SPARRA: Scottish Patients At Risk of Re-admission and Admission

Louis J. M. Aslett (louis.aslett@durham.ac.uk)

Department of Mathematical Sciences
Durham University

& The Alan Turing Institute

Department of Health and Social Care
OR Talk
10 February 2022

Introduction

Outline

- Introduction
  - Brief biography, overview of research interests, introduce team
- SPARRA Project
  - Motivation, history, objectives, data and methodological approach
- Results
  - Highlights of performance and some insights provided by the model
- Updating Paradox
  - Important theoretical challenge raised by SPARRA

Brief Biography

- 1998–2005
  - Founder & Technical Director, 6 Internet Limited
- 2005–2008
  - BA (Mod) Mathematics, Trinity College Dublin
- 2008–2013
  - PhD Mathematical Statistics, Trinity College Dublin
- 2013–2017
  - Postdoctoral researcher, Department of Statistics, University of Oxford
  - Junior Research Fellow, Corpus Christi College, Oxford
- 2017–
  - Assistant (17–20)/Associate(20–) Professor of Statistics, Department of Mathematical Sciences, Durham University
- 2018–
  - Secondment / Health Programme Fellow, The Alan Turing Institute

Louis Aslett

Research Interests

- Privacy & cryptography in statistics
- Statistical & machine learning
  - Health applications
  - Privacy preservation
- Computational statistics
  - Markov chain Monte Carlo
  - Multilevel Monte Carlo
  - Statistical genetics
  - High performance computing
- Reliability theory

The team

Core team

James Liley
Durham

Sam Emerson
Durham

Catalina Vallejos
Edinburgh

Louis Aslett
Durham

Further Turing team

Gergo Bohner
Nathan Cunningham
Ioanna Manolopoulou
Bilal Mateen
Sebastian Vollmer
Katrina Payne
Chris Holmes

Public Health Scotland team

Rachel Porteous
David Carr
Simon Rogers (NSS)
Katie Borland
Sam Oduro
Stephen Riddell
Keith Moffat
Jill Ireland
Susan Frame
Scott Heald

SPARRA Project

Background

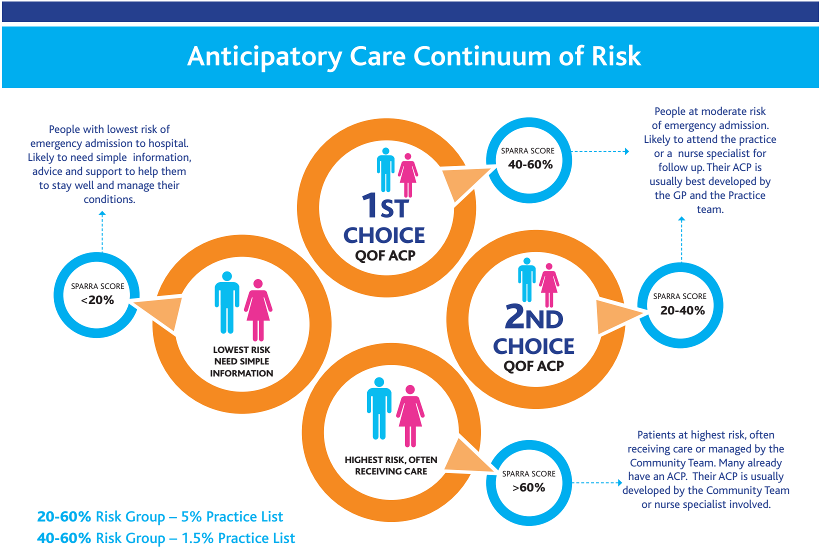
"The NHS should work with other public services and with patients and carers to provide continuous, ***anticipatory*** *care to ensure that, as far as possible,* ***health care crises are prevented from happening****."*

— Kerr Report, NHS Scotland, 2005

Admission to an emergency department (EA)

- breakdown of health control
- transition from primary (preventative) to secondary (curative) care
- increased morbidity and mortality risk
- more expensive and specialised healthcare services

SPARRA Motivation



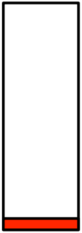
Source: NHS Scotland Anticipatory Care Planning and Polypharmacy Review

SPARRA History

A brief history of SPARRA ...

2006

Version 1
> 65 years old
EA in last 3 yr

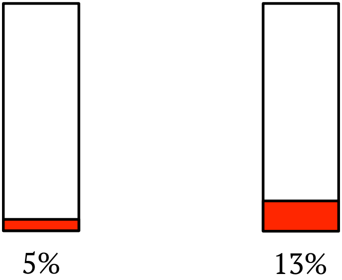


5%

SPARRA History

A brief history of SPARRA ...

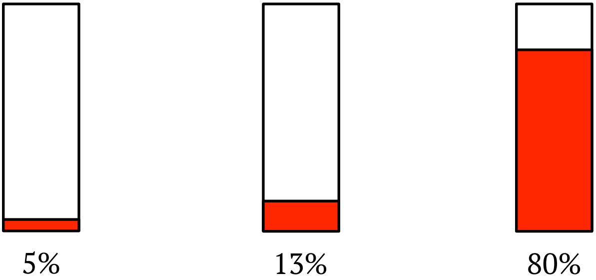
2006	2008
Version 1	Version 2
> 65 years old	Any age
EA in last 3 yr	EA in last 3 yr



SPARRA History

A brief history of SPARRA ...

