clinical practice in Finland was recognized, potentially resulting in a higher likelihood of false negatives and fewer false positives (Isohanni et al., 1997). However, the accuracy of diagnostic practices in primary or secondary care mental health services has not been evaluated recently in Finland. Diagnostics in psychiatry is a complex process with clinical and administrative considerations (First et al., 2018). Accordingly, the unspecified category (F29) emerged as the most prevalent diagnosis among all the disorders within the ICD-10 sub-chapter Schizophrenia, schizotypal and delusional disorders before and after the processing of the data.

This study provides insight for efforts to enhance the precision of register-based metrics, both in Finland and internationally. While the scripts provided in this study are specific to Finnish registers, the underlying principles are applicable in other contexts where temporal overlapping entries may exist in the data. While some countries' register authorities pre-process the register data before delivery to researchers and thus ensure comparability within the country, it's useful to recognize that pre-processing includes specific decisions that can have an observable impact on the data and should be understood by researchers. Overall, the issue of partly overlapping register entries and preliminary diagnoses remains relatively underexplored in register-based research. Some researchers have addressed this by excluding diagnoses from emergency departments and by eliminating duplicate entries (Kerkelä et al., 2021; Vernal et al., 2018). Nevertheless, the provision of open-source methodologies is instrumental for achieving reproducible and unequivocal analyses and plays a role in promoting open science (Crüwell et al., 2019; Open Science Coordination in Finland, 2023).

Finally, understanding the complexities of the Finnish healthcare system and its influences to the collection of register data is beneficial for interpretation of the current results. Finland's tax-funded healthcare system offers universal access, which includes mental healthcare services. Since 2023, health care services are increasingly centralized under the administration of 22 Well-being Service Counties, a reform aimed at integrating primary and secondary levels of care, among other things (Tynkkynen et al., 2023). This may impact the recognition of specific service modalities within the registers.

Psychiatric inpatient care in Finland is publicly organized and regulated by specific legislation and there are no private psychiatric hospitals. However, general practice inpatient units in health centers may address mental health issues under the supervision of GPs, without provisions for compulsory care. These primary care inpatient treatments are recorded alongside secondary care inpatient care, leading to potential overlaps in the registers.

Outpatient psychiatric services are primarily provided by psychiatric outpatient departments of county psychiatric units and may be located at the hospitals or in separate units. Additionally, primary care health centers offer outpatient mental healthcare, supplemented by psychiatric nurse, psychologist, and consultant psychiatrist services. It is possible that these services may be recorded in both secondary care and primary care registers. Notably, private, self or employer-paid mental health outpatient care, alongside publicly funded rehabilitation services and higher education students' health services provided by the Finnish Student Health Services, are integral components of the Finnish healthcare system, increasingly covered in national healthcare registers. Wide coverage of services is a clear strength of the Finnish healthcare registers compared to some other Nordic countries (Weye et al., 2023). While the focus of this study does not lie in analyzing administrative structures, the provided scripts can handle all types of services included in the Finnish registers. As registers grow more complex and comprehensive, open data pre-processing becomes increasingly essential for further development.

4.1 | Limitations

First, this study solely relied on register data and lacked clinical or other reference data for performance comparison among the models. While all models successfully identify overlapping days, we lacked the means to determine the superiority of any model in recognizing hospital transfers resulting from early readmissions. Second, individual patients may have intricate inpatient care pathways involving multiple medical specialties. The general procedures outlined in this study may not offer sufficient detail if the focus is on diagnosing the length of stay within specific specialties in such cases. Instead, further development of methods is warranted. Third, present results are based on Finnish psychiatric register data and further studies are needed to replicate the present findings in other settings.

4.2 | Conclusions

Register-based research benefits from constant methodological development. Technical intricacies, including those related to partly overlapping register entries, may prove complex and labor-intensive. While the issues tied to partly overlapping register entries may be central to certain research questions, the actual scientific interest may often lay elsewhere. The application of open science principles and collaborative development of readily available solutions can significantly save time and enhance research quality by improving comparability and reproducibility of results.

AUTHOR CONTRIBUTIONS

Kimmo Suokas: Conceptualization; writing - original draft; methodology; software; formal analysis. **Mai Gutvilig:** Software; writing - review & editing; methodology. **Sonja Lumme:** Supervision; writing - review & editing. **Sami Pirkola:** Supervision; writing - review & editing. **Christian Hakulinen:** Supervision; writing - review & editing.

ACKNOWLEDGMENTS

Kimmo Suokas was supported by the Jalmari and Rauha Ahokas Foundation and the Finnish Psychiatric Association. Christian Hakulinen was supported by the Academy of Finland (354237) and the European Union (ERC, MENTALNET, 101040247). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

CONFLICT OF INTEREST STATEMENT

We declare no competing interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the National Institute of Health and Welfare (www.thl.fi) and Statistics Finland (www.stat.fi). Restrictions apply to the availability of these data, which were used under license for this study. Inquiries about secure access to data should be directed to data permit authority Findata (findata.fi/en). The method described in this article has been made publicly available and contain supplementary description of each step of the process (https://github.com/kmmsks/hilmo_identify_episodes/).

