Health Translation Queensland

Data Management and Analytics for Healthcare

Rudolf Schnetler
Townsville Institute of Health Research and Innovation

Sankalp Khanna
CSIRO Australian e-Health Research Centre

26th March 2024

Queensland Government | Townsville Hospital and Health Service | CSIRO Australian e-Health Research Centre

Australian Government National Health and Medical Research Council | NHMRC ACCREDITED RESEARCH TRANSLATION CENTRE

healthtranslationqld.org.au

+61 7 3346 4637 info@healthtranslationqld.org.au

Acknowledgement of **Country**

Health Translation Queensland acknowledges the Traditional Owners and their custodianship of the lands on which we meet.

We pay our respects to their Ancestors and their descendants, who continue cultural and spiritual connections to Country.

We recognise their valuable contributions to Australian and global society.

Agenda

- Data management planning in the early stages of research projects
- The importance of documentation for data reproducibility and sharing
- Using analytics and Artificial Intelligence to improve healthcare
- Applying appropriate scientific rigor in planning and executing data science projects

Data management planning in the early stages of research projects

The importance of documentation for data reproducibility and sharing

Queensland Government
Townsville Hospital and Health Service
CSIRO
Australian e-Health Research Centre

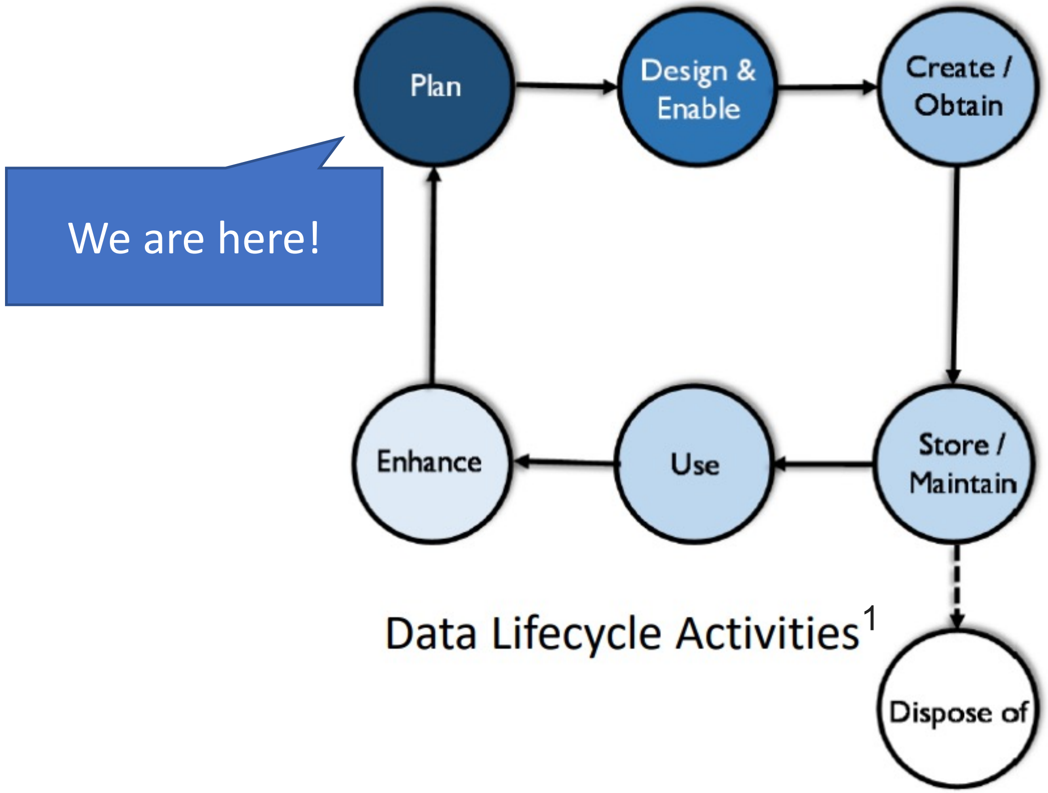
Data Management Planning

- Data Lifecycle

- Early Planning:
  - Protocol and HREA

Excluding data governance

Data Lifecycle



[1]International, D. (2017). DAMA-DMBOK: data management body of knowledge. Technics Publications, LLC.

Data Management Planning

- Start at the protocol!
  - A data management plan should be developed as early as possible
  - Data management considered throughout approval steps
  - Guides the data lifecycle throughout the project
  - Starting Guide: National Statement on Ethical Conduct in Human Research[2]
    - Data Management: Section 3.1.43 – 3.1.49

[2]National Health and Medical Research Council, Australian Research Council and Universities Australia (2023). National Statement on Ethical Conduct in Human Research. Canberra: National Health and Medical Research Council.

Queensland Government | Townsville Hospital and Health Service | CSIRO | Australian e-Health Research Centre

Key HREA Considerations

- 3.1.43: Agreement of data custodian
- 3.1.44: How the data is collected, access, used, disseminated and disposed
- 3.1.45: Information security agreements to meet privacy risks
- 3.1.46: Researchers to comply with all legal and regulatory requirements (data governance)
- 3.1.47: Preservation of biological samples
- 3.1.48: Data, information and biospecimen disposal in legal manner
- 3.1.49: Data should be reusable for future research

Queensland Government
Townsville Hospital and Health Service
CSIRO
Australian e-Health Research Centre

Documentation for Reproducibility

- Significant resource investment in collecting, cleaning and structuring data
- Data sharing is difficult to navigate in health research
- Enable reproducibility within a constrained environment

