MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231084

Challenges in Interpreting Norwegian Child and Adolescent Mental Health Records

Kaban KOOCHAKPOUR[a,1], Frida SOFIE SOLHEIM[a,g], Øystein NYTRØ[a], Carolyn CLAUSEN[b], Thomas FRODL[c], Roman KOPOSOV[e], Bennett LEVENTHAL[f], Dipendra PANT[a], Thomas BROX RØST[h], Line STIEN[b], Odd SVERRE WESTBYE[b,d] and Norbert SKOKAUSKAS[b]

[a]*Dept. Computer Science, The Norwegian University of Science and Technology (NTNU), Trondheim, Norway*
[b]*Reg. Centre for Child and Youth Mental Health and Child Welfare, NTNU, Trondheim, Norway*
[c]*Dept. of Psychiatry, Psychotherapy and -somatic, Univ. Hosp, RWTH Aachen, Germany*
[d]*Department of Child and Adolescent Psychiatry, St. Olav's Univ. Hosp, Trondheim, Norway*
[e]*Reg. Centre for Child and Youth Mental Health and Child Welfare, Arctic Univ. of Norway*
[f]*Dept. of Psychiatry, Division of Child and Adolescent Psychiatry, UCSF, USA*
[g]*Bekk Consulting AS, Oslo, Norway*
[h]*Vivit AS, Trondheim, Norway*
ORCiD ID: Kaban Koochakpour https://orcid.org/0000-0001-7214-8651

Abstract. The Electronic Health Record system BUPdata served Norwegian Child and Adolescent Mental Health Services (CAMHS) for over 35 years and is still an important source of information for understanding clinical practice. Secondary usage of clinical data enables learning and service quality improvement. We present some insights from explorative data analysis for interpreting the records of patients referred for hyperkinetic disorders. The major challenges were data preparation, pre-analysis, imputation, and validation. We summarize the main characteristics, spot anomalies, and detect errors. The results include observations about the patient referral diversity based on 12 different variables. We modeled the activities in an individual episode of care, described our clinical observations among data, and discussed the challenges of data analysis.

Keywords. Data analysis challenges, electronic health record interpretation, episode of care, hyperkinetic disorders

1. Introduction

Mental health problems and disorders among children and adolescents can negatively affect individuals, families, and communities. This motivates researchers and service providers to study and understand clinical processes and patient outcomes and seek solutions that can potentially improve mental health care services [1]. These solutions depend highly on complete, rich, and consistent clinical data.

Our goal is to demonstrate the potential in analyzing clinical data and review and mitigate data quality challenges. We describe data from an Electronic Health Record (EHR) system, BUPdata, that initially served all Norwegian CAMH clinics, and some for over 35 years [2]. Our cohort from one university hospital CAMHS has 22,643 distinct patients. The earliest referral was made in 1982 and the last in 2018. The database

---

[1] Corresponding Author: Kaban Koochakpour, email: kaban.koochakpour@ntnu.no.

has 30,938 unique referrals, 41,411 registered episodes of care, and 1,840,045 distinct database records (multiple per individual contact). About 1/3 of patients had more than one referral for services, and 32,051 diagnoses were given. There are slightly more males (51.9%) than females (47.8%). The age distribution ranged from 0-18 (patients above 18 are not eligible for CAMHS) at the time of referral with referral rates increasing, with age to a peak at ages 14-16.

The available information about an episode of care covers the time from when the referral was received at CAMHS until the case was closed. From 2008, Norwegian CAMHS has used the WHO multi-axial coding system. This study focused on ICD-10 Classification of Mental and Behavioral Disorders [3], diagnosis of hyperkinetic disorders, the F90-group, coded on axis 1.

1.1. Referral Process

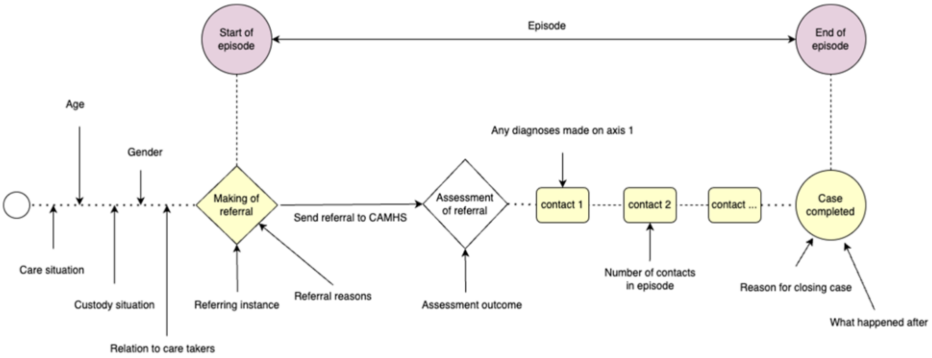

Figure 1. Episode of care in CAMHS.

