MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Centre for Health Analytics

@kidsanalytics

Using EMR Data for Clinical Research: Opportunities and Challenges

Centre for Health Analytics

The Royal Children's Hospital, Melbourne

AIDH IMIA

What is the Centre for Health Analytics?

Melbourne Children's
A world leader in child and adolescent health

The Royal Children's Hospital Melbourne
murdoch children's research institute
THE UNIVERSITY OF MELBOURNE

Supported by The Royal Children's Hospital Foundation

- Established in 2020 at the Melbourne Children's Campus (The Royal Children's Hospital, Murdoch Children's Research Institute and The University of Melbourne)
- Our mission is to make the Melbourne Children's Campus an internationally leading paediatric centre in the use of data to improve all aspects of patient care, operations, education and research
- We achieve this by putting data into the hands of clinicians, researchers and other decision-makers.

What are our roles?

Alice Voskoboynik

Peter Summers

James Scandol

Ahuva Segal

Objectives

By the end of the workshop, you will be able to:

- Explain the concept and importance of secondary use of EMR data for clinical research;
- Appreciate the opportunities, challenges, and considerations associated with EMR data;
- Describe the impact of data types and clinical workflows on data quality (confounding, biases and missing data)

Before we begin…

- Place your hands up if you are:
  - A clinician?
  - A clinician researcher?
  - A data scientist /clinical data analyst?
  - An epidemiologist?
  - Anyone else..?
- Does your workplace have an EMR (Electronic Medical Record)?
- Is the information discretely stored?
- Do you have experience with data extracted from the EMR?

MEDINFO23
8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Centre for Health Analytics

@kidsanalytics

Secondary Use of EMR Data

Peter Summers

Data Scientist

What is secondary use of EMR data?

- **Health information** – defined in your jurisdiction's *Health Records Act*
- **Primary purpose of health information** is the purpose for which it was collected – in the EMR, to facilitate the delivery of clinical care to individual patients.
- **Secondary purposes of health information** are purposes other than the primary:
    - to measure and improve healthcare quality
    - for clinical and translational research
    - for public health surveillance
- **The primary use is what the EMR was designed for; the secondary uses are not.**

Moral considerations for secondary use

This is not data on people's shopping habits;
it is data associated with human beings at some of the most difficult and vulnerable times of their lives.

Should we be using patient data for this purpose?

We need to very clear on what the purpose is.

Using EMR data for secondary use

- All use of EMR data for research must be reviewed and approved by a Human Research Ethics Committee (HREC). Approval must be obtained prior to commencing the research.
- Expedited approval pathways exist for quality assurance and audits, but the ethical principles remain.
- Staff accessing data for operational purposes must gain approval for Head of Department or delegate.

Group discussion

What opportunities and challenges have you found with working with EMR data?

Opportunities (at RCH)

- Data provide insights; insights provide solutions.
- The EMR contains large volumes of clinical, health service and patient journey information.
- The EMR is a complete record of most patient activity at the hospital.
- Much (but not all) of the data is captured is a readily reportable form and can be extracted electronically.

Challenges (at RCH)

- Information has been collected primarily for clinical treatment and care, not for secondary purposes.
- Significant amounts of information is captured as free text or in scanned documents.
- The underlying data model of the EMR is very complex.
- Biases and censoring exist in EMR-captured data.
- The EMR configuration can change over time.
- All observational data has inferential limitations.

MEDINFO23
8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Centre for Health Analytics

@kidsanalytics

Using EMR Data for Clinical Research

James Scandol

Lead Health Informatician

Data in the EMR

Even at its best, raw electronic medical record data is disorganized and full of uncodified variables

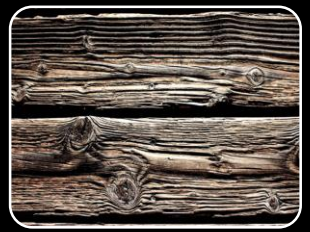
EMRs works well for taking care of patients but wasn't developed with researchers in mind!

