Australian Clinical Trials Alliance

Extracting data from eMRs to develop virtual registries: a case study in Acute Coronary Syndrome

Charmaine Tam, PhD

Analytics Translator & Senior Research Fellow

Northern Clinical School, Faculty of Medicine and Health
Centre for Translational Data Science, The University of Sydney

ACTA gratefully acknowledges operational funding from the Australian Government's Medical Research Future Fund

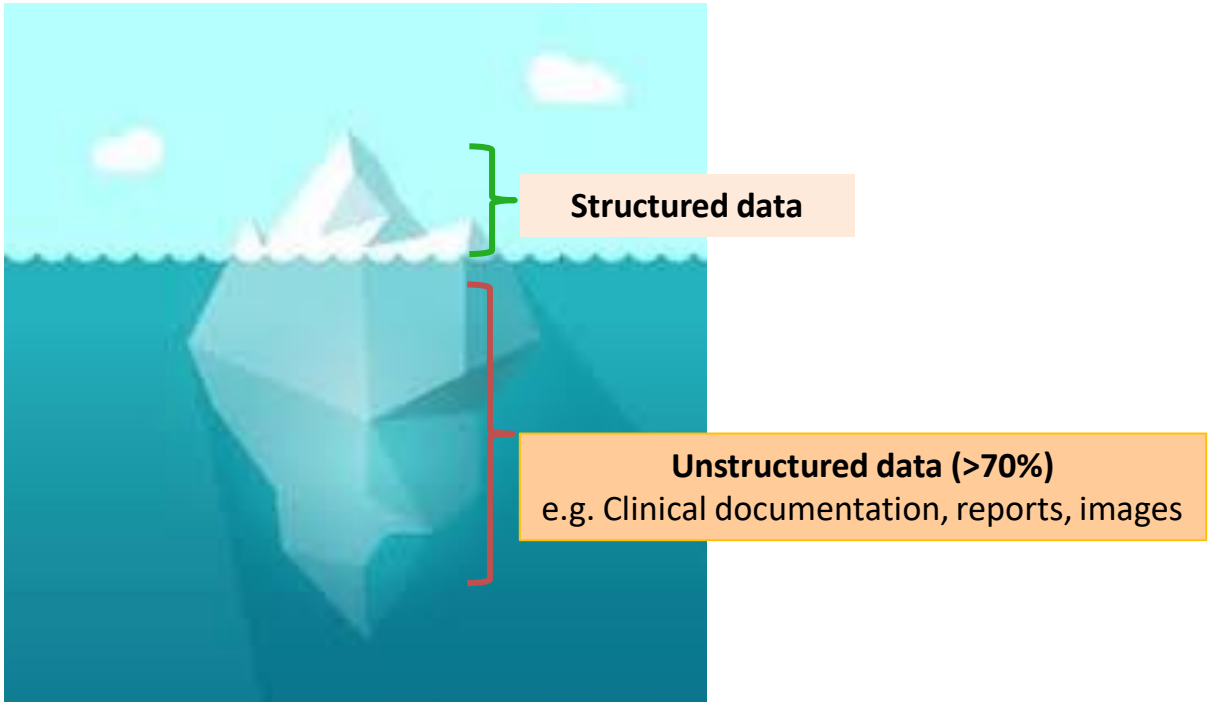
@charmainecodes
@ctds_usyd

Overview of today's talk

1) Findings from a proof-of-concept study using eMR to measure quality and processes of care in patients with acute coronary syndrome

2) Software developed to achieve the project goals

3) Type of skillsets required to extract data from eMRs for performing clinical analytics



SPECIAL ARTICLE ARCHIVE

Electronic Data Processing of Medical Records

Henry W. Baird, M.D.†, and Joseph M. Garfunkel, M.D.‡

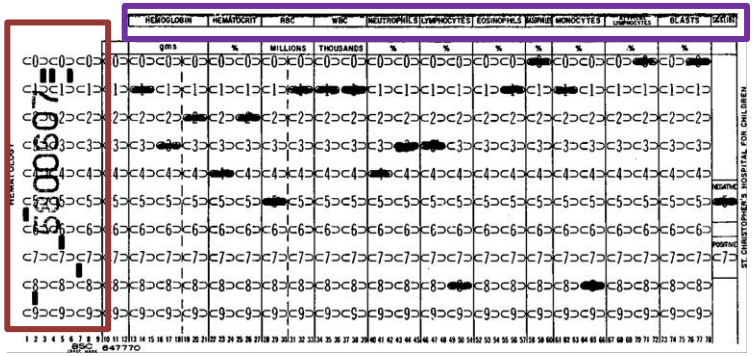
THE NEED FOR BETTER MEANS OF RECORDING AND RETRIEVING DATA FOR MEDICAL records has become increasingly apparent.[1 2 3] The tremendous accumulation of data reflects the use of more sophisticated laboratory procedures, the changing emphasis of medicine from acute to chronic problems and the increased attention to detailed recording of clinical observations. At St. Christopher's Hospital for Children, Philadelphia, over a period of approximately one year, certain clinical observations and laboratory results were recorded via electronic data-processing equipment and thence transcribed for the patients' charts.

It was hoped that a data-processing system would provide an organized, chronologic record of the factual information usually accumulated during a patient's hospital stay. The identifying information found on the "head sheet" of the chart, the child's height, weight and head circumference, the diagnostic impression on admission, the results of various laboratory studies (hematology, urinalysis, chemistry and bacteriology), the requests for special tests and consultations, operative procedures and diagnosis at discharge were thought to be essential.

It was expected that the medical student, the resident, attending consultants and referring physicians would benefit from having these data available on a daily basis throughout the child's hospital stay and complete on the day of discharge. Moreover, we believed that storage of the data in easily accessible and legible form was certain to facilitate research projects, hospital reports and the like.

- Tremendous accumulation of data
- Organise into patient-centric view
- Benefits for clinicians, quality improvement and research

Example of a Mark-Sense card



MRN and date of birth

Haemoglobin= 13.2gm/100ml; **Hematrocrit**: 42%;
Red Blood Count, White Blood Count, Neutrophils, Lymphocytes…

Summary of the day before discharge

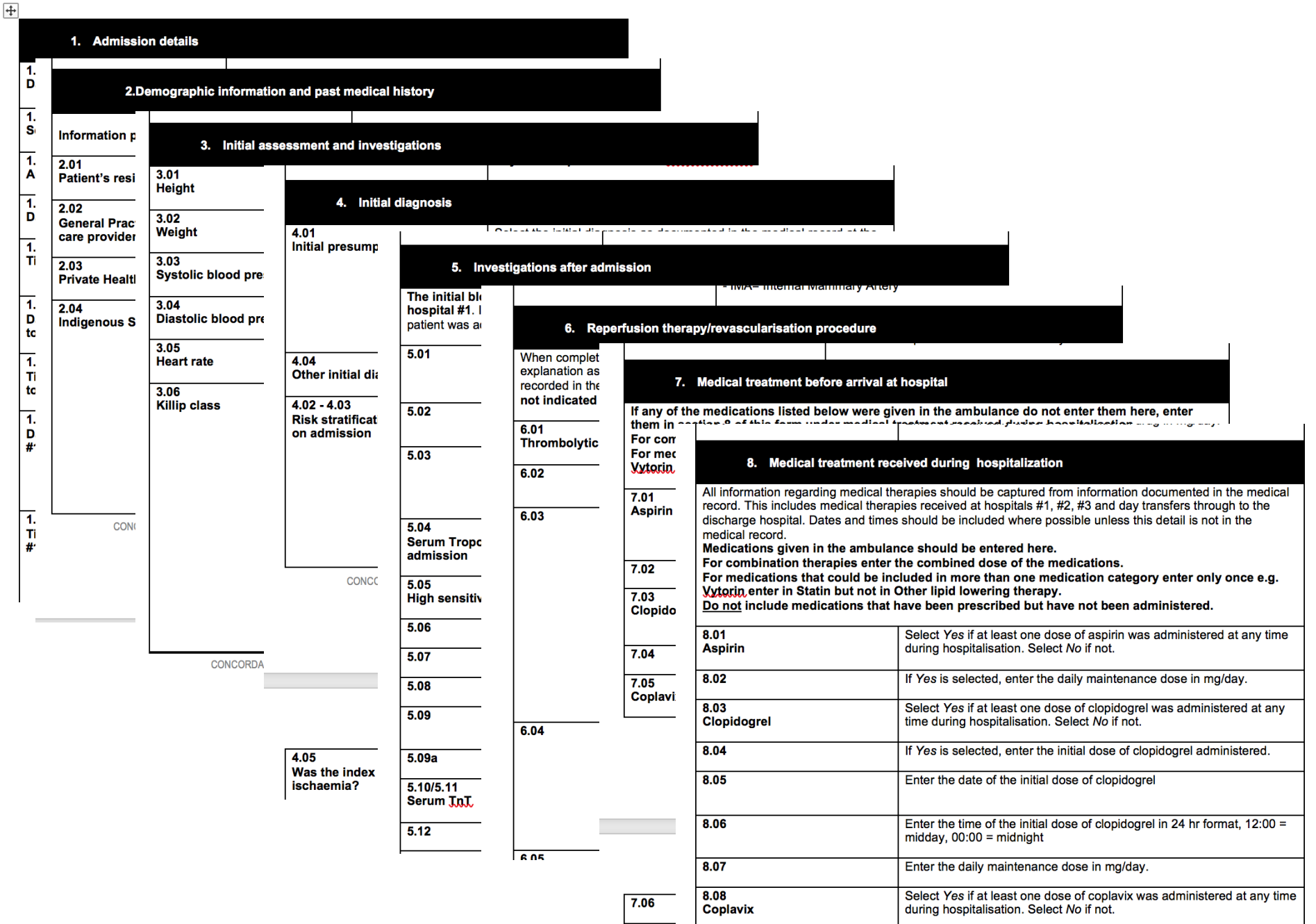
```
                6300983   ANDRE [redacted]                          2
                   NEW  ADMISSION
12/30/63           ADM HT-24 WT-13 LBS HEAD CIRC 17 3/4                    456
12/30/63  920-063  HYDROENCEPHALY                                          456
                   PROCEDURES
      BLOOD.       HB HCT RBC WBC NE LY EO B  MO ATYP BLAST SLK
12/30/63           115    .   06 40 57       03                  .    N
01/06/64           138 39 .                                      .    -
01/06/64           103 35 .                                      .    -
      BLOOD.       BDS MET MYEL NOR PL  RETIC SR BT COAG LE E

      URINE.       PH SP1.0 AL S AC WBC RBC SD CELL CRY CST BAC
12/31/63           7   11   0 0 0    0   0 0   1   1    0 N
12/31/63   11.     OCCASIONAL EPITHELIAL CELLS
12/31/63   111.    OCCASIONAL URIC ACID      CRYSTALS
12/31/63   121.    OCCASIONAL AMORPHOUS      CRYSTALS

      CHEM .       PH  CO2  CL   NA   K   PROT  AL   GLOB  SUN
      CHEM .       S  CHOL OSML CA   P     ALK PE 0  CREATININE
      CHEM .       PCO2 S-B  BIL.T BIL.D TT TF CF SGOT SFP  CFS
01/02/64   106.0    MG 0/0 VENTRICULAR FLUID PROTEIN

01/02/64 2   .     XRAY HYDROCEPHALUS OBST AQUEDUCT SYLVIUS
01/02/64           XRAY VENTRICULOGRAPHY

12/30/63           CBF          BACT RUN NUMBER      038
12/30/63   502.    NO GROWTH IN 72 HOURS
12/30/63                          WBC   OLYMPH   OPOLY     RBC
12/30/63                        NUMEROUS RBC
01/02/64           VENTRICULAR  BACT RUN NUMBER     057
01/02/64                          WBC   OLYMPH   OPOLY   0 RBC
01/04/64*          EYE          BACT RUN NUMBER     138
01/04/64*  502.    NO GROWTH IN 72 HOURS

01/06/64*          ABNORMAL                                    EEG
```