The patient trajectories followed the pattern outlined in Figure 1, demonstrating the activities in an individual episode of care. The information gathered at the time of referral includes the date, referring professional(s), up to three reasons for the referral (e.g., symptoms), up to three reasons for the referral regarding the child's environment, and patient information. An episode is the period from when the referral is received and accepted by CAMHS for assessment, treatment, and rehabilitation until the case is closed, and follow-up is completed. Only the first episode of care for a specific disorder was considered for our analysis, even if the patient may have more than one episode of care related to a specific diagnosis or disorder.

2. Methods

To investigate the data, summarize the main characteristics, spotting anomalies and detecting errors, an explorative data pre-analysis was carried out. The aim was to demonstrate the potential of clinical data by presenting insights on children and adolescents with hyperkinetic problems, as well as addressing some of the data challenges, by running pre-analysis statistics and using data visualization methods.

All records contained structured and coded data. Data for the analysis was restricted to including valid codes. Variables with many missing values were excluded to maintain a sufficiently informative dataset for analysis. Data only includes patients referred based on suspected of hyperkinetic disorder (ADHD), suspected of defiance/conduct disorder, and/or previously received a diagnosis in the F90-group on axis 1. Pre-analysis was then used to probe the health data of the remaining 4,201 patients based on 12 variables in their first episode. For the given cohort, the selected variable fields (number of unique values) include: age [Min (1), Mean (9.5), Max (18)], gender (2), care situation (9), custody (4), relation to care takers (10), referring professionals (24), referral reasons (21), assessment done by referring party (3), number of contacts in the first episode (Min (1), Mean (1.36), Max (24) days), Axis 1 diagnosis assessment (56), reason for case closing (9), and status after episode (5).

Regarding data pre-processing, exclusion and imputation are used as mitigation strategies. Variables with out-of-range values were excluded to ensure data consistency and quality. In data cleaning, missing or erroneous entries were handled by replacing the value with a potential or mean value. Changing coding guidelines may yield intra-patient EHR inconsistencies. To mitigate this, old codes translated to new ones or abstracted to a restricted, more general set of categories. This strategy was enforced to replace referral reasons with current referral reasons in 2009-2010.

3. Results

3.1. Initial insights into the Cohort

Pre-analysis provided initial insights on the referral and care of the patients in the selected dataset. Most patients lived with their biological mother and father. Boys had a peak age of referral at 8-9 years, while girls had 10 years. On average the patients were referred 5-7 years earlier than those with other disorders, and boys outnumbered girls 2.4:1.

In the dataset, the most common primary referral reasons are suspicion of hyperkinetic disorder (ADHD), suspicion of defiance/conduct disorder, suspicion of depression, suspicion of anxiety, and other (unspecified referral reason). Suspicion of autism and Tourette's syndrome were both more frequently recorded for boys. Suspicion of eating disorders and suicide risk were more frequently recorded for girls. Learning difficulties had a later onset for girls. As many girls as boys are referred for suspicion of depression [4].

Diagnosis code F90.0 was most frequently recorded, followed by none given, Z032, 1999, and F90.1. Axis 1 codes may have been assigned in other parts of specialist care, e.g., a psychiatrist working in a general pediatric department. Girls had their first episode on average two years later than boys. Many girls were assigned F90.0 and Z032 diagnoses, but the F90-group was more frequently recorded for boys (F90.1, F90.8, and F90.9). For F90.0 and F90.8, boys outnumber girls 2:1, while F90.1 and F90.9 are closer to 4:1 [4].

Most of the referrals came from General Physician (GP), the Educational Psychological Service, and the Child Welfare Services. GPs had the most rejected referrals, even though GPs made more referrals in total. Rejected referrals made up 10.8% of the entire referral cohort (68.4% boys; 31.6% girls), with a median age of 9, 56.4% of the rejected patients are referred by a GP, and 17.0% by the Help Service for Children and Young People. The most common primary reason for rejected referrals were suspicion of defiance/conduct disorder [4].

3.2. Challenges

During data preparation and pre-analysis, several cohort limitations were identified: Firstly, many records were found to have missing values and/or invalid codes that did not represent an existing value. Record keeping in CAMHS depends on finite code lists

[5]. For instance, 65% of the patients did not have any registration for the variable which represents the reason for referral related to child's environment. Keeping fields like these in the dataset will reduce the quality of the data rather than enhancing it. Therefore, they were excluded. Additional variables were also excluded for this reason. Another reason for not including some variables is missing maps for the most frequently used codes. For example, the variable legal basis for cases, most frequently coded as 11, had no mapping (i.e., the code map list only spans from 1-8) and thus was excluded. Furthermore, duplicates of what should have been unique codes required exclusion of variables as proper mapping could not be determined, like the code representing the CAMHS professional, which has tens of associated professional titles in the associated table. This is due to lack of unique constraints in the EHR system. Furthermore, duplicates of what should have been unique codes required these variables to be excluded as their proper mapping could not be decided, e.g., the code representing the CAMHS professional, which has tens of associated professional titles in the associated table. This is because of changed coding systems and could have been resolved by further investigation. Moreover, several code list replacements during the data coverage time (e.g., referral reasons) introduced data consistency issues. The Norwegian CAMHS referral reasons were replaced in 2009-2010, ten years before the discontinuation of the BUPdata. This issue also applies to guidelines of diagnostic codes, like the use of diagnostic code Z032 [6]. Inconsistencies like these bring noise into the data because the change of coding guidelines may not be accounted for in the later analysis.

