MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

NTNU

Norwegian University of Science and Technology

Challenges in interpreting Norwegian child and adolescent mental health records

Kaban Koochakpour
PhD candidate
Norwegian University of Science and Technology

Frida Solheim
Graduate
Norwegian University of Science and Technology

AIDH IMIA

Aim

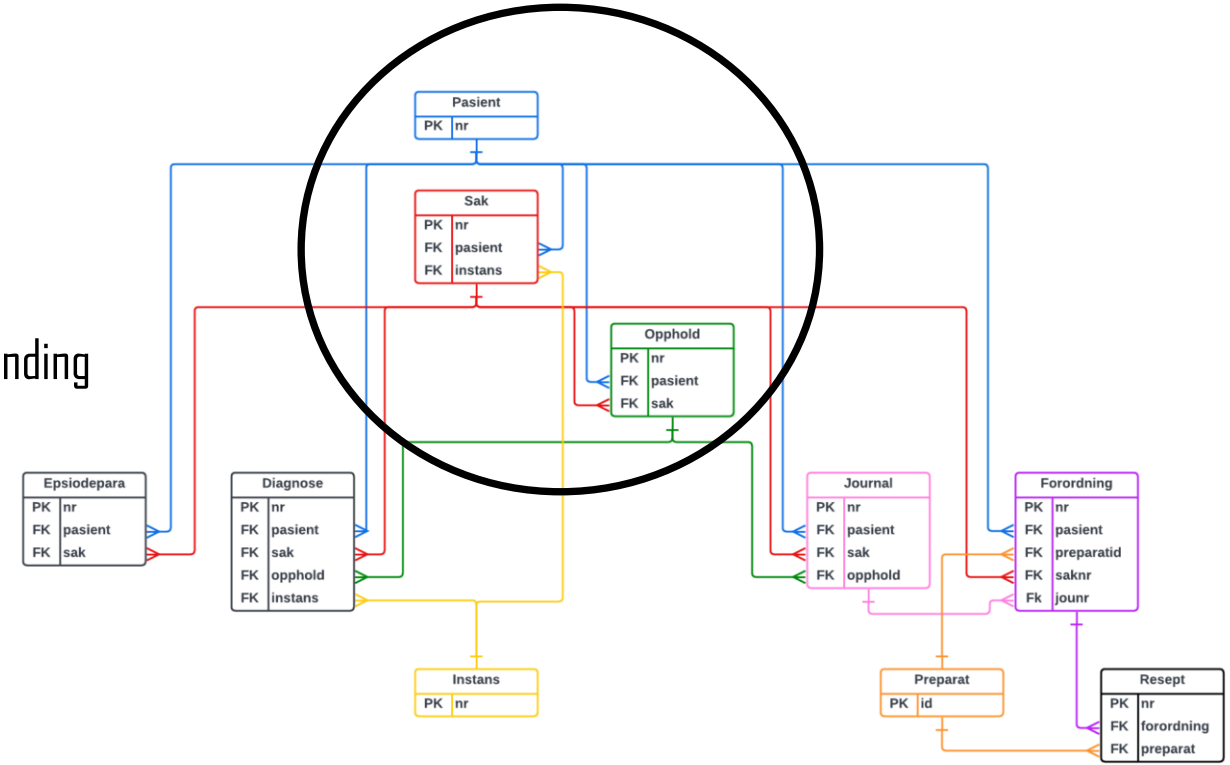
- Insights from explorative data analysis
- Data quality
- Challenges in data analysis and interpreting EHR
- The way we faced the challenges

Motivation

- Mental health disorders and negative effects
- Improve the mental health care process and services
- Understand clinical processes
- Complete and consistent clinical data analysis