Queensland Government

Townsville Hospital and Health Service

CSIRO

Australian e-Health Research Centre

Documentation for Reproducibility

- FAIR Principles[3]
  - Findable
  - Accessible
  - Interoperable
  - Reusable

[3]Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Queensland Government

Townsville Hospital and Health Service

CSIRO

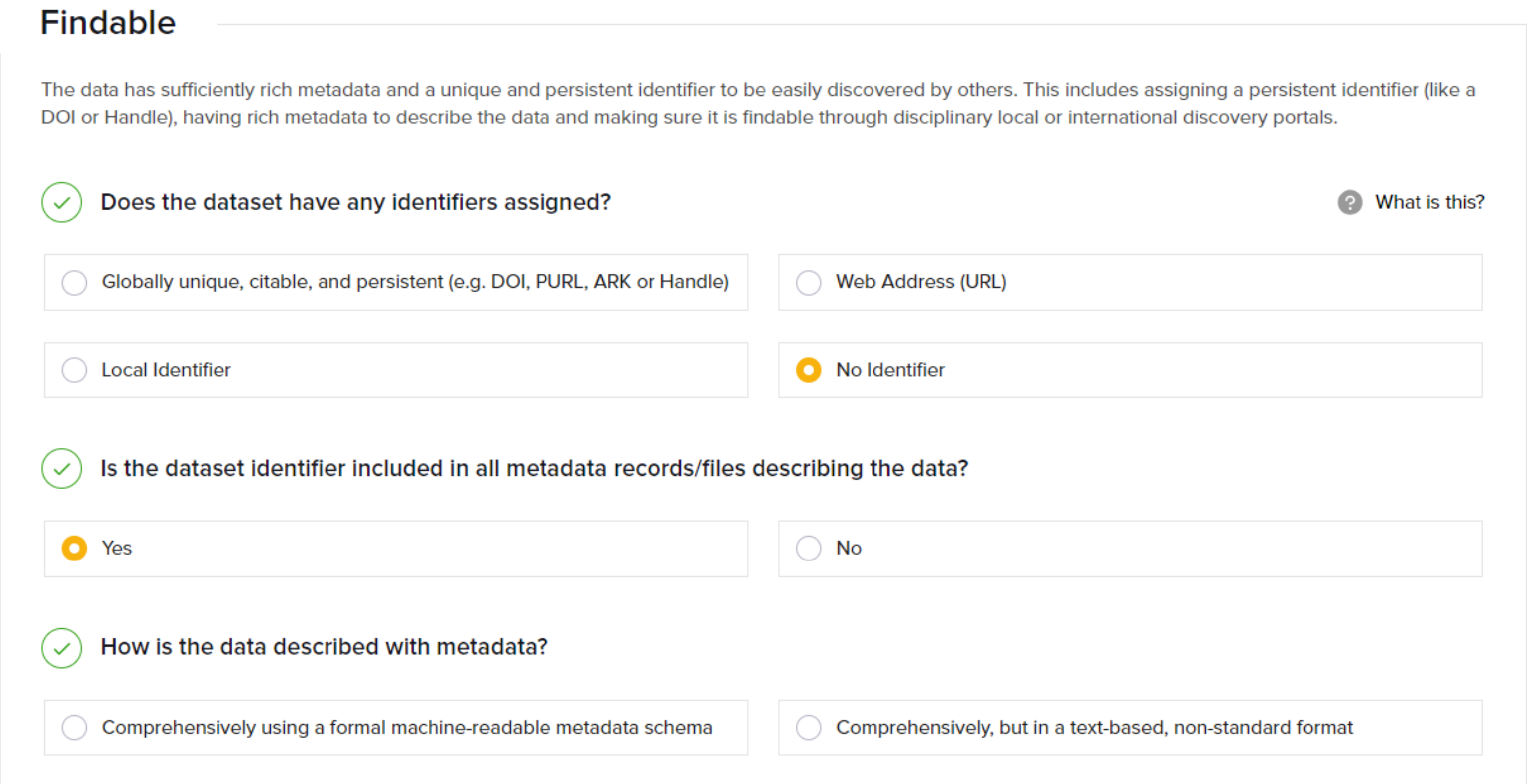
Australian e-Health Research Centre

Documentation for Reproducibility

- Using FAIR Principles for Reproducibility
  - Identifiers
  - Described with Metadata
  - Metadata and data accessibility
    - Retention of the metadata record
  - Metadata and data format
  - Metadata linkage
  - License
  - Provenance information

Documentation for Reproducibility

- ARDC FAIR Assessment Tool[4]



[4]FAIR Data Self Assessment Tool | ARDC. (2023, November 22). ARDC. https://ardc.edu.au/resource/fair-data-self-assessment-tool/

Documentation for Reproducibility

- ARDC FAIR Assessment Tool[2]
- Explore the Questionnaire early

[4]FAIR Data Self Assessment Tool | ARDC. (2023, November 22). ARDC. https://ardc.edu.au/resource/fair-data-self-assessment-tool/

Documentation for Reproducibility

- ARDC FAIR Assessment Tool[2]
- Explore the Questionnaire early
- Assessment:

[4]FAIR Data Self Assessment Tool | ARDC. (2023, November 22). ARDC. https://ardc.edu.au/resource/fair-data-self-assessment-tool/

Queensland Government
Townsville Hospital and Health Service
CSIRO
Australian e-Health Research Centre

Using Analytics and Artificial Intelligence to Improve Healthcare

Applying appropriate scientific rigor in planning and executing data science projects

Queensland Government
Townsville Hospital and Health Service
CSIRO
Australian e-Health Research Centre

Motivation: The Australian Health System

Over a Decade (2008-09 – 2018-19)

Health Spending: 71% (raw) 42% (adj)

Population Growth: 17%

Reference: Australia's health 2018, https://www.aihw.gov.au/reports/australias-health/australias-health-2018/

Queensland Government

Townsville Hospital and Health Service

CSIRO

Australian e-Health Research Centre

Motivation: Australian Patients are at Risk

Queensland Government
Townsville Hospital and Health Service
CSIRO
Australian e-Health Research Centre

The Solution – Digital Health

The health system will shift focus...

...from treating patient illness to managing consumer health and wellbeing

...from accepting one-size-fits-all to precision health solutions

...from a reactive system to a holistic and predictive approach

...from extending life to improving quality of life over a lifetime

Five enabling themes

Each enabling theme is designed to assist in Australia's health system shift, with the aim of making continual improvements to Australia's overall health outcomes

Empowering consumers

Consumers are an underutilised resource in the health sector. Consumers can be empowered to better prevent illness and manage their health via increased information access and consumer focussed health solutions.

Addressing health inequity

Supporting groups that have inequitable health outcomes to move up the health curve will provide greater social and economic returns than just extending the lifespan of those most advantaged.

Unlocking the value of digitised data

Behavioural change is needed by all health stakeholders to ensure the growing volume of personal health data is securely shared, collated, analysed, interpreted, and paired with action for improved health and wellbeing.

Supporting integrated and precision health solutions

Greater systems integration and precision health solutions must be underpinned by improved predictive analytics, an outcomes-based mindset, and a new set of skills for health professionals.

Integrating with the global sector

Improved global integration will help the sector connect and contribute to world leading health and management solutions and encourage the development of novel and globally exportable solutions in Australia.

Government, industry, researchers and the community must collaborate to create the **value** inherent in this shift

Improved health outcomes and equity for all Australians.

Greater system efficiencies that flatten the cost curve of health financing.

More impactful and profitable business models.

Creation of new industries based on precision and preventative health.

More sustainable and environmentally friendly healthcare practices.

