MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Queen Mary
University of London

Challenges in Health Informatics:

Knowledge, AI and Data Science Perspectives

Seminar Roadmap

Session Outline

- Presenter Introductions
- Introduction & background
- **Research Areas:** Our work in Health Informatics

Presenter Introductions

Dr Kuda Dube

Dr Scott McLachlan

Dr Evangelia Kyrimi

Introduction and background

Pattern Recognition

The Problem

- Humans are good at pattern recognition
- Many approaches to Machine Learning in healthcare are based on simple pattern recognition
- Pattern recognition can easily identify when something is or isn't the desired target

But…

We need to go beyond simple pattern recognition in order to realise the goals of Learning Health Systems and Precision Medicine

ANSWER: (from top to bottom) green clew; cat; yellow sample; purple clew; green sample; cat looking left; yellow clew.

Health Data

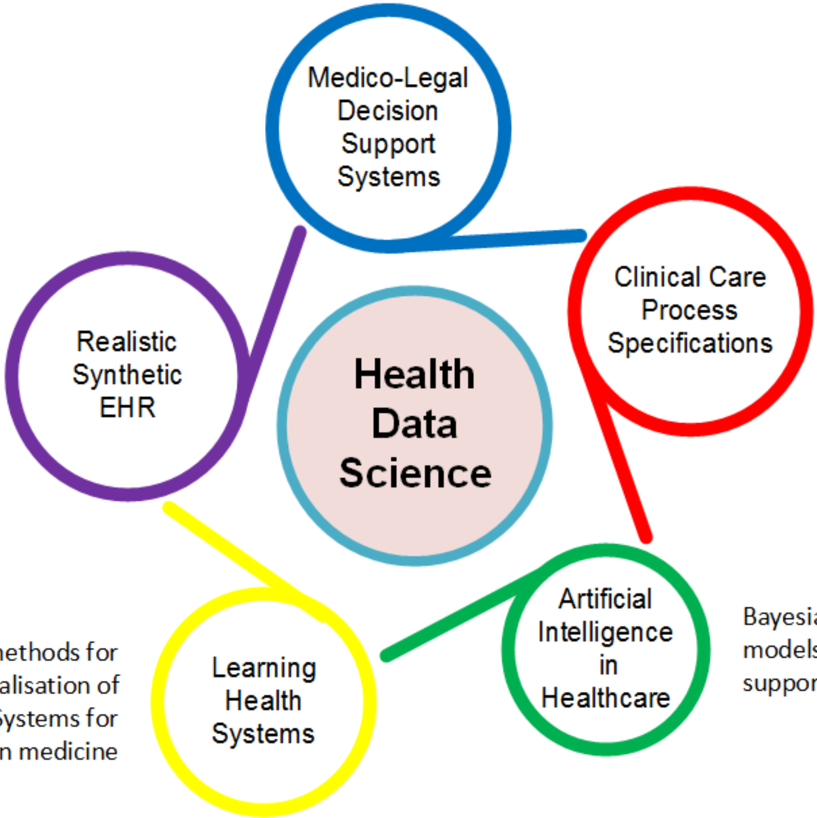
The Health Record

- Predate the personal computer
- Most personal and highly sensitive type of data: *privacy laws and ethical issues*
- Adoption of Electronic Healthcare Record (EHR) has been slow
- Secondary use of EHR for research and development is still problematic
- EHRs still have many unresolved issues
- Opportunities to capture and link other data, such as physiometric data from sensors through the Medical Internet of Things (MIoT) paradigm, exist but mostly remain unexplored

While EHR continue to lack standardisation in form, function and how the data is captured and stored, it is difficult for them to support Learning Health Systems and Precision Medicine

Overview of Research Framework

Our Health Data Science Related Research



Models, methods and mechanisms for supporting legal decision-making in healthcare

Documentation standards for clinical care processes based on best practice and scientific evidence

Knowledge-intensive methods and algorithms for generating realistic synthetic EHR for secondary uses (including research)

Bayesian network-based models for clinical decision support

Models and methods for facilitating the realisation of Learning Health Systems for supporting precision medicine

Research Areas

Dr Kuda Dube

Realistic Synthetic Electronic Health Records

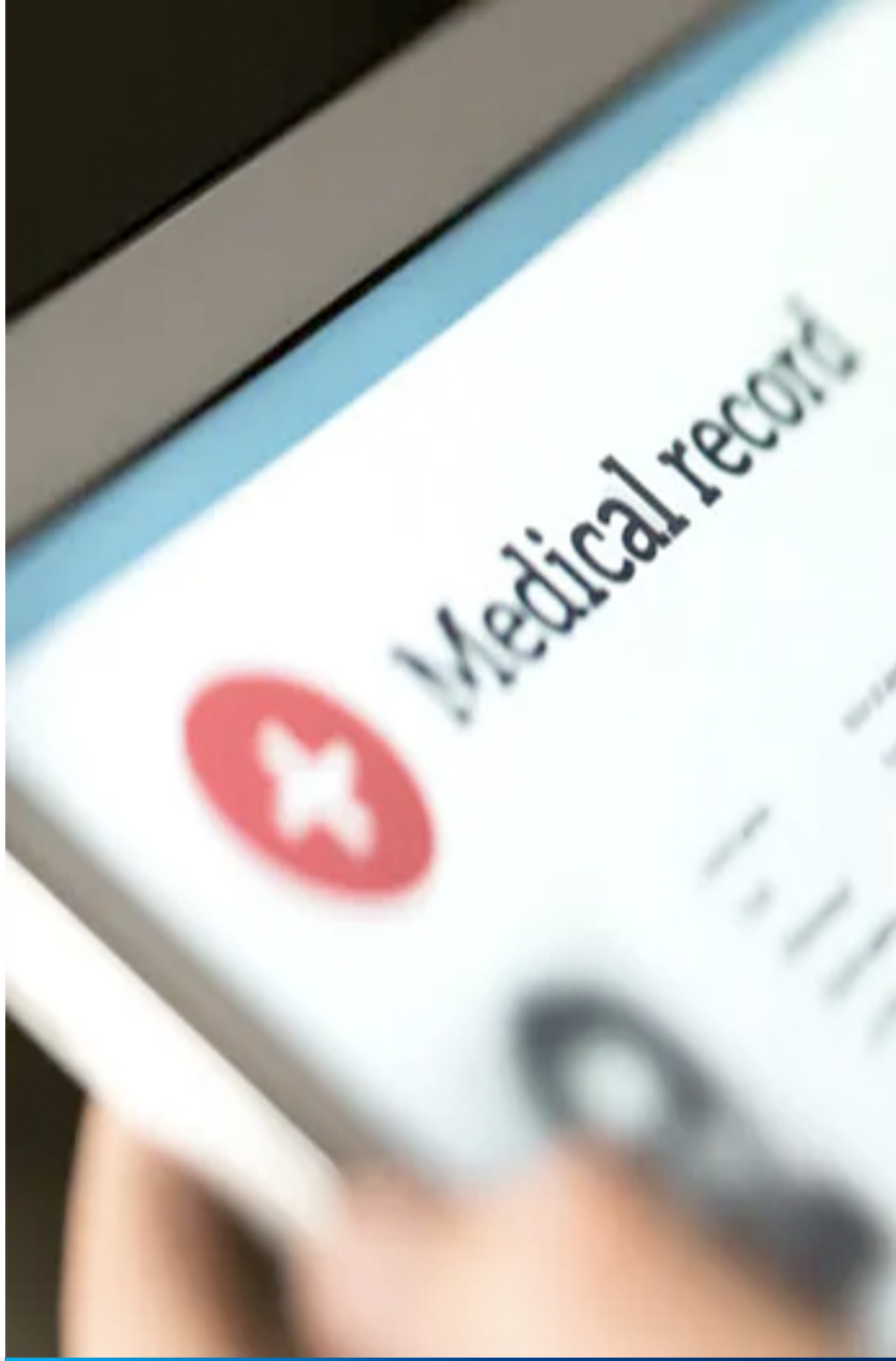
Knowledge-intensive methods and algorithms for generating the "realistic synthetic" EHR

Key research challenges

- Generation of synthetic patient avatars for use as control cohorts in development and simulation;
- Generation of "realistic" synthetic EHR: *birth-to-grave, birth-to-current age*
- Validation of "realism" in synthetic EHR

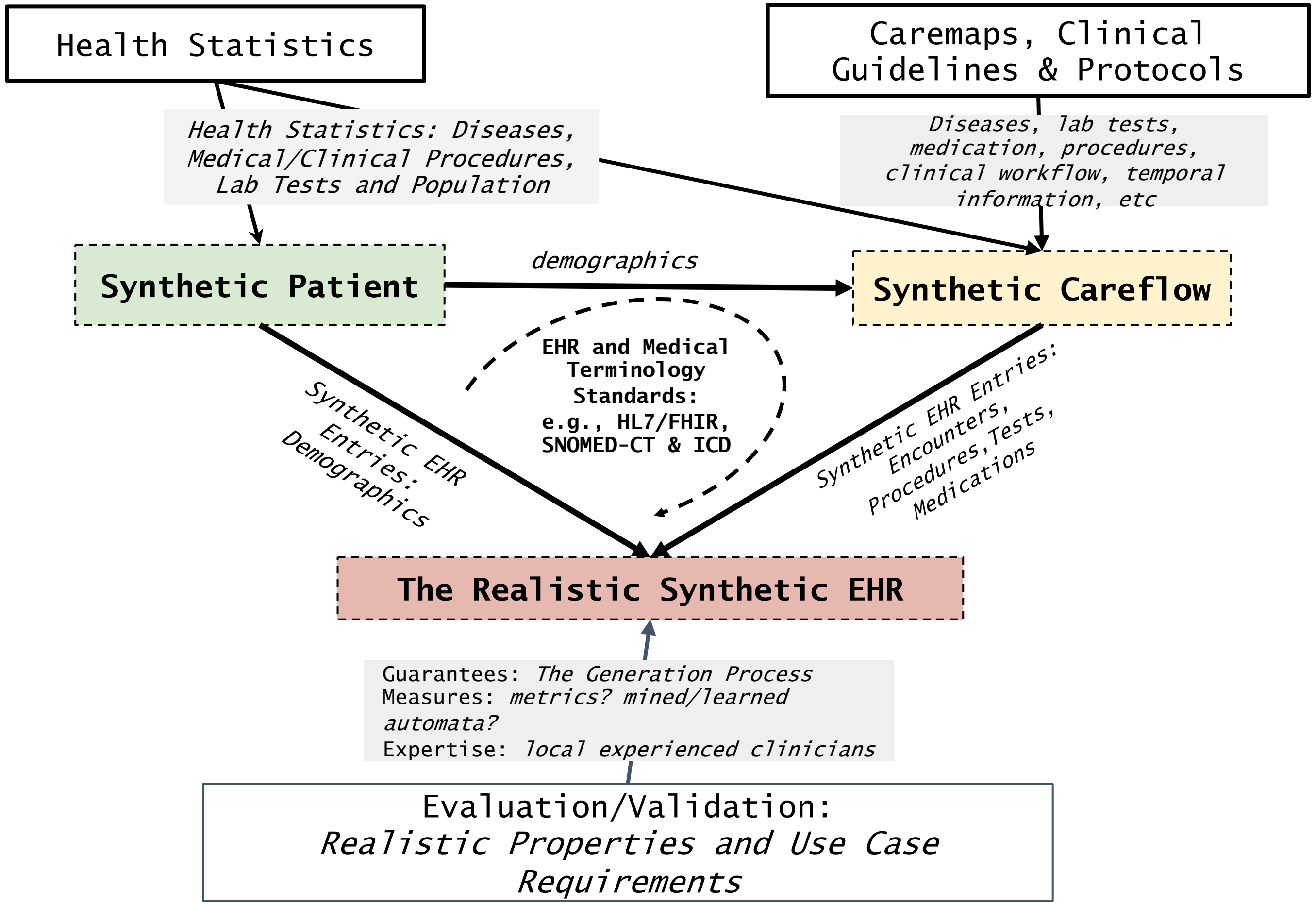
Work Done to Date

- The CoMSER Method for generating the RS-HER
- The ATEN Framework for realism in RS-EHR
- Contributed to **Synthea** with Mitre Corporation, State of Massachusetts, USA



The Realistic Synthetic Patient and EHR

Knowledge-intensive Approach and Method



Health Statistics

Caremaps, Clinical Guidelines & Protocols

Health Statistics: Diseases, Medical/Clinical Procedures, Lab Tests and Population

Diseases, lab tests, medication, procedures, clinical workflow, temporal information, etc

Synthetic Patient

demographics

Synthetic Careflow

EHR and Medical Terminology Standards: e.g., HL7/FHIR, SNOMED-CT & ICD

Synthetic EHR Entries: Demographics

Synthetic EHR Entries: Encounters, Procedures, Tests, Medications

The Realistic Synthetic EHR

Guarantees: *The Generation Process*
Measures: *metrics? mined/learned automata?*
Expertise: *local experienced clinicians*

