Digital Health and the curse of Missing Data

David Wong, University of Leeds

Vital Signs – a brief intro

Vital signs are recorded by nurses in hospitals

- Heart rate, breathing rate, blood pressure etc.

Deviations in vital signs are associated with poor outcomes*

In practice, deviations are quantified using Early Warning Scores

Overview

1.) Missing and Unreliable Data

We will introduce the concept of missing and unreliable data, how it has caused problems for predictive modelling using three clinicaly examples.

2.) Overcoming Missing Data

Briefly, describe how methods from stats and signal processing can mitigate some problems with missing data. Discuss how problem can be solved by fixing processes.

Examples of research using cleansed data

Examples of large scale research that can be conducted after fixing data collection processes including:

- Evaluation of national standards
- Personalised deterioration models

“While it’s hard to predict the future, we know artificial intelligence, digital medicine and genomics will have an enormous impact on improving efficiency and precision in healthcare

-Eric Topol

1.

MISSING and UNRELIABLE DATA

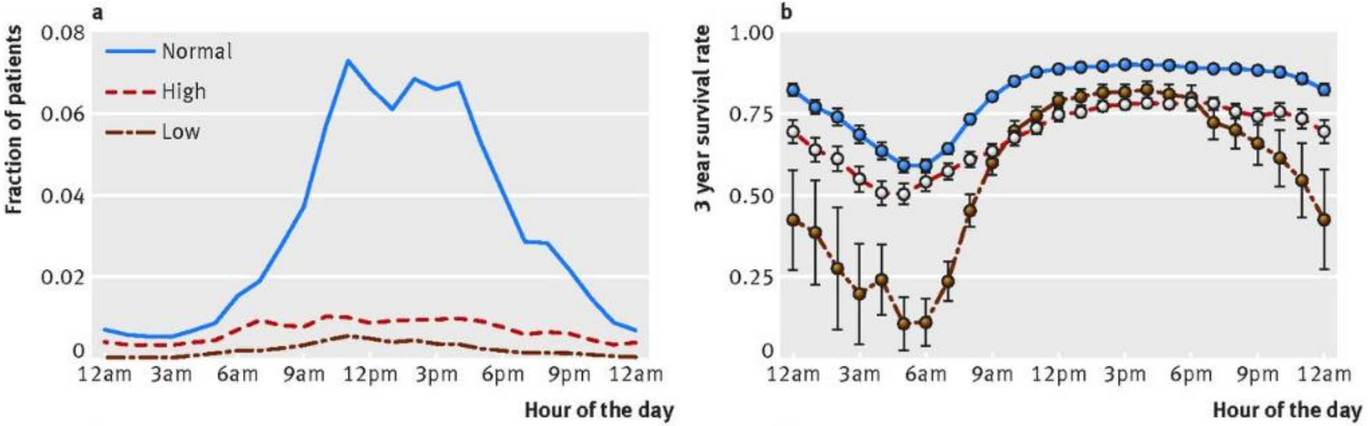
What is it and why does it matter?

1.) Prediction of Mortality

- Attempted to predict survival based on lab test results
- Presence of test, was a significant predictor in 233/272 tests
- Time of test was **more** predictive than test results in 118/174 tests

Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. Agniel et al. BMJ, 2018

1.) Prediction of Mortality



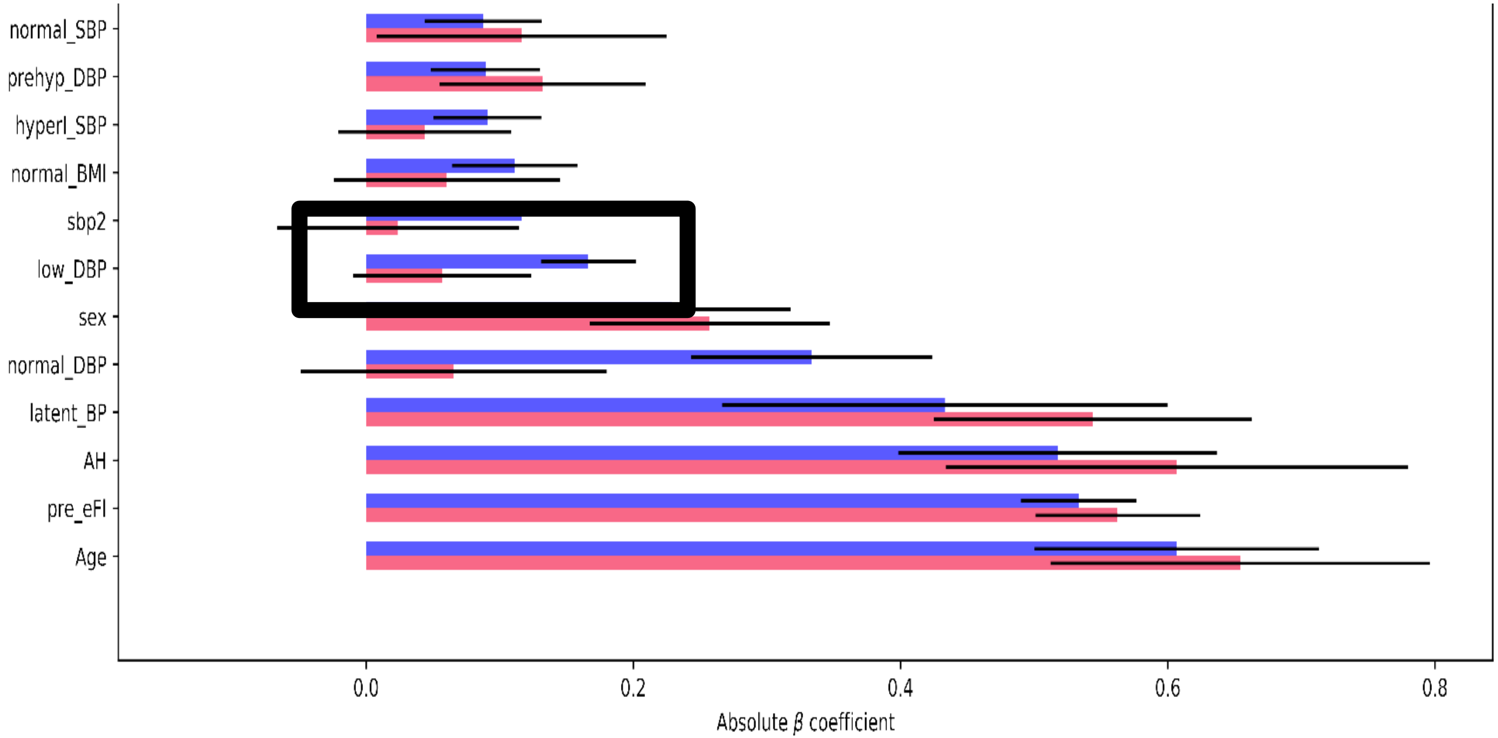
Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. Agniel et al. BMJ, 2018

2.) Prediction of Mortality

- Attempted to predict survival based on primary care variables
- Low Diastolic Blood Pressure was a significant factor
- Impact of low Diastolic Pressure insignificant after accounting for missing readings

Identification of clinical factors associated with poor surgical outcomes in a large primary care data set. Narganes et al. in Proc BioMedEng, 2018

2.) Prediction of Mortality



3.) Symptom Checker Predictions

The Babylon Triage and Diagnostic System – a new implementation after the previous generation [Middleton et al., 2016] – is based on a Probabilistic Graphical Model (PGM) [Koller and Friedman, 2009] of primary care medicine, which models the prior probabilities of diseases and the conditional dependencies between diseases, symptoms and risk factors via a directed acyclic graph.