- Milinovich and Kattan (2018)

EMR Data Types – a materials analogy

Structured Data

- Discretely captured
- Mandatory entry for treatment, care and administration
- Examples - demographics, ADT, medication orders, vitals, time stamps, encounter diagnosis



Structured with Complexity

- Subject to third parties (e.g. pathology labs)
- Category values changeable by providers
- Mixture of text and discrete values
- Examples – patient problem list, procedures, med admin, lab results

Unstructured Data

- Information exists but not well ordered nor easy to find
- Information requires a lot of effort and expertise to extract
- Can be very difficult to de-identify
- Examples – family history, physical examinations, images, notes

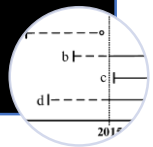
Data quality - with a statistical lens

- Optional fields that are not completed
- Important data stored elsewhere

Missing Data

- Exact value or event time is not known or not observed

Censored Data

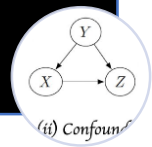


- Data does not represent the information required for study question

Bias

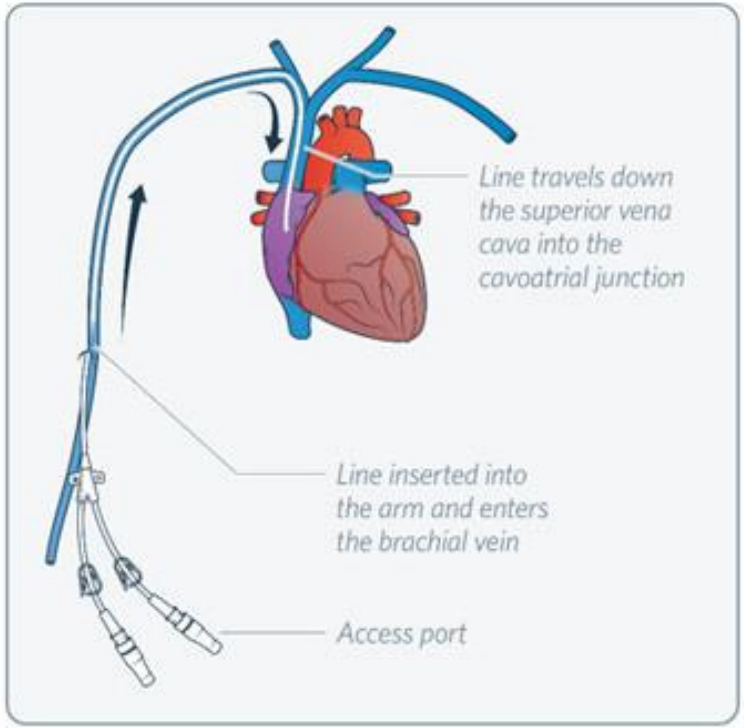
- Additional context to the data that effects interpretation is not available

Confounding



Breakout groups

EMR Data – An example with CVAD complications



www.rch.org.au/kidsinfo/fact_sheets/CVAD__Peripherally_Inserted_Central_Catheter_PICC/

CVAD complications

Progress Notes by [redacted], Resident at [redacted] 6:42 PM

Author: [redacted], Resident
Filed: [redacted]
Status: Addendum
Related Notes: Original Note by [redacted], Resident (Resident) filed at [redacted]

Specialty: After Hours Cover
Date of Service: [redacted] 6:42 PM
Editor: [redacted], Resident (Resident)