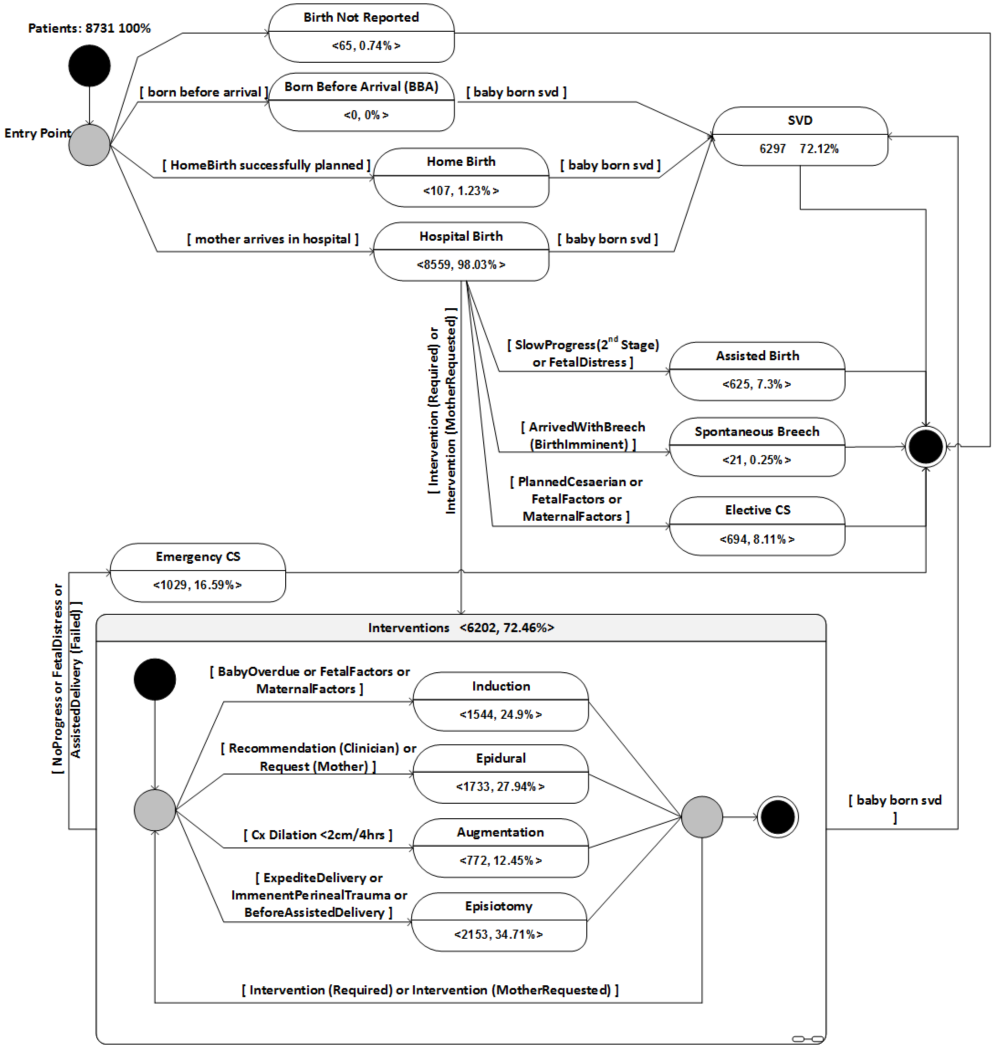
Evaluation/Validation:
Realistic Properties and Use Case Requirements

Practical computational challenge:

Generate a synthetic patient population and its synthetic EHR equivalents to those of the entire state of Massachusetts.

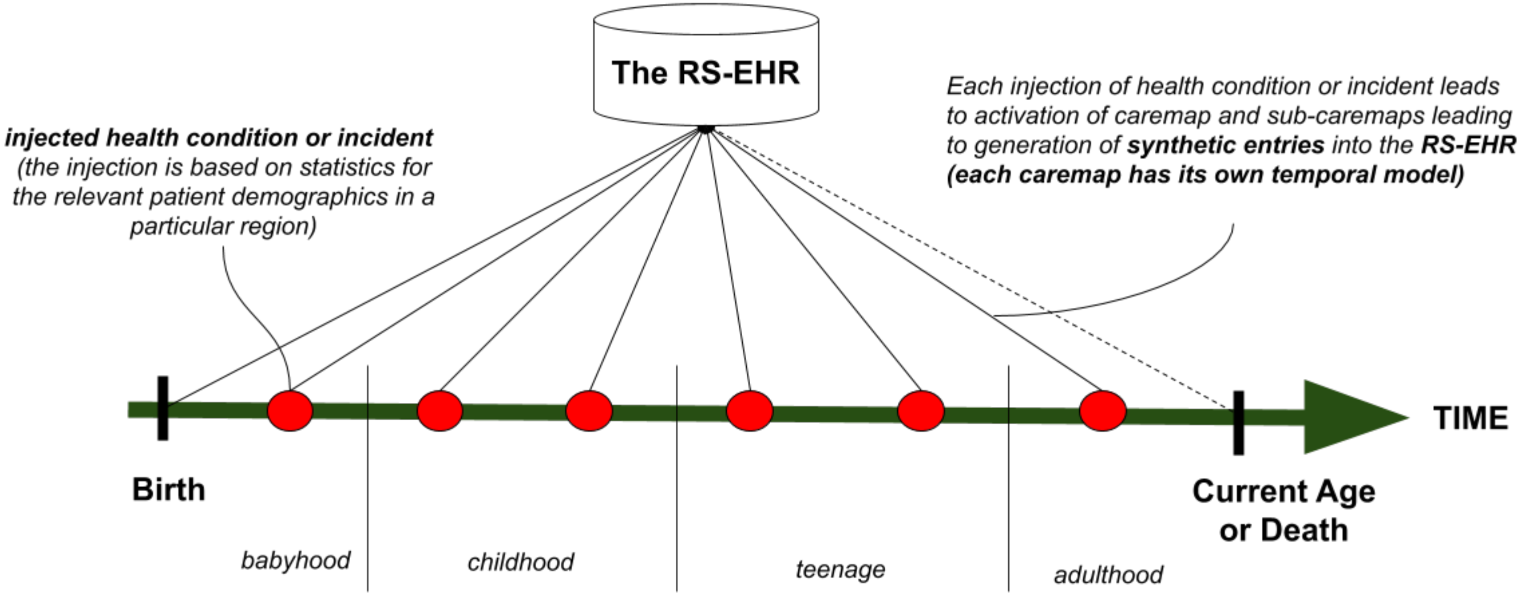
The RS-EHR: *The CoMSER Proof-of-Concepts System*

The Synthetic EHR for the Labour and Births Events



The RS-EHR: Current and ongoing challenge

Computational challenges for generating **birth-to-present** or **birth-to-current-age** RS-EHR



lifetime phases for the synthetic patient
(these phases should conform to phases found in national/international health statistics)

Issues to grapple with:

1. Complex medical reasoning, logic and variables;
2. Data and knowledge-intensive – non-existent required statistical data;
3. Computationally intensive as required population increase;
4. Ensuring “realism” in result RS-EHR

Language Technologies for Healthcare

Research Challenges

- Extend use of NLP tools for indigenous languages to the healthcare setting
- Development of new language models, algorithms and NLP tools that will exploit these to support Bantu languages of Southern Africa and similar languages around the world

Patient-clinician encounters in non-English language contexts:

The case for adaptive agglutinative languages of Southern Africa

Language Technologies for Healthcare

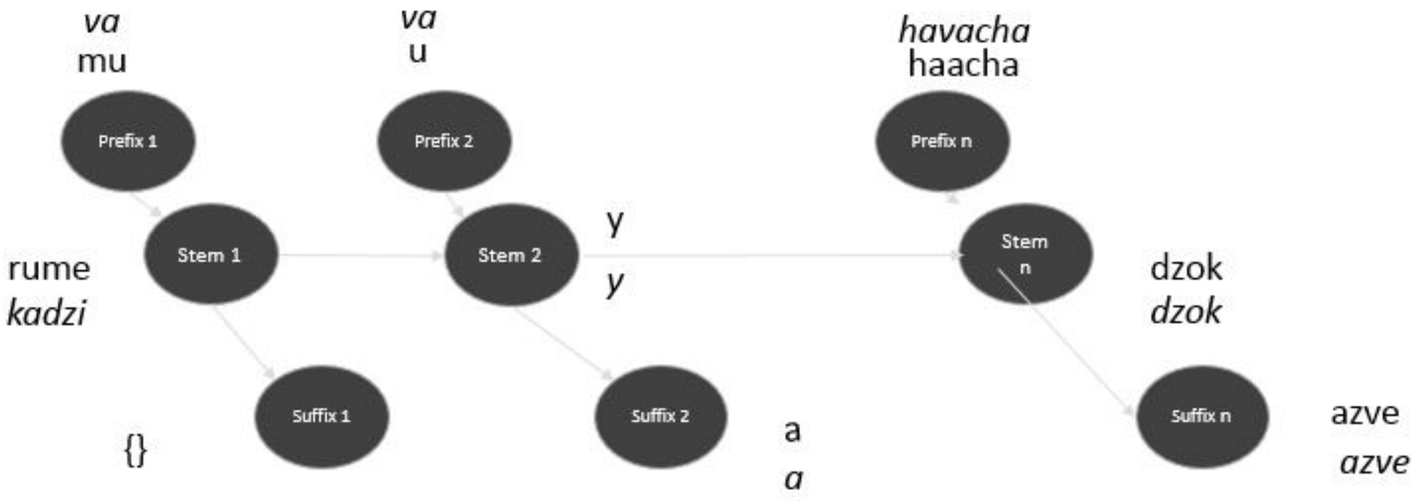
The case for Adaptive Agglutinative Languages of Southern Africa (Bantu)

- Communication and Language: patient-clinician encounter and the EHR;
- Problem: Existing language models and computational tools cannot be used to perform even simple tasks like checking spelling and grammatical analysis.
- Example consequence: In Zimbabwe, clinicians speak in Shona to patients, but the Health Record is in English

Patient-clinician encounters in non-English language contexts:

The case for adaptive agglutinative languages of Southern Africa

Example: *Simple Shona Sentence and Words*



Murume uya haachadzokazve
Vakadzi vaya havachadzokazve

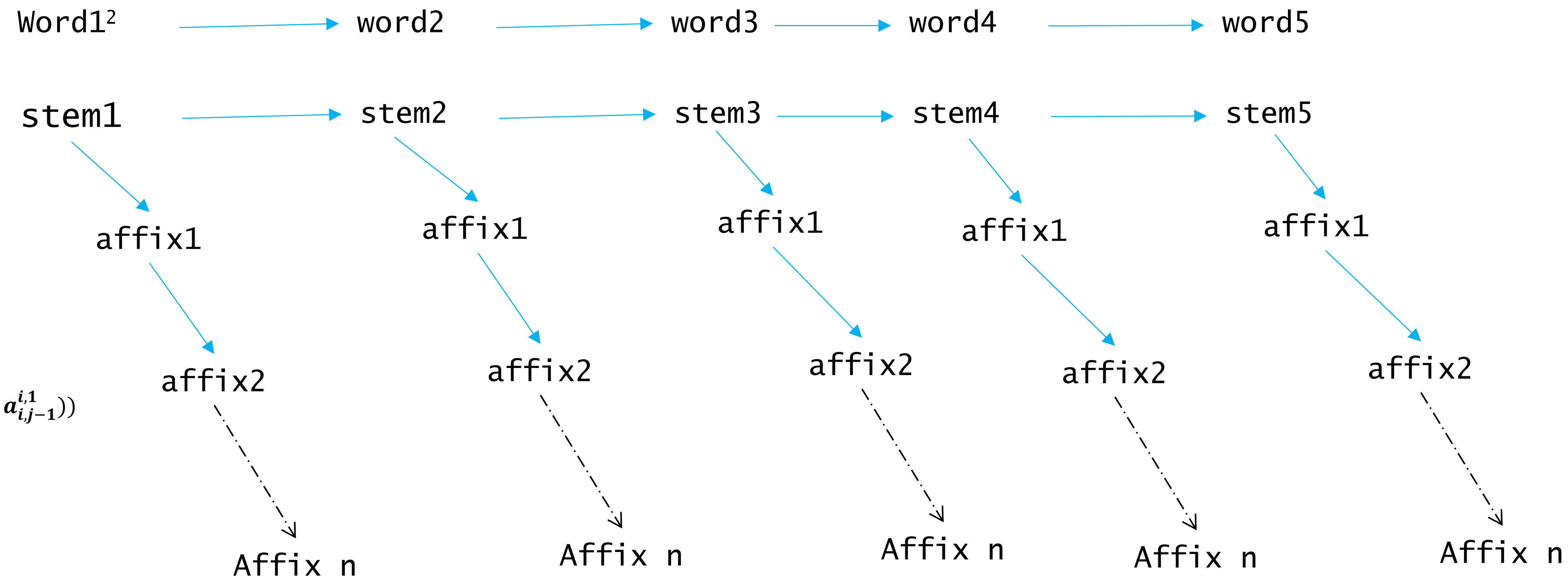
Spell-checking, syntactic and semantic analysis is not easy and require new methods to achieve functional accuracy.

Complex syntactic composition of Shona words

Main Approaches to Language Models

Linear Models[1]

"Graphical" Models[3]



Word1[2] word2 word3 word4 word5

stem1 stem2 stem3 stem4 stem5

affix1 affix1 affix1 affix1 affix1

affix2 affix2 affix2 affix2 affix2

Affix n Affix n Affix n Affix n Affix n

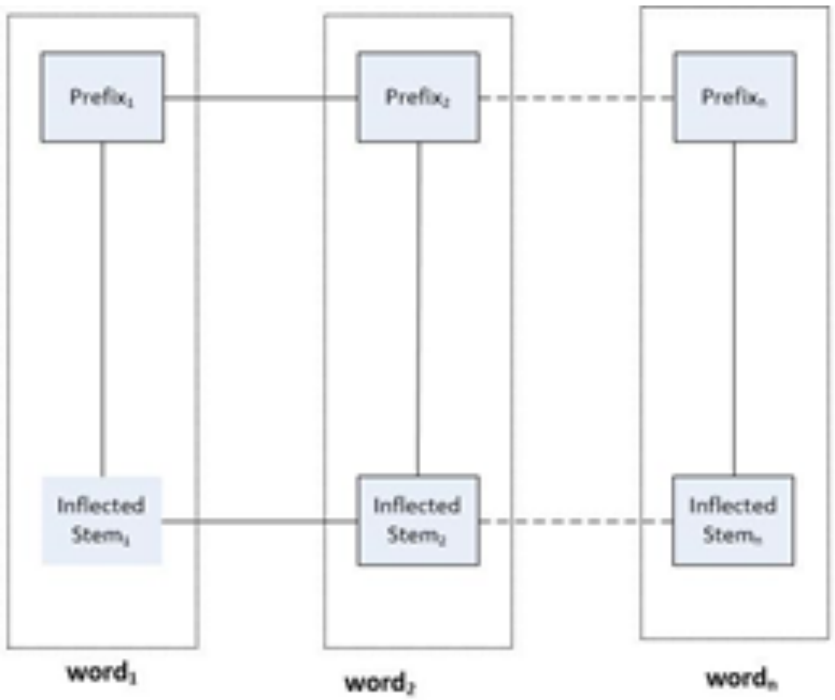
$$P(w_n^1) = \prod_i (P(s_i|s_{i-1}^{i-n}) \prod_j P(a_{i,j}|s_i, a_{i,j-1}^{i,1}))$$

Neural Models[4] Various architectures – mainly RNN, more specifically LSTMs

1: Most common definition of LMs. See Jurasky and Martin, 2018, for example
2: words can be replaced by syllables or characters in this model
3: See Xuehelaiti et al, 2013 for details of this approach
4. See Dengliang Shi, 2017

Our early experiments: N-gram methods further improved by smoothing and interpolation

Way Forward



$$P(\text{word}_n) = P(\text{prefix}_n|\text{prefix}_{n-1}).P(\text{infl. stem}_n|\text{infl. stem}_{n-1})$$

Proposed Language Model:

- Instead of modelling Shona language sentences at the word level, we do it at the sub-word level.

