Health Data Reuse: ‘Do nothing’ is not an option

IMIA EU Summit on Trustworthy Reuse of Health Data, Brussels, 14th May 2012

Iain Buchan
Troublesome Threes

• 3 Ingredients
  – Data; Models; Expertise

• 3 Myths
  – Big data warehouses are the solution
  – Science provides the models to utilise the data
  – Clinicians will continue to be the main source of data

• 3 Pipelines
  – R&D; Quality Improvement; Payor & Public Health
Body Mass Index (BMI) trend in Wirral 3y-olds from 1988 to 2003

Three-monthly rolling average BMI SDS

Clues

Actions

SDS = standard deviation score from 1990 British Growth Reference charts – adjusts for age and sex of the child
Easily Disregarded Data

Trends in liver function sought in patients with long-term conditions.

Most did not have regular repeat tests. Complex time structure to model.
Improved Liver Function on Starting Glitazone

Decrease (pooled) = 0.15 (se=0.009), p<1x10^{-59}

Similar irrespective of co-prescriptions {G = 0.12; GM = 0.16; GS = 0.16; GSM = 0.15
G = glitazone, M = metformin, S = sulphonylurea}
Rosiglitazone to Pioglitazone: No Liver Signal

Lines almost join $\rightarrow$ no evidence of a change in ALT
Trapped Data

• Narrative
  – Vioxx risk of MI detectable pre-2004
    • By text mining 154k RA patient notes
    • Not by ICD-9 codes from same notes
    • Repeated for 2 other risks

• Rubrics
  – Code for type 2 diabetes annotated “DM r/o”
  – Access blocked by some ethics committees
UK Diabetes Prevalence from Different Databases

Estimated UK Prevalence of Diabetes

- GPRD - diagnostic codes
- GPRD - all codes
- THIN
- Q-research
- QOF

Data from Tim Doran, University of Manchester
Anaemia at lower levels of kidney impairment than commonly thought

Clinical (audit) question leading to scientific finding: required local metadata (assay change) not in national datasets

Anaemia at lower levels of kidney impairment than commonly thought
Data Warehouses are not the Solution

• Need contextual metadata continuously adding value to big data

• Need engaged communities adding the metadata

• OMOP lessons
  – Current big data only for crude effects
  – Different answers from different datasets & methods
Big Data in Context

Datasets
(+ models)
(searched by experts)

Data ∪ Models ∪ Expertise
“sense-making network”
Fragmented Evidence & Workflows

Patient is the union, not the sum of pathways
Need usefully complex models of health & care

Weight control
Physical activity

Nephrology: Blood pressure focus

Primary Care

Public Health

Secondary Care

Self Care

Diabetology: Glucose focus

Specialist A

Specialist B

Weight → Blood pressure
Health Records & Knowledge Silos

Data-intensive Paradigm Shift

Open Unifying Modelling: Across mechanisms and contexts

Expertise

Health Avatars & Dynamic Models

○ models = Avatar

Electronic Health Records (eHR)

Unified Graphical Model

Data-intensive Paradigm shift

Multi-scale & Multi-system Health:
• Research
• Policy
• Care

Large scale inference

Model refinement

Data

• Expertise

E.g. Chronic obstructive pulmonary disease

E.g. Coronary heart disease

E.g. Lung cancer
Not About Mining Big Data

Problem Space  Observation Space  Data Space

...like squinting at an image through a doyley and prism

Need to harness networks of reasoning about models

Not just structure in big data

\[ y = b_1 x_1 + b_2 x_2 + b_3 x_3 + c \]
Model-based Machine Learning

• Suspected myth: false division of children into allergic tendency (atopy) or not

• Life-course data: birth cohort of 1,000 children from Manchester with careful measurements

• Approach: unsupervised search for patterns of sensitisation \( \rightarrow \) shape hypotheses
Unsupervised Clustering of Allergic Sensitisation Development

Switch Between Sensitization Classes

Sensitization Class

Allergens

Probabilities:
- Probability of Sensitization at Age 1
- Probability of Gaining Sensitization
- Probability of Retaining Sensitization
- 3 intervals (1-3, 3-5, 5-8)

Acquired Sensitization at Age 1
- Skin Test at Age 1
- IgE Test at Age 1

Acquired Sensitization at Age 3
- Skin Test at Age 3
- IgE Test at Age 3

Acquired Sensitization at Age 5
- Skin Test at Age 5
- IgE Test at Age 5

Acquired Sensitization at Age 8
- Skin Test at Age 8
- IgE Test at Age 8

Children (1053)

To Infer

Probabilities:
- Probability of Positive Skin Test Given Sensitized
- Probability of Positive Skin Test Given NOT Sensitized
- Probability of Positive IgE Test Given Sensitized
- Probability of Positive IgE Test Given NOT Sensitized
From 2 to 5 Useful Classes of Atopy
Predicted Real-world Outcomes