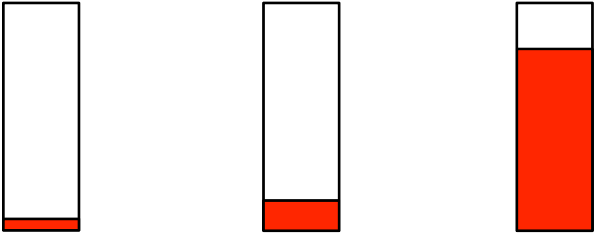
2006	2008	2009
Version 1 > 65 years old EA in last 3 yr	Version 2 Any age EA in last 3 yr	Version 3 Any age Any NHS use



SPARRA History

A brief history of SPARRA ...

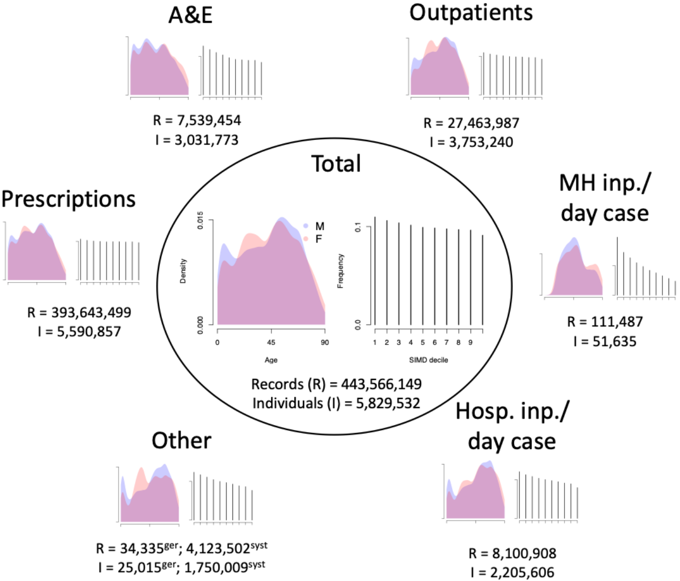
2006	2008	2009	2022
Version 1	Version 2	Version 3	Version 4
> 65 years old	Any age	Any age	Any age
EA in last 3 yr	EA in last 3 yr	Any NHS use	Any NHS use



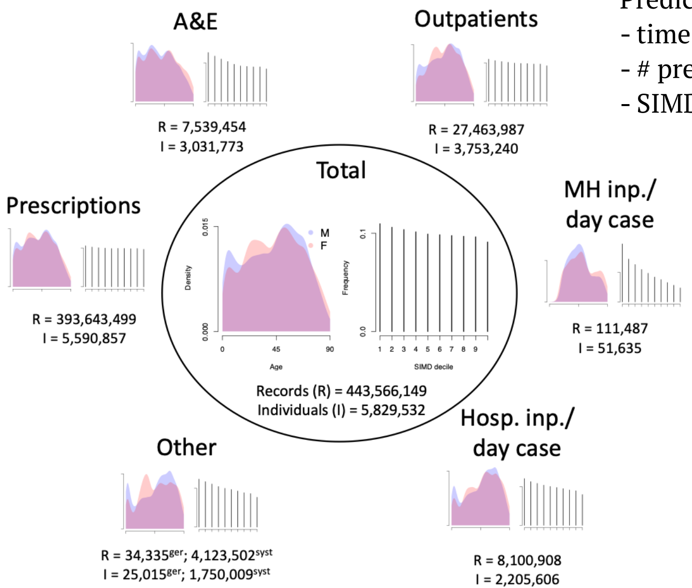
5% 13% 80%

Exploit modern:
Feature engineering
Machine learning
Model validation
Reproducibility

Data sources

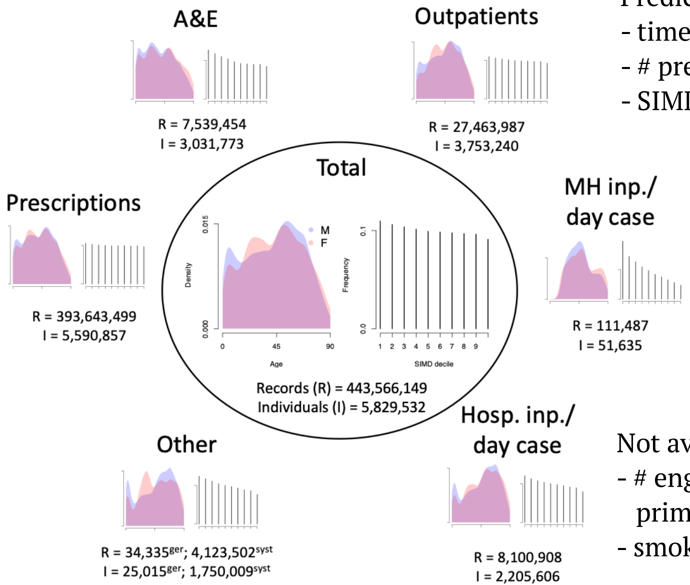


Data sources



Predictors (eg):
- time since last EA (if any)
- # prescriptions filled
- SIMD (deprivation)

Data sources

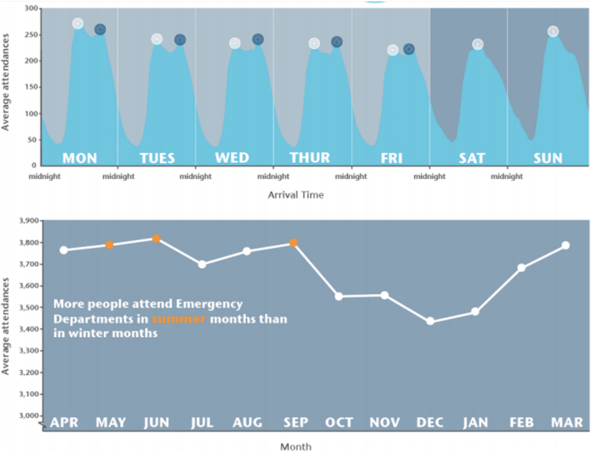


Predictors (eg):
- time since last EA (if any)
- # prescriptions filled
- SIMD (deprivation)

Not available (eg):
- # engagements with primary care
- smoking, marital status, ..

