



Pitfalls in health data science

Who am I?



Data Scientist with Orion / PDH



Work with variety of hospital based clinicians



Interrupt me at any time

What do I
work on?

Interpretability

Surgical
outcomes

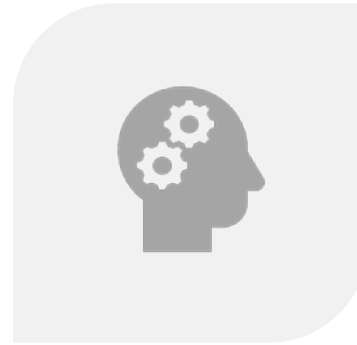
MSD + Orion
consulting

Optimization

What am I
talking about
today



NMDS – A VALUABLE NZ
DATA SOURCE



EXAMPLE OF IT'S USE FOR
RESEARCH




LESSON'S & PITFALLS
FROM CLINICAL
COLLABORATION

NMDS – a valuable NZ data source

“The National
collection
information,

**Routinely collected
data was not
collected for what
you want to use it for**

is a national
discharge
or inpatients



What can
you find?

Operations – ACHI coding

Diagnosis coding – ICD codes

Patient demographics

Location information

Mortality linked

Readmissions

Things that could be important to billing

Sometimes clinical information lacking

Example research - NZRISK

- Validation in a NZ context
- Improved, NZ specific tool
- Two years complete data used, ~360,000 patients
- Functional tool TODAY at www.nzrisk.com
- Implementation for triaging at ADHB



PDH | NZRISK in action

Browser address bar: <https://www.nzrisk.com/#calculate>

Navigation: Home About Calculate Contact

Calculate


User notes

ASA-PS (American Society of Anaesthesiology – Physical Status) Score

1. Normal healthy patient
2. Patient with mild systemic disease
3. Patient with severe systemic disease
4. Patient with severe systemic disease that is a constant threat to life
5. Patient who is moribund and not suspected to survive without the operation

Active malignancy

Cancer that is being actively treated, recurrent, metastatic or inoperable. This definition excludes squamous skin cancer and basal cell carcinoma.

Age (in years, 18 or above)	<input type="text" value="Enter age in years"/>
Gender	<input type="radio"/> Male <input type="radio"/> Female
Ethnicity	<input type="text" value="-- Select --"/>
ASA	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5
Acuity	<input type="checkbox"/> Tick if acute
Cancer	<input type="checkbox"/> Tick if cancer present
Specialty	<input type="text" value="-- Select --"/>
Sub	<input type="text" value="Select Specialty first"/>
Procedure	<input type="text" value="Select Specialty first"/>
Please complete	<input type="checkbox"/> I'm not a robot 

Contact

PDH | NZRISK in action

The screenshot shows a web browser at the URL <https://www.nzrisk.com/#calculate>. The website has a dark blue header with the 'nzRISK' logo and navigation links for 'Home', 'About', 'Calculate', and 'Contact'. The main content area has a red background for the 'About' section and a white background for the 'Calculate' section.

About

Prior to surgery, patients and clinicians decide whether to proceed, partly determined by the balance of risk and benefits. In addition to clinical assessment, risk assessment tools can help to inform these decisions.

nzRISK has been developed to be a simple and easy to use risk calculator based on the New Zealand population. It has been compiled using data from over 270,000 patients aged 18 or over undergoing non-cardiac surgery. It includes eight risk factors that can be entered below to give a 30 day, one and two year estimate of mortality. Each risk factor is associated with mortality and improves the performance of the calculator.

For more information see the publication in the British Journal of Surgery. The nzRISK tool is available on this website for personal use only.

Development

This work was developed by clinicians at Auckland City Hospital, through the Precision Driven Health research partnership with the support of the Perioperative Mortality Review Committee, Ministry of Health, New Zealand.

Calculate

Mortality

33.6% at 30 days

78.2% at 1 year

88.1% at 2 years

In the next 30 days mortality would be expected in 33.6 of every 100 similar people having this procedure, as shown in the right-hand graphic. At one year death would be expected in 78.2 of every 100 similar people having this procedure, rising to 88.1 at two years.

[New Calculation](#)

Time Period	Mortality Rate (%)
30 days	33.6%
1 year	78.2%
2 years	88.1%

Contact

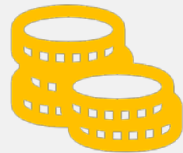
NMDS – how do you get access?



Email and communicate with MoH



Need appropriate ethics beforehand



Cost involved but not huge

\$400 for 5 years data (for us)

Lessons and pitfalls from clinical collaboration

Collaboration turns administration data

```
graph TD; A[Collaboration turns administration data] --> B[Clinically meaningful information]; B --> C[Data science];
```

Clinically meaningful information

Data science

What trips up a data scientist?



Real world meaning of data



Terminology



One word – many meanings in clinical context



How does what you see in the data relate to a patient? Does it make sense clinically?

What does a data scientist need to remember?



What does missing data mean? In this dataset + to a clinician



Statistics isn't everything for selection



Conclusions need to have clinical plausibility



Solve the problem collaboratively, not in your way



Lots of time figuring out how to apply concepts to your data

Things
not as
obvious
to a
clinician?



How to think programmatically



What can a computer do fast?



Difficulty differences for
problems.

WHEN A USER TAKES A PHOTO,
THE APP SHOULD CHECK WHETHER
THEY'RE IN A NATIONAL PARK...

SURE, EASY GIS LOOKUP.
GIMME A FEW HOURS.

... AND CHECK WHETHER
THE PHOTO IS OF A BIRD.

I'LL NEED A RESEARCH
TEAM AND FIVE YEARS.



IN CS, IT CAN BE HARD TO EXPLAIN
THE DIFFERENCE BETWEEN THE EASY
AND THE VIRTUALLY IMPOSSIBLE.

What people
sometimes
forget about
data?



Why was the data collected and by who?



Not everything that should be recorded is



Data is usually written by non-experts



What information is meaningful for
modelling

Pitfalls of translating personal data



Appropriate level for communication



Communication of correlation vs causation - sensitive subjects



Respect other peoples ideas – multiple ethnicity example



Who is and isn't collected / represented

Working together



Set up frame works that operate in each other's thought spaces



Let clinical lead define the variables first from a data set



Data Scientist can iterate over versions of variables



Keep it visual and explainable