E-Scaling Science for Public Health

National eScience Forum, Edinburgh, 10th March 2011

Iain Buchan
Professor of Public Health Informatics
& CSO NorthWest e-Health
This Talk

• Problem:
  Path of least