Target definition

Prediction target: Emergency Admission (EA) or death within 1 year after time cutoff



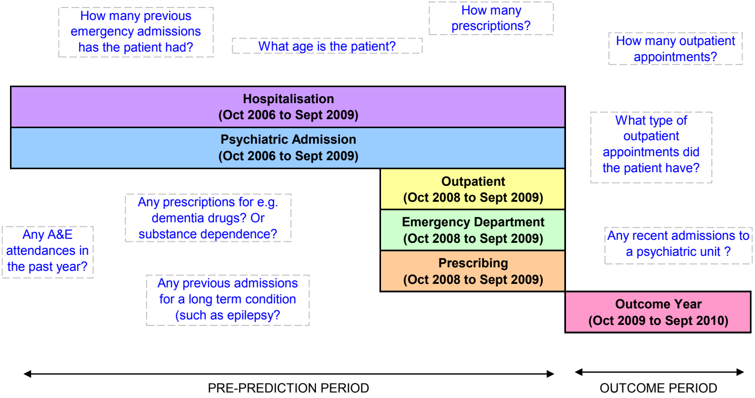
Source: NHS Scotland Emergency Care Report

Motivation

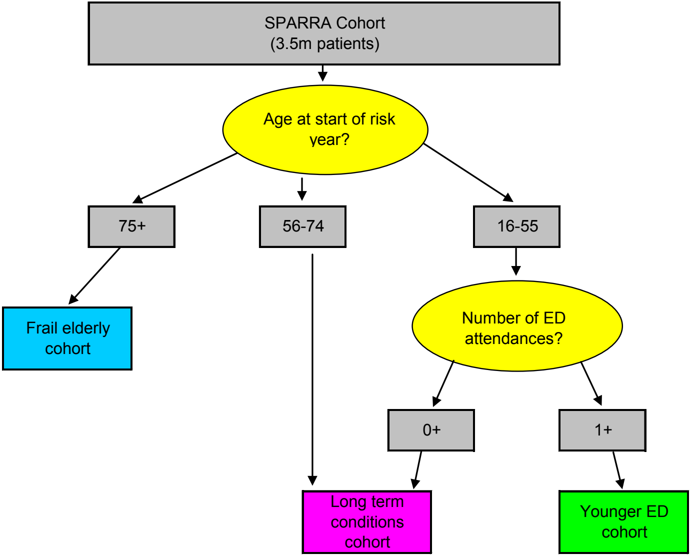
- Do not consider seasonal, weekly or daily variation in risk
- Consider death as similar to EA in implication (may be true in younger people)
- Does not include obstetric admissions

Probabilistic estimate of occurrence.

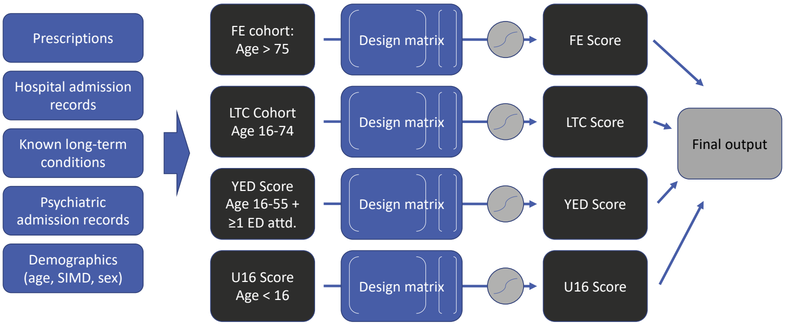
SPARRA v3 details (I)



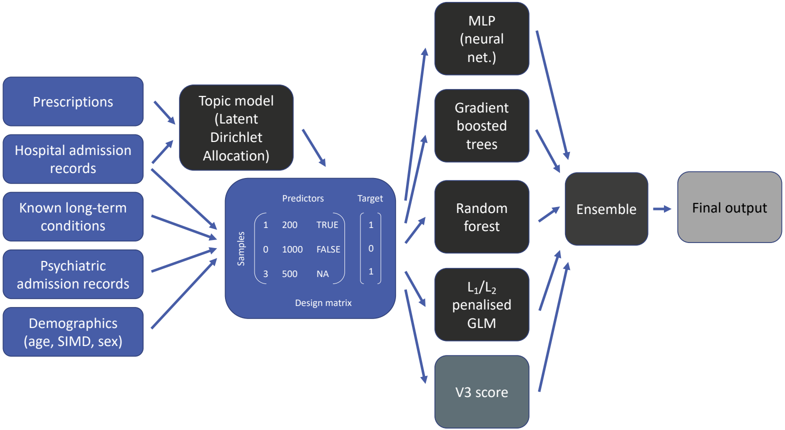
SPARRA v3 details (II)