Word Analysis:

- Word as a composite of a prefix and an inflected stem (stem + sufixes),
- Inflected stem is the root and the derivational suffix.

Knowledge Incorporation

Challenges from how knowledge is incorporated into computational models

Even well-engineered solution models incorporate knowledge, here is simple example:

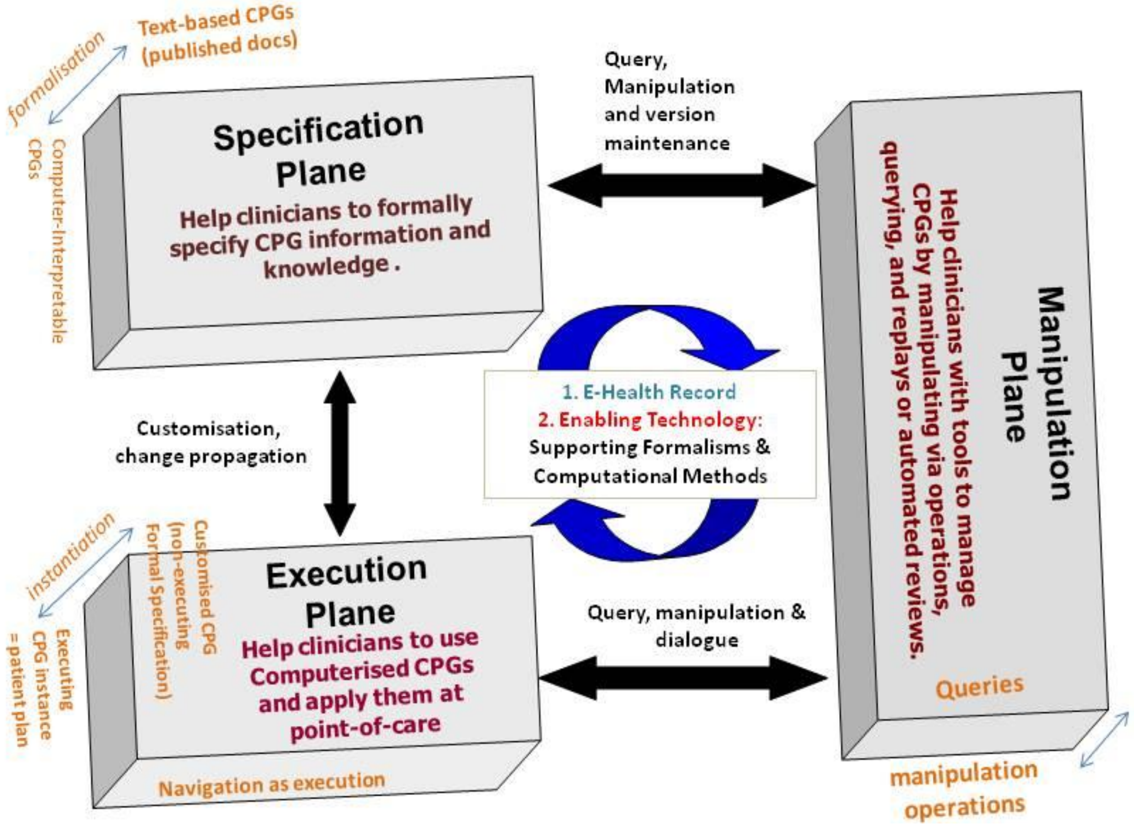
Simple Example: Generic *sorting algorithmic processes* incorporate knowledge about objects to be sorted: *precedence relations*

Complex Example: Health monitoring system that generate alerts incorporates clinical care and medical knowledge about managing the condition being monitored and the patient.

Key challenges:

1. To what extent can we examine/query and modify/update the knowledge that is incorporated in Health IT?
2. To what extent can we customize the resulting Health IT to suit different patients, circumstances, cultures, and regions of the world?

Knowledge Incorporation



Text-based CPGs (published docs)

formalisation

Computer-Interpretable CPGs

Specification Plane

Help clinicians to formally specify CPG information and knowledge .

Query, Manipulation and version maintenance

Manipulation Plane

Help clinicians with tools to manage CPGs by manipulating via operations, querying, and replays or automated reviews.

Queries

manipulation operations

Customisation, change propagation

1. E-Health Record

2. Enabling Technology: Supporting Formalisms & Computational Methods

instantiation

Customised CPG (non-executing Formal Specification)

Executing CPG instance = patient plan

Execution Plane

Help clinicians to use Computerised CPGs and apply them at point-of-care

Navigation as execution

Query, manipulation & dialogue

Conceptual framework for our work on:

1. Food, nutrition and lifestyle decision-support
2. Clinical guideline modelling and computerisation
3. Clinical Care Process Specifications *(Dr Scott McLachlan)*

Research Areas

Dr Scott McLachlan

Research Interests (by domain)

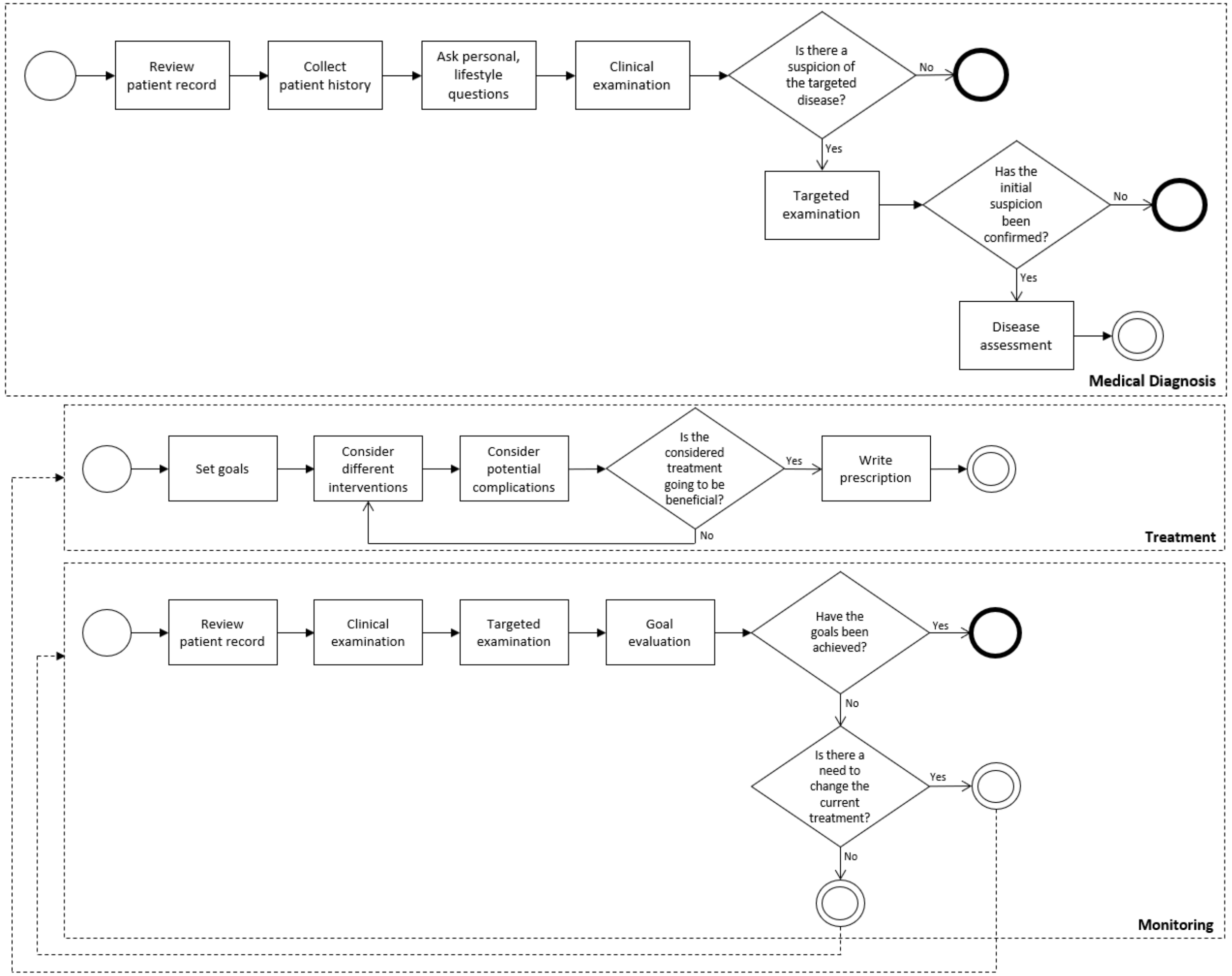
Clinical Care Process Specifications (CCPS)

Existing Work

1. Taxonomy for CCPS
2. TaSC approach for standardisation of development, structure and content of clinical caremaps

Research Interests (by domain)

Caremaps with Decision Points



Medical Diagnosis

Review patient record
Collect patient history
Ask personal, lifestyle questions
Clinical examination
Is there a suspicion of the targeted disease?
No
Yes
Targeted examination
Has the initial suspicion been confirmed?
No
Yes
Disease assessment

Treatment

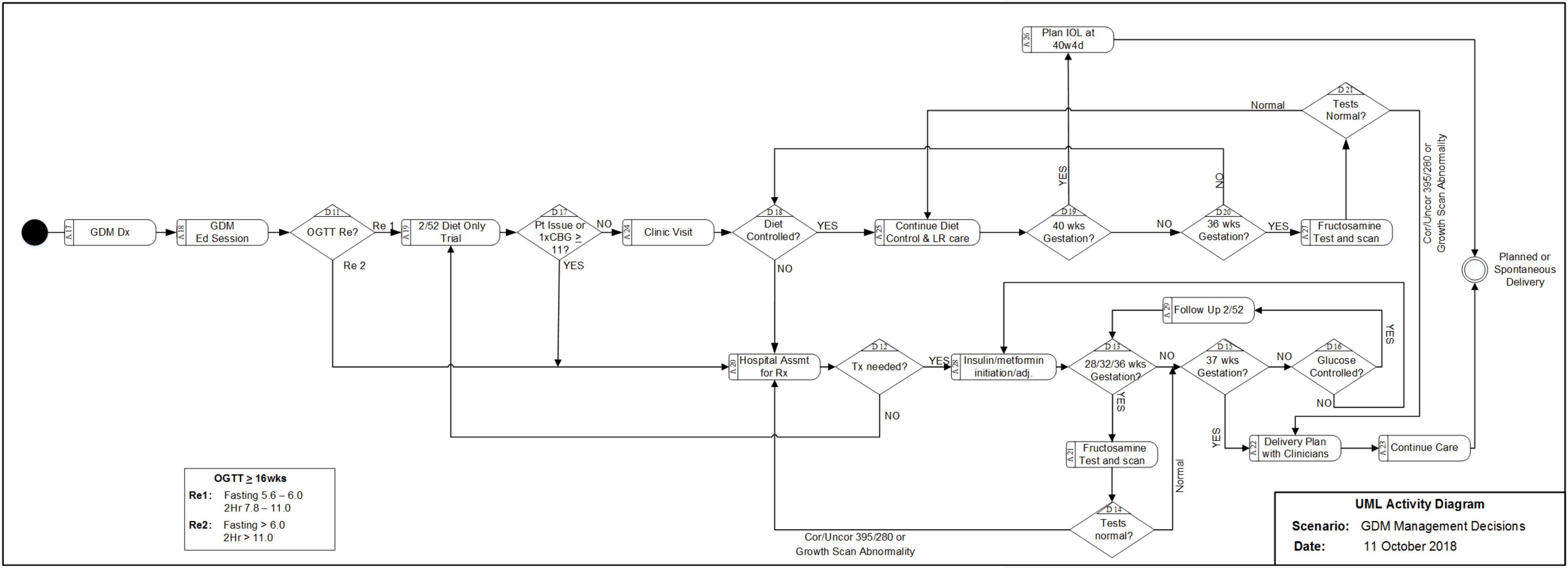
Set goals
Consider different interventions
Consider potential complications
Is the considered treatment going to be beneficial?
Yes
No
Write prescription

Monitoring

Review patient record
Clinical examination
Targeted examination
Goal evaluation
Have the goals been achieved?
Yes
No
Is there a need to change the current treatment?
Yes
No

Research Interests (by domain)

Gestational Diabetes Mellitus (Patient Management)



OGTT ≥ 16wks

Re1: Fasting 5.6 – 6.0
2Hr 7.8 – 11.0

Re2: Fasting > 6.0
2Hr > 11.0

UML Activity Diagram

Scenario: GDM Management Decisions

Date: 11 October 2018

Research Interests (by domain)

Clinical Care Process Specifications (CCPS)

Future Work

Investigation and/or development of:

1. Standardisation approaches for other CCPS;
2. A validation and evaluation approach to ensure standards conformity for CCPS (and later, EHR);

Research Interests (by domain)