Diagnosis provided by clinician #456

Eye culture showed no growth in 72 hours from the date of collection

Baird and Garfunkel, NEJM, June 10, 1965

Example of a case report form



1. Admission details

2.Demographic information and past medical history

Information p

2.01 Patient's resi

2.02 General Prac care provider

2.03 Private Healt

2.04 Indigenous S

3. Initial assessment and investigations

3.01 Height

3.02 Weight

3.03 Systolic blood pre

3.04 Diastolic blood pre

3.05 Heart rate

3.06 Killip class

4. Initial diagnosis

4.01 Initial presump

4.04 Other initial dia

4.02 - 4.03 Risk stratificat on admission

4.05 Was the index ischaemia?

5. Investigations after admission

The initial bl hospital #1. I patient was a

5.01

5.02

5.03

5.04 Serum Tropc admission

5.05 High sensitiv

5.06

5.07

5.08

5.09

5.09a

5.10/5.11 Serum TnT

5.12

6. Reperfusion therapy/revascularisation procedure

When complet explanation as recorded in the not indicated

6.01 Thrombolytic

6.02

6.03

6.04

6.05

7. Medical treatment before arrival at hospital

If any of the medications listed below were given in the ambulance do not enter them here, enter them in

For con

For mec

Vytorin

7.01 Aspirin

7.02

7.03 Clopido

7.04

7.05 Coplavi

7.06

8. Medical treatment received during hospitalization

All information regarding medical therapies should be captured from information documented in the medical record. This includes medical therapies received at hospitals #1, #2, #3 and day transfers through to the discharge hospital. Dates and times should be included where possible unless this detail is not in the medical record.

Medications given in the ambulance should be entered here.

For combination therapies enter the combined dose of the medications.

For medications that could be included in more than one medication category enter only once e.g. Vytorin enter in Statin but not in Other lipid lowering therapy.

Do not include medications that have been prescribed but have not been administered.

8.01 Aspirin	Select *Yes* if at least one dose of aspirin was administered at any time during hospitalisation. Select *No* if not.
8.02	If *Yes* is selected, enter the daily maintenance dose in mg/day.
8.03 Clopidogrel	Select *Yes* if at least one dose of clopidogrel was administered at any time during hospitalisation. Select *No* if not.
8.04	If *Yes* is selected, enter the initial dose of clopidogrel administered.
8.05	Enter the date of the initial dose of clopidogrel
8.06	Enter the time of the initial dose of clopidogrel in 24 hr format, 12:00 = midday, 00:00 = midnight
8.07	Enter the daily maintenance dose in mg/day.
8.08 Coplavix	Select *Yes* if at least one dose of coplavix was administered at any time during hospitalisation. Select *No* if not.

SPEED-EXTRACT (STEMI Patient ElEctronic Data EXTRACTion)

Proof-of-concept funded by MoH (July 18-Dec 2019) using a 3 month historical data extract in 2 Local Health Districts

Primary Aim

To demonstrate the feasibility of accurately identifying (>90%) patients with ST Elevation Myocardial Infarction (STEMI) from existing suspected Acute Coronary Syndrome data, that reside in electronic medical record systems (EMR) from one quaternary and two feeder hospitals from NSLHD within the Sydney Health Partners collaborative.

Specific aims

1. To develop a process and method for standardised extraction of eMR data to identify patients with a discharge diagnosis of (STEMI)
2. To determine the extent to which extracted data elements can be used to describe clinical quality measures
3. To share this identified cohort and proposed quality measures with practicing clinicians to ensure 'face validity' of the extracted data

Information required for a STEMI Diagnosis

1) Clinical presentation
2) Troponin changes
3) ECG findings

Project timeline

Aug '17 | Jan-April '18 | Mar '18 | July-Dec '18 | Mar '19 | Oct '19 | Jan '20 | May '20

Submitted ethics to **NSLHD**

Ethics approval (NSLHD & CCLHD)

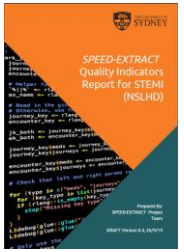
SSA approval (NSLHD)

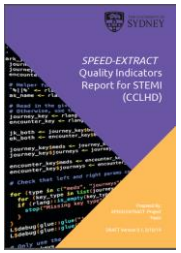
SSA approval (CCLHD)

Test data extracts from Cerner and McKesson systems in NSCCLHDs

(1st April- 30th June 2017)