ETHICS STATEMENT

The Research Ethics Committee of the Finnish Institute for Health and Welfare approved the study protocol (decision #10/2016\$751).

PATIENT CONSENT STATEMENT

Informed consent is not required for register-based studies in Finland.

ORCID

Kimmo Suokas  <https://orcid.org/0000-0001-6296-6343>

Sonja Lumme  <https://orcid.org/0000-0001-9979-0038>

REFERENCES

- Bromet, E. J., Kotov, R., Fochtmann, L. J., Carlson, G. A., Tanenberg-Karant, M., Ruggero, C., & Chang, S. w. (2011). Diagnostic shifts during the decade following first admission for psychosis. *American Journal of Psychiatry*, *168*(11), 1186–1194. <https://doi.org/10.1176/appi.ajp.2011.11010048>
- Cheng, T. M. (2015). Reflections on the 20th anniversary of taiwan's single-payer national health insurance system. *Health Affairs*, *34*(3), 502–510. <https://doi.org/10.1377/hlthaff.2014.1332>
- Coulter, C., Baker, K. K., & Margolis, R. L. (2019). Specialized consultation for suspected recent-onset schizophrenia: Diagnostic clarity and the distorting impact of anxiety and reported auditory hallucinations. *Journal of Psychiatric Practice*, *25*(2), 76–81. <https://doi.org/10.1097/PRA.0000000000000363>
- Crüwell, S., Doorn, J. van, Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., Orben, A., Parsons, S., & Schulte-Mecklenbeck, M. (2019). Seven easy steps to open science. *Zeitschrift für Psychologie*, *227*(4), 237–248. <https://doi.org/10.1027/2151-2604/a000387>
- First, M. B., Rebello, T. J., Keeley, J. W., Bhargava, R., Dai, Y., Kulygina, M., Matsumoto, C., Robles, R., Stona, A., & Reed, G. M. (2018). Do mental health professionals use diagnostic classifications the way we think they do? A global survey. *World Psychiatry*, *17*(2), 187–195. <https://doi.org/10.1002/wps.20525>
- Isohanni, M., Mäkiyö, T., Moring, J., Räsänen, P., Hakko, H., Partanen, U., Koironen, M., & Jones, P. (1997). A comparison of clinical and research DSM-III-r diagnoses of schizophrenia in a Finnish national birth cohort. *Social Psychiatry and Psychiatric Epidemiology*, *32*(5), 303–308. <https://doi.org/10.1007/BF00789044>
- Kajantie, M., Manderbacka, K., McCallum, A., Notkola, I., Arffman, M., Forssas, E., Karvonen, S., Kortteinen, M., Leyland, A., & Keskimäki, I. (2006). How to carry out register-based health services research in Finland?: Compiling complex study data in the REDD project. <https://www.julkari.fi/handle/10024/77679>
- Katschnig, H., & Straßmayr, C. (2017). Comparative effectiveness research on psychiatric hospitalisation by record linkage of large administrative data sets (CEPHOS-LINK). Final scientific report for objectives 1, 2 and 3.
- Katschnig, H., Straßmayr, C., Endel, F., Berger, M., Zauner, G., Kalseth, J., Sfetcu, R., Wahlbeck, K., Tedeschi, F., & Šprah, L. (2019). Using national electronic health care registries for comparing the risk of psychiatric re-hospitalisation in six European countries: Opportunities and limitations. *Health Policy*, *123*(11), 1028–1035. <https://doi.org/10.1016/j.healthpol.2019.07.006>
- Kerkelä, M., Gyllenberg, D., Gissler, M., Sillanmäki, L., Keski-Säntti, M., Hinkka-Yli-Salomäki, S., Filatova, S., Hurtig, T., Miettunen, J., Sourander, A., & Veijola, J. (2021). Cumulative incidences of hospital-treated psychiatric disorders are increasing in five Finnish birth cohorts. *Acta Psychiatrica Scandinavica*, *143*(2), 119–129. <https://doi.org/10.1111/ACPS.13247>
- Köhler-Forsberg, O., Antonsen, S., Pedersen, C. B., Mortensen, P. B., McGrath, J. J., & Mors, O. (2023). Schizophrenia spectrum disorders in Denmark between 2000 and 2018: Incidence and early diagnostic transition. *Acta Psychiatrica Scandinavica*, *148*(2), 90–198. <https://doi.org/10.1111/acps.13565>
- Laugesen, K., Ludvigsson, J. F., Schmidt, M., Gissler, M., Valdimarsdottir, U. A., Lunde, A., & Sørensen, H. T. (2021). Nordic health registry-based research: A review of health care systems and key registries. *Clinical Epidemiology*, *13*, 533–554. <https://doi.org/10.2147/CLEP.S314959>
- Lichtenberg, P., Kaplan, Z., Grinshpoon, A., Feldman, D., & Nahon, D. (1999). The goals and limitations of Israel's psychiatric case register. *Psychiatric Services*, *50*(8), 1043–1048. <https://doi.org/10.1176/ps.50.8.1043>
- Ludvigsson, J. F., Andersson, E., Ekblom, A., Feychting, M., Kim, J. L., Reuterwall, C., Heurgren, M., & Olausson, P. O. (2011). External review