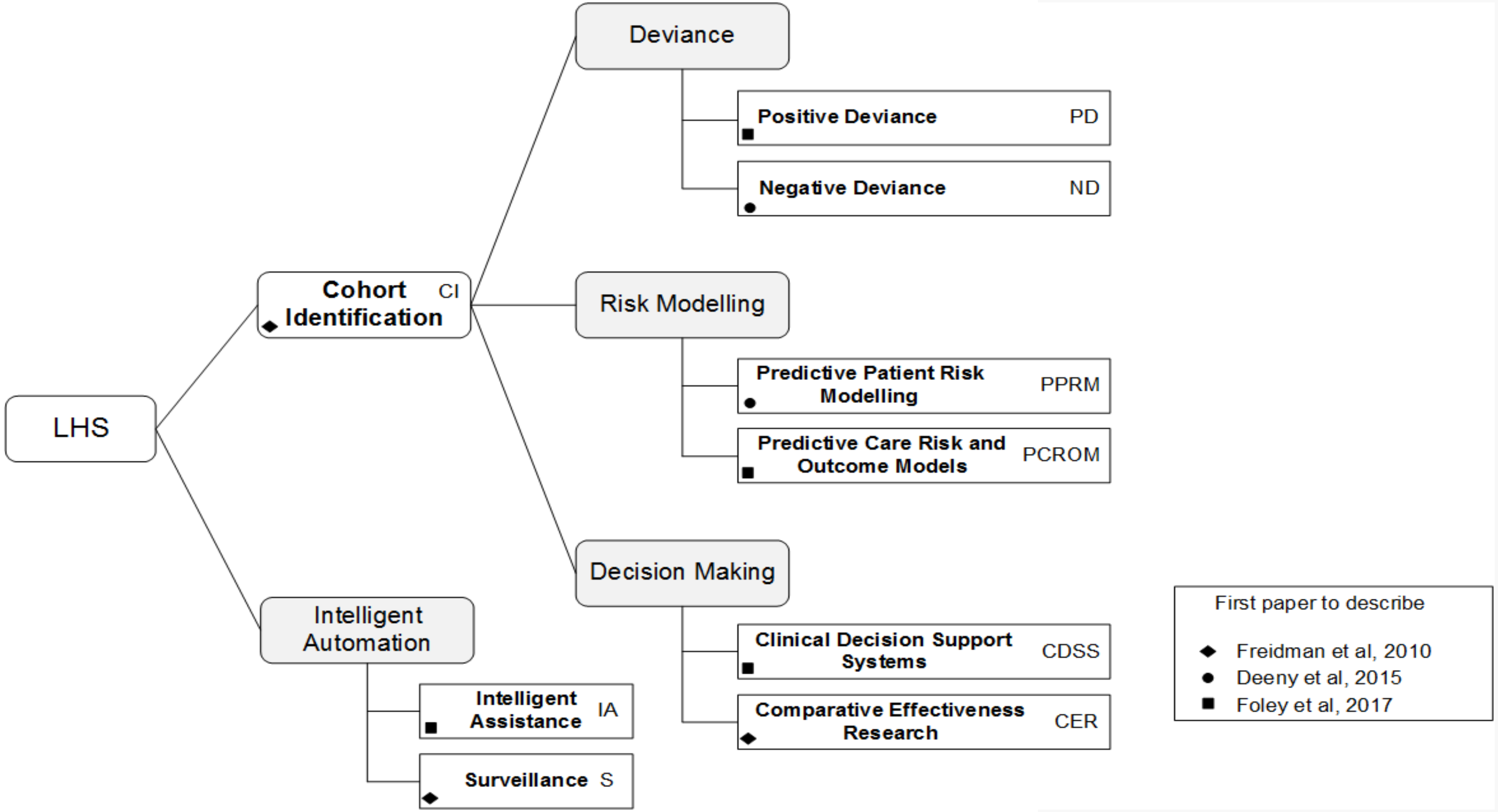
Learning Health Systems (LHS)

Existing Work

1. The LHS Community Awareness Challenge
2. The Heimdall Taxonomy and Framework for LHS
3. The ITPOSMO-BBF model for evaluating barriers, benefits and facilitating factors (adoption) of LHS

Research Interests (by domain)

Taxonomy for Learning Health Systems



Research Interests (by domain)

Learning Health Systems (LHS)

Future Work

Investigation and/or development of:

1. Existing and potential methodologies for developing LHS
2. Near- and real-time LHS tools/mobile apps
3. Human and non-human factors acting as barriers or facilitating factors
4. Solutions for combining smart sensors and AI (e.g. smart clothing for medical monitoring of astronauts and military personnel)

Research Interests (by domain)

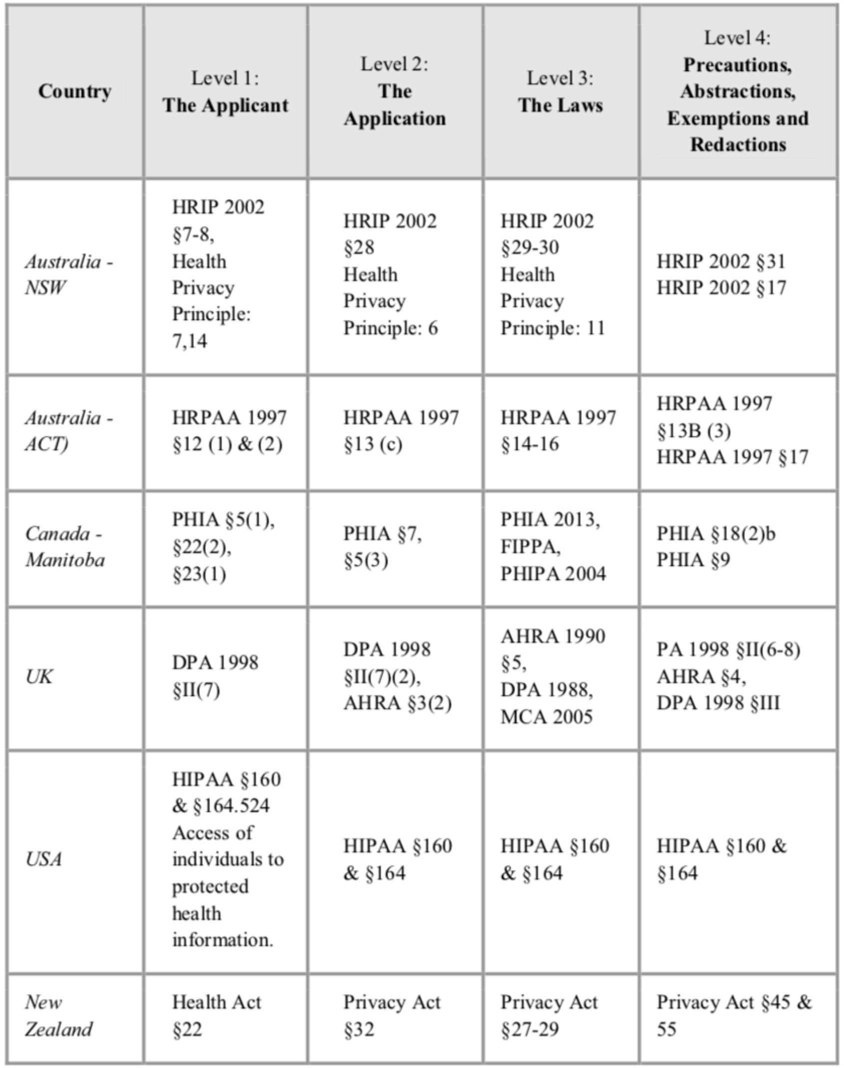
Medico-Legal Decision Support Systems (MLDSS)

Completed Work

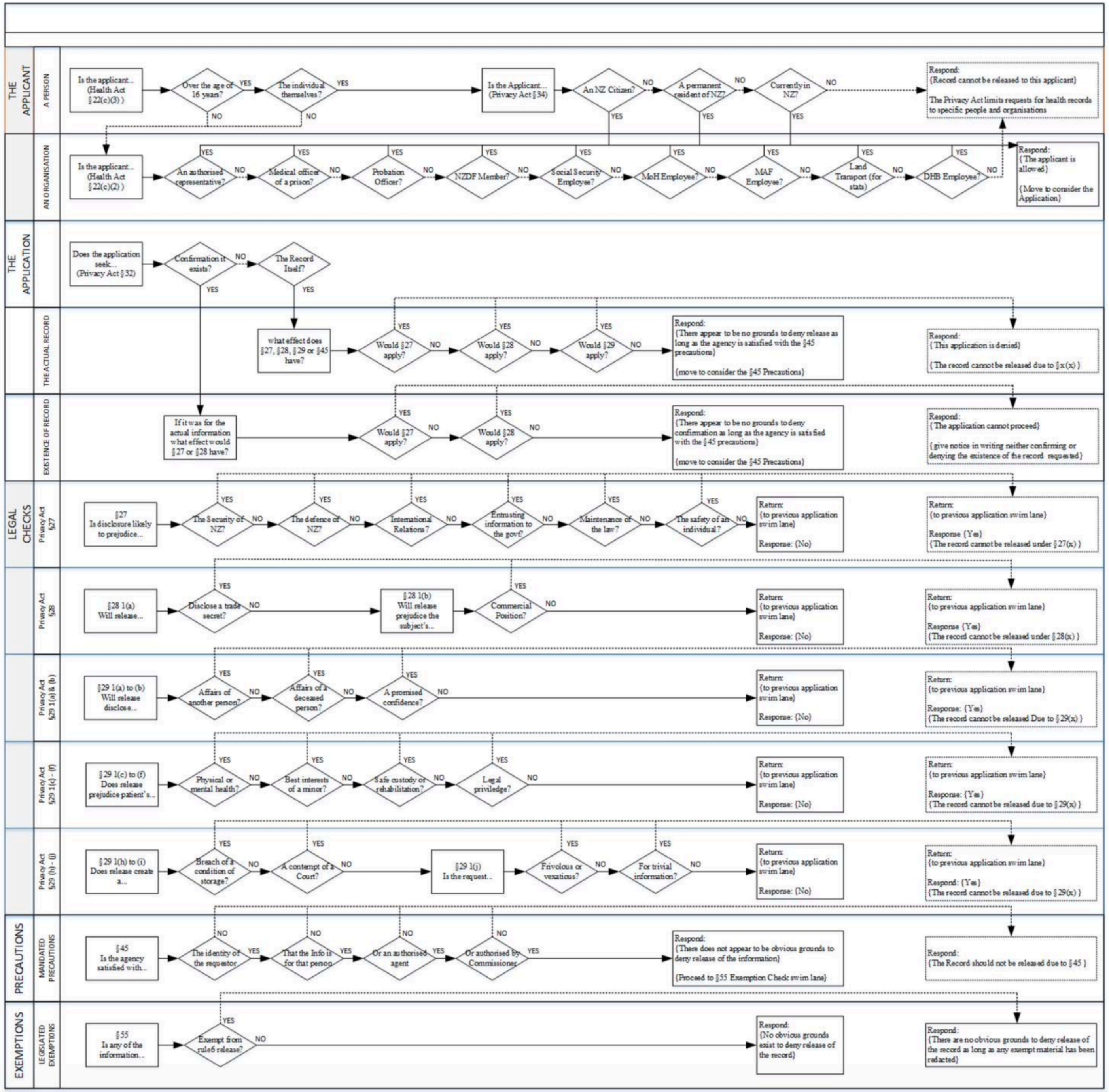
1. Request for Access Tool (RAT)
2. Medical negligence litigation prediction tool
3. Consent models for use in secondary use research using RCHD

Research Interests (by domain)

RAT Tool – Legal Decision Flowchart



Country	Level 1: **The Applicant**	Level 2: **The Application**	Level 3: **The Laws**	Level 4: **Precautions, Abstractions, Exemptions and Redactions**
Australia - NSW	HRIP 2002 §7-8, Health Privacy Principle: 7,14	HRIP 2002 §28 Health Privacy Principle: 6	HRIP 2002 §29-30 Health Privacy Principle: 11	HRIP 2002 §31 HRIP 2002 §17
Australia - ACT)	HRPAA 1997 §12 (1) & (2)	HRPAA 1997 §13 (c)	HRPAA 1997 §14-16	HRPAA 1997 §13B (3) HRPAA 1997 §17
Canada - Manitoba	PHIA §5(1), §22(2), §23(1)	PHIA §7, §5(3)	PHIA 2013, FIPPA, PHIPA 2004	PHIA §18(2)b PHIA §9
UK	DPA 1998 §II(7)	DPA 1998 §II(7)(2), AHRA §3(2)	AHRA 1990 §5, DPA 1988, MCA 2005	PA 1998 §II(6-8) AHRA §4, DPA 1998 §III
USA	HIPAA §160 & §164.524 Access of individuals to protected health information.	HIPAA §160 & §164	HIPAA §160 & §164	HIPAA §160 & §164
New Zealand	Health Act §22	Privacy Act §32	Privacy Act §27-29	Privacy Act §45 & 55



Research Interests (by domain)

Medico-Legal Decision Support Systems (MLDSS)

Future Work

Investigation and/or development of:

1. Flexible ongoing consent approaches
2. Integration of consent collection and reporting
3. Automatic extraction, aggregation, anonymisation/pseudonymisation and knowledge extraction methods

Research Areas

Dr Evangelia Kyrimi

Research Interests (by domain)

Artificial Intelligence in Healthcare

Outline

1. Need for causal models
2. Existing work on AI in healthcare
3. Future directions

Research Interests (by domain)