3.) Symptom Checker Predictions

The structure of the graph (i.e., the connections between diseases, symptoms and risk factors) is created by medical experts and reviewed from a modelling perspective.

3.) Symptom Checker Predictions

Are disease priors and dependencies enough?

- Location
- Time


Home of the Daily and Sunday Express

HOME NEWS SHOWBIZ & TV SPORT COMMENT FINANCE

STYLE LIFE HEALTH CARS GARDEN PROPERTY FOOD TECH DIETS SATURDAY

Home > Life & Style > Health > Measles outbreak: Should your child be given the MMR vaccine as me

'FIRE OR FRYING PAN!' Sturgeon reveals plot to SINK BREXIT and...

BREXIT BREAKTHROUGH: Brexit deal '90 percent' done says Irish...

Brett Kavanaugh wife: Who is Ashley Estes Kavanaugh? Is she...

Measles outbreak: Should your child be given the MMR vaccine as measles cases soar

MEASLES is making a resurgence throughout Europe, as cases of the once eradicated disease skyrocket amid anti-vaccination popularity. One of the best ways to combat the spread is to ensure your child is given the MMR vaccine.

Recap

1.) Mortality

1.) missingness of test result affects the predictive model – those with test at higher risk

2.) presence of test results is conditioned on latent information (doctor's nous)

2.) Mortality (primary care)

1.) missingness of blood pressure measurement affects mortality prediction

3.) Symptom

1.) missingness of 'superfluous' variables, location and time, could affect disease prediction.