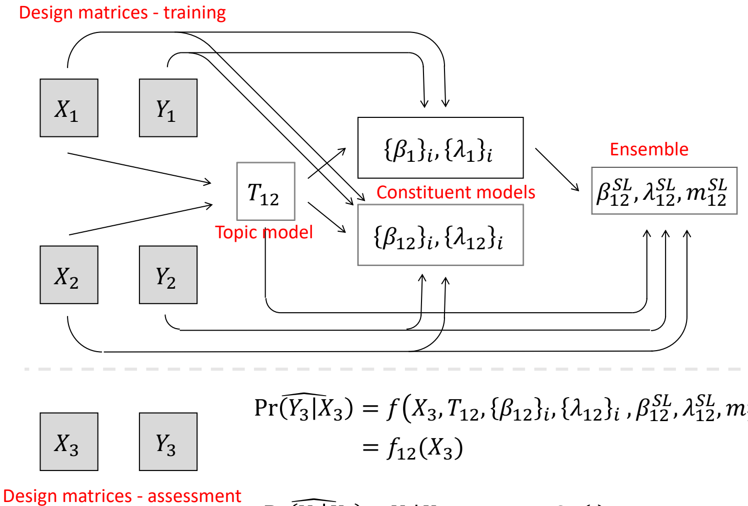
SPARRA v3 details (III)



SPARRA v4 overview



SPARRA v4 cross-validation framework



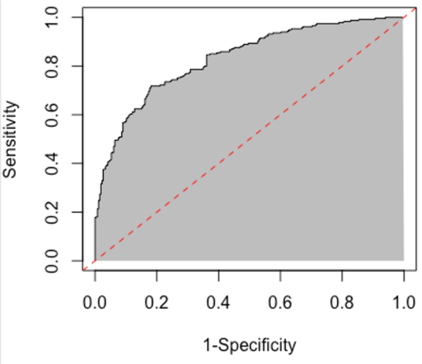
$$\widehat{\Pr(Y_3|X_3)} = f\left(X_3, T_{12}, \{\beta_{12}\}_i, \{\lambda_{12}\}_i, \beta_{12}^{SL}, \lambda_{12}^{SL}, m_{12}^{SL}\right)$$
$$= f_{12}(X_3)$$

$\widehat{\Pr(Y_3|X_3)} \perp\!\!\!\perp Y_3|X_3$ $\qquad f_{12}(\cdot) \perp\!\!\!\perp X_3, Y_3$

Results

ROC and Calibration refresher

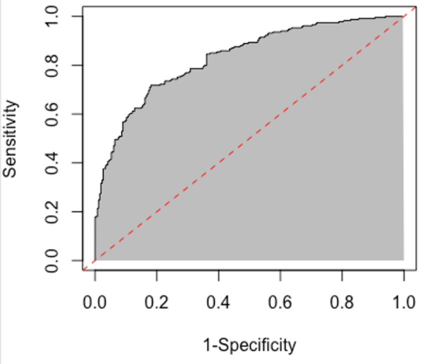
Receiver-operator characteristic (ROC)



Do predictions differentiate individuals who did have an emergency admission from those who did not?

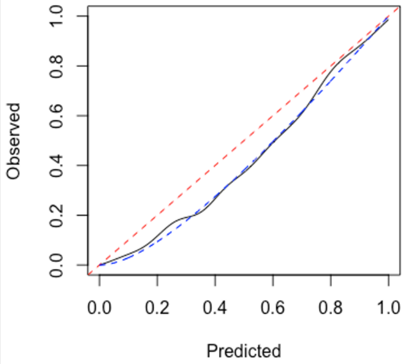
ROC and Calibration refresher

Receiver-operator characteristic (ROC)



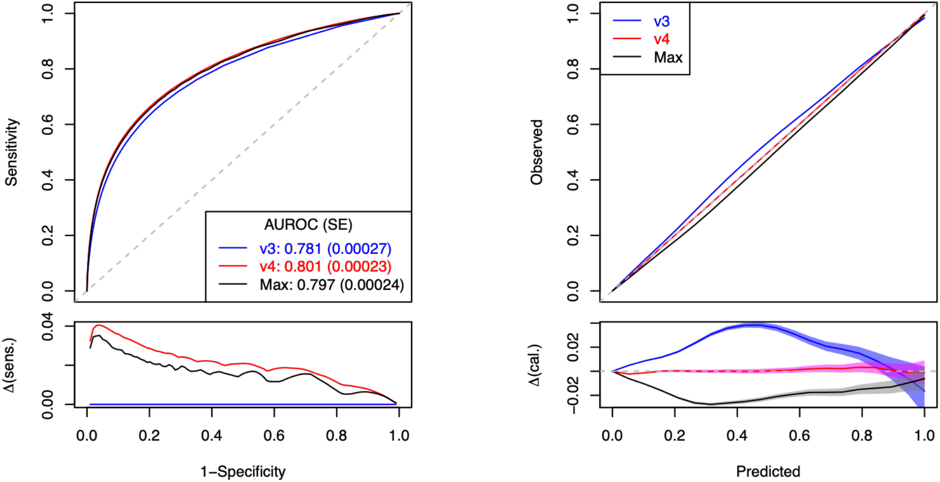
Do predictions differentiate individuals who did have an emergency admission from those who did not?