More productive workers leading to increased job satisfaction and improved work-life balance.

https://www.csiro.au/en/work-with-us/services/consultancy-strategic-advice-services/csiro-futures/future-health

Queensland Government

Townsville Hospital and Health Service

CSIRO

Australian e-Health Research Centre

AEHRC - CSIRO's National Digital Health Research Program

HEALTH INFORMATICS

Improving health system performance & productivity from electronic health data

How: Meaningful data interoperability and analysis for decision support, analytics, modelling and reporting

BIOMEDICAL INFORMATICS

Biostatistics, imaging and genomics based -clinical workflows

How: Leveraging operational & clinical data through analytics, modelling, decision support & automation

HEALTH SERVICES

Improving access to services & management of chronic diseases

How: Service delivery models utilising telehealth, mobile health & remote monitoring

Queensland Government

Townsville Hospital and Health Service

CSIRO

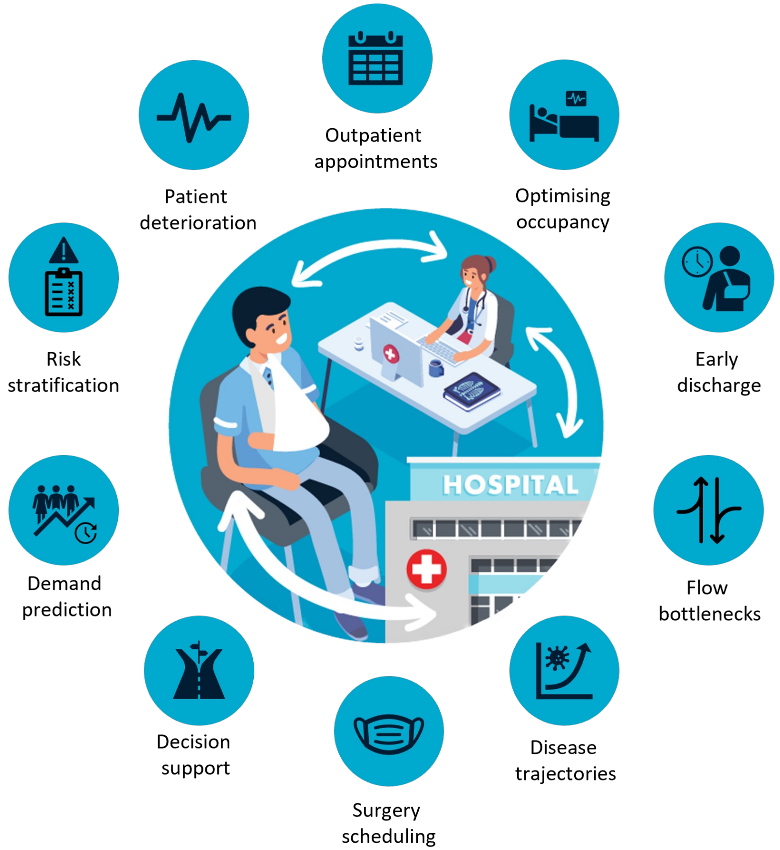
Australian e-Health Research Centre

Health Intelligence @ CSIRO AEHRC

Evidence Based Healthcare

Health System Productivity & Operational Decision Support

Clinical Decision Support



Queensland Government

Townsville Hospital and Health Service

CSIRO Australian e-Health Research Centre

Evidence Based Healthcare

Hospital Access and Patient Flow

- Bottlenecks and strategies for improving flow
- Workflows and health system KPIs
- ED, inpatient and outpatient patient journeys

Clinical Insights and Reporting

- Clinical care & treatment pathways
- Reporting and benchmarking

Contributing Statistical Expertise

- Advanced statistical support for internal and external collaborators

"A major multi-agency study is underway to investigate contributing factors to rising waiting times in public hospital emergency departments in Queensland, and to help find solutions.

Source: Emergency Medicine Foundation website

COVID-19 Critical Care Consortium
Incorporating ECMOCARD

63 countries across 6 continents
26,427 patients across 441 collaborative centres

Fig: Reach of the International ECMOCARD collaborative

Queensland Government
Townsville Hospital and Health Service
CSIRO
Australian e-Health Research Centre

Operational Decision Support

The Goals

- Proactive management instead of reactive
- Data-driven instead of "this is how we do it"

The Research

- Time series forecasting and machine learning
- Simulation, Optimisation and mathematical models
- Digital twins

The Solutions we are Building

- Predicting demand for services and resources
- What-if scenario-based decision making
- Surveillance, outbreak & aberrance detection
- Optimising outpatient and surgery workflows
- Digital twin of a statewide patient flow control room

Austin HEALTH

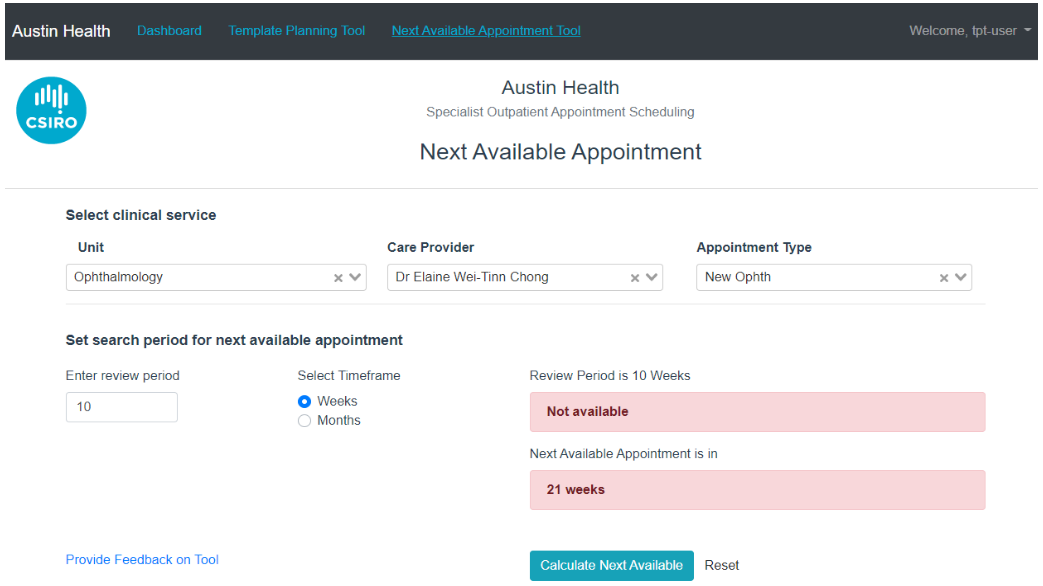


Fig: Next Available Appointment tool developed for trial at Austin Health

Queensland Government

Townsville Hospital and Health Service

CSIRO

Australian e-Health Research Centre

Clinical Decision Support

The Goals

- Precision decision support at the point of care
- Predictions that clinicians can trust/use

The Research

- High-performance explainable machine learning
- Predict "patients risk profile" based on outcomes of interest (generally adverse) to inform care planning
- Compliance with Regulatory requirements and internal Quality Management Systems (QMS) efforts

The Outcomes we are Predicting

- Preventable hospitalisation in primary care
- Preventable hospitalisations in acute care
- Clinical deterioration (Adults)
- Neonatal adverse outcomes including Sepsis
- Central line associated bloodstream infection (CLABSI)
- Non-alcoholic Fatty Liver Disease (NAFLD)
- Post-operative Hypotension