DATA

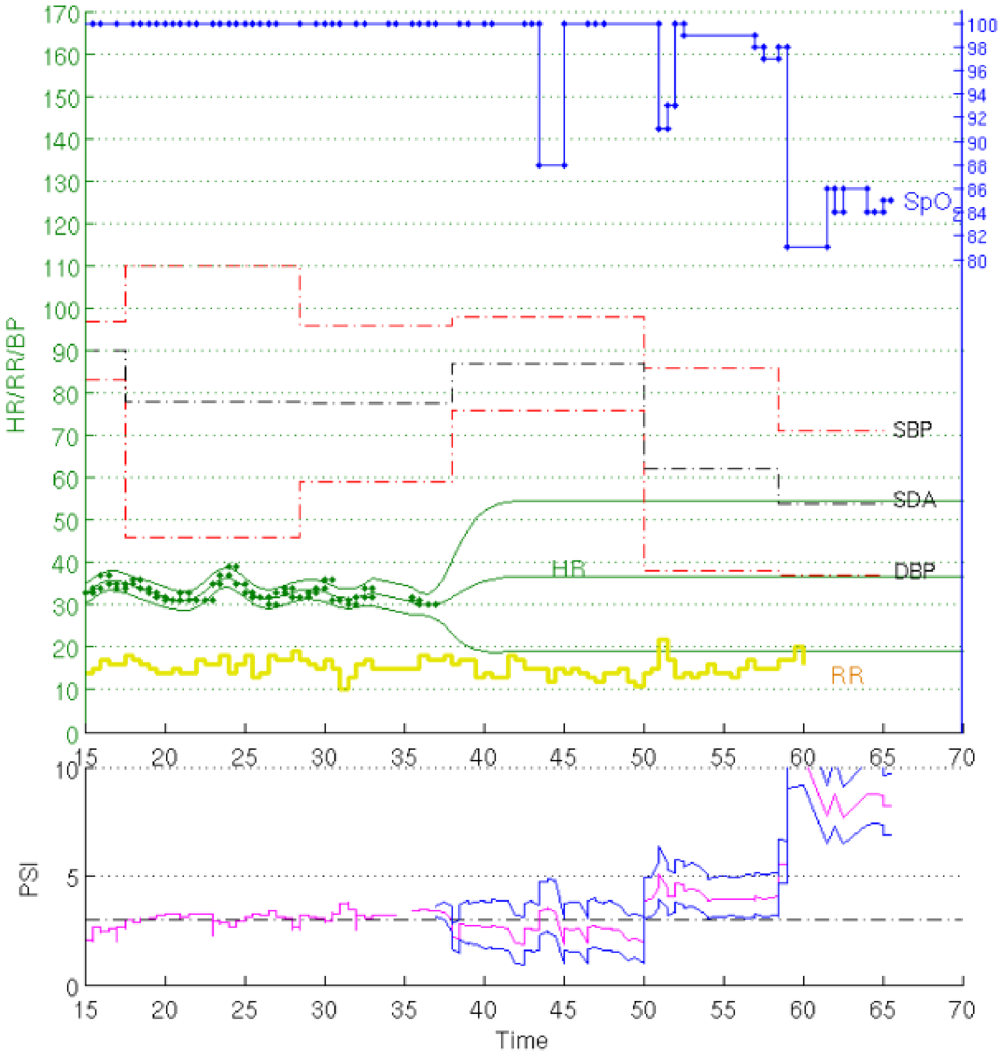
are not agnostic to human bias

2.

DEALING WITH MISSING DATA

3.)Temporal (GPs)

- Model time series as multivariate Gaussian
- Time covariance matrix is based on historic data
- Estimate missing values



The problem: it's hard.

1.) Even for well-defined problems, like multiple imputation, work is computationally intensive

2.) We are making the best of a bad situation – trying to squeeze out a tiny bit more information.

FIX THE INPUT

Make the process of collecting data easier, to get the right variables and reduce likelihood of bias

Common Clinical Opinions

- Grassroots dissatisfaction with clinical systems (https://github.com/dhinet/csus-2016-data)
- General understanding that good record-keeping is important
- Summary – willingness to improve data collection, if it doesn't impact workload.

SEND – System for Electronic Notification and Documentation

Vital Signs – the problem

- C. 2011 – vital sign observations recorded on paper
  - Often illegible
  - Only accessible in one place
  - Goes missing (up to 30%!)

- Early Warning Scores calculated incorrectly in approx. 20% cases

- Opportunity to fix a clinical problem AND fix our data problem

SEND

- Redesign process to make data capture a natural part of the process
- Automate where possible

SEND

- Redesign process to make data capture a natural part of the process
- Automate where possible

SEND

Change Patient
ZZZTEST, CHARLOTTE (F)
01 JAN 1990 (25 years old)
NHS No: Not Recorded
Hospital No: 10152425
DATE
TIME
T&T
TEMP (°C)
HEART RATE (bpm) & BLOOD PRESSURE (mmHg)
NO OBS FOR ≥24 HRS
RESP RATE (/min)
SATS
OXYGEN
AVPU
GCS
Concern
FRI 04/09/2015
TIME
now
TEMPERATURE
BLOOD PRESSURE
HEART RATE
RESPIRATORY RATE
OXYGEN SATS
OXYGEN THERAPY
CONSCIOUSNESS
NURSE CONCERN
T&T SCORE
View Chart
Save obs
Temperature
C
patient refused
delete
Δ Previous
Next ∇
Chris FULTON is currently logged in
Log Out

SEND –clinical advantages

- Documentation – at least as fast as paper
- Information reviewable remotely – critical outreach teams can view data from anywhere
- Automation of regularly audits

SEND – missing data advantages

- Multi-hospital data set.
- Data validated according to our own input rules
- Design process has helped in identifying further potential biases