>300k ECGs extracted from NSCCLD Cerner





(2013-14 extract) SEF Meeting Additional funds from MoH for Jan-June 2020

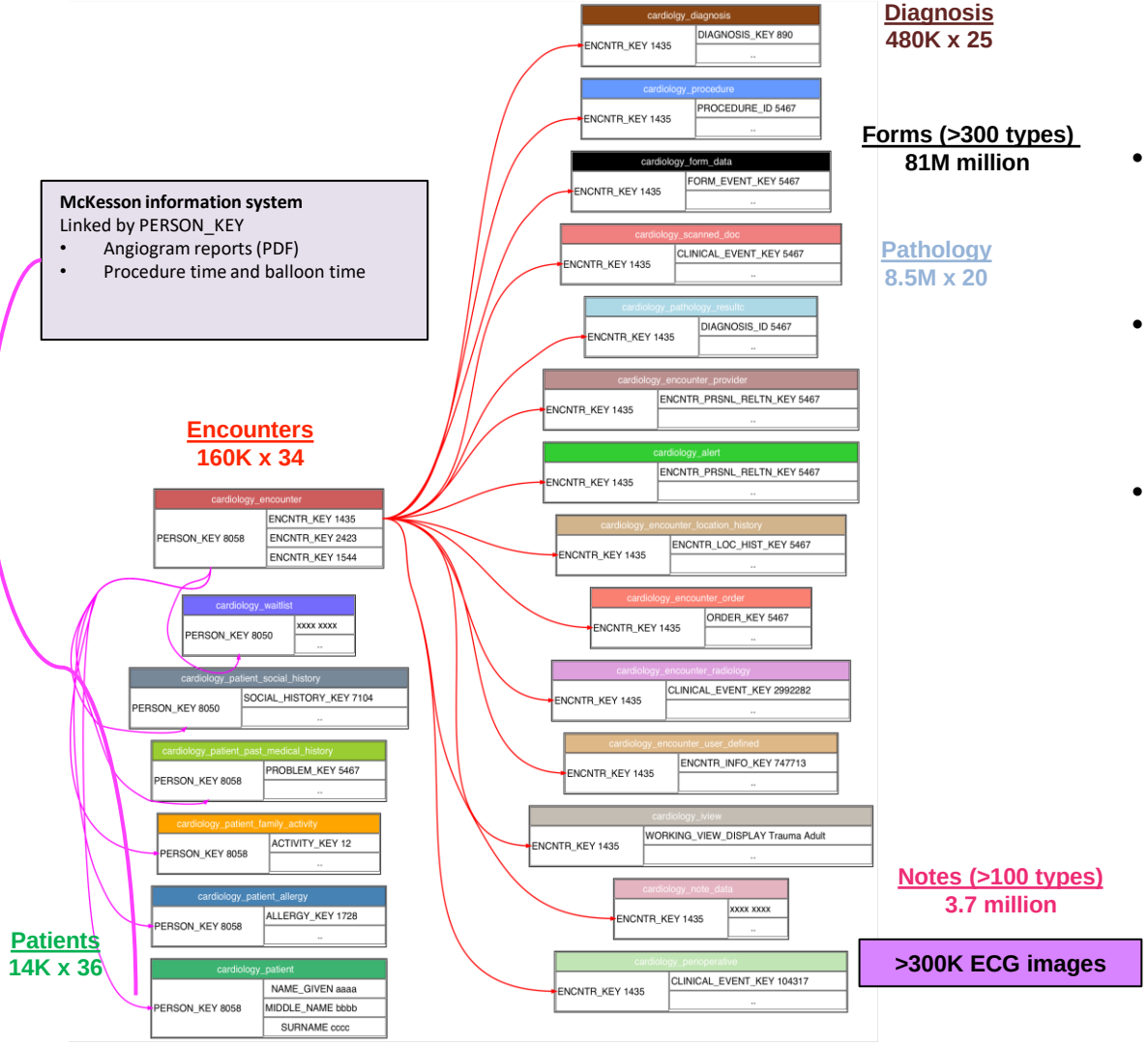
(2013-18 extract, excluding PowerNotes)

Ethics approved and shared CCL code with **SLHD**

1 day test data

CCL code shared with **WSLHD**

Raw Electronic Medical Records

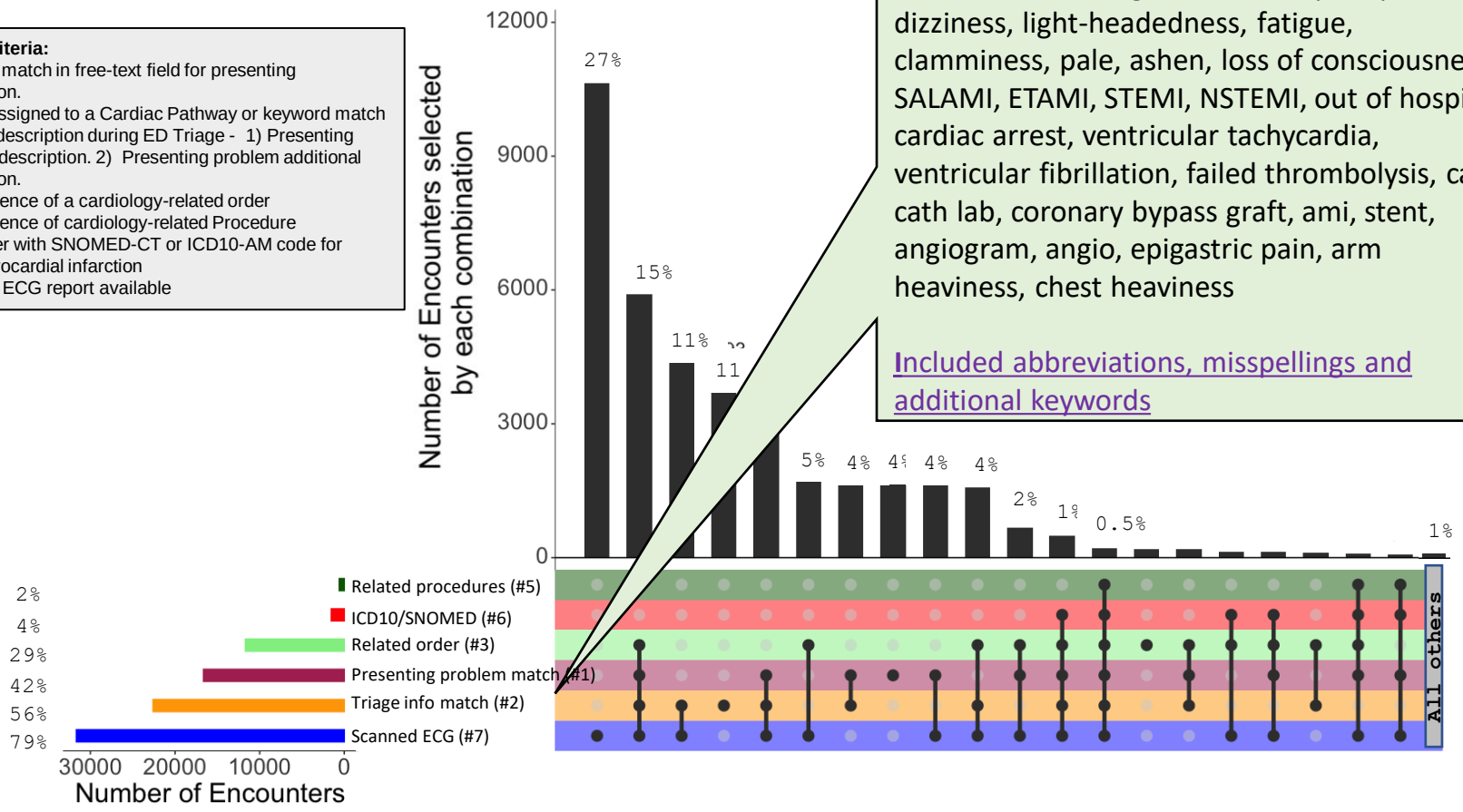


- The data resides within the CERNER Millenium eMR system as well as some specific databases
- The data resides in structured and un-structured formats with the majority being free-text
- Tables are linked by encounter or patient keys

Identifying all presentations of suspected acute coronary syndrome

Inclusion Criteria:

1. Keyword match in free-text field for presenting information.
2. Patient assigned to a Cardiac Pathway or keyword match in either description during ED Triage - 1) Presenting Problem description. 2) Presenting problem additional information.
3. The existence of a cardiology-related order
5. The existence of cardiology-related Procedure
6. Encounter with SNOMED-CT or ICD10-AM code for acute myocardial infarction
7. Scanned ECG report available



Cardiac keywords and symptoms

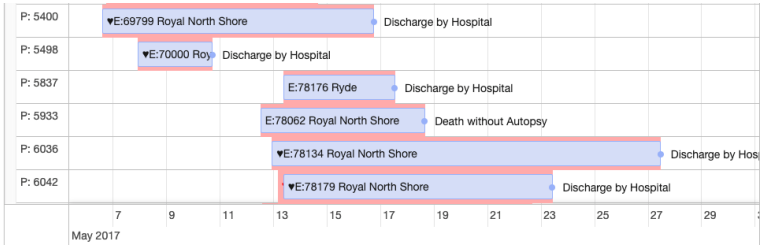
Chest pain, chest tightness, shortness of breath, dyspnoea, weakness, nausea, vomiting, palpitations, syncope, presyncope, unwell, cardiac arrest, indigestion, sweaty, diaphoresis, dizziness, light-headedness, fatigue, clamminess, pale, ashen, loss of consciousness, SALAMI, ETAMI, STEMI, NSTEMI, out of hospital cardiac arrest, ventricular tachycardia, ventricular fibrillation, failed thrombolysis, cath, cath lab, coronary bypass graft, ami, stent, angiogram, angio, epigastric pain, arm heaviness, chest heaviness

Included abbreviations, misspellings and additional keywords

Diversity of episodes of care

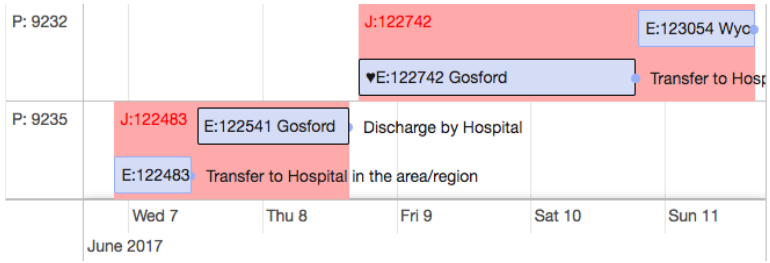
Data in the EMR is captured as encounters. To enable patient-centric analyses, this data was converted into episodes of care. This is especially important for transferred patients where data is captured as ≥2 encounters

Single Episodes of Care



ICD10= STEMIs that present to and are discharged from a single facility.