Artificial Intelligence in Healthcare

Need for causal models

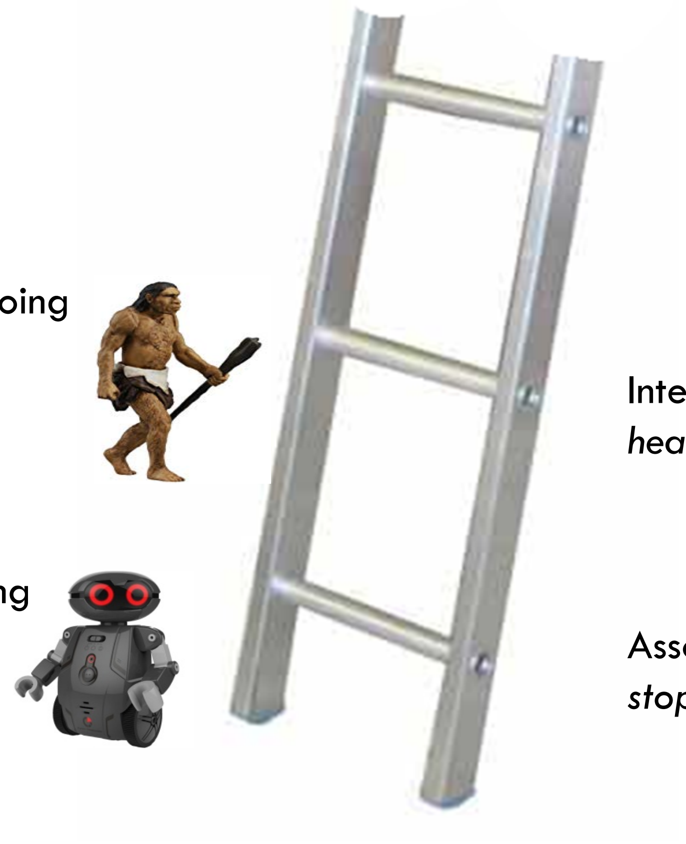
Seeing

Association: “*From trials data is drug x effective at stopping headaches?*”

Research Interests (by domain)

Artificial Intelligence in Healthcare

Need for causal models



Doing

Seeing

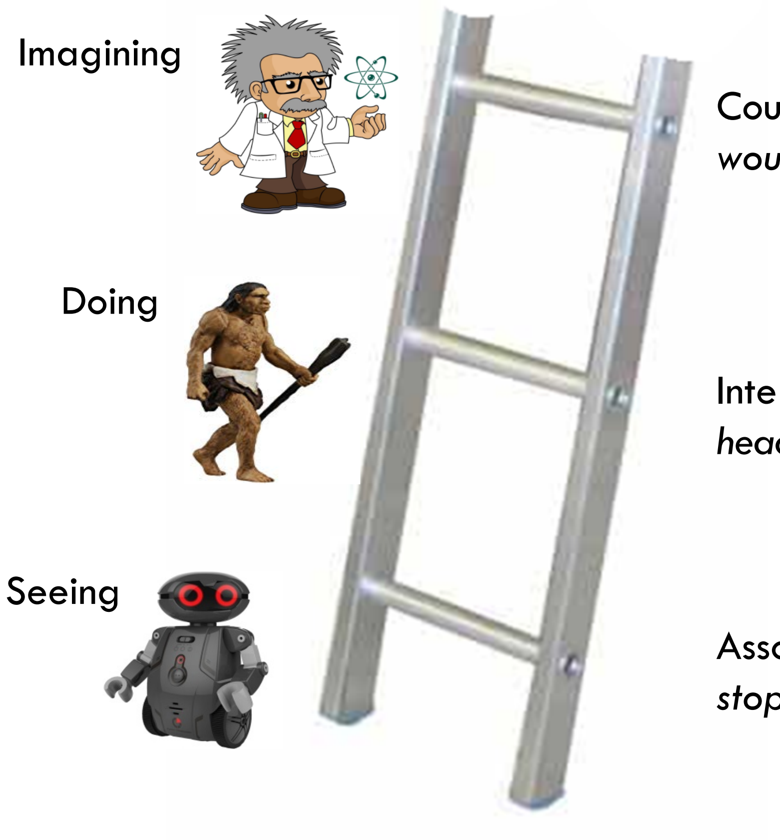
Intervention: “*If I take drug x, will it stop my headache?*”

Association: “*From trials data is drug x effective at stopping headaches?*”

Research Interests (by domain)

Artificial Intelligence in Healthcare

Need for causal models



Counterfactuals: “*If I hadn’t taken drug x, would my headache have stopped?*”

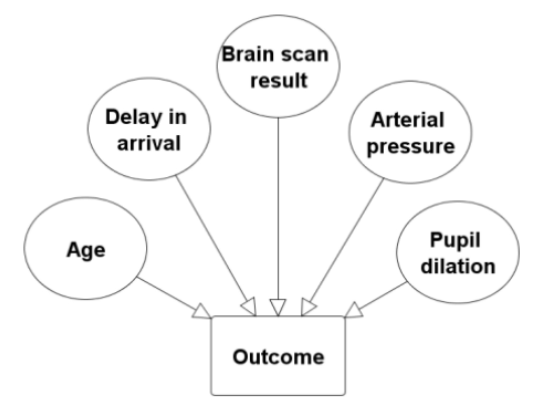
Intervention: “*If I take drug x, will it stop my headache?*”

Association: “*From trials data is drug x effective at stopping headaches?*”

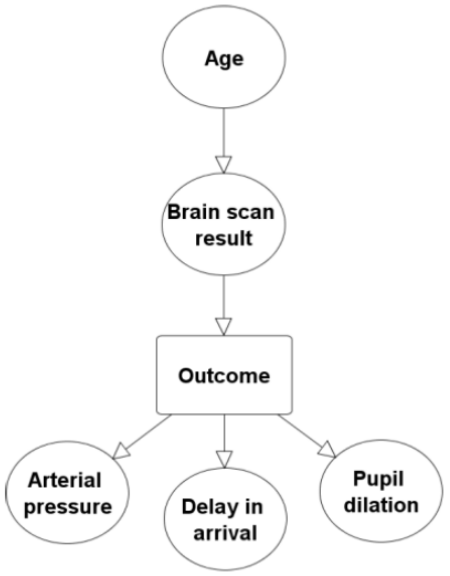
Research Interests (by domain)

Artificial Intelligence in Healthcare

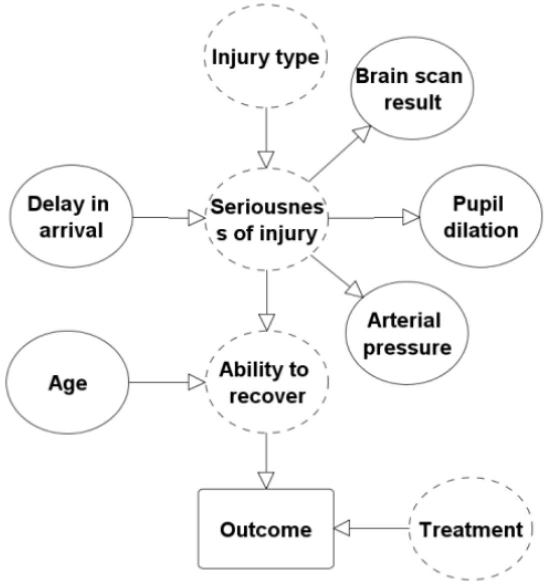
Need for causal models



a) Standard statistical regression model learnt from data: Outcome= *f*(inputs)



b) 'Causal' model learnt purely from data (Sakellaropoulos & Nikiforidis, 1999)



c) Sensible causal model with missing/ unobserved variables

Research Interests (by domain)

Artificial Intelligence in Healthcare

Trauma care

- RIM Group
- The Royal London Hospital
- US Army Inistitute of Surgical Research

Queen Mary
University of London

RIM
Risk and Information Management

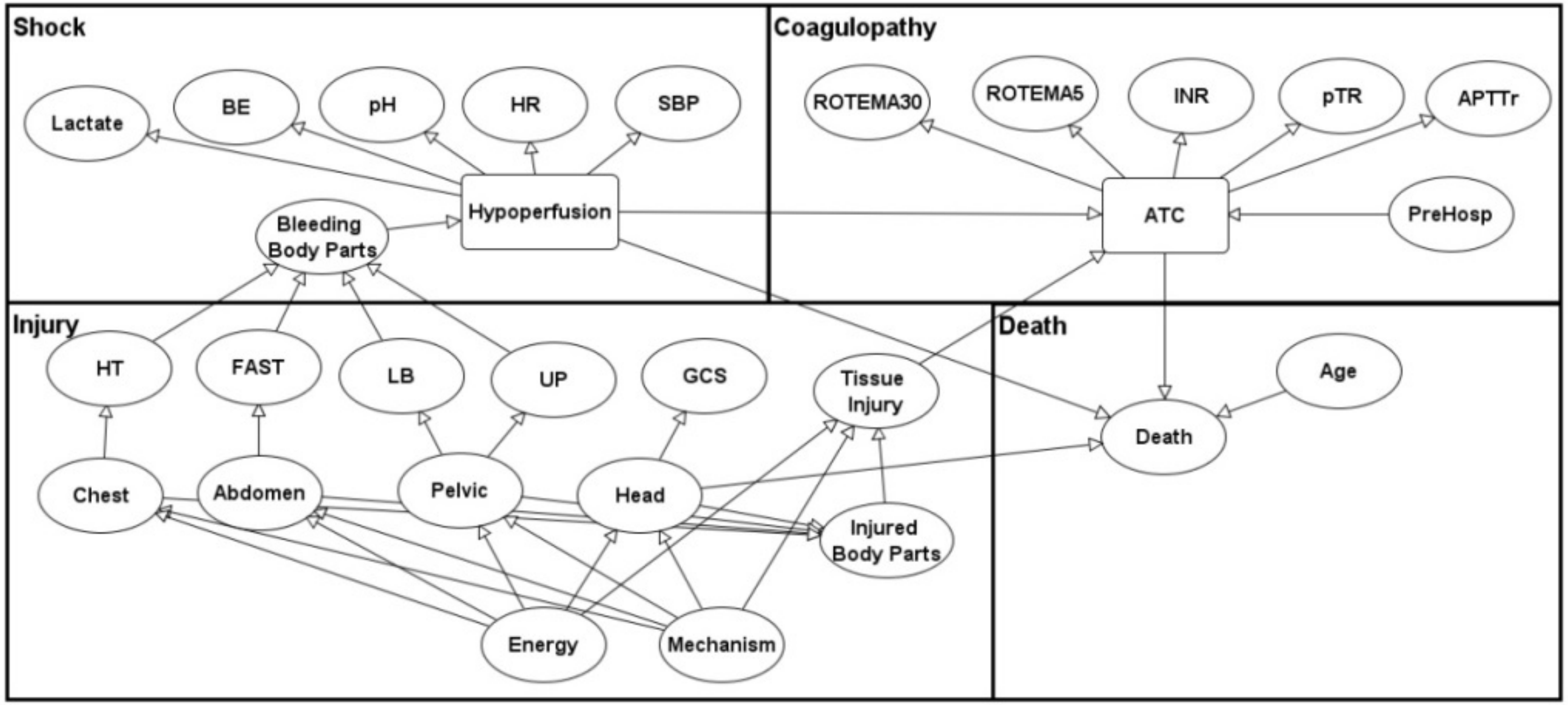
Barts and The London
TRAUMA
SCIENCES

TRAUMA SERVICE
THE ROYAL LONDON HOSPITAL

Research Interests (by domain)

Artificial Intelligence in Healthcare

Existing work - Trauma care



Queen Mary
University of London

RIM
Risk and Information Management

Barts and The London
TRAUMA
SCIENCES

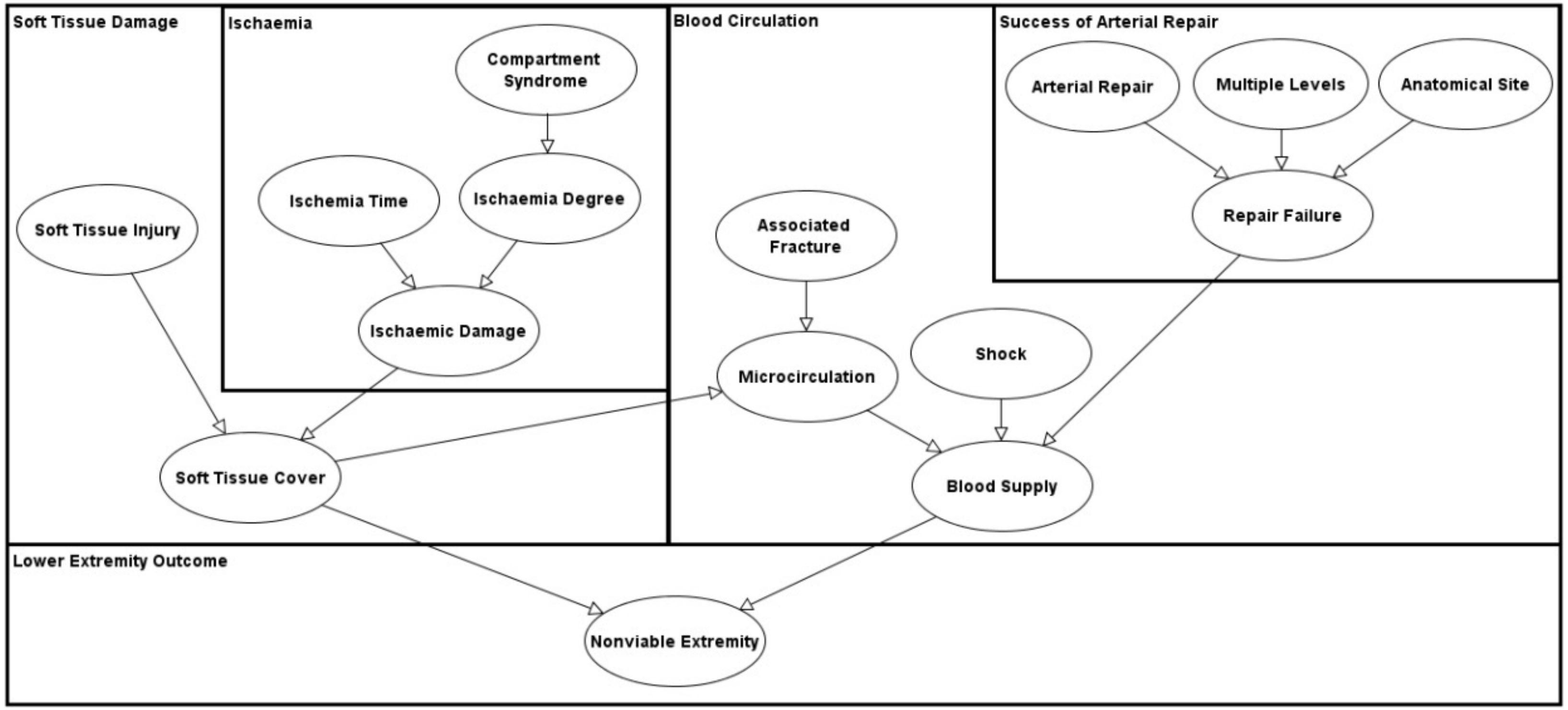
TRAUMA SERVICE
THE ROYAL LONDON HOSPITAL