4. Discussion

Most of the challenges encountered are due to relatively poor data quality, like mistyping and missing entries. Many data items from referring, assessing, and treating patients were left optional, thus lacking proper validation during registration. This introduces potential errors at later stages. Therefore, registrations in an electronic record system should enforce entry validation and the use of local codes should be limited. In addition, improved coding quality and reduction of local code variants is also important for secondary use of data.

Replacing missing values [7] requires a contextual and statistical understanding of the domain. Data imputations can prevent information loss from record or patient exclusion. However, this is a trade-off between feature depth and quality versus cohort size. Better documentation of features such as waiting times, family history, referral reasons, and professionals involved in the patient trajectories, as well as the specific diagnosis would be helpful for analytical purposes [8]. This also emphasizes the need for field validation during entry, and fewer fields being optional, or introducing explicit codes for non-observations or default codes. Features from the care process, such as waiting times, family history, referral reasons, and caretaker roles, must be validated both with guidelines and local practice since they may change over time. Externally validated patient trajectories can be helpful for analysis and service improvement [8]. Data consistency can be problematic when the aim is to analyze the same metrics for long patient history. It can be challenging to translate old codes into new ones or simplify them to general categories to eliminate inconsistencies. Some referral reasons could not be mapped directly as they did not have similar counterparts. The mapping or feature simplification process should be tailored to the analysis and research objectives' needs. Using exact mapping or generalization strategies across different research studies is not advisable. When coding practice and guidelines are introduced or changed for a patient and cohort, it is crucial to observe these shifts, have insight into clinical guidelines used at various times, and validate them with documentation to extract the clinical practice from the structured record data. The importance of collaboration with experienced clinicians to understand the changing practice behind the coded data is not ignorable.

5. Conclusions

We could analyze referrals for hyperkinetic disorders after data preparation that uncovered surprising findings such as most rejected referrals which belongs to the GPs, invalid use of Z032 on axis 1, which is a procedure code, where none or a disorder code was expected, and the marked difference in referrals/diagnosis of girls and boys with hyperkinetic disorders. And finally, we came across this fact that secondary use of even well-structured clinical requires high data quality, and known coding guidelines, but also more posterior ability to detect and correct errors and missing content.

The paper elaborates on some suggestions for enhancing data quality and relevant techniques but still can be further extended. Also, it can be interesting and useful to evaluate the effects of low-qualified registrations in health systems, based on results of different patients' care.

Acknowledgements

The IDDEAS project is funded by The Norwegian Research Council (grant no. 269117) and the Norwegian University of Science and Technology (NTNU). The project has received clearance from the regional ethical oversight committee and approval from the health service for this research on an anonymized extract of the health record.

References

[1] Røst TB, Clausen C, Nytrø Ø, Koposov R, Leventhal B, Westbye OS, Bakken V, Flygel LH, Koochakpour K, Skokauskas N. Local, early, and precise: designing a clinical decision support system for child and adolescent mental health services. Front Psychiatry. 2020 Dec;11:564205, doi: 10.3389/fpsyt.2020.564205.

[2] Koochakpour K, Nytrø Ø, Westbye OS, Leventhal B, Koposov RA, Bakken V, Clausen C, Røst TB, Skokauskas N. Success factors of an early EHR system for child and adolescent mental health: Lessons learned for future practice data-driven decision aids. Stud Health Technol Inform. 2022;290:182-186, doi:10.3233/SHTI220057.

[3] World Health Organization. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. World Health Organization; 1992 Jan. https://hdl.handle.net/11250/3024707.

[4] Solheim FS. Characterising Patients Referred on Suspicion of ADHD and Behavioral Difficulties: An Exploratory Cluster Analysis of Norwegian Electronic Health Records (master's thesis, NTNU). 2022.

[5] Norwegian Association for Child and Youth Psychiat, Kodebok for BUP, 1999.

[6] St. Olavs Hospital, "KORTVERSJON AV ICD-10 FOR BRUK VED BARNE- OG UNGDOMSPSYKIATRISK KLINIKK, ST. OLAVS HOSPITAL HF," docplayer.me, Trondheim, 2016.

[7] Glas CAW, Missing Data. In: Peterson PL, Baker E, editor. International Encyclopedia of Education (Third Edition), Oxford, Elsevier; 2010; p. 283-8, doi: 10.1016/B978-0-08-044894-7.01346-4.

[8] Surén P, Thorstensen AG, Tørstad M, Emhjellen PE, Furu K, Biele G, Aase H, Stoltenberg C, Zeiner P, Bakken IJ, Reichborn-Kjennerud T. Diagnosis of hyperkinetic disorder among children in Norway. Tidsskr Nor Legeforen. 2018 Nov;138(20), doi: 10.4045/tidsskr.18.0418.