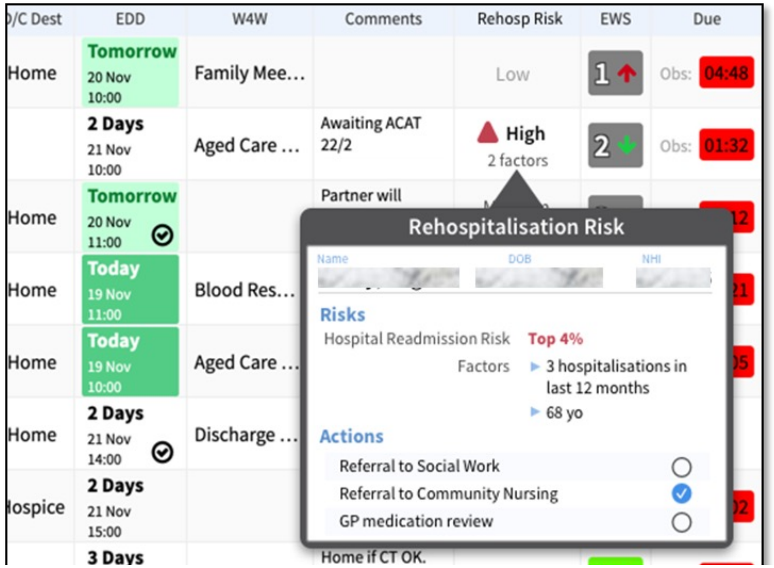


Fig: CSIRO rehospitalisation risk algorithm embedded within Alcidion's Miya flow dashboard

Fig: CSIRO's hospitalisation risk algorithm embedded within Pen CS's population health portal

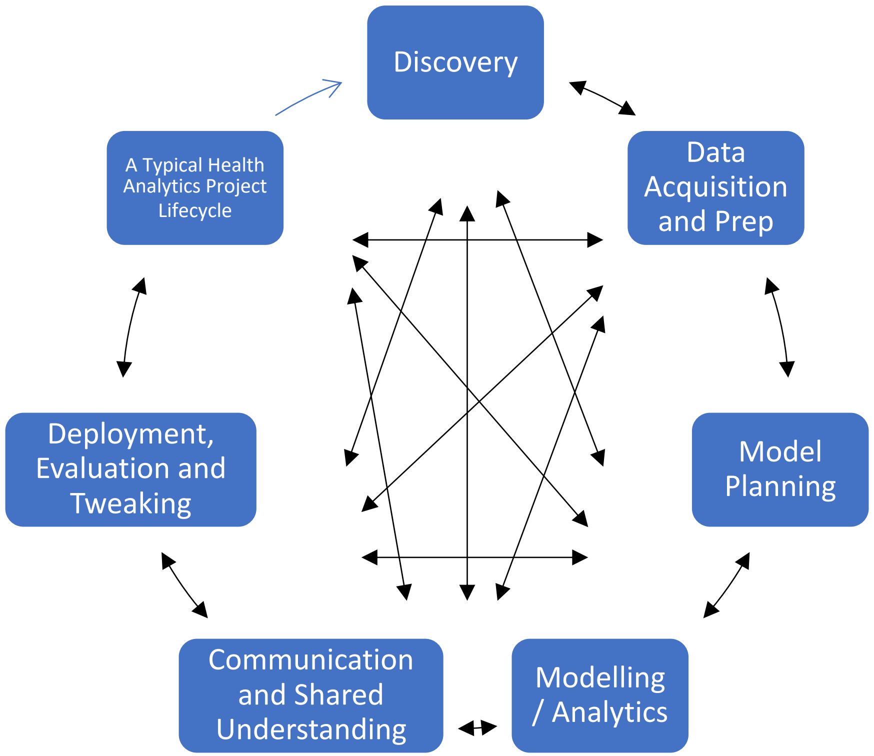
Queensland Government

Townsville Hospital and Health Service

CSIRO

Australian e-Health Research Centre

A Typical Health Analytics Project Lifecycle



Queensland Government
Townsville Hospital and Health Service
CSIRO
Australian e-Health Research Centre

Barriers to Development, Deployment & Scalability

Barriers:

- Data quality issues
- Poor structural interoperability
- Poor semantic interoperability

Consequence:

- 80% time spent on getting the data ready
- Bespoke model development – i.e. build for the specific use case
- Poor transferability/scalability of developed solutions