Progress Note

Consultant Present?: No

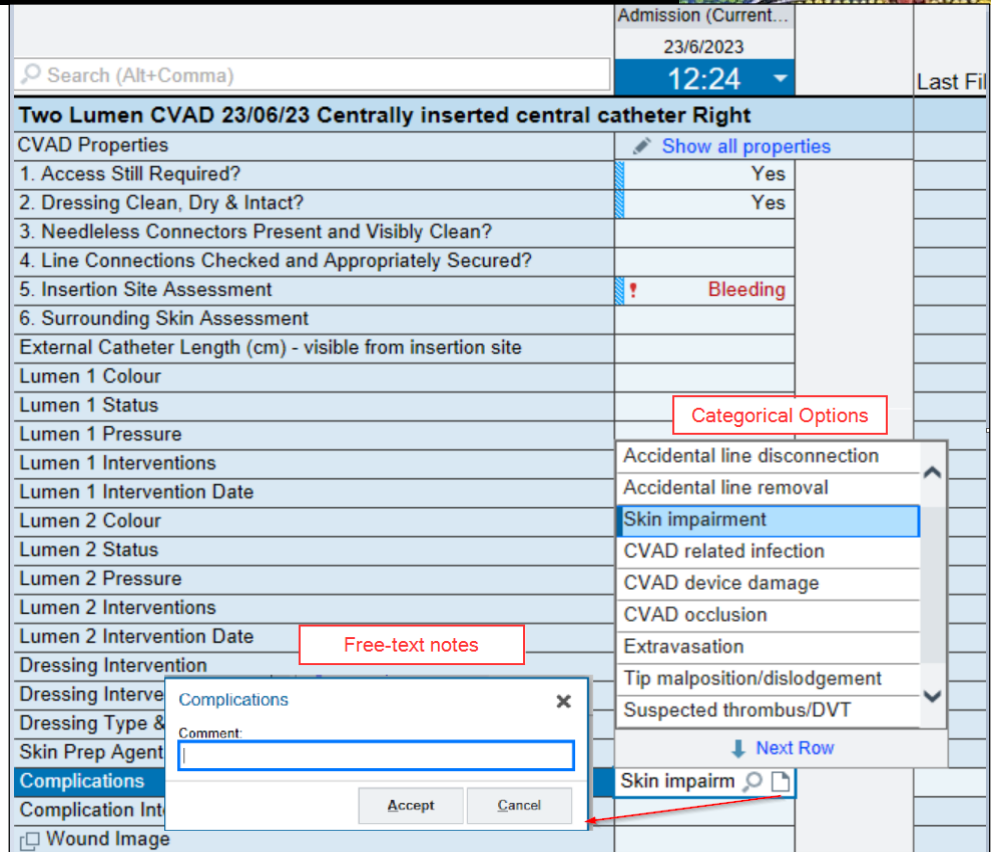
[redacted] 6:43 PM

Resident review ([redacted])
Asked to review re:
1. Brief period of numbness to CVC site (resolved)
2. Change of antibiotic as per IFD note from fellow

[redacted] admitted CVC site suppurative infection with [redacted]

Issues:
New CVC line inserted today [redacted]
- Line inserted under general anaesthetic this afternoon
- Afebrile since line change
- Brief period of numbness over new line insertion site. At time of review this had completely resolved. No numbness/tingling or other sensory changes.
- No pain over line line
- Line appeared clean, no surrounding discharge or erythema.
- No shortness of breath
- Line currently flushing well without issue

Antibiotic plan



CVAD complications

- Complications drop-down list *does not* contain option of 'No Complication'
    - And the Complications drop-down is not required to be completed
- Therefore, if a complication is not documented, it is not clear if
    - Patient has had a complication, *and it has not been documented OR*
    - Patient doesn't have a complication, *and it has not been documented*

Calculations will have a misleading/incorrect numerator (i.e., # of CVADs with complications)

- *Possible Solution: Add another row 'Complications – Yes/No' (default No) which would be used for calculations of CVAD complication prevalence*

Additional considerations for CVAD data

- Need to correct/account for:
    - Number and type of CVADs
    - Length of time in situ
    - Double documentation of discrete data and notes
    - Introduction of default response of 'No' introduces additional bias
        - Change to 'Yes' can result in display of complications row

Centre for Health Analytics

@kidsanalytics

Final Comments

Opportunities

- Data ⟶ Information ⟶ Knowledge ⟶ Wisdom
- No data, no insights; no insights, no solutions
- The EMR contains large volumes of clinical, health service and patient journey information
- The EMR is a complete record of most patient activity at the hospital
- Much of the data is captured is a readily reportable form and can be extracted electronically

Challenges

- Information primarily collected for clinical treatment and care, not secondary purposes
- Significant amounts of information captured as free text or in scanned documents
- Underlying EMR data model very complex
- Missing data, biases and censoring in EMR captured data
- The EMR configuration can change over time
- Inferential limitations of observational data

Acknowledgements

- The Royal Children's Hospital Foundation
- Royal Children's Hospital
- Murdoch Children's Research Institute
- The University of Melbourne Department of Paediatrics
- All of the staff at the Centre for Health Analytics
- The Melbourne Children's Campus staff who have participated in our workshops