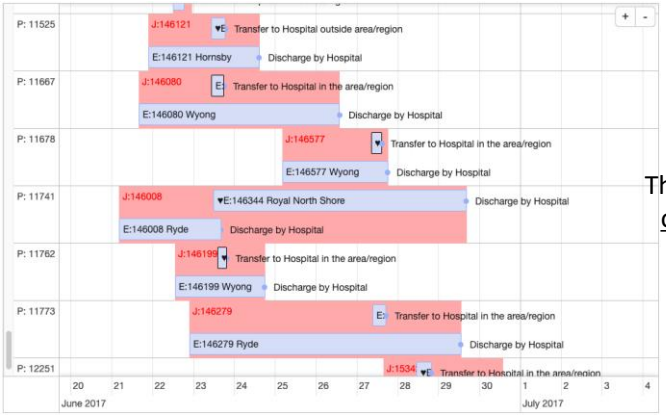
Adjacent Episodes of Care



Patient 9232 (ICD10=STEMI) that presents to Gosford, goes to the cath lab, and is transferred to Wyong which is seen as 2 separate encounters. The time stamps do not overlap.

Patient 9235 presents to Wyong and is transferred to Gosford

Overlapping Episodes of Care



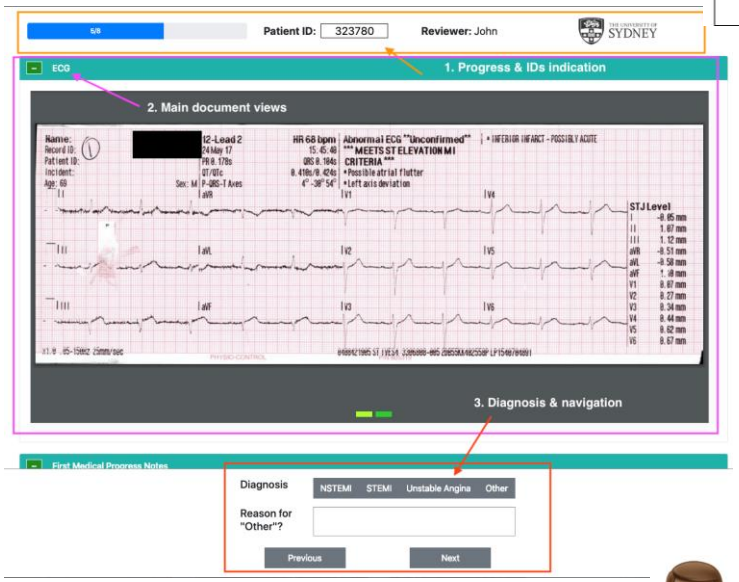
The time stamps do overlap between encounters

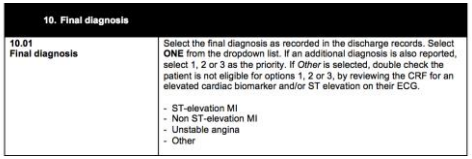
Each row depicts a single patient. The pink shading represents the new journeyID that's been created. Blue boxes represent encounters.
♡ is a proxy for cath lab (admission to or procedure performed in cath lab)

Validation study of ICD10 coded STEMI

- **Rationale: ICD10 codes can be used to identify STEMI but are not entirely reliable and are only available after the episode of care**

- Designed and built a user interface where 4 advanced trainees can easily sight all relevant aspects of *N* patient records (one at a time) and select a diagnosis. Data includes:
  - ECGs
  - First medical note
  - Blood tests (incl. hsTroponin)
  - Angiogram report
  - Discharge letter

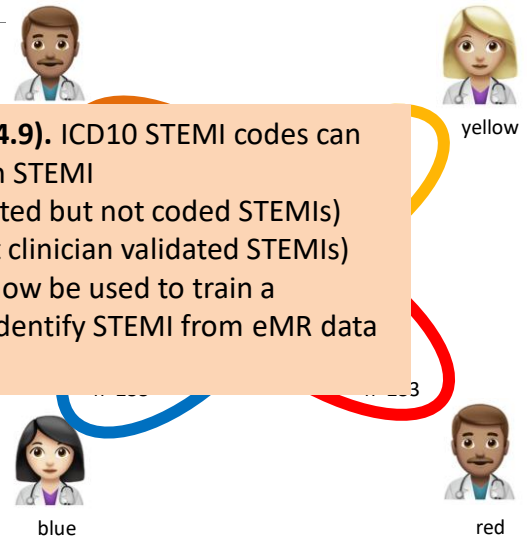




Design
- 4 raters
- 35 shared cases
- 4 diagnostic classes: STEMI, NSTEMI, Unstable Angina, Other

Outcome
Labelled dataset that can be used to train algorithm(s) to identify "real" STEMIs

- **ICD-10 STEMI sensitivity was 93.5% (±4.9).** ICD10 STEMI codes can accurately identify (>90%) patients with STEMI
- 5 false negative STEMIs (clinician validated but not coded STEMIs)
- 10 false positive STEMIs (coded but not clinician validated STEMIs)
- In principle, ICD10 STEMI labels could now be used to train a machine learning model to accurately identify STEMI from eMR data with a high level of accuracy



Clinical Care Standards

Acute Coronary Syndromes Clinical Care Standard

Indicators that are potentially measurable using eMR data

1 A patient presenting with acute chest pain or other symptoms suggestive of an acute coronary syndrome receives care guided by a documented chest pain assessment pathway.

2 A patient with acute chest pain or other symptoms suggestive of an acute coronary syndrome receives a 12-lead electrocardiogram (ECG) and the results are analysed by a clinician experienced in interpreting an ECG within 10 minutes of the first emergency clinical contact.

3 A patient with an acute ST-segment-elevation myocardial infarction (STEMI), for whom emergency reperfusion is clinically appropriate, is offered timely percutaneous coronary intervention (PCI) or fibrinolysis in accordance with the time frames recommended in the current National Heart Foundation of Australia/Cardiac Society of Australia and New Zealand *Guidelines for the Management of Acute Coronary Syndromes*.[a]

In general, primary PCI is recommended if the time from first medical contact to balloon inflation is anticipated to be less than 90 minutes, otherwise the patient is offered fibrinolysis.

4 A patient with a non-ST-segment-elevation acute coronary syndrome (NSTEACS) is managed based on a documented, evidence-based assessment of their risk of an adverse event.

5 The role of coronary angiography, with a view to timely and appropriate coronary revascularisation, is discussed with a patient with a non-ST-segment-elevation acute coronary syndrome (NSTEACS) who is assessed to be at intermediate or high risk of an adverse cardiac event.

6 Before a patient with an acute coronary syndrome leaves the hospital, they are involved in the development of an individualised care plan. This plan identifies the lifestyle modifications and medicines needed to manage their risk factors, addresses their psychosocial needs and includes a referral to an appropriate cardiac rehabilitation or another secondary prevention program. This plan is provided to the patient and their general practitioner or ongoing clinical provider within 48 hours of discharge.

a. Acute Coronary Syndromes Guidelines Working Group. Guidelines for the management of acute coronary syndromes 2006. Medical Journal of Australia. 2006; 184(8):S1-S30.

Quality statement 2 – Early Assessment
- *Indicator 2b*: ECG performed within 10 minutes of arrival of ambulance
- *Indicator 2c*: ECG performed and interpreted within 10 minutes of arrival to ED

✓

Quality statement 3 – Timely Reperfusion
- *Indicator 3a*: STEMI patients receiving fibrinolysis or PCI
- *Indicator 3b*: STEMI patients receiving fibrinolysis within 30 minutes of hospital arrival
- *Indicator 3c*: PCI patients with STEMI with door-to-device within 90 minutes