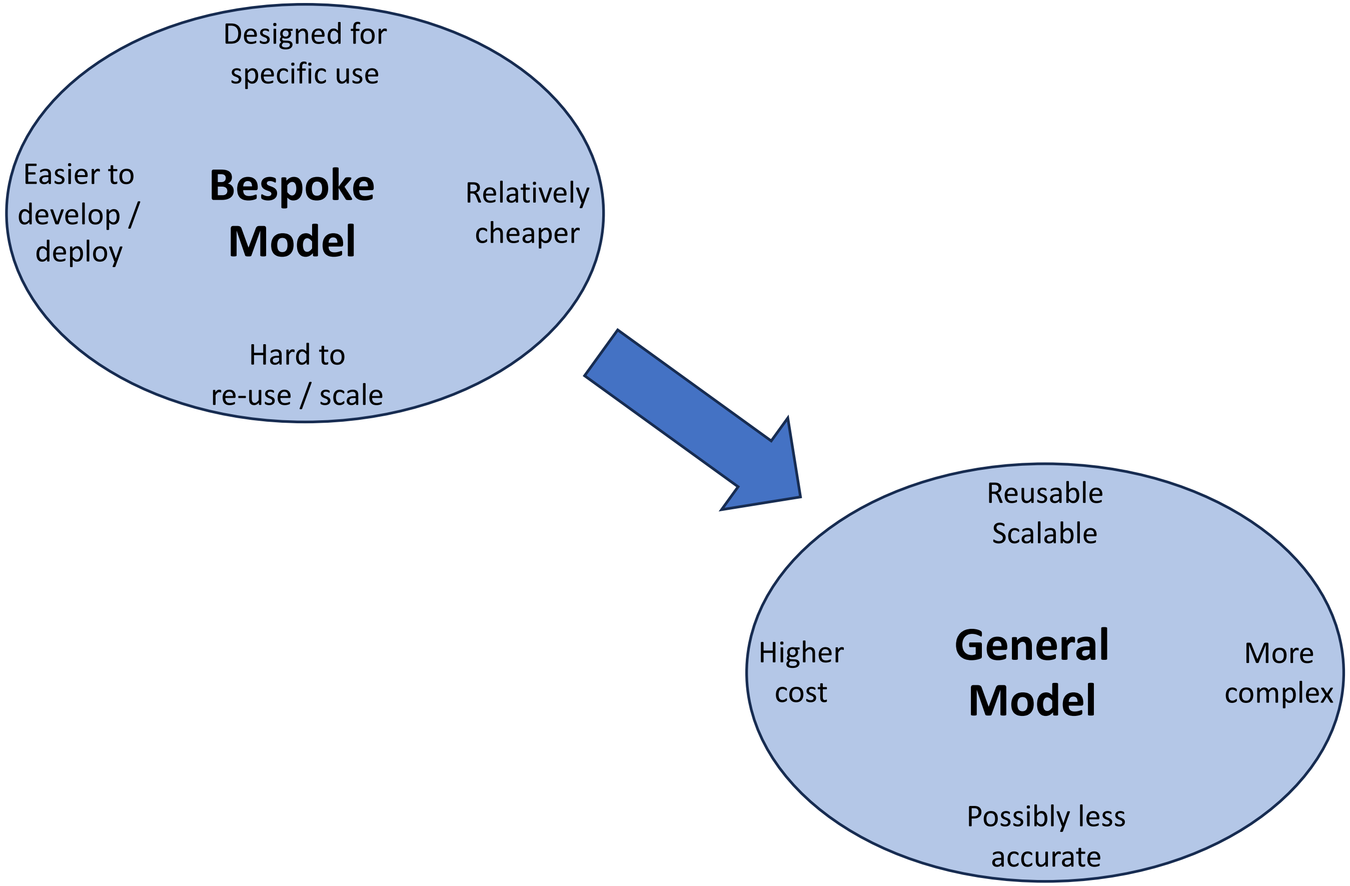
DOB: 02/12/21

Weight: 12.6

Medicine Name: Pyrmont Pharmacy, 4b/60 Union St, Pyrmont NSW 2009

Best Practice Analytics and Model Development

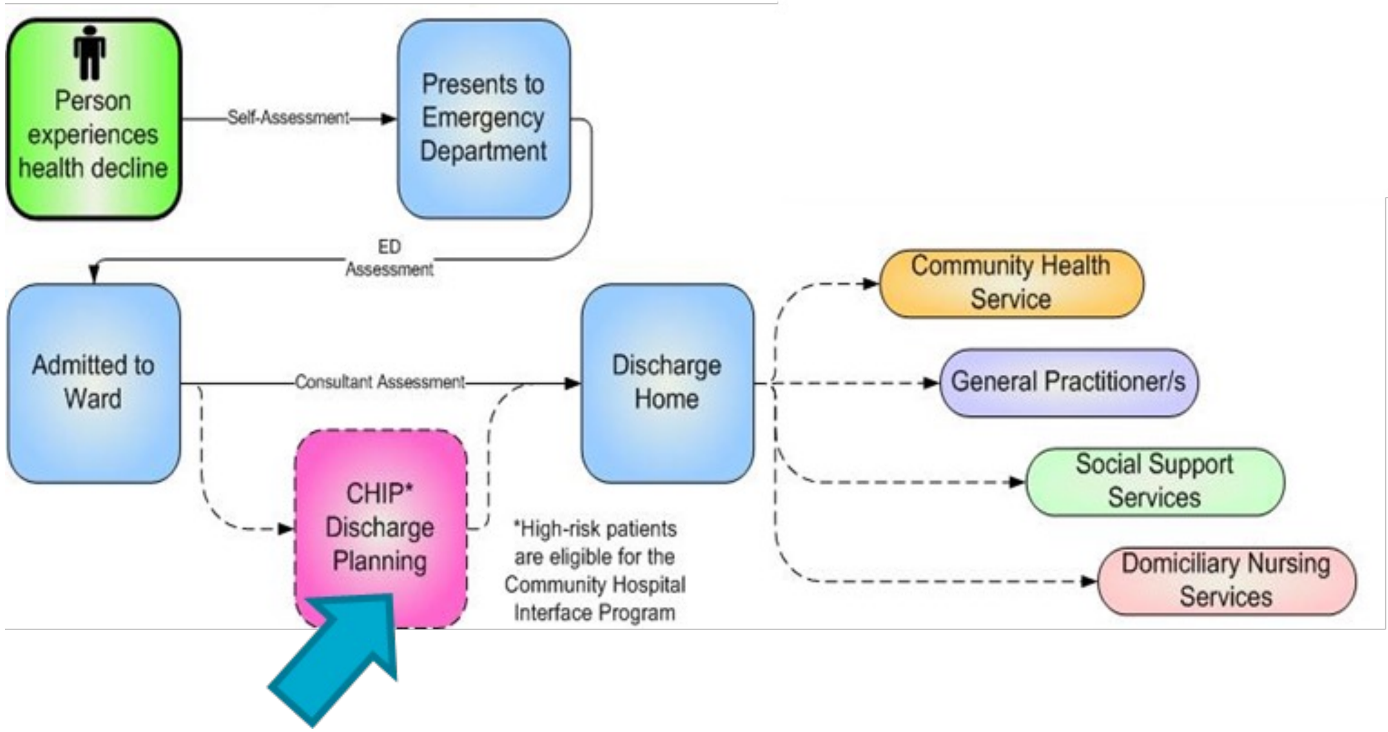
- Current efforts are built for specific use – hard to re-use/scale
- You want models that are re-usable and scalable
- Solution – a standards-based approach



Queensland Government
Townsville Hospital and Health Service
CSIRO
Australian e-Health Research Centre

Use Case: Clinical Decision Support – Acute Care

- Predictive Algorithm Driven Risk Stratification to inform in-hospital care and discharge planning
- 24 Month trial at QLD Metropolitan Hospital - April 2018 – March 2020
- Explainable Machine Learning employed to help interpret risk scores
- Redeveloped on state-wide data for implementation in QLD – June 2021



Queensland Government
Townsville Hospital and Health Service
CSIRO
Australian e-Health Research Centre

Background Work in Risk Stratification

Patient Cohort Coverage

- 2 hospitals from a lower socio-economic area in QLD
- Include surrounding hospitals
- All QLD hospitals

Admissions of Interest
(Performance measured as Area under the ROC curve)

- All admissions : 80-95%
- All except dialysis admissions : 65-78%
- Emergency admissions : 50-68%
- What else do we remove ??

Response Variable

- 28 days Vs 30 days
- Readmission Vs Emergency Readmission Vs Representation to ED Vs Either

Diagnosis Code Block	Description
E11*	Type 2 Diabetes Mellitus
I25*	Chronic Ischaemic Heart Disease
I50*	Heart Failure
I60*	Subarachnoid Haemorhage
I61*	Intracerebral Haemorrhage
I62*	Other Nontraumatic Intracranial Haemorrhae
I63*	Cereral Infarction
I64*	Stroke, Not Specified as Haemorrhage or Infarction
J44*	Other Chronic Obstructive Pulmonary Disease
J45*	Asthma
J46*	Status Asthmaticus
N18*	Chronic Kidney Failure
Z49*	Care Involving Dialysis

Fig: List of ICD-10 diagnosis codes used to identify chronic disease patients

Queensland Government
Townsville Hospital and Health Service
CSIRO
Australian e-Health Research Centre