Calibration plot



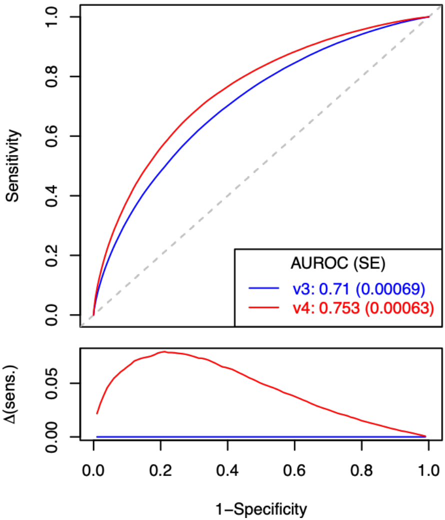
Amongst individuals with a given probability of emergency admission, was the probability correct?

SPARRA v4 overall results

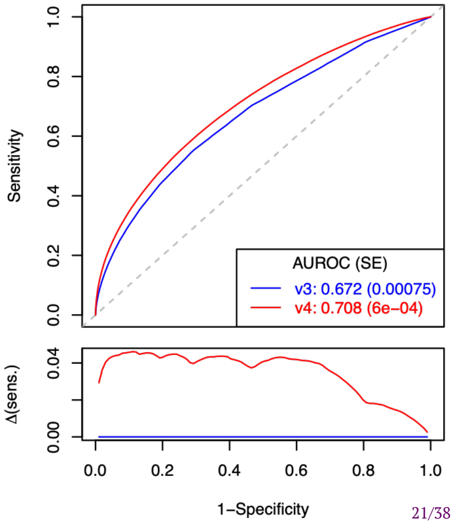


SPARRA v4 challenging cohorts

High-risk individuals
(age>80)

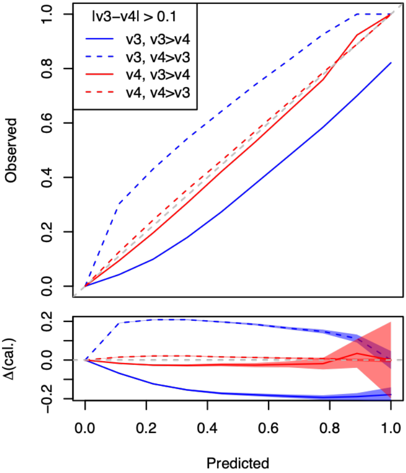


Low-risk individuals
(age 20-70, no previous EA)

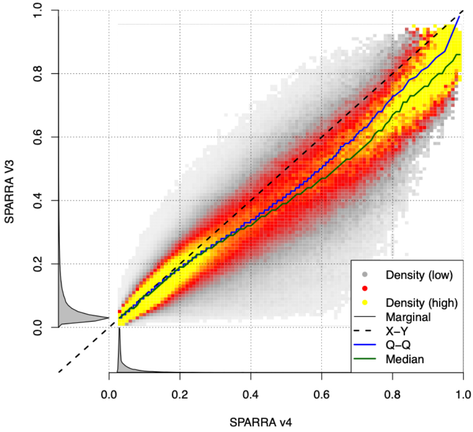


SPARRA v3/v4 direct comparison

Differential risk scores

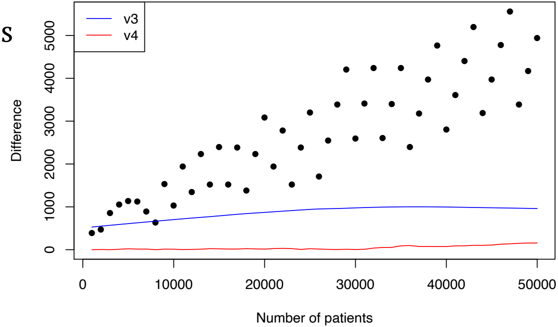


Bivariate density



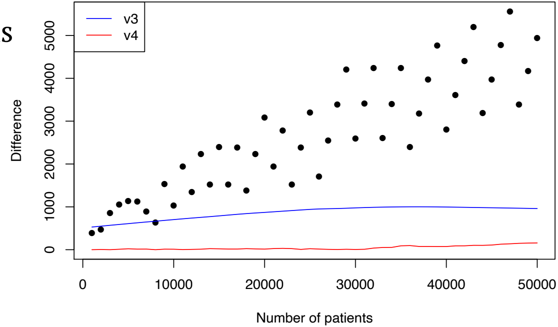
SPARRA v4 interpretable impacts (I)

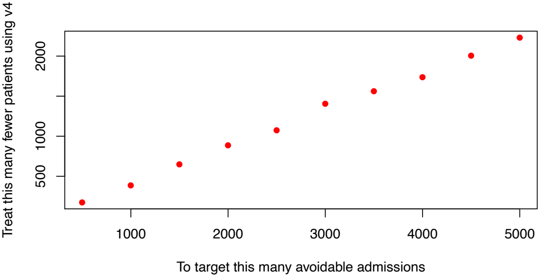
Number of actual admissions among *N* predicted to be most at risk



SPARRA v4 interpretable impacts (I)