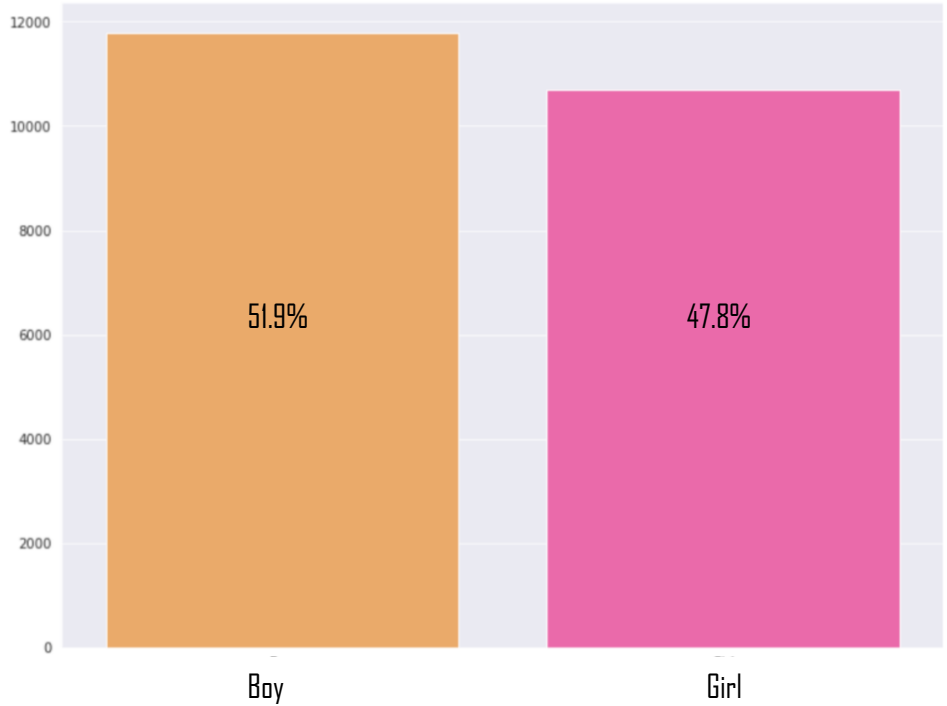
Cohort

- BUPdata (35+ years)
- CAMHS
- Important source for understanding clinical practice