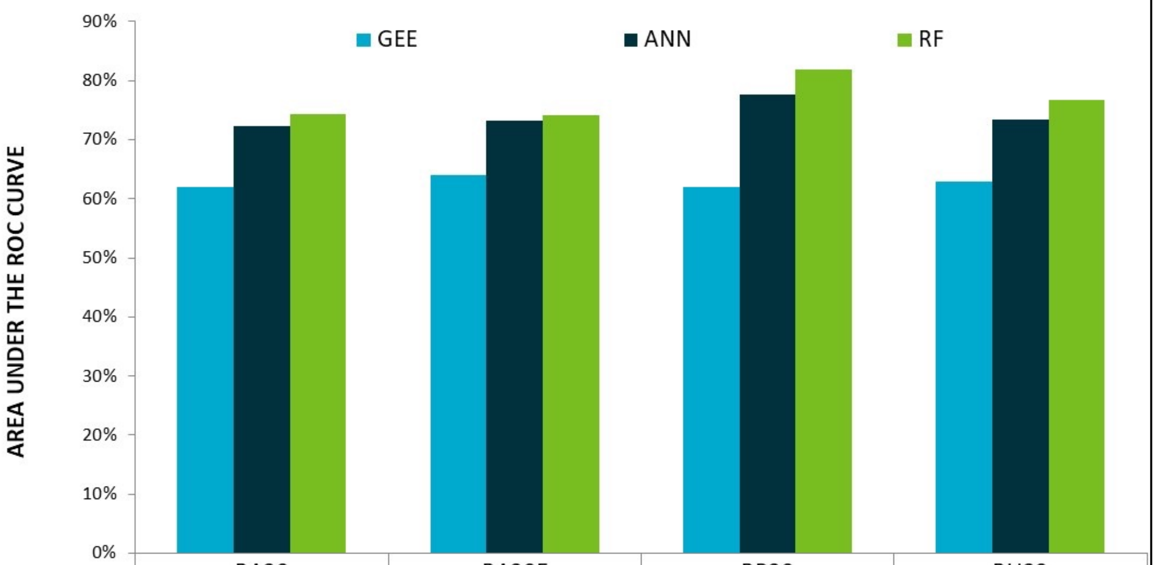
Background Work in Risk Stratification

Cohort : Statewide data for all patients who presented at the original 2 hospitals with at least one Chronic Disease admission over 5 years

Exclusions :
- Routine admissions
- Obstetric admissions
- Index admissions
- Episodes resulting in inpatient death

4 Response Variables :
- RA30 - Readmitted within 30 days
- RA30E - Readmitted within 30 days through ED
- RP30 - Represented to ED within 30 days
- RU30 - Return to hospital within 30 days

3 Algorithms
- Generalised Estimating Equations (GEE)
- Artificial Neural Networks (ANN)
- Random Forests (RF)



	RA30	RA30E	RP30	RU30
GEE	62%	64%	62%	63%
ANN	72%	73%	78%	74%
RF	74%	74%	82%	77%

Fig: Comparing model performance

S. Khanna, N. Good, and J. Boyle, “Predicting Unplanned Return to Hospital for Chronic Disease Patients,” *Studies in health technology and informatics*, vol. 227, pp. 67–73, 2016.

Queensland Government

Townsville Hospital and Health Service

CSIRO

Australian e-Health Research Centre

Next Steps : A Trial

What did we predict?

- Unplanned re-admission within 30 days of discharge from hospital
- Unplanned ED re-presentation within 30 days of discharge from hospital

Research Questions : Does the algorithm:

- Improve the process of identifying patients at high risk of unplanned re-hospitalisation?
- Reduce re-hospitalisation rates?
- Provide information to staff not readily available at the time of discharge planning?

Trial Phase 2	Apr 2019 to March 2020
Mid-Trial Surveys & Focus Group	During and at the end of trial
Post-Trial Evaluation	Apr 2020 to Jun 2020

Chronic Disease Patient Admitted to Hospital

↓

Risk Score generated overnight

↓

Risk score used by care teams for appropriate interventions and care/discharge planning

Queensland Government

Townsville Hospital and Health Service

CSIRO

Australian e-Health Research Centre

Web-based Decision Support Tool

RISK Overview Tools

Overview

Today All | Search | Export

Patient	UR	Admission Date	Last Discharge Date	Age	Ward	ED 31 days	LOS 180 days	Readmit RISK	ED RISK
		09 May 2019	02 May 2019	22	2B	2	80	Top 5%	Top 5%
		19 May 2019	15 May 2019	67	3C	4	6	Top 5%	Top 5%
		19 May 2019	04 May 2019	69	3A	7	16	Top 5%	Top 5%
		19 May 2019	07 May 2019	64	3A	3	19	Top 10%	Top 5%
		20 May 2019	15 May 2019	54	EDCDU	3	25	Top 10%	Top 10%
		20 May 2019	13 May 2019	59	3C	3	37	Top 10%	Top 5%
		20 May 2019	11 May 2019	53	MAPU	3	11	Top 10%	Top 5%
		20 May 2019	30 Apr 2019	69	3B	2	16	Top 10%	Top 20%
		20 May 2019	04 Mar 2019	39	2K	2	14	Top 20%	Top 5%
		08 May 2019	14 Apr 2019	17	2A	2	21	Top 20%	Top 10%
		19 May 2019	10 May 2019	91	3C	3	10	Top 20%	Top 10%
		20 May 2019	19 May 2019	22	EDCDU	1	10	Top 20%	Top 10%
		19 May 2019		80	AMU	2	18	Top 20%	Top 40%
		20 May 2019	29 Apr 2019	54	EDCDU	3	4	Top 20%	Top 5%
		20 May 2019	05 May 2019	28	EDSSW	6	9	Top 20%	Top 10%
		18 May 2019	16 May 2019	30	2D	3	2	Top 20%	Bottom 50%
		16 May 2019		39	2C	0	20	Top 20%	Top 20%
		13 May 2019	10 May 2019	49	2C	4	68	Top 20%	Top 20%
		11 May 2019	02 May 2019	15	2A	2	56	Top 20%	Top 20%
		16 May 2019	28 Mar 2019	28	3C	1	1	Top 20%	Top 20%