Number of actual admissions among N predicted to be most at risk





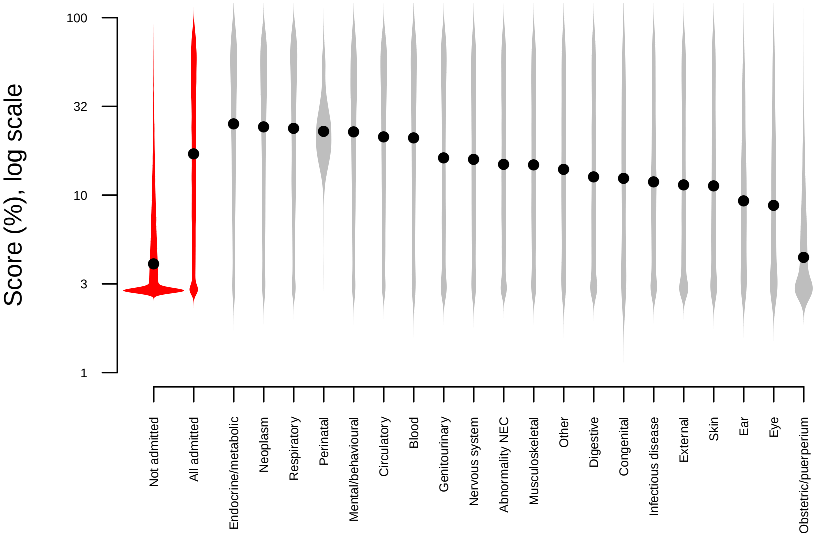
Reductions in targeted intervention required

SPARRA v4 recap so far ...

In other words:

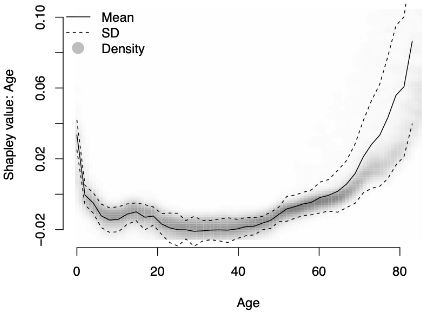
- improvements to calibration in high risk score region
- higher accuracy in challenging cohorts
- upon matching at-risk cohort size to SPARRA v3's top 50,000:
    - recommended follow-up for an extra $\approx$ 4,000 patients who did later undergo emergency admission
    - $\approx$ 4,000 fewer incorrect follow-up recommendations to GPs
    - significant opportunity for improved patient outcomes and NHS cost savings

SPARRA v4 effectiveness by admission type

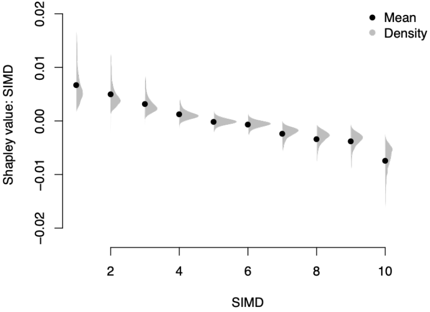


SPARRA v4 Shapley values

Importance of age

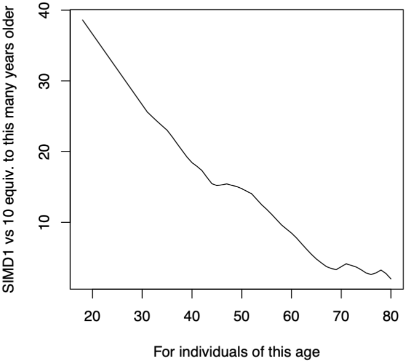


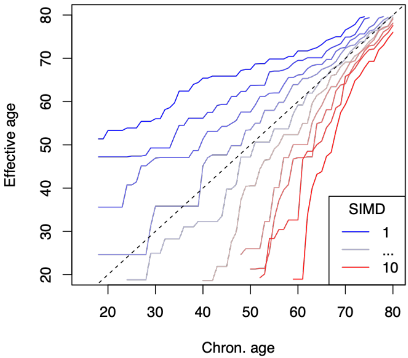
Importance of deprivation



SPARRA v4 age/deprivation equivalence

Using Shapley value to explore age equivalent effect of deprivation levels:





Some thoughts about these results …

- Emergency admissions can be predicted to a potentially useful degree from routinely collected healthcare data on a population scale in Scotland.

Some thoughts about these results ...

- Emergency admissions can be predicted to a potentially useful degree from routinely collected healthcare data on a population scale in Scotland.
- Apparent opportunities for improved patient outcomes and NHS cost savings.

Some thoughts about these results …

- Emergency admissions can be predicted to a potentially useful degree from routinely collected healthcare data on a population scale in Scotland.
- Apparent opportunities for improved patient outcomes and NHS cost savings.
- Contemporary machine learning methods enable meaningfully more accurate prediction on this scale.

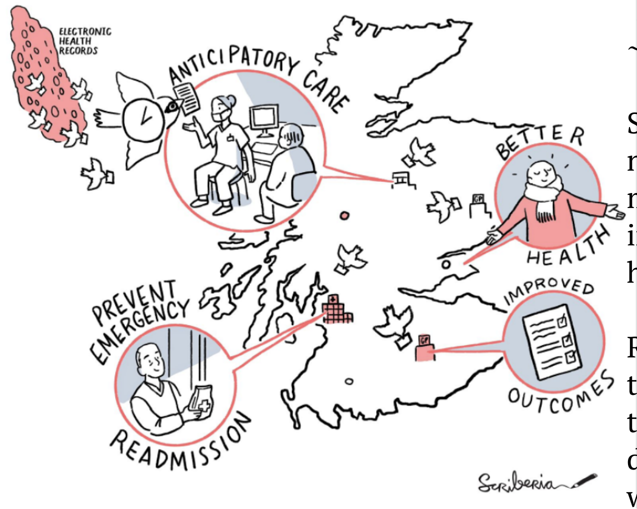
Some thoughts about these results ...