65s*

'Average' time saving per observation

2m vital sign sets

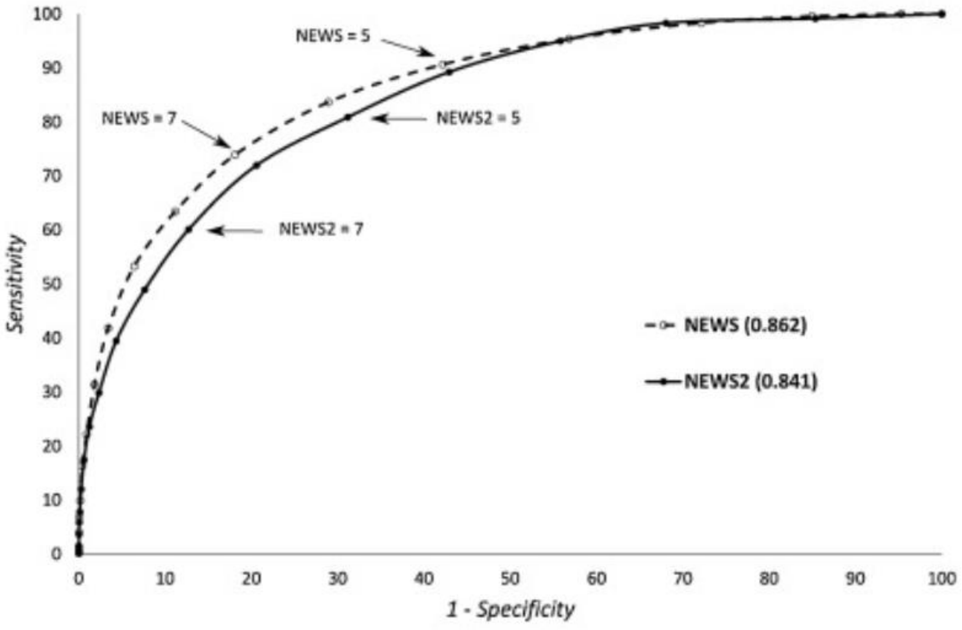
Across two hospital trusts

A ward-based time study of paper and electronic documentation for recording vital sign observations. Wong et al. JAMIA, 2017
SEND: A system for electronic notification and documentation. Wong et al. BMC MIDM, 2015

3.

APPLICATION OF SEND DATA

EVIDENCE-BASED EVALUATION OF NATIONAL STANDARDS



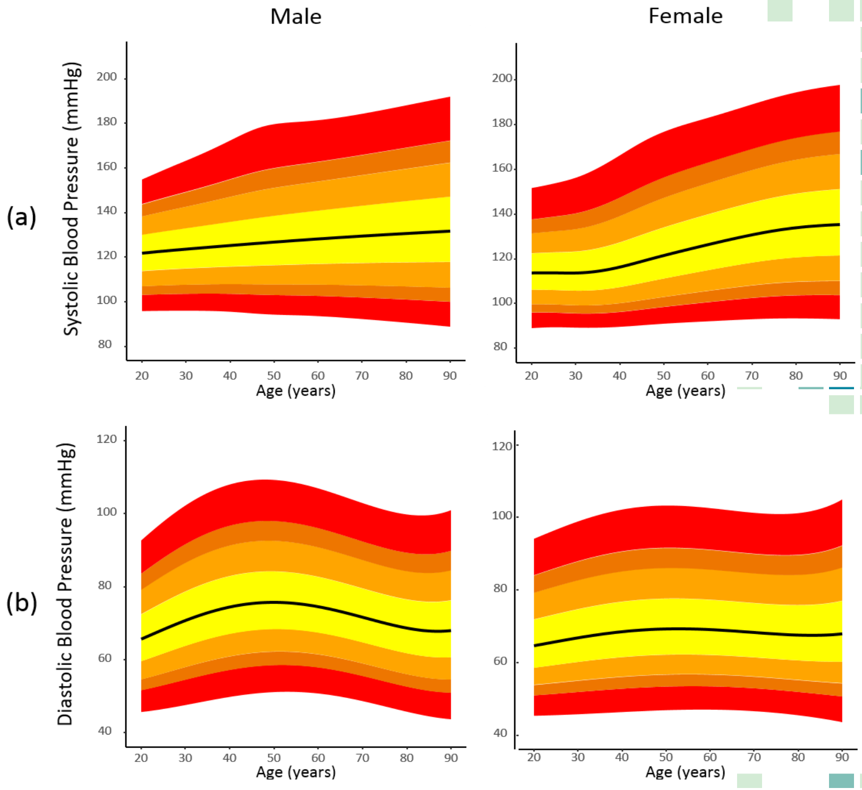
A comparison of the ability of the National Early Warning Score and the National Early Warning Score 2 to identify patients at risk of in-hospital mortality: a multi-centre database study