<< < 1 2 3 4 5 6 7 8 9 > >>

Queensland Government

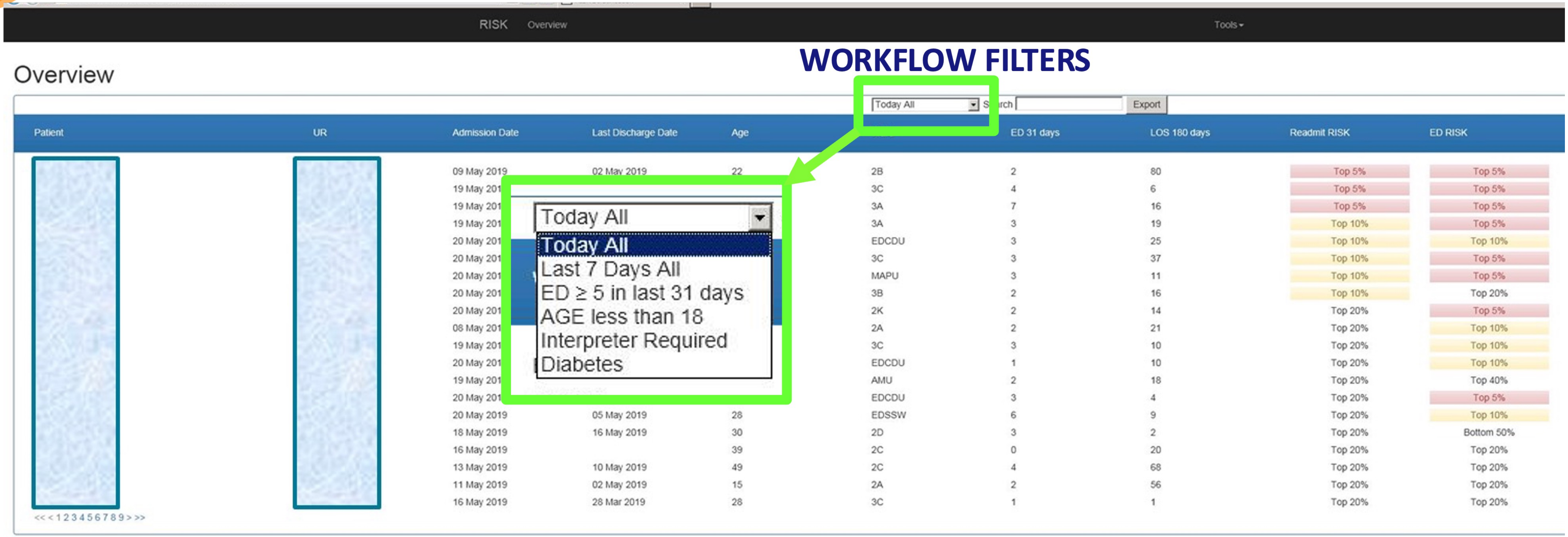
Townsville Hospital and Health Service

CSIRO

Australian e-Health Research Centre

Web-based Decision Support Tool

WORKFLOW FILTERS



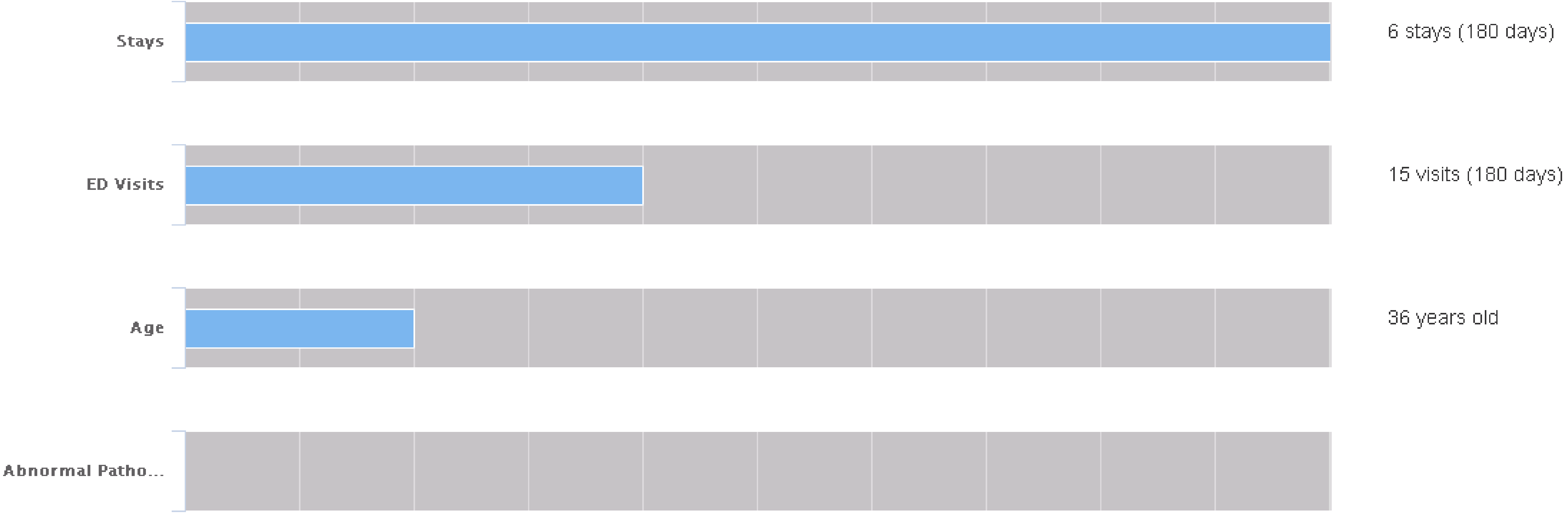
Queensland Government

Townsville Hospital and Health Service

CSIRO

Australian e-Health Research Centre

Supporting Clinician Decision Making

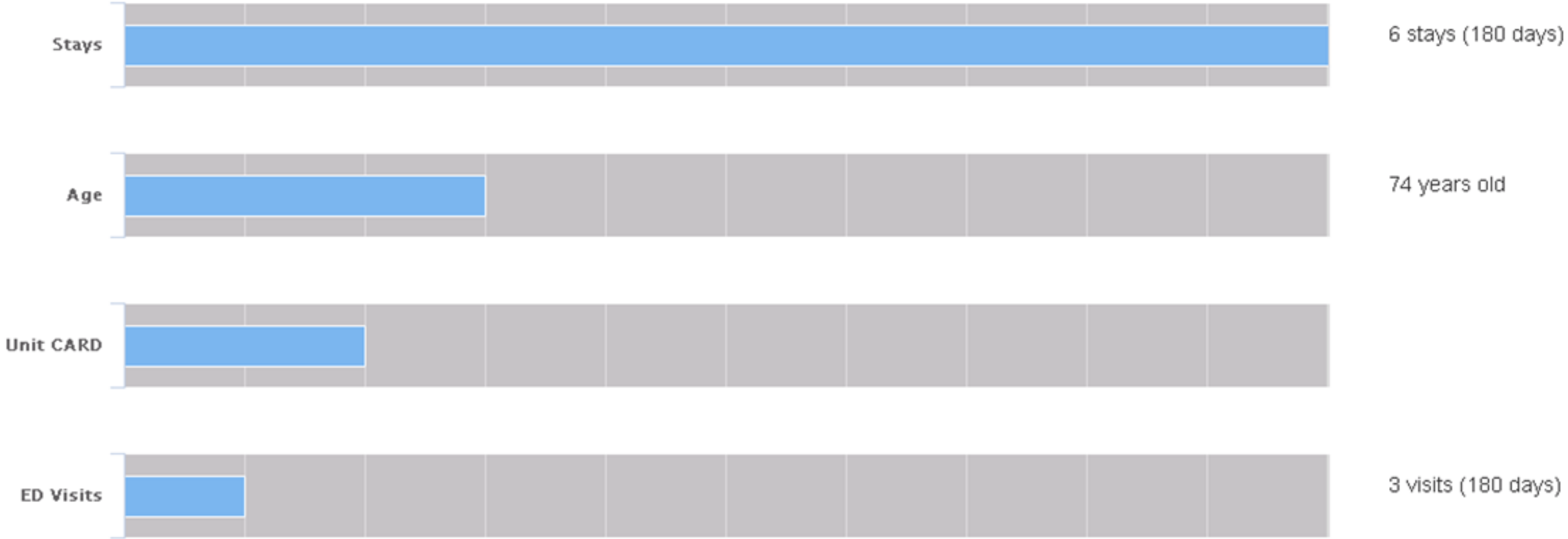


Queensland Government

Townsville Hospital and Health Service

CSIRO Australian e-Health Research Centre

Supporting Clinician Decision Making



Queensland Government

Townsville Hospital and Health Service

CSIRO

Australian e-Health Research Centre

Next Step: Statewide Validation

Cohort: Patients with at least one chronic disease ED presentation / hospital admission at any Queensland public hospital over 5 years.