Admitted at Any Age

First Admitted > 3 Years Old
(remove early virus wheeze)
Mirage of Evidence

Reality: feedback is too little too late

Algorithms may be out of date by the time they are “validated”

Note EU Directive 2007/47
Predicting Mortality from Cardiac Surgery

In-hospital mortality proportion

- Observed in UK (SCTS audit)
- Expected via EuroSCORE I model

Drifting EuroSCORE Calibration

Source: G. Hickey 2012
Re-inventing Reuse

• 3 Pipelines
  – Research and Development
  – Provider Quality Improvement
  – Payor & Public Health Evidence

• 1 Bottle-neck
  – Human resource to make sense of data
We Could Borrow Strength

**Clinical Trial:**
1 vs. 2 per day dosing → adherence (* deprivation)

**Clinical Audit:**
Depression vs. readmission

**Public Health:**
Recalibrated deprivation score

“Users who selected the variables in your basket also selected these variables and these models...”
Health Intelligence ‘e-Lab’ from/for a Population

Anonymised linked records

Firewall 1

Analysis

2nd Level Privacy Control

Firewall 2

Unified Health Intelligence Platform

External collaborator/client
Data queries: From this...

```sql
```
...to this
Clinical Study Feasibility Assessment

MASCOT

Options
- Create new study
- Create copy of this study
- Cancel Study
- Return to list of studies

Protocol Templates
- Create Template

Study Overview

State
Feasibility

Study Members
- lucy (Creator)
- bruce (Viewer)
- gary (Editor)
- markdelderfield (Editor)
- normanstein (Editor)

Baseline

Demographics
- 885 people

Additional Criteria
- Include: Has Asthma
  - Asthma
  - AND Asthma attack diagnosis made less than 12 Months ago
  - <5 people
- Include: Asthma Medications
- Exclude: BMI - Overweight

Recruitment Estimate
- <5 people
Clinical Study Recruitment

Options
- Create new study
- Create copy of this study
- Cancel Study
- Return to list of studies

MASCOT

Protocol Templates
- Create Template

Study Overview

State
- Approved

Study Members
- lucy (Creator)
- bruce (Viewer)
- gary (Editor)
- markdelderfield (Editor)
- normanstein (Editor)
- pat (Editor)
- sarahthew (Editor)
- steph (Viewer)

Add patient consent form
- Choose File
  - No file chosen

and patient information sheet (if separate)
- Choose File
  - No file chosen

Add ethics approval letter
- Choose File
  - No file chosen

Patient invitation letter template

```
[[GPSurgeryName]]
[[GPAddressLine1]]
[[GPTown]]
Dear [PatientFirstName] [PatientLastName]

We would like to invite you
```

Patient

Address Line 1 Address Line 2 Address Line 3 Town County Postcode Title Firstname Lastname

GP

Surgery Name Address Line 1 Address Line 2 Address Line 3 Town County Postcode Firstname Lastname Qualifications
Preserving Consent for Consent

Integrated Health Record

Local Health Community

Study Recruitment Tool

Clinical Care Boundary

Study Protocol Designer

Clinical Care Boundary

Patient-Clinician Relationships

Essential Local Metadata

e-Lab

SQL

Anonymised Repository

HTTPS

Query Objects

Clinician

Researcher / Manager

Bulk ETL with Anonymisation

email

HTTPS

SQL

Patient-Clinician Relationships

Clinical Care Boundary
HBA1C by ward

Mean HBA1C values by ward

Legend Title
- 6.63 - 6.81
- 6.81 - 6.97
- 6.97 - 7.09
- 7.09 - 7.12
- 7.12 - 7.40

Value
Mean: 7.106
Sample Size: 16
e-Lab Currency: Work Object

Visible Value-Adding

Socially Stimulating Science & Service Intelligence, In-silico
Data Sharing: Contextualised

Local Community Integrated Health Record

Depersonalised records

Collaboration with other trusted e-Labs

Local uses

Shared Work object

Each locality can police the passage of objects at its boundary.

Changes to objects are fully auditable.
“Borrowing Strength” along Service Buses

Federation of e-Lab communities shares work objects without remote data warehousing.

Strength is borrowed and costs reduced by pooling expertise.
Toward m-Health: n-of-1 Gateway

• Dark areas
e.g. adherence vs. antibiotic resistance

• Current signal
“How did you get on with your medicine?”

• m-Health
  – Smartphone app. with drug
    • Quality of life scores
    • Position and acceleration
  – Watch/plaster
    • Temperature
    • Pulse

Patient/Citizen

Co-produced Health Record

Algorithm/ System

Clinician/ Health Ally
Exploit EU Heterogeneity

Diversity of risk & treatment environments across EU could be harnessed to tease out evidence with smaller overall samples faster than at present

Conclusion

• Current healthcare evidence predicts less than 30% of true outcomes

• Smart networking of data, models and expertise across EU environments would provide a world leading engine for health system adaptation

• Do nothing is not an option

Rousseau: “Social Contract”