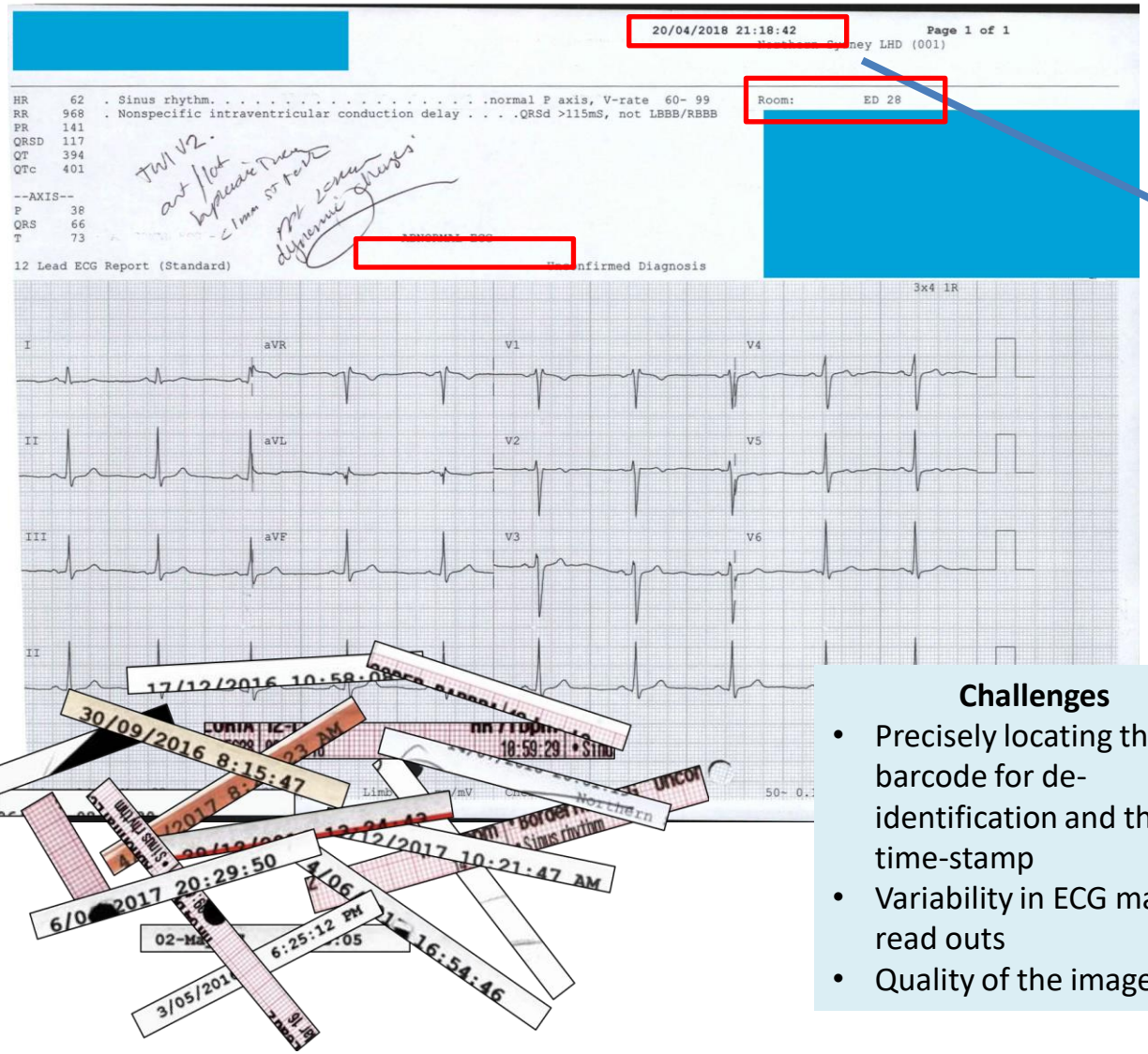
✓

Quality statement 6 – Individualised Care Plan
- *Indicator 6b*: Patients discharged on aspirin or dual antiplatelet therapy

✓

Extracting information from ECGs

4.06 **Date of index ECG**	Record the date of the index ECG in DD/MM/YYYY format.
4.07 **Time of index ECG**	Record the time of the index ECG in 24 hour clock format, 12:00 = midday, 00:00 = midnight. Be aware of the time recorded during daylight saving times. Correct for automated ECG machine time recordings if the settings have not been adjusted.



Optical Character Recognition
Tesseract v4

20/04/2018 21:18:42

20/04/2018
21:18:42

- ABNORMAL ECG -

- ABNORMAL ECG -

Room: ED 28

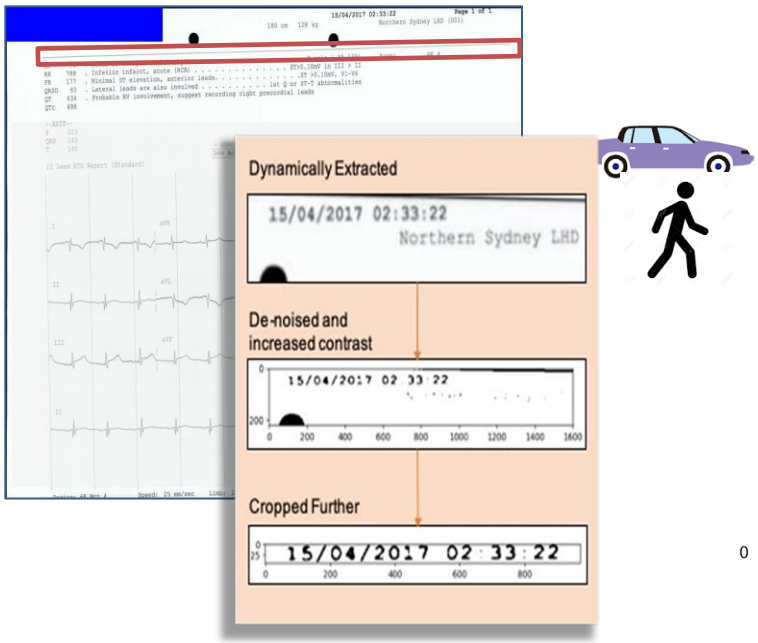
Room: ED 28

Challenges

- Precisely locating the barcode for de-identification and the time-stamp
- Variability in ECG machine read outs
- Quality of the image

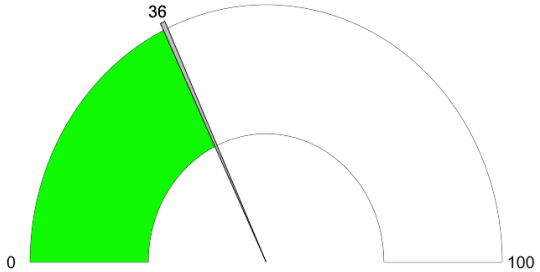
Early Assessment. *Indicator 2c: Time to ECG*

Software



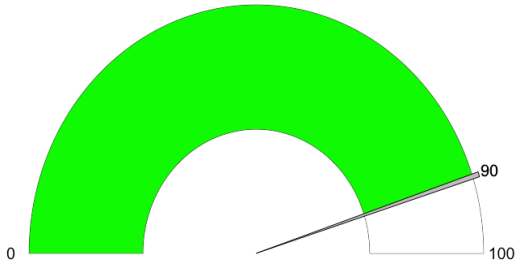
'15/04/2017 02:33:22'

Proportion of self-presenting patients who received an ECG within 10 minutes of arrival



36%

Proportion of patients arriving by ambulance who received an ECG in the ambulance or an ECG within 10 minutes



90%

Results are affected by incomplete data extraction of ECG reports and triage category

Results are in ICD10-AM=STEMI in NSLHD

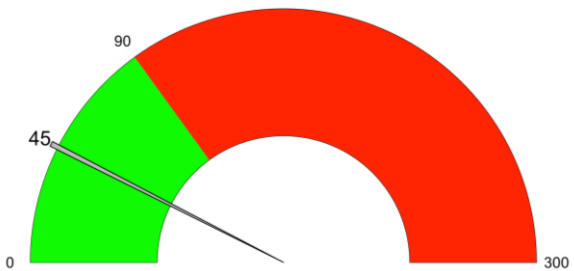
5.78	Enter the date of the **first** cardiac catheterisation.
5.79	Enter the time when the first cardiac catheterisation was performed using the 24hr clock, 12:00 = midday, 00:00 = midnight. The start time should be entered as the time either lignocaine or heparin was administered in the cardiac catheter laboratory. If the time of commencement is unknown, please estimate the time and document this for audit purposes on the patient contact sheet.

Evidence of variation

Timely reperfusion. *Indicator 3c: Door to balloon time*

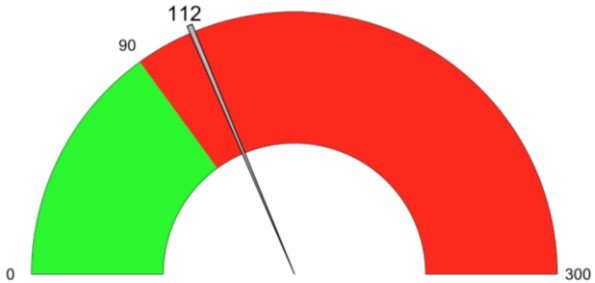
Cerner and other information systems

Patients that first presented a Tertiary facility



45 (28-69)
Median (IQR) (minutes)
in RNSH only

Patients that were transferred from a District to a Tertiary Facility



112 (97-140)
Median (IQR) (minutes)
in transfers to RNSH

Results are in ICD10-AM=STEMI in NSLHD

Individualised care plan.
Indicator 6b: Medications at discharge

9. Medical treatment at discharge or death.	
Enter the medication prescribed for the patient at discharge from hospital. It is not necessary to enter medication at discharge for those patients who died during their hospital admission. **For combination therapies enter the combined dose of the medications.**	
9.01 Aspirin	Select *Yes* if aspirin was prescribed at discharge. Select *No* if not.
9.02	If *Yes* is selected enter the daily dose of aspirin prescribed in mg/day.
9.03	If *No* is selected, select *Yes* if there a documented contraindication to aspirin in the medical records. Select *No* if not.
9.04 Clopidogrel	Select *Yes* if clopidogrel was prescribed at discharge. Select *No* if not.