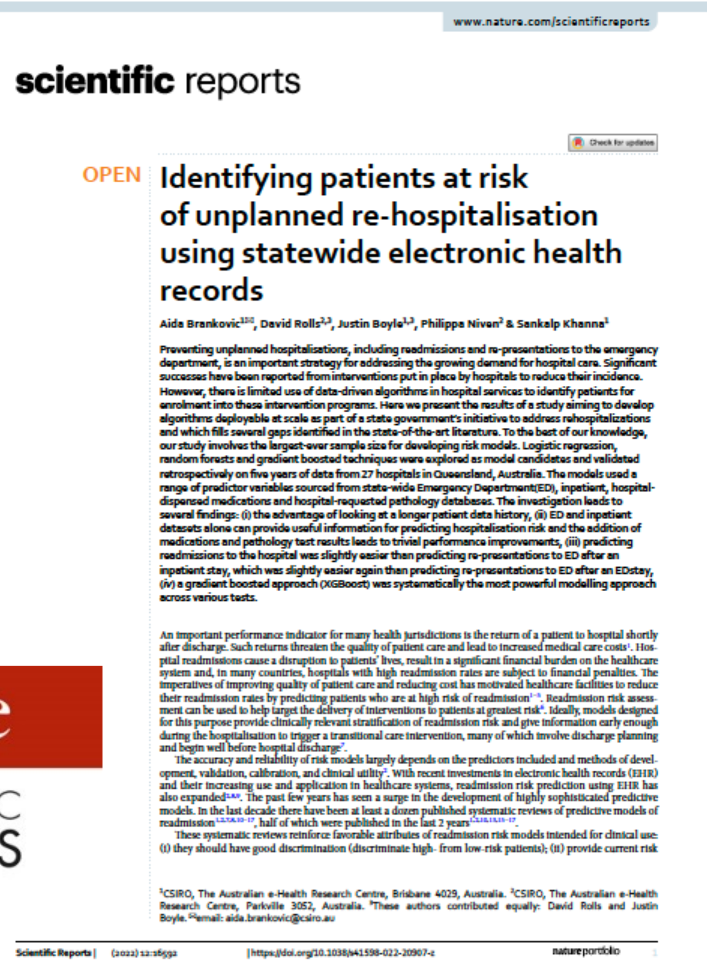
Data Employed: Emergency Presentations (EDC), Inpatient Admissions (ePADT), Death data (death registry), In-hospital Pathology (Auslab), In-hospital Medications (eLMS)

Unplanned Hospitalisation metrics of interest:

- 30-day unplanned readmission (RA30)
- 30-day unplanned ED re-presentation (RP30)
- 30-day unplanned ED re-presentation (ED discharge) (RP30E)

Hospital peer groups

- Principal referral
- Childrens hospital
- Public acute



www.nature.com/scientificreports

scientific reports

OPEN

Identifying patients at risk of unplanned re-hospitalisation using statewide electronic health records

Aida Brankovic[1✉], David Rolls[2,3], Justin Boyle[1,3], Philippa Niven[2] & Sankalp Khanna[1]

Preventing unplanned hospitalisations, including readmissions and re-presentations to the emergency department, is an important strategy for addressing the growing demand for hospital care. Significant successes have been reported from interventions put in place by hospitals to reduce their incidence. However, there is limited use of data-driven algorithms in hospital services to identify patients for enrolment into these intervention programs. Here we present the results of a study aiming to develop algorithms deployable at scale as part of a state government's initiative to address rehospitalizations and which fills several gaps identified in the state-of-the-art literature. To the best of our knowledge, our study involves the largest-ever sample size for developing risk models. Logistic regression, random forests and gradient boosted techniques were explored as model candidates and validated retrospectively on five years of data from 27 hospitals in Queensland, Australia. The models used a range of predictor variables sourced from state-wide Emergency Department(ED), inpatient, hospital-dispensed medications and hospital-requested pathology databases. The investigation leads to several findings: (i) the advantage of looking at a longer patient data history, (ii) ED and inpatient datasets alone can provide useful information for predicting hospitalisation risk and the addition of medications and pathology test results leads to trivial performance improvements, (iii) predicting readmissions to the hospital was slightly easier than predicting re-presentations to ED after an inpatient stay, which was slightly easier again than predicting re-presentations to ED after an EDstay, (iv) a gradient boosted approach (XGBoost) was systematically the most powerful modelling approach across various tests.

An important performance indicator for many health jurisdictions is the return of a patient to hospital shortly after discharge. Such returns threaten the quality of patient care and lead to increased medical care costs[1]. Hospital readmissions cause a disruption to patients' lives, result in a significant financial burden on the healthcare system and, in many countries, hospitals with high readmission rates are subject to financial penalties. The imperatives of improving quality of patient care and reducing cost has motivated healthcare facilities to reduce their readmission rates by predicting patients who are at high risk of readmission[1–5]. Readmission risk assessment can be used to help target the delivery of interventions to patients at greatest risk[6]. Ideally, models designed for this purpose provide clinically relevant stratification of readmission risk and give information early enough during the hospitalisation to trigger a transitional care intervention, many of which involve discharge planning and begin well before hospital discharge[7].

The accuracy and reliability of risk models largely depends on the predictors included and methods of development, validation, calibration, and clinical utility[3]. With recent investments in electronic health records (EHR) and their increasing use and application in healthcare systems, readmission risk prediction using EHR has also expanded[2,8,9]. The past few years has seen a surge in the development of highly sophisticated predictive models. In the last decade there have been at least a dozen published systematic reviews of predictive models of readmission[1,2,7,8,10–17], half of which were published in the last 2 years[1,2,10,13,15–17].

These systematic reviews reinforce favorable attributes of readmission risk models intended for clinical use: (i) they should have good discrimination (discriminate high- from low-risk patients); (ii) provide current risk

[1]CSIRO, The Australian e-Health Research Centre, Brisbane 4029, Australia. [2]CSIRO, The Australian e-Health Research Centre, Parkville 3052, Australia. [3]These authors contributed equally: David Rolls and Justin Boyle. ✉email: aida.brankovic@csiro.au

Scientific Reports | (2022) 12:16592 | https://doi.org/10.1038/s41598-022-20907-z nature portfolio

Brankovic A, Rolls D, Boyle J, Niven P, Khanna S. Identifying patients at risk of unplanned re-hospitalisation using statewide electronic health records. Sci Rep. 2022 Oct 5;12(1):16592.

In Summary: The Path to Value/Impact

- **E**ngage and codesign with stakeholders
- **F**ocus on understanding translational needs early on
- **F**rameworks, Standards and Governance
- **E**nsure statistical rigor and account for challenges
- **C**hoose the right outcome measures
- **T**ackle social, ethical and regulatory matters
- **I**nnovation can help resolve traditional challenges
- **V**alue of understanding the domain and data
- **E**mpathise with pain points of problem owners

Health Translation Queensland

Questions ?

Rudolf Schnetler - Rudolf.Schnetler@health.qld.gov.au
Townsville Institute of Health Research and Innovation

Sankalp Khanna – Sankalp.Khanna@csiro.au
CSIRO Australian e-Health Research Centre