- Emergency admissions can be predicted to a potentially useful degree from routinely collected healthcare data on a population scale in Scotland.
- Apparent opportunities for improved patient outcomes and NHS cost savings.
- Contemporary machine learning methods enable meaningfully more accurate prediction on this scale.
- Certain types of admissions can be predicted differentially well: cancer and endocrine-related admissions are predicted well, eye/ear and traumatic admissions poorly.

Some thoughts about these results ...

- Emergency admissions can be predicted to a potentially useful degree from routinely collected healthcare data on a population scale in Scotland.
- Apparent opportunities for improved patient outcomes and NHS cost savings.
- Contemporary machine learning methods enable meaningfully more accurate prediction on this scale.
- Certain types of admissions can be predicted differentially well: cancer and endocrine-related admissions are predicted well, eye/ear and traumatic admissions poorly.
- SIMD has a substantial effect on EA probability, with the difference between SIMD1 and SIMD10 equivalent to 20-40 additional years of age.

SPARRA v4 deployment



SPARRA v4 deployment ~ Q2, 2022

Scores to be deployed nationwide to GPs and may be used to guide intervention or public health actions.

Reproducibility has been taken seriously throughout and final deployed code/models will be open sourced.

Updating Paradox

The setting

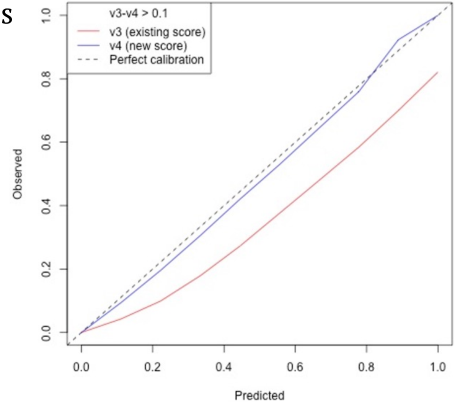
SPARRA v4

- 80% of Scottish population
- Modern machine learning methods
- Up-to-date

SPARRA v3

- 80% of Scottish population
- Logistic regression
- Fitted 2012 and ***in use ever since***
- Can overestimate risk – ***why?***

Healthcare system might have just improved (***concept drift***)



Model updating paradox: what’s happening?

H. Simpson, 45

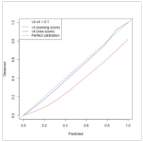
Prev. admissions: 12

Meds: statins, insulin

LTCs: T2 diabetes

True risk now 20%

v3 score: 40%



H. Simpson, 45

Prev. admissions: 12

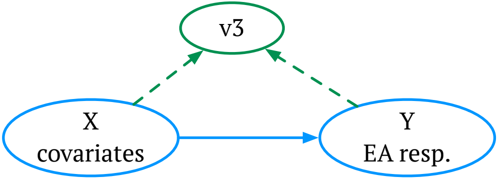
Meds: statins, insulin

LTCs: T2 diabetes

v4 training data

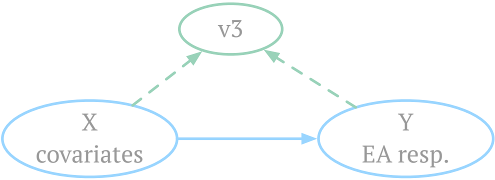
True risk 20%

Model updating paradox: why?

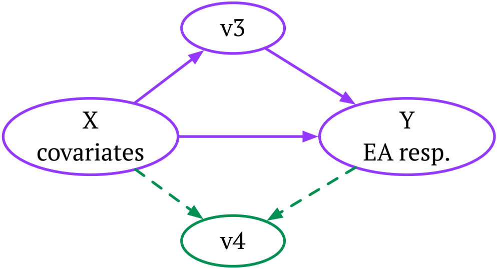


SPARRA v3 trained to blue system

Model updating paradox: why?

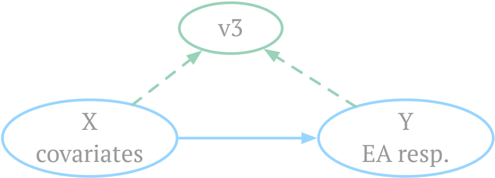


SPARRA v3 trained to blue system

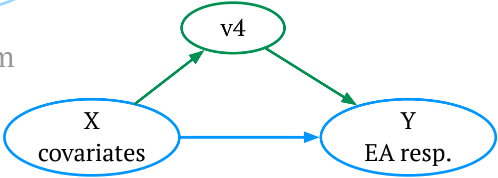


SPARRA v4 trained to purple system

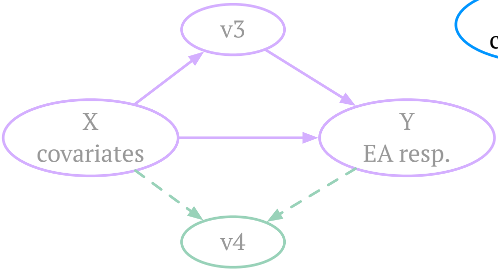
Model updating paradox: why?



SPARRA v3 trained to blue system



SPARRA v4 used for predictions in blue system, which is ***not*** the system it was trained on!



SPARRA v4 trained to purple system

Model updating paradox: is it bad, really?