- and validation of the Swedish national inpatient register. *BMC Public Health*, 11(1), 450. <https://doi.org/10.1186/1471-2458-11-450>
- Maret-Ouda, J., Tao, W., Wahlin, K., & Lagergren, J. (2017). Nordic registry-based cohort studies: Possibilities and pitfalls when combining Nordic registry data. *Scandinavian Journal of Public Health*, 45(17_suppl), 14–19. <https://doi.org/10.1177/1403494817702336>
- Open Science Coordination in Finland. (2023). Federation of Finnish Learned Societies. Open research data and methods. National policy and executive plan by the higher education and research community for 2021–2025. <https://edition.fi/tsv/catalog/book/669>
- Pedersen, C., Mors, O., Bertelsen, A., Waltoft, B. L., Agerbo, E., McGrath, J. J., Mortensen, P. B., & Eaton, W. W. (2014). A comprehensive nationwide study of the incidence rate and lifetime risk for treated mental disorders. *JAMA Psychiatry*, 71(5), 573–581. <https://doi.org/10.1001/jamapsychiatry.2014.16>
- Perälä, J., Suvisaari, J., Saarni, S. I., Kuoppasalmi, K., Isometsä, E., Pirkola, S., Partonen, T., Tuulio-Henriksson, A., Hintikka, J., Kieseppä, T., Härkänen, T., Koskinen, S., & Lönnqvist, J. (2007). Lifetime prevalence of psychotic and bipolar I disorders in a general population. *Archives of General Psychiatry*, 64(1), 19–28. <https://doi.org/10.1001/archpsyc.64.1.19>
- Pirkola, S., & Sohlman, B. (2005). Atlas of mental health: Statistics from Finland. <https://www.julkari.fi/handle/10024/110627>
- Sund, R. (2012). Quality of the Finnish hospital discharge register: A systematic review. *Scandinavian Journal of Public Health*, 40(6), 505–515. <https://doi.org/10.1177/1403494812456637>
- Suokas, K., Hakulinen, C., Sund, R., Kampman, O., & Pirkola, S. (2022). Mortality in persons with recent primary or secondary care contacts for mental disorders in Finland. *World Psychiatry*, 21(3), 470–471. <https://doi.org/10.1002/WPS.21027>
- Suokas, K., Kurkela, O., Nevalainen, J., Suvisaari, J., Hakulinen, C., Kampman, O., & Pirkola, S. (2023). Geographical variation in treated psychotic and other mental disorders in Finland by region and urbanicity. *Social Psychiatry and Psychiatric Epidemiology*, 59(1), 37–49. Published online June 13. <https://doi.org/10.1007/s00127-023-02516-x>
- THL. (2021). Psykiatrin erikoissairaanhoito 2020: Psykiatristen hoitajaksojen kesto lyheni yhä. Retrieved December 14, 2021. <https://www.julkari.fi/handle/10024/143510>
- THL. (2022). Mental health visits in primary health care per 1000 inhabitants. <https://sotkanet.fi/sotkanet/en/taulukko/?indicator=s3YutNZNNQAA®ion=s07MBAA=&year=sy5zsjbR0z>
- UEAA==&gender=t&abs=f&color=f&buildVersion=3.0-SNAPSHOT&buildTimestamp=202109301228
- Tynkkynen, L. K., Keskimäki, I., Karanikolos, M., & Litvinova, Y. (2023). Finland: Health system summary. Retrieved December, 2023. <https://apps.who.int/iris/handle/10665/366710>
- Vernal, D. L., Stenström, A. D., Staal, N., Christensen, A. M. R., Ebbesen, C., Pagsberg, A. K., Correll, C. U., Nielsen, R. E., & Lauritsen, M. B. (2018). Validation study of the early onset schizophrenia diagnosis in the Danish Psychiatric Central Research Register. *European Child & Adolescent Psychiatry*, 27(8), 965–975. <https://doi.org/10.1007/s00787-017-1102-z>
- Weye, N., McGrath, J. J., Lasgaard, M., Momen, N. C., Knudsen, A. K., Musliner, K., & Plana-Ripoll, O. (2023). Agreement between survey- and register-based measures of depression in Denmark. *Acta Psychiatrica Scandinavica*, 147(6), 581–592. <https://doi.org/10.1111/acps.13555>
- WONCA International Classification Committee (WICC). (2005). *ICPC-2-r: International classification of primary care*. Oxford University Press Inc.
- World Health Organization. (1994). Division of mental health. In *The ICD-10 classification of mental and behavioural disorders: Conversion tables between ICD-8, ICD-9 and ICD-10*. <https://apps.who.int/iris/handle/10665/62577>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Suokas, K., Gutvilig, M., Lumme, S., Pirkola, S., & Hakulinen, C. (2024). Enhancing the accuracy of register-based metrics: Comparing methods for handling overlapping psychiatric register entries in Finnish healthcare registers. *International Journal of Methods in Psychiatric Research*, e2029. <https://doi.org/10.1002/mpr.2029>