Cohort

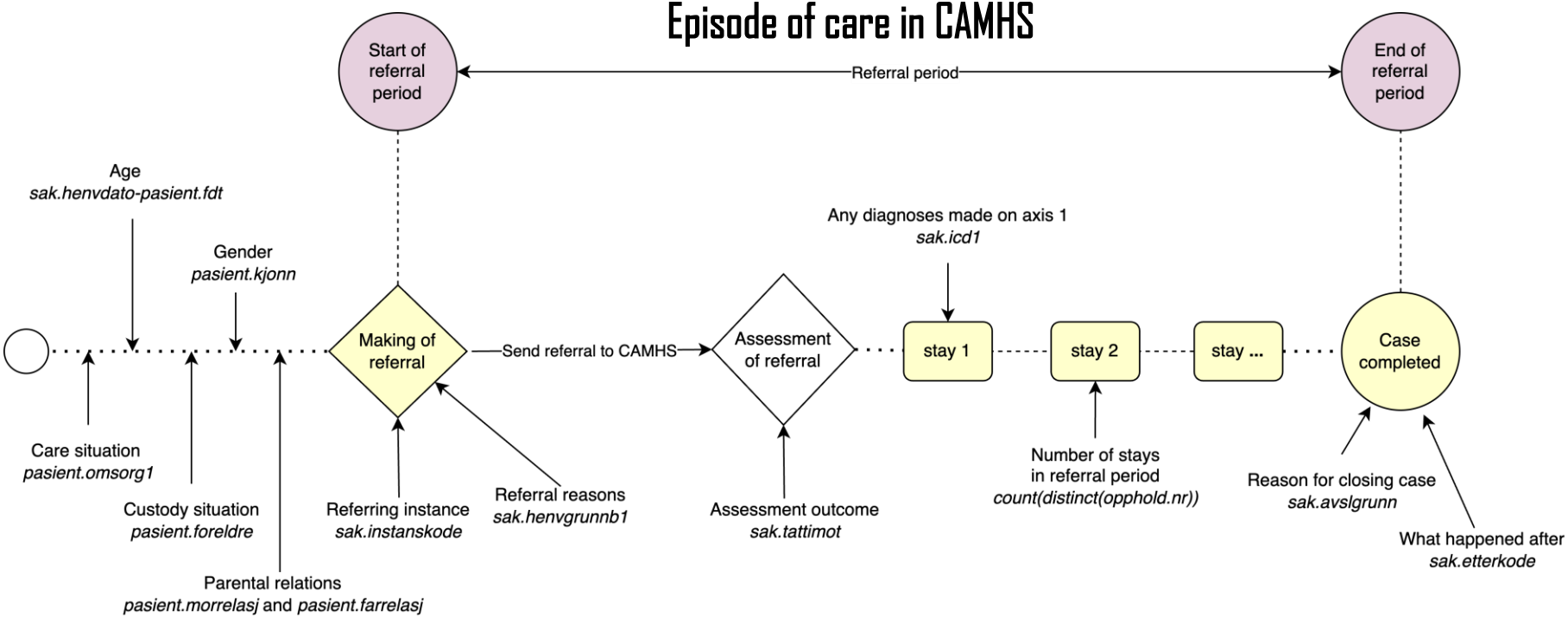
- 22,643 distinct patients
- 1982-2018



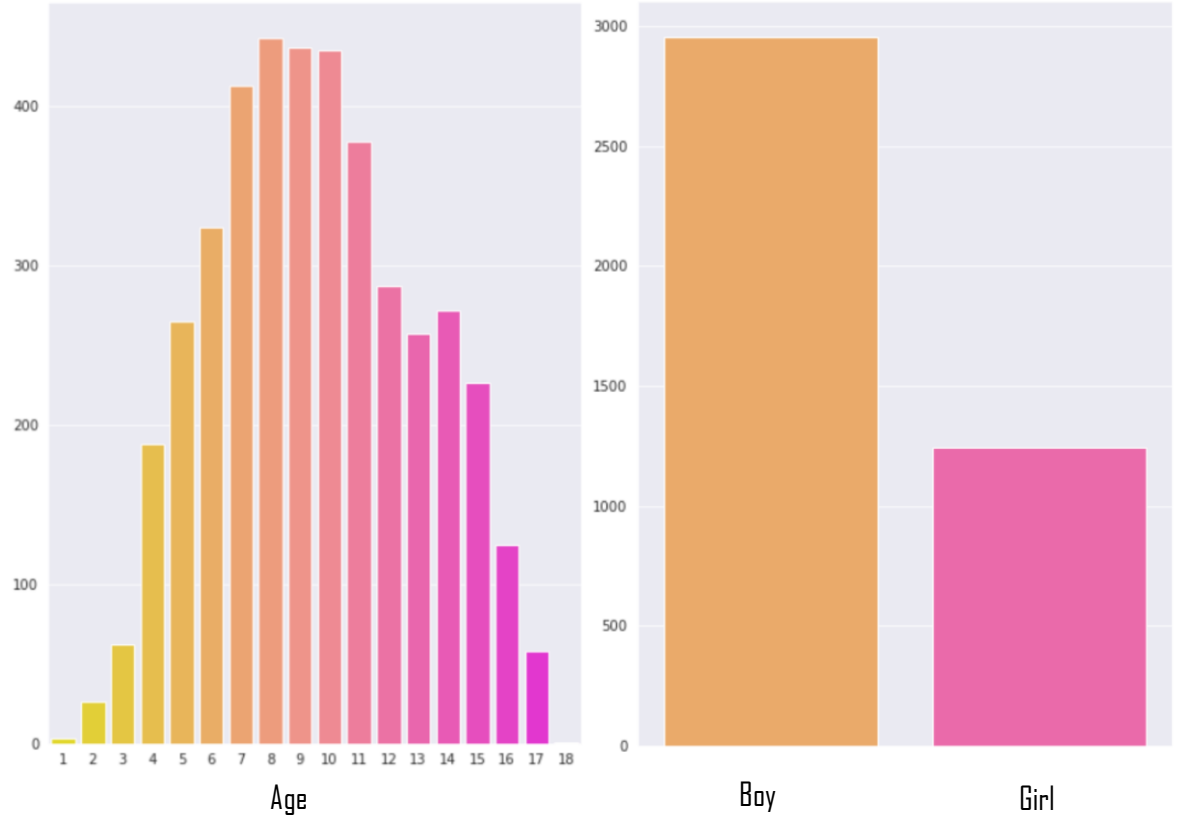
Data Set of study

- Multi-axial: F90-group in axis 1 (referred or diagnosed)
- ICD-10 Classification of Mental and Behavioural disorders
- First referral period to CAMHS
- 4,201 patients, 12 variables

Episode of care in CAMHS



Data characteristics



Explorative Data Analysis

- Demonstrate potential of clinical data

- Data characteristics, anomalies and errors
- Data visualization
- Descriptive statistics
- Clustering

Data pre-processing

Exclusion and imputation

- Invalid codes or out-of-range values → imputed or excluded the record (row)
- Changing coding system → Mapping strategy
- Missing values → Missing Value Prediction, Most Frequent Value, Mean value,…
- Variables with many missing values → excluded the variable (column)

Findings

- Patients were referred 5-7 years earlier than those with other disorders
- Girls had their first referral on average two years later than boys
- Most frequent diagnosis: ADHD (F900), none given, Z032, 1999, conduct disorder (F901)
- 56.4% of rejected referrals were from GPs

Challenges

- Null values
- Missing values
- Invalid/out-of-range values
- Missing maps
- Duplicates of unique codes
- Code list replacements

Examples:

(referral reason related to child's environment)

(referral date)

(invalid not existing diagnostic code)

(code value 11, the legal basis variable)

(tens of associated professional titles)

(referral reasons replacement in 2010)

Discussion

- Poor data quality
- Optional registration fields
- Undocumented local codes

- Process and context understanding
- Validation with local practice & guidelines
- Trade-off between feature depth and cohort size

Learnings

As system designer:

- Improve registration (validation techniques and control optional fields)

As system user:

- Make documentation, limit local codes
- Keep data consistency, mapping guidelines

As data analyst:

- Avoiding data loss, data imputation
- Keep data efficiency