H. Simpson, 45

Prev. admissions: 1

Meds: ☠toxomycin ☠

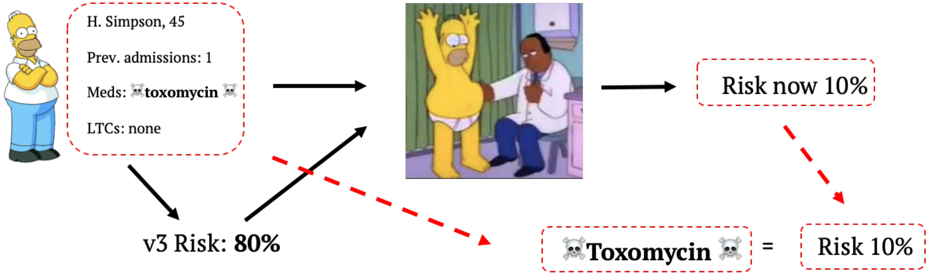
LTCs: none

v3 Risk: **80%**

Risk now 10%

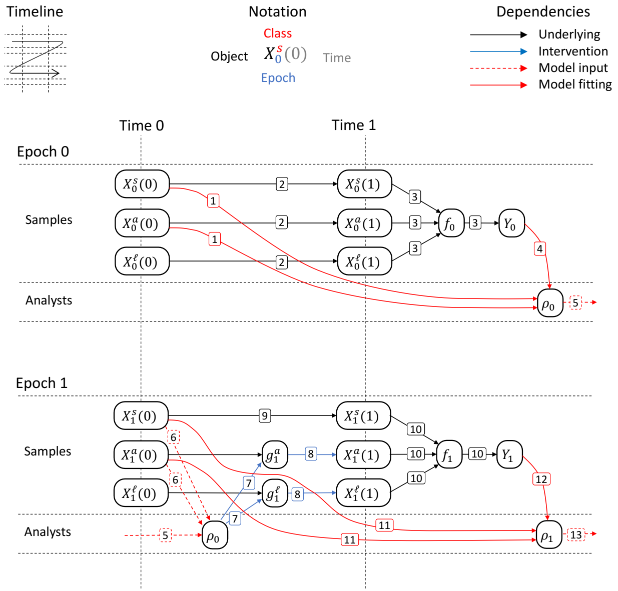
☠**Toxomycin** ☠ = Risk 10%

Model updating paradox: is it bad, really?

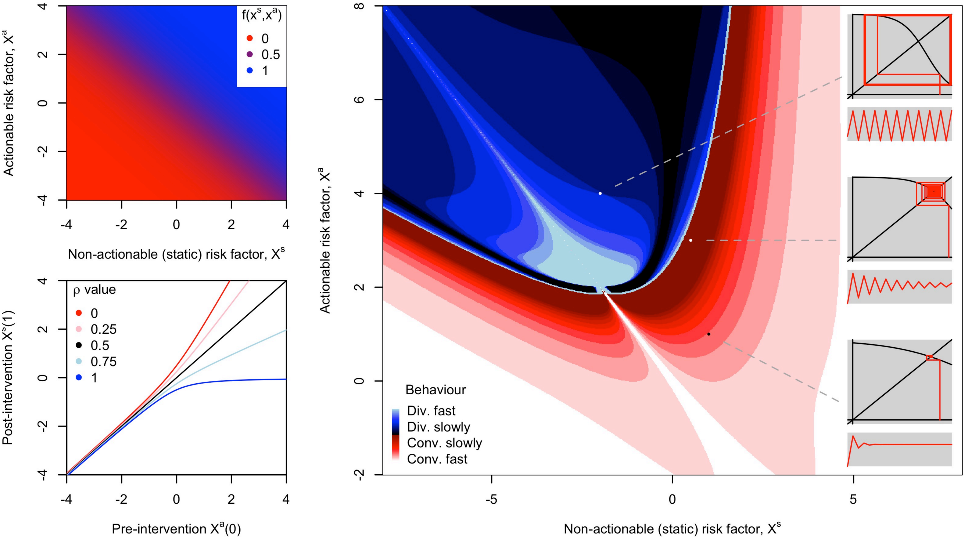


- This effect has been observed in real life (Caruana, 2015)
- This is a problem right now! USFDA 2019 working paper notes RCTs expensive: posits avoiding repeating each time a model is updated.
- The more the score is used, the more it exacerbates the problem
- Can prove the better the model is, the worse subsequent updates will perform!

Naïve model updating + updating framework



Results


f(x^s,x^a)
0
0.5
1
Actionable risk factor, X^a
Non-actionable (static) risk factor, X^s
ρ value
0
0.25
0.5
0.75
1
Post-intervention X^a(1)
Pre-intervention X^a(0)
Behaviour
Div. fast
Div. slowly
Conv. slowly
Conv. fast
Actionable risk factor, X^a
Non-actionable (static) risk factor, X^s

Is equilibrium a bad thing?

H. Simpson, 45

Prev. admissions: 1

Meds: ☠ **toxomycin** ☠

LTCs: none

Score: **80%**

→

→

Risk now 10%

H. Simpson, 45

Prev. admissions: 1

Meds: ☠ **toxomycin** ☠

LTCs: none

Score: **45%**

→

→

Risk still 45%

Possible resolutions?

- Model full causal structure and interventions (practicality?)
- Holdout set (work forthcoming)
- Stacked interventions (J Liley)

Possible resolutions?

- Model full causal structure and interventions (practicality?)
- Holdout set (work forthcoming)
- Stacked interventions (J Liley)

Thank you!