UNITED STATES ARMY
INSTITUTE OF SURGICAL RESEARCH
Optimizing Combat Casualty Care
U.S.ARMY

Research Interests (by domain)

Artificial Intelligence in Healthcare

Existing work - Trauma care



Queen Mary University of London

RIM Risk and Information Management

Barts and The London TRAUMA SCIENCES

TRAUMA SERVICE THE ROYAL LONDON HOSPITAL

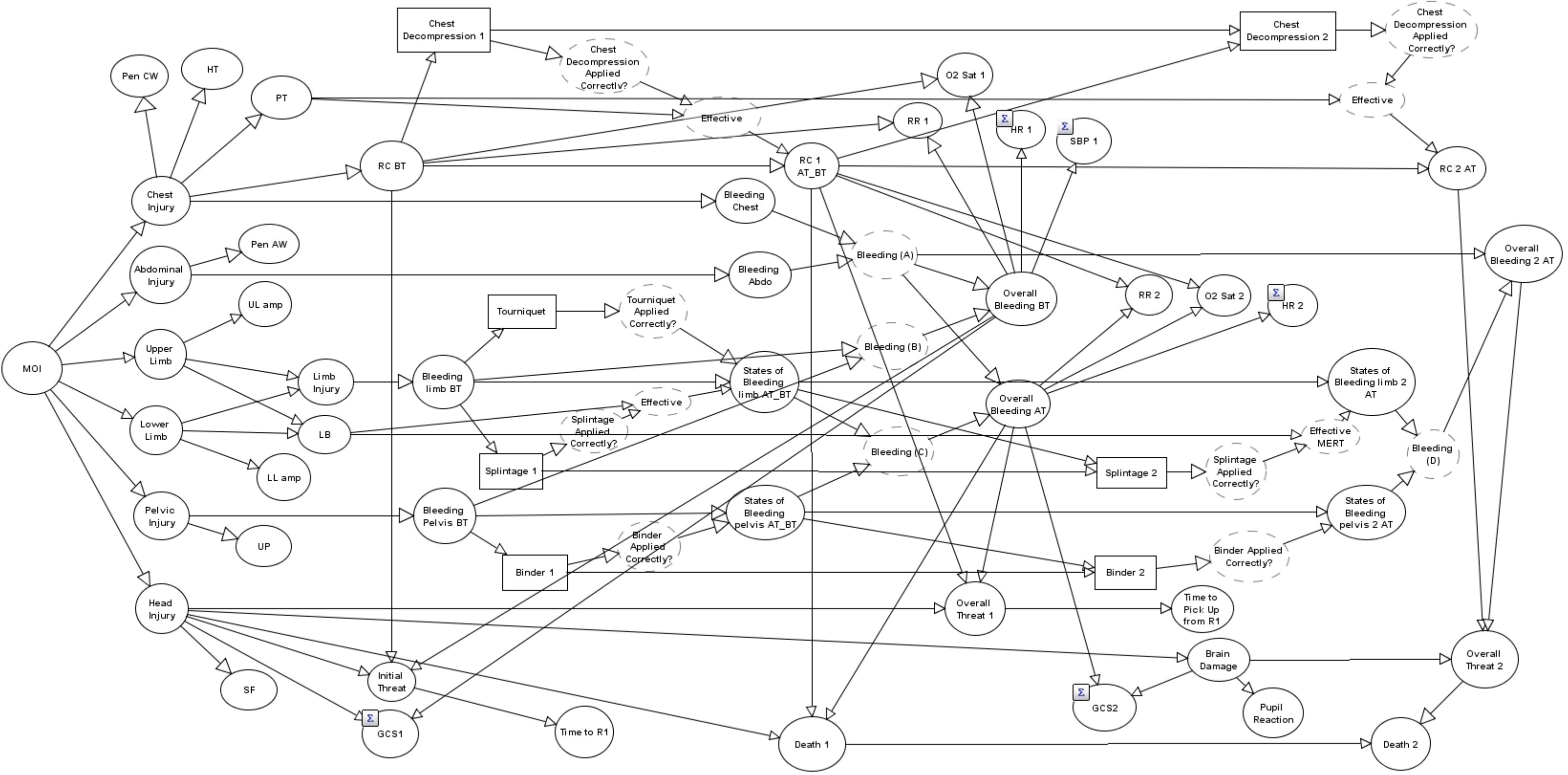
UNITED STATES ARMY INSTITUTE OF SURGICAL RESEARCH *Optimizing Combat Casualty Care*

U.S.ARMY

Research Interests (by domain)

Artificial Intelligence in Healthcare

Existing work - Trauma care



Queen Mary
University of London

RIM
Risk and Information Management

Barts and The London
TRAUMA SCIENCES

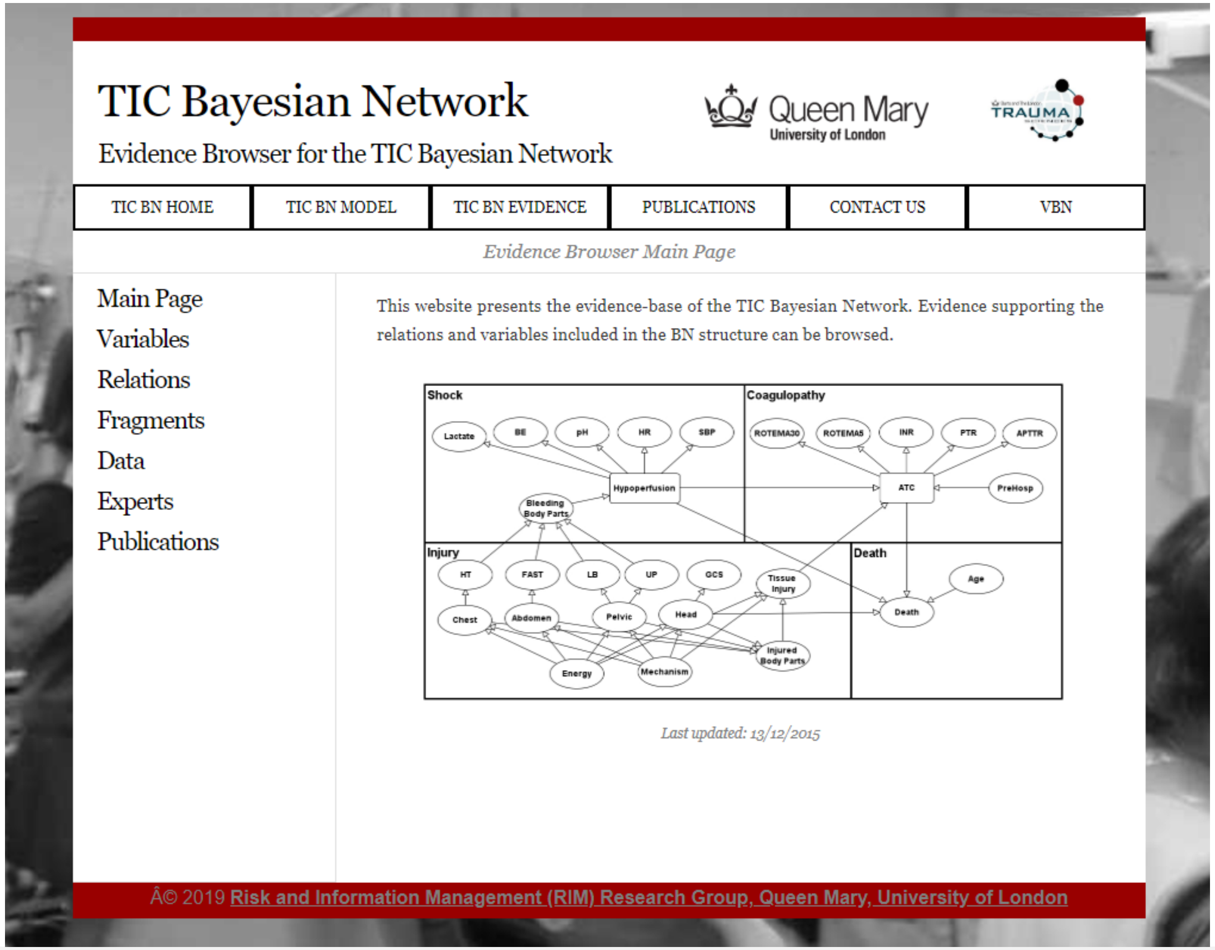
TRAUMA SERVICE
THE ROYAL LONDON HOSPITAL

UNITED STATES ARMY
INSTITUTE OF SURGICAL RESEARCH
Optimizing Combat Casualty Care
U.S.ARMY

Research Interests (by domain)

Artificial Intelligence in Healthcare

Existing work - Trauma care (www.traumamodel.com)



TIC Bayesian Network

Evidence Browser for the TIC Bayesian Network

TIC BN HOME | TIC BN MODEL | TIC BN EVIDENCE | PUBLICATIONS | CONTACT US | VBN

Evidence Browser Main Page

- Main Page
- Variables
- Relations
- Fragments
- Data
- Experts
- Publications

This website presents the evidence-base of the TIC Bayesian Network. Evidence supporting the relations and variables included in the BN structure can be browsed.

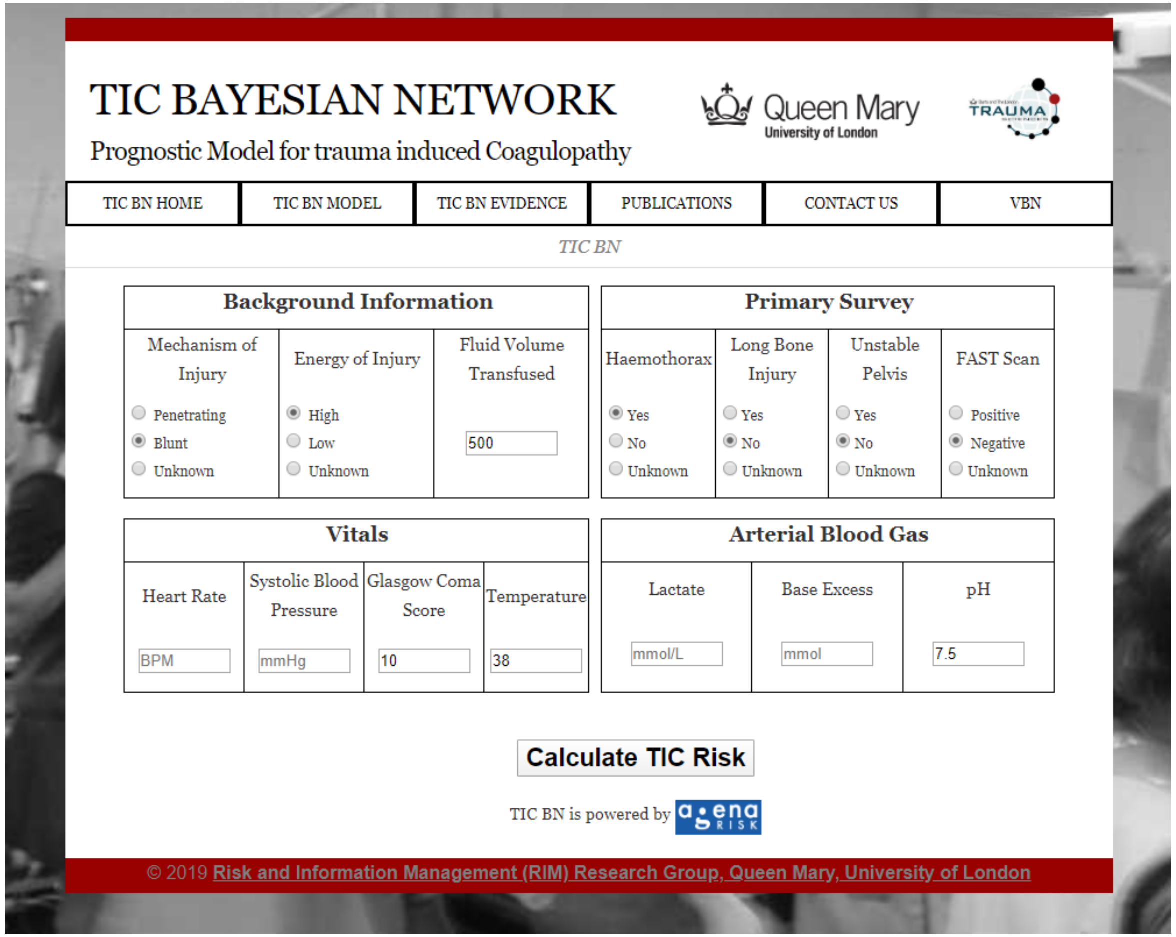
Last updated: 13/12/2015

Â© 2019 Risk and Information Management (RIM) Research Group, Queen Mary, University of London

Research Interests (by domain)

Artificial Intelligence in Healthcare

Existing work - Trauma care



TIC BAYESIAN NETWORK

Prognostic Model for trauma induced Coagulopathy

TIC BN

Background Information

Mechanism of Injury	Energy of Injury	Fluid Volume Transfused
○ Penetrating ● Blunt ○ Unknown	● High ○ Low ○ Unknown	500

Primary Survey

Haemothorax	Long Bone Injury	Unstable Pelvis	FAST Scan
● Yes ○ No ○ Unknown	○ Yes ● No ○ Unknown	○ Yes ● No ○ Unknown	○ Positive ● Negative ○ Unknown