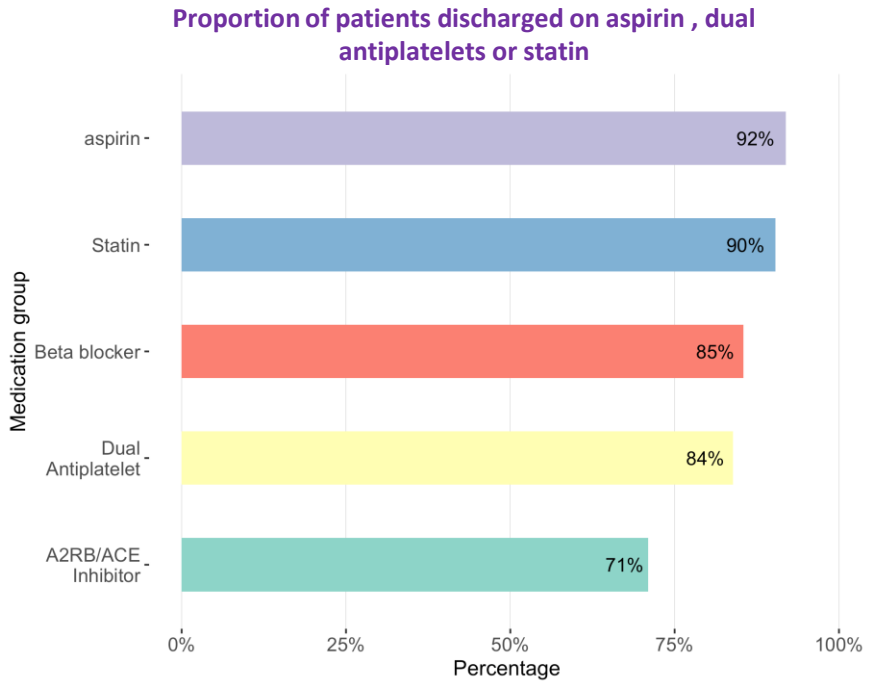
Software

Extract 12666 unique medication names (~350K medication results) in SPEED-EXTRACT

Example from a discharge letter

Discharge Medications:

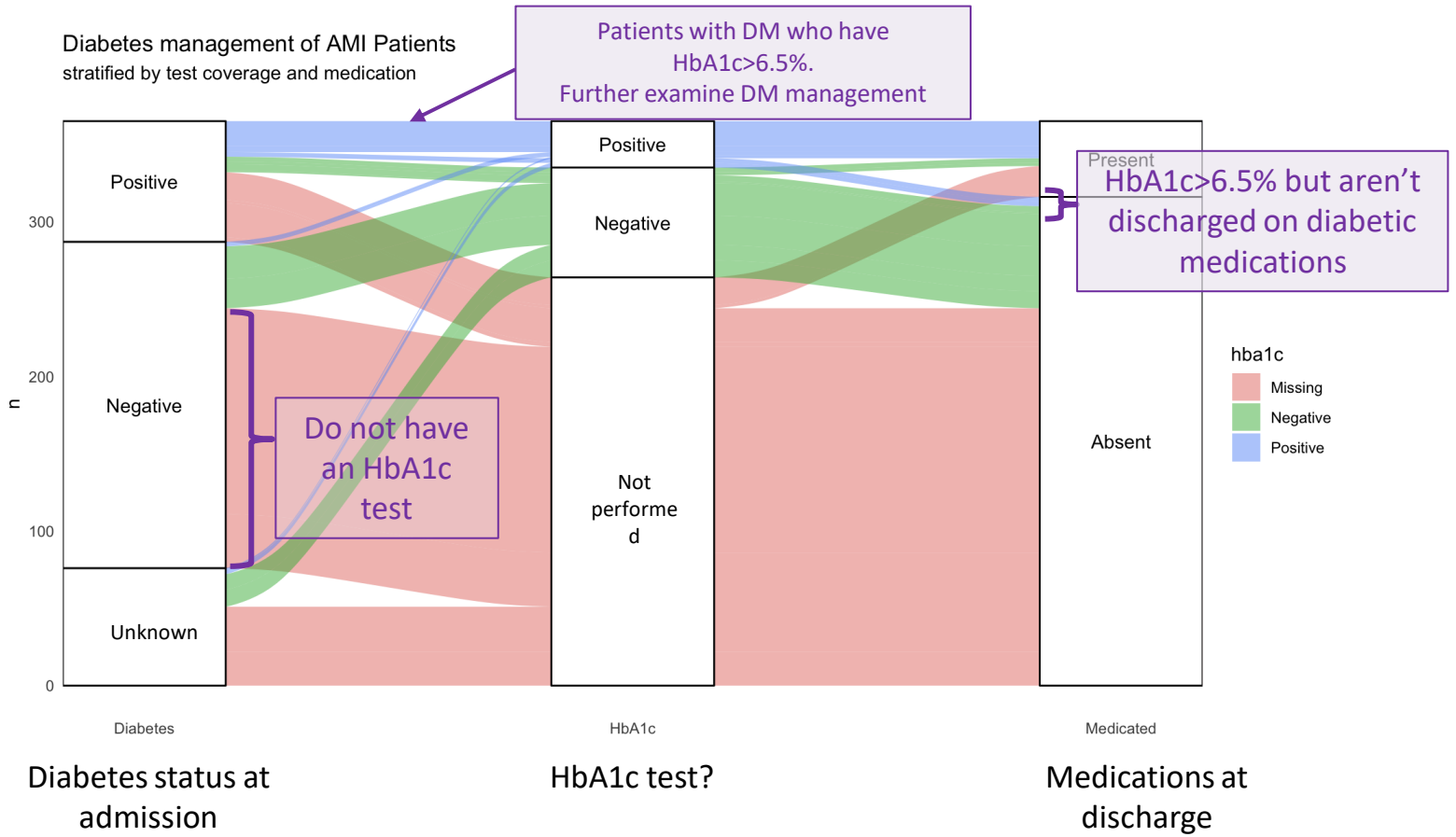
Medication Name	Dose	Freq	Route
ATORVASTIN	**40mg**	**Night**	**Oral**
Status:	New medication		
Last Updated:	13/04/2017 09:00		
Medication Name	**Dose**	**Freq**	**Route**
BISOPROLOL	**1.25mg**	**Morning**	**Oral**
Status:	New medication		
Last Updated:	13/04/2017 09:00		
Medication Name	**Dose**	**Freq**	**Route**
ASPIRIN	**100mg**	**Morning**	**Oral**
Status:	New medication		
Last Updated:	13/04/2017 09:00		
Medication Name	**Dose**	**Freq**	**Route**
TICAGRELOR	**90mg**	**twice daily**	**Oral**
Status:	New medication		
Last Updated:	13/04/2017 09:00		
Medication Name	**Dose**	**Freq**	**Route**
RAMIPRIL	**10mg**	**Night**	**Oral**
Status:	Medication continued - dose unchanged		
Last Updated:	13/04/2017 09:00		
Medication Name	**Dose**	**Freq**	**Route**
ARIMADEX	**1mg**	**Morning**	**Oral**
Status:	Medication continued - dose unchanged		
Last Updated:	13/04/2017 09:19		



Patient with NSTEMI was discharged on statin, β-blocker, aspirin, other antiplatelet and ACE inhibitor

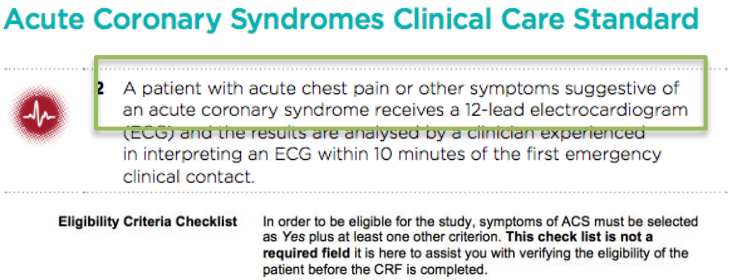
Results are in ICD10-AM=STEMI in NSLHD

Screening for and management of diabetes in patients with AMI

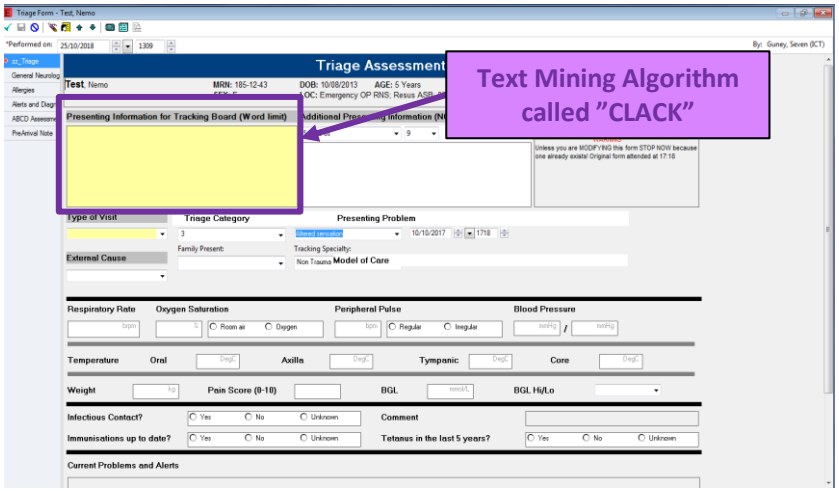


Results are in ICD10-AM=STEMI, NSTEMI in NSLHD

What symptoms do patients with ACS present with?