Pimentel et al, Resuscitation, 2018

Large cross-sectional studies with reliable data

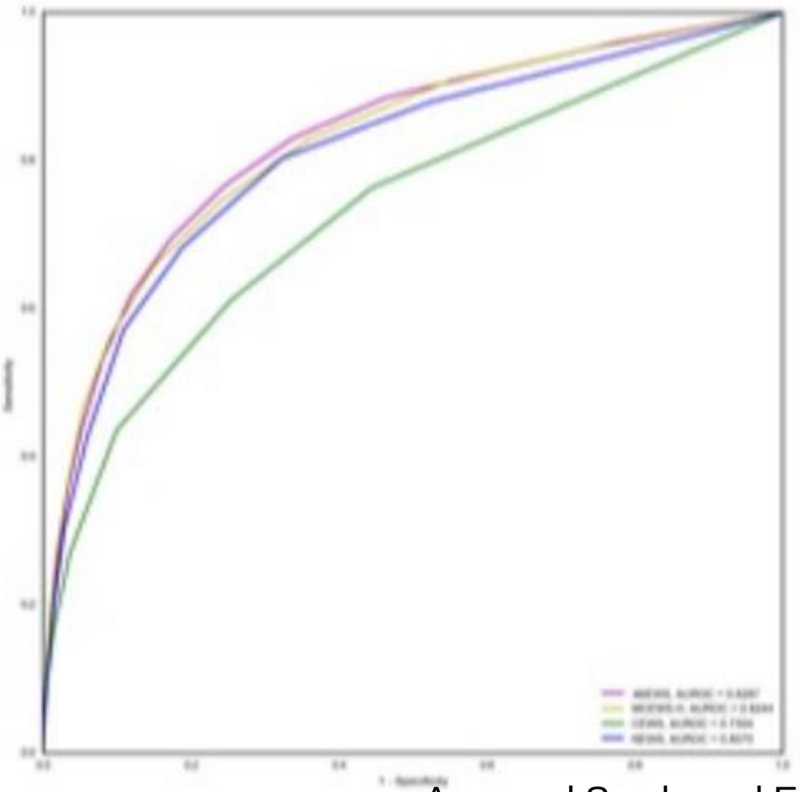
Blood Pressure Centiles

- Generation of blood pressure centiles by age and sex.
- Sub-group analysis by:
  - Emergency vs non-emergency
  - Hyper- vs normo-tensive



Cross-sectional Centiles of Blood Pressure by Age and Sex in a Large Hospital Population

Personalised Models of Deterioration



Inclusion of age-based centiles into Early Warning Score

Performance exceeds National Early Warning Score

Performance particularly better for younger adults (not shown here)

Age- and Sex-based Early Warning Scores. Shamout et al. [In Prep]

Overview

1.) Missing and Unreliable Data

We will introduce the concept of missing and unreliable data, how it has caused problems for predictive modelling using three clinicaly examples.

2.) Overcoming Missing Data

Briefly, describe how methods from stats and signal processing can mitigate some problems with missing data. Discuss how problem can be solved by fixing processes.

3.) Examples of research using cleansed data

Examples of large scale research that can be conducted after fixing data collection processes including:

- Evaluation of national standards
- Personalised deterioration models

THANKS!

Any questions?

You can find me at:

@drdavecwong

d.c.wong@leeds.ac.uk