Vitals

Heart Rate	Systolic Blood Pressure	Glasgow Coma Score	Temperature
BPM	mmHg	10	38

Arterial Blood Gas

Lactate	Base Excess	pH
mmol/L	mmol	7.5

TIC BN is powered by agena risk

© 2019 Risk and Information Management (RIM) Research Group, Queen Mary, University of London

Queen Mary University of London

RIM Risk and Information Management

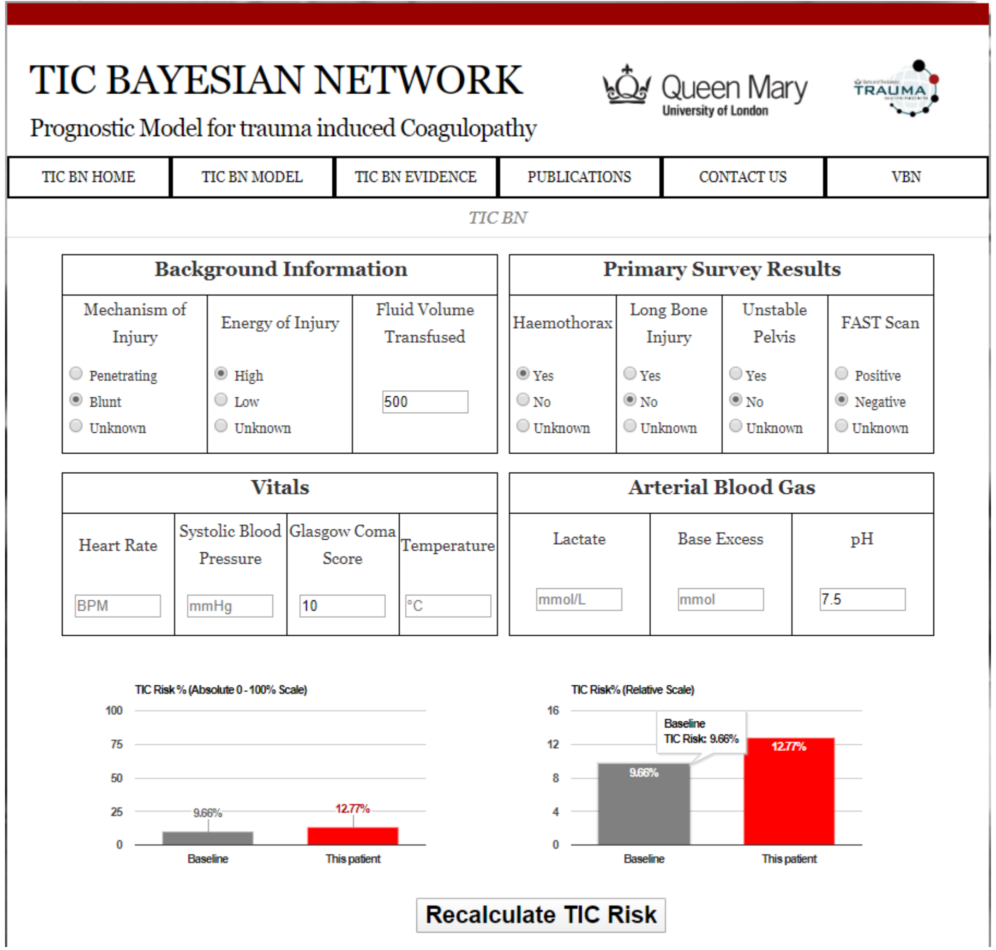
Barts and The London TRAUMA SCIENCES

TRAUMA SERVICE THE ROYAL LONDON HOSPITAL

Research Interests (by domain)

Artificial Intelligence in Healthcare

Existing work - Trauma care



TIC BAYESIAN NETWORK

Prognostic Model for trauma induced Coagulopathy

Queen Mary University of London

TIC BN HOME | TIC BN MODEL | TIC BN EVIDENCE | PUBLICATIONS | CONTACT US | VBN

TIC BN

Background Information

Mechanism of Injury: ○ Penetrating ● Blunt ○ Unknown

Energy of Injury: ● High ○ Low ○ Unknown

Fluid Volume Transfused: 500

Primary Survey Results

Haemothorax: ● Yes ○ No ○ Unknown

Long Bone Injury: ○ Yes ● No ○ Unknown

Unstable Pelvis: ○ Yes ● No ○ Unknown

FAST Scan: ○ Positive ● Negative ○ Unknown

Vitals

Heart Rate: BPM

Systolic Blood Pressure: mmHg

Glasgow Coma Score: 10

Temperature: °C

Arterial Blood Gas

Lactate: mmol/L

Base Excess: mmol

pH: 7.5

Queen Mary University of London

RIM Risk and Information Management

Barts and The London TRAUMA SCIENCES

TRAUMA SERVICE THE ROYAL LONDON HOSPITAL

UNITED STATES ARMY INSTITUTE OF SURGICAL RESEARCH Optimizing Combat Casualty Care

U.S.ARMY

Research Interests (by domain)

Artificial Intelligence in Healthcare

Future work – COMBAT-AID project

1. Refine and validate existing models in military environment
2. Develop an overflight triage model
3. Investigate the interpretability and usability of the developed models

US Department of Defense awards £1m to Queen Mary University of London for AI research on treating injured soldiers

Medical care of injured soldiers could improve with new Artificial Intelligence (AI) tools designed for the battlefield and the hospital following a grant from the US Department of Defense for research at Queen Mary University of London.

1 October 2019

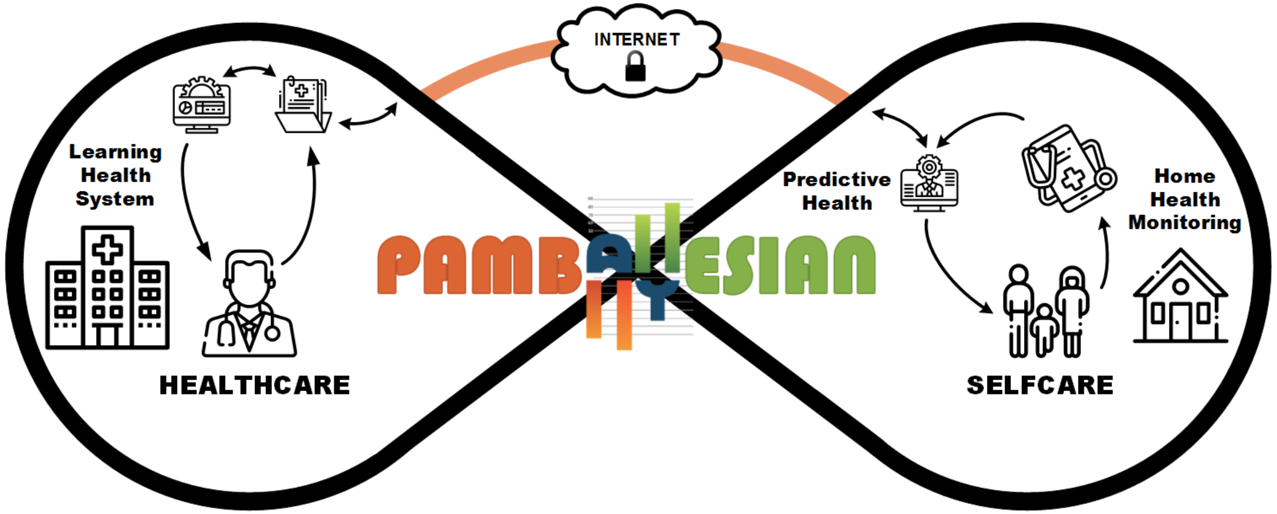
The US Department of Defense has awarded the Centre for Trauma Sciences (C4TS) at Queen Mary a $1.2 million (£976.500) grant to develop AI tools that could help save the lives of badly injured soldiers.

It is aimed at developing and validating a suite of accurate prediction models and Clinical Decision Support (CDS) tools that clinicians can use to treat wounded soldiers on the battlefield, travelling to

Research Interests (by domain)

Artificial Intelligence in Healthcare

Existing Work – PAMBAYESIAN project



- Create new generation of intelligent medical decision support systems for direct patient use with real-time monitoring for chronic conditions
- Increase patient independence and decrease reliance on direct consultation
- Allow more autonomous care at home and reduce associated health care cost

Research Interests (by domain)

Artificial Intelligence in Healthcare

Existing Work – PAMBAYESIAN project

1. Scoping review of BNs in Healthcare (6 papers)

Research Interests (by domain)

Artificial Intelligence in Healthcare

Existing Work – PAMBAYESIAN project

1. Scoping review of BNs in Healthcare (6 papers)
2. Standardisation of the process for eliciting expert knowledge using caremaps

Research Interests (by domain)

Artificial Intelligence in Healthcare

Existing Work – PAMBAYESIAN project

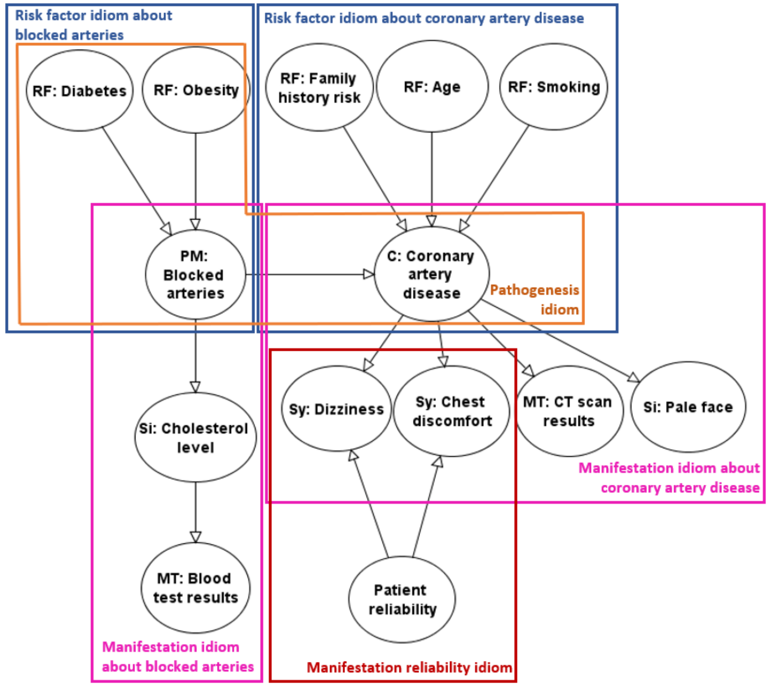
1. Scoping review of BNs in Healthcare (6 papers)
2. Standardisation of the process for eliciting expert knowledge using caremaps
3. Standardise the process of developing medical BNs using an idiom based approach

Research Interests (by domain)

Artificial Intelligence in Healthcare

Existing Work – PAMBAYESIAN project

1. Scoping review of BNs in Healthcare (6 papers)
2. Standardisation of the process for eliciting expert knowledge using caremaps
3. Standardise the process of developing medical BNs using an idiom based approach



Research Interests (by domain)

Artificial Intelligence in Healthcare

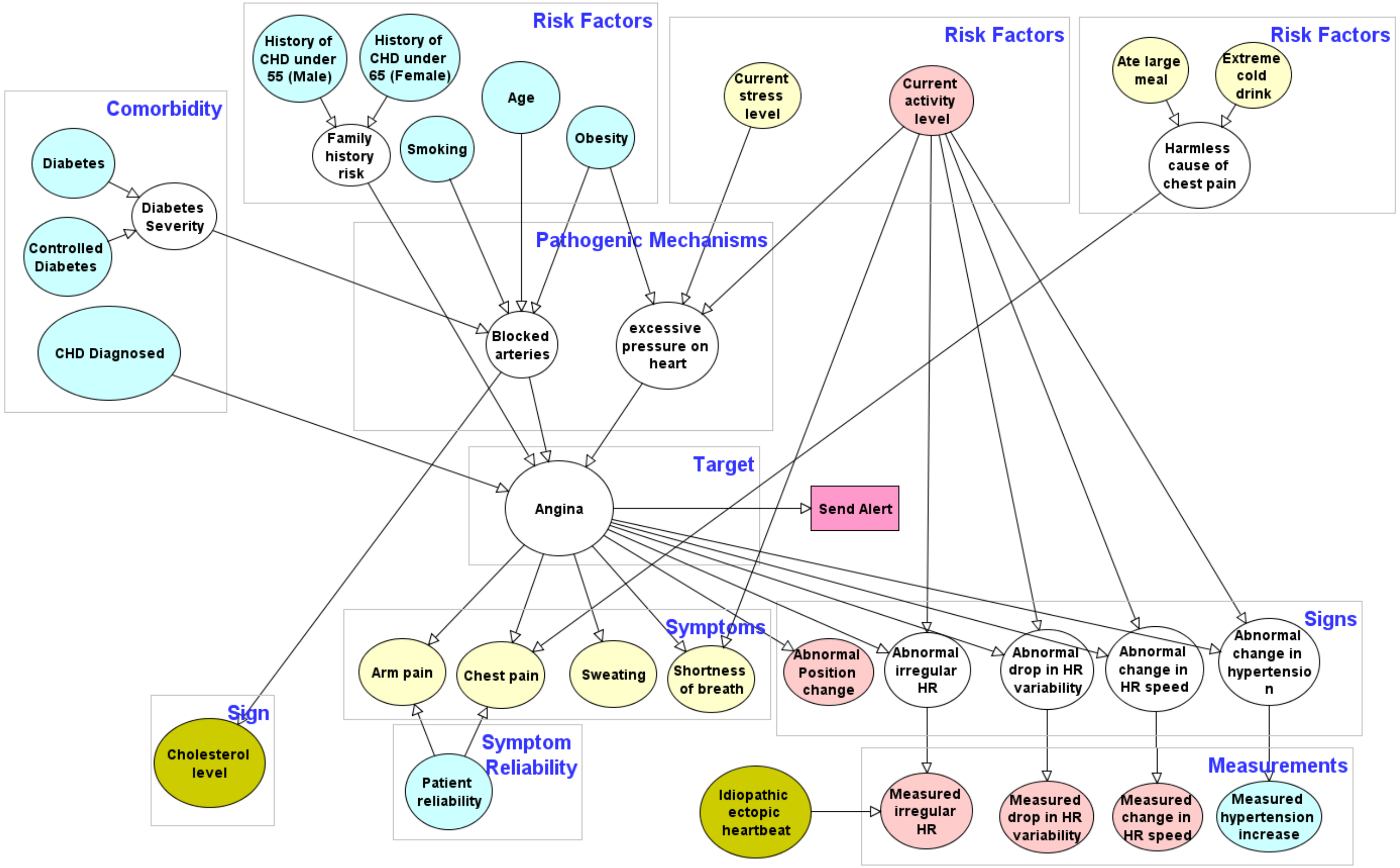
Existing Work – PAMBAYESIAN project

1. Scoping review of BNs in Healthcare (6 papers)
2. Standardisation of the process for eliciting expert knowledge using caremaps
3. Standardise the process of developing medical BNs using an idiom based approach
4. Web-based application using BNs for Diagnosis and management of chronic conditions (GDM, RA)