ED Triage Assessment Interface



Aim

1) Identify patients presenting with acute chest pain or other symptoms associated with ACS
2) Classify encounters into those with high, intermediate, low or no likelihood of cardiac-related chest pain

Source

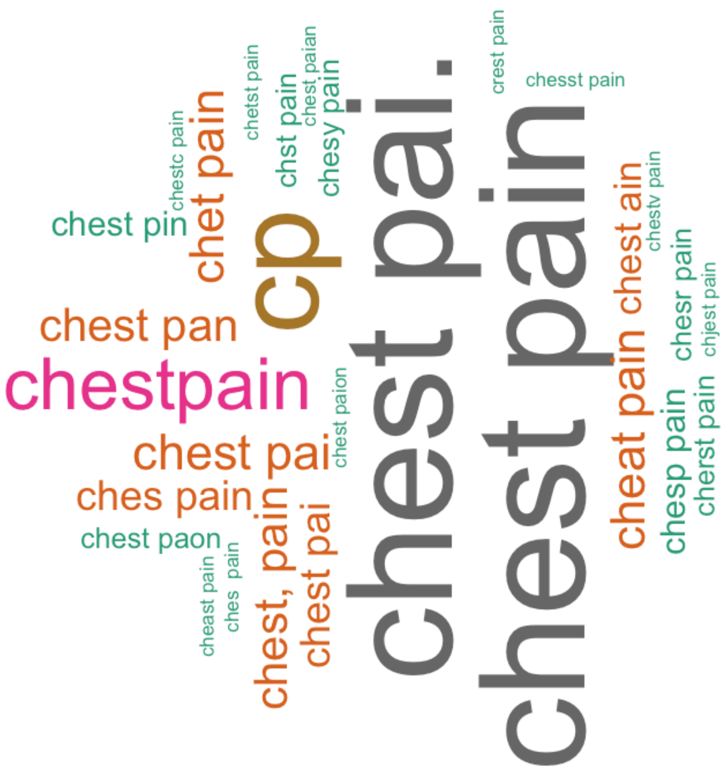
43000 ED Triage forms from 1st April- 30th June 2017

Features of cardiac-related chest pain

i. **Nature** (crushing, heavy, weight, pressure)
ii. **Location of the pain and radiation** (central, typically left sided, up to the jaw, retrosternal, epigastric, radiating to the throat)
iii. **Associated features** (e.g. sweating, nausea, shortness of breath)
iv. **Exacerbating and relieving factors** (e.g, chest pain that is worse with respiration is less likely to be cardiac-related)
v. **Timing** (prioritising current symptoms, also taking into account resolving symptoms)

Step 1: Text Mining – Identifying Symptoms And Keywords

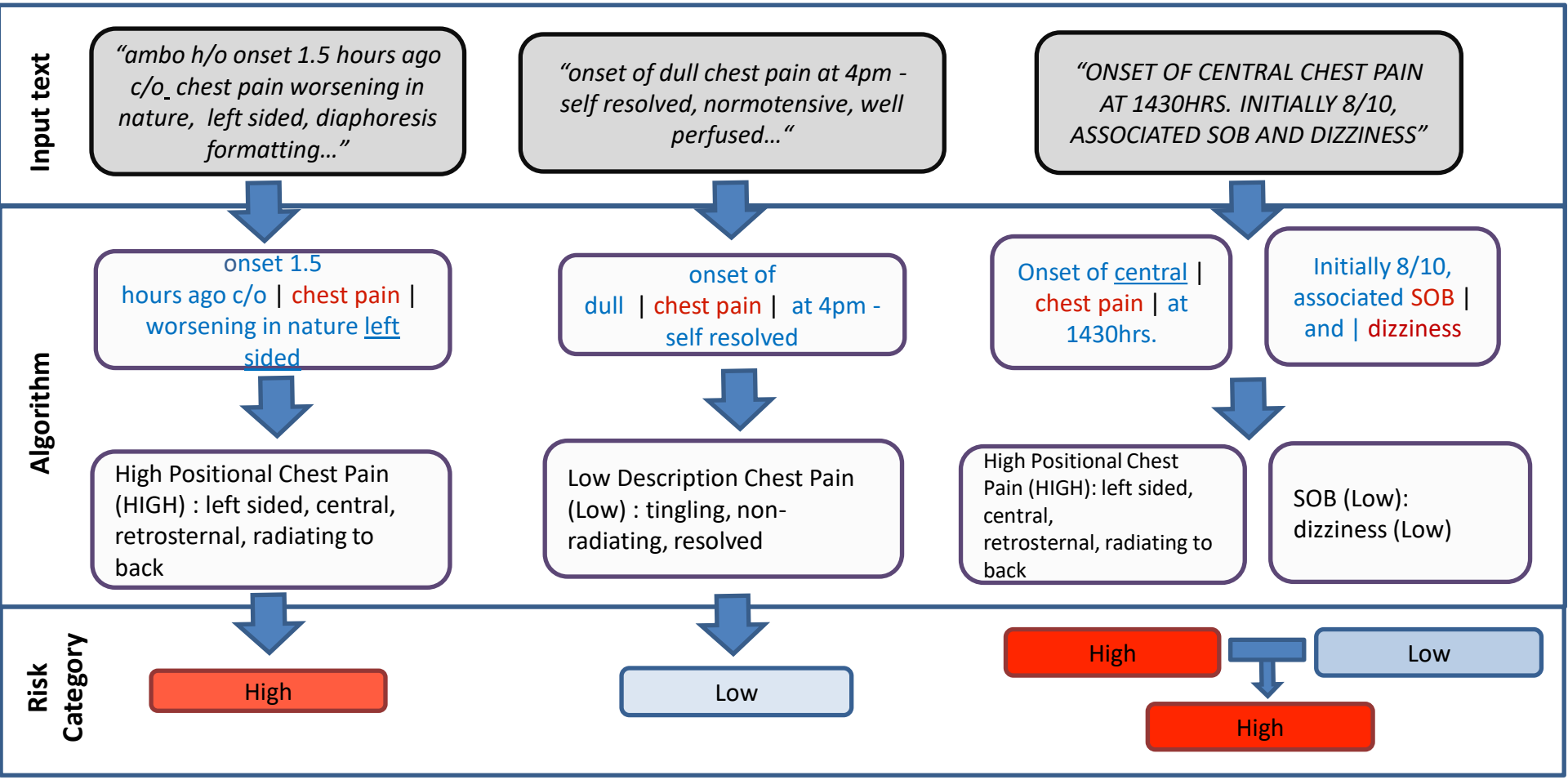
Example: Chest Pain



Fuzzy Text Matching

- Counts minimum number of insertions, deletions and substitution it takes to turn one word to another
- Deletion: Cherst Pain -> Chest Pain – requires 1 deletion (r)
- Insertion: Ches Pain -> Chest Pain – requires 1 insertion (t)
- Substitution: Chast Pain -> Chest Pain – requires 1 substitution (a -> e)
- Allow upto 2 of any deletions, insertions, substitution
  - **32 different ways "chest pain" has been spelt**
- Better identification of symptoms on sparse ED Triage Form data

Step 2: Text Mining Contextual Analysis Examples

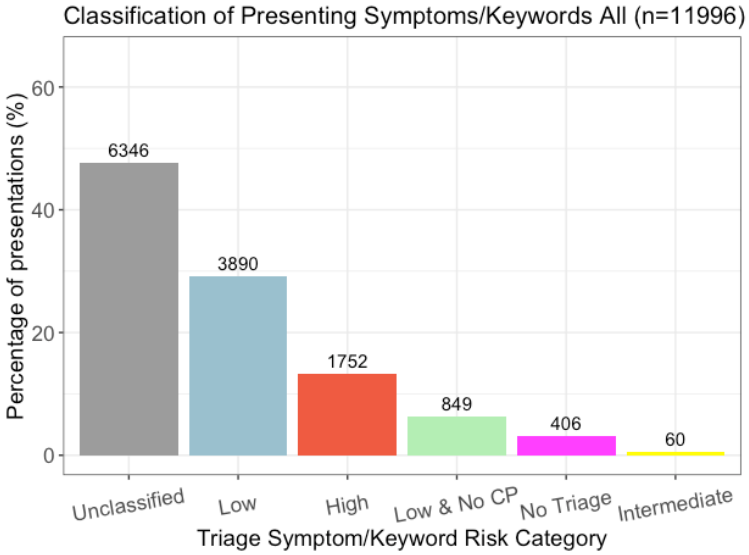


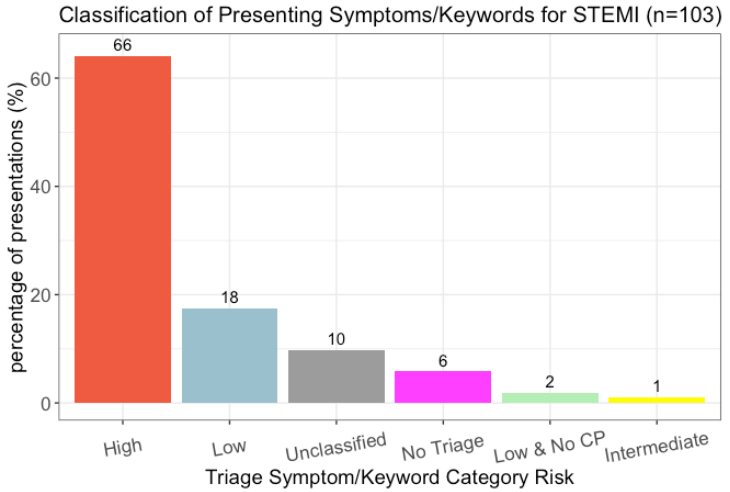
Symptom/Keyword Grouping and Risk Category

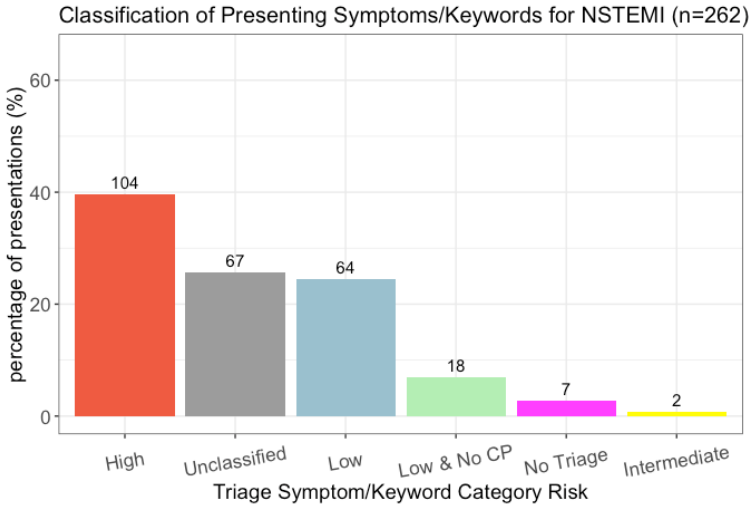
Symptom/Keyword Group	Risk Category
high_positional_chest_pain	High
chest tightness/heaviness	High
check_coronary_artery_symptom	High
high_description_chest_pain	High
check_squeezing_constricted_banding	High
check_ecg_changes	High
check_etami	High
check_l_axilla_pain	High
check_retros_heaviness	High
intermediate_description_chest_pain	Intermediate
check_clopidogrel	Intermediate
check_pain_on_exertion	Intermediate
shortness_of_breath	Low
dizzy	Low
vomiting	Low
nausea	Low
pale	Low
palpitation	Low
no chest pain	No

- Developed text mining algorithm in the entire cohort of patients with suspected acute coronary syndrome
- **36** different symptom/keyword searches which accommodates for abbreviations and misspellings
- The encounter is assigned the highest risk category associated with the symptom/keyword group identified in the text
- **Our text mining tool called "CLACK" can take any input text source and examine strings of interest eg. risk factors, evidence of thrombolysis, past medical history**

Risk category at presentation







In presentations of suspected acute coronary syndrome

- 50% did not feature any of the 36 symptom/keyword groupings
- 15% features symptom/keywords of high risk category – greater in both ICD10 STEMI and NSTEMI cohorts

High risk presentations compose 63% and 40% of ICD10 STEMI and NSTEMI cohort, respectively

Text mining and natural language processing approaches in medical notes

1. **Identify patients presenting with specific symptoms and classify their severity**
   - **Low to high risk of cardiac-related chest pain**

2. **Describe the patient population**
   - Risk factors- diabetes, hypertension, dyslipidemia, smoking, obesity
   - Medical history
   - Family history- prior PCI, prior CABG, prior AF
   - Cognitive impairment/dementia

3. **Inform on the mode of presentation**
   - field triage
   - presented with cardiac arrest/OHCA

4. **Determine whether investigations and treatment h**
   - Thrombolysis
   - Seen by a specialist

5. **Identify in-hospital clinical events**
   - Major bleeding
   - Worsening chronic heart failure
   - Stroke

6. **Examine plans at discharge**. Eg. referred for cardiac r

Ongoing considerations and challenges

- Rule-based approaches are imperfect
- Algorithms generally require gold-standard labelled datasets
- Unique nature of medical notes (eg. duplication, variation in formatting)
- Sparsity of true cases within large datasets
- >300 different types of notes in the eMR (eg. ED Assessment, Discharge summaries, progress notes, etc) with different structures

Finding unique section headings in ED Assessments and Discharge Summaries

- There were >132K different section headings across 64K ED Assessments and 119K Discharge Summaries
- Some popular section headings included:

Section Heading	Frequency	Section Heading	Frequency
Health Status	104653	Facility:	24469
Visit Information	101262	Plan	24043
Discharge Information	84921	**Allergies**	22614
Summary of Care	82946	Allergies	22614
Medications	70466	Histories	22270
Results Review	63457	* Final Report *	20640
Discharge Medications:	35160	Impression and Plan	18544
Chief Complaint	32149	Impression	17871
Other Results	27634	Physical Examination	17199

- Small differences can cause new section headings to be identified. For example,
  - "Medications" versus "Discharge Mediations:"
  - "Plan", "Impression" and "Impression and Plan"
  - "** Allergies **" vs "Allergies"

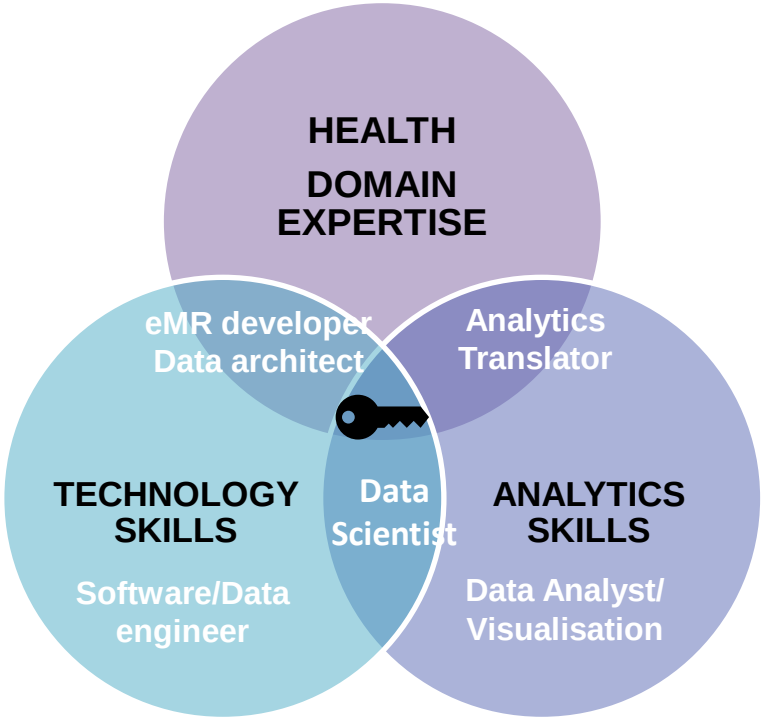
Standardising section headings would help…

- For example, “Chief Complaint” or “Presenting Complaint” are likely the same heading.
- There were >200 different ways presenting complaint was described. Using these section headings, we can identify presenting symptoms in 85% of all ED Assessments
- Some examples include:

Section Heading	Freq	Section Heading	Freq
Chief Complaint	32153	History of Presenting Concern	90
History of Presenting Complaint	17000	Presenting Problem:	50
Presenting Complaint	7734	Complaint:	38
History of Presenting Problem	1307	History of Presenting	23
History of Present Illness:	912	History of Presentation	22
History of Presenting Complaint:	621	.Presenting Complaint	21
History of Presenting Illness	530	Main complaint	15
Presentation:	300	Presenting Issue	13
Hx of Presenting Complaint	7	Presenting with	11
Primary complaint	6	History of Presenting Complain	9
This presentation:	4	Presenting:	9
Emergency Presentation	7	Presented with:	8

Relevant for identifying and aggregating sections where presenting symptoms are documented

Clinical analytics requires a multi-disciplinary partnership model and a broad analytics skillset



Role	Function
Analytics translator	Ensures analytics solve critical operational and clinical problems
eMR developer	Extracts data from information systems
Data architect	Ensures quality and consistency of present and future data flows
Data engineer	Build software tools to structure, process and analyse data
Data/visualisation analyst	Exploratory data analysis, visualise data and build reports and dashboards
Data scientist	Develop statistical models and algorithms

Some of these skillsets already reside within Departments within jurisdictions (eg. Performance and Analysis, ICT, eMR Team)

Adapted from: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/ten-red-flags-signaling-your-analytics-program-will-fail#

Towards virtual clinical registries

Real-time decision support

Audit and benchmarking

Close to real time audit and feedback

Data repository for research and pragmatic clinical trials

SPEED-EXTRACT Team

SPEED-EXTRACT- NSW
SEF meeting, 28/2/20

Partnership NSW Health/eHealth
Michelle Cretikos
Marianne Gale
Wilson Yeung

Cardiology clinical expertise (SHP)
David Brieger**
Janice Gullick
Steve Vernon
Gemma Figtree
Clara Chow

Other clinical leaders and contributors with data expertise
Jonathan Morris**
Angus Ritchie
Seven Guney

External domain expertise
MKM Health

Data / informatics Team

Lou

External funders
Sydney Health Partners
ACI
NSW Ministry of Health

PoC (eHealth/NSLHD)
Rapid Data Ingestion
Goal:
Demonstrate the feasibility and value of a rapid extraction and ingestion cloud platform

Accord for the Sharing of Health Data across Sydney Health Partners

** Principal investigators

USYD Centre for Translational Data Science

USYD Sydney Informatics Hub

@charmainecodes
@ctds_usyd