Research Interests (by domain)

Artificial Intelligence in Healthcare

Existing Work – PAMBAYESIAN project



Background factor entered once by clinicians at the first clinic visit and potentially updated in a later visit

Observation on current state that may be entered by patients after the initial alert is activated

Measurement from sensor and patients specific monitoring

Information from health records entered direcly

Not directly observable. Only enter observations to test the model

Research Interests (by domain)

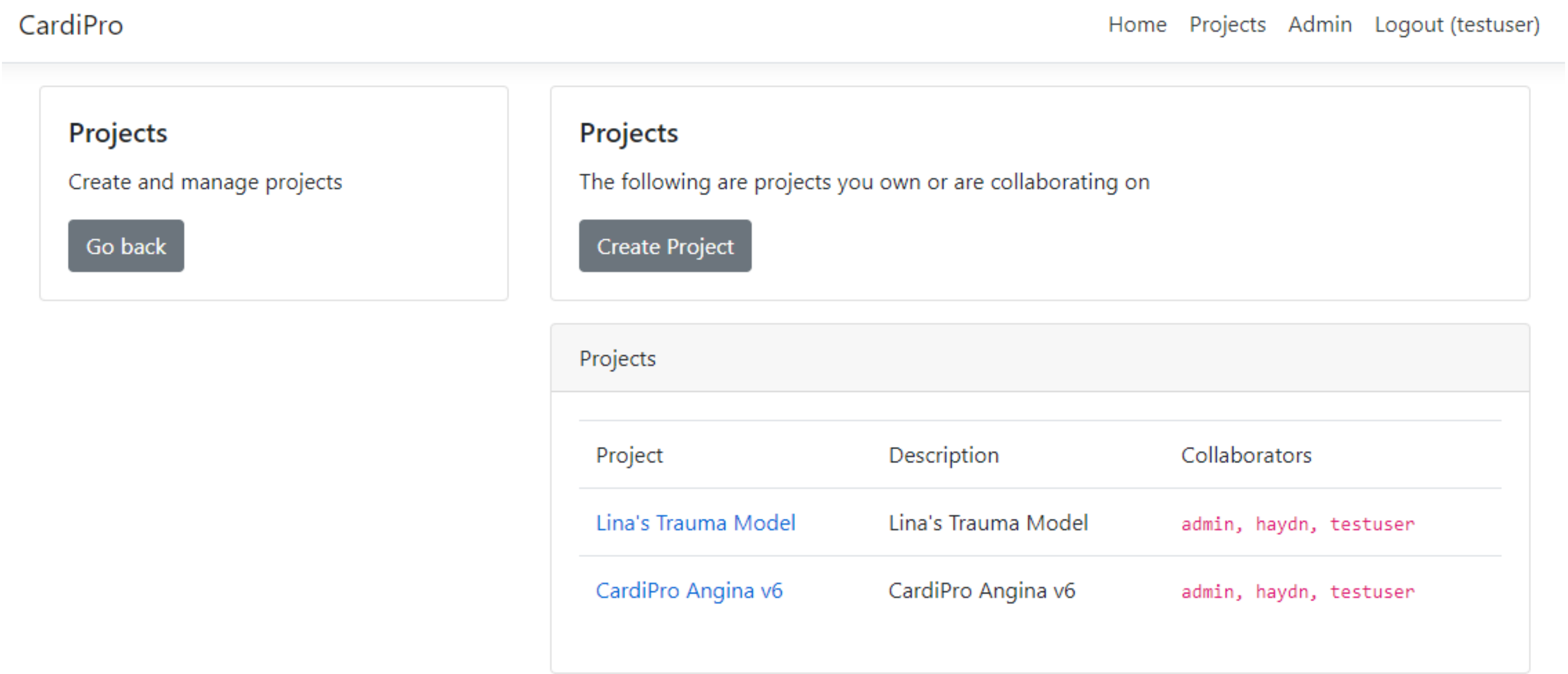
Artificial Intelligence in Healthcare

Existing Work – PAMBAYESIAN project

https://cardipro.mclachlandigital.com

Username: testuser

Password: password



Projects

Create and manage projects

Projects

The following are projects you own or are collaborating on

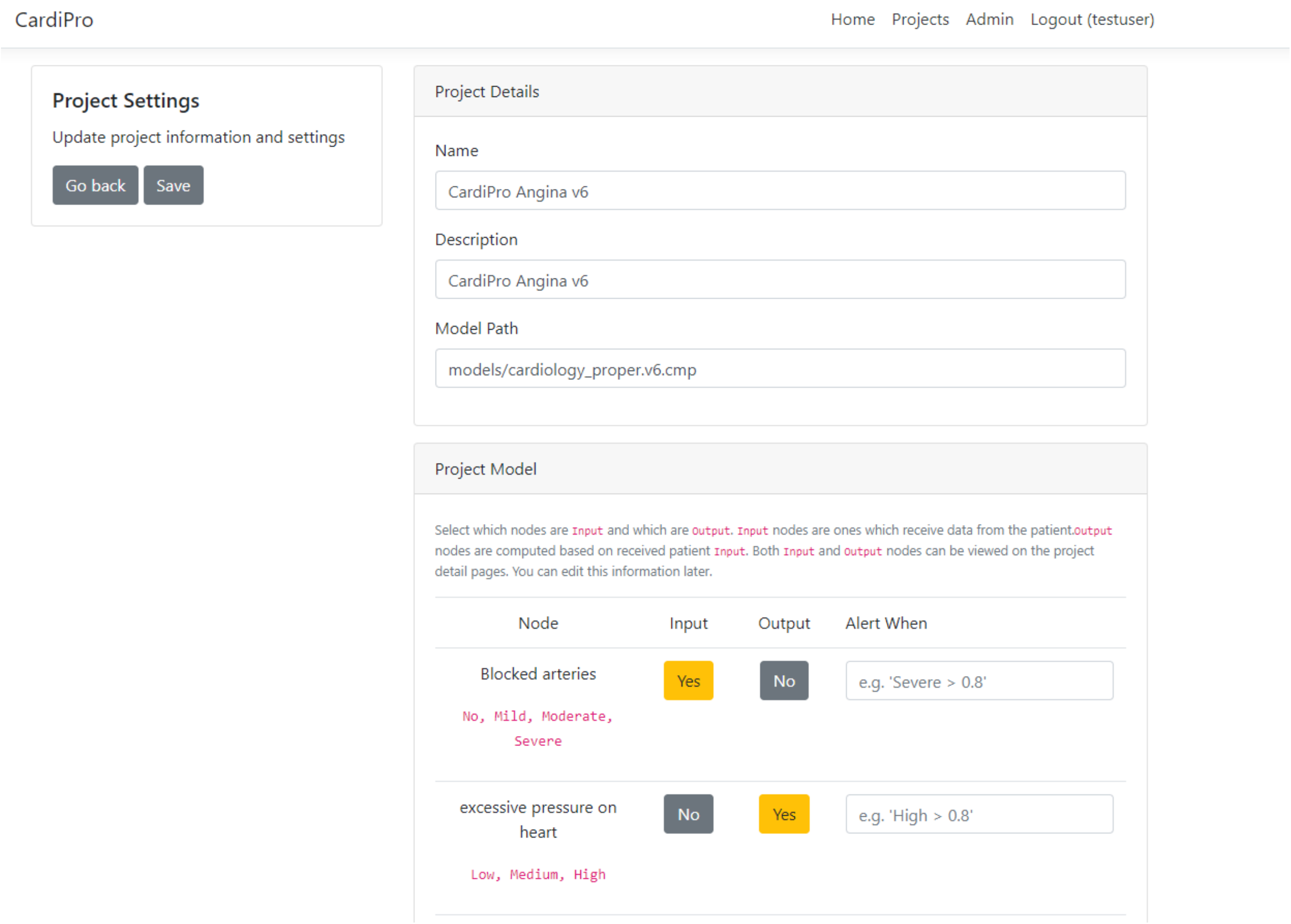
Projects

Project	Description	Collaborators
Lina's Trauma Model	Lina's Trauma Model	`admin, haydn, testuser`
CardiPro Angina v6	CardiPro Angina v6	`admin, haydn, testuser`

Research Interests (by domain)

Artificial Intelligence in Healthcare

Existing Work – PAMBAYESIAN project



Project Settings

Update project information and settings

Project Details

Name: CardiPro Angina v6

Description: CardiPro Angina v6

Model Path: models/cardiology_proper.v6.cmp

Project Model

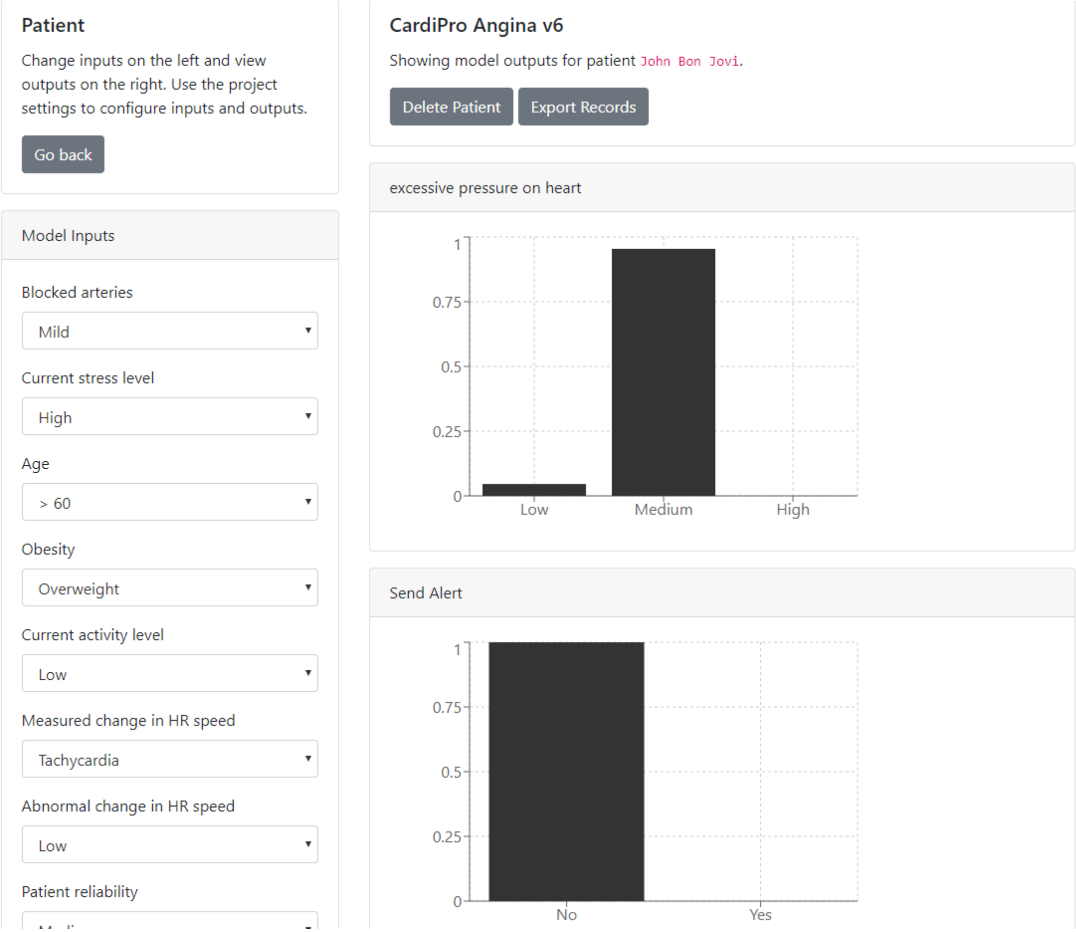
Select which nodes are `Input` and which are `Output`. `Input` nodes are ones which receive data from the patient.`Output` nodes are computed based on received patient `Input`. Both `Input` and `Output` nodes can be viewed on the project detail pages. You can edit this information later.

Node	Input	Output	Alert When
Blocked arteries `No, Mild, Moderate, Severe`	Yes	No	e.g. 'Severe > 0.8'
excessive pressure on heart `Low, Medium, High`	No	Yes	e.g. 'High > 0.8'

Research Interests (by domain)

Artificial Intelligence in Healthcare

Existing Work – PAMBAYESIAN project



Patient

Change inputs on the left and view outputs on the right. Use the project settings to configure inputs and outputs.

Model Inputs

Blocked arteries: Mild

Current stress level: High

Age: > 60

Obesity: Overweight

Current activity level: Low

Measured change in HR speed: Tachycardia

Abnormal change in HR speed: Low

Patient reliability

CardiPro Angina v6

Showing model outputs for patient `John Bon Jovi`.

Research Interests (by domain)

Artificial Intelligence in Healthcare

Future Work

1. Expedited methods for development of medical BNs using medical idioms, ontologies and caremaps;
2. Abstractions for modelling of chronic diseases;
3. Methods to enhance the adoption of BNs in clinical practice.

Thank you

Kudakwashe Dube

✉ K.Dube@massey.ac.nz

Scott McLachlan

✉ s.mclachlan@qmul.ac.uk

Evangelia Kyrimi

✉ e.kyrimi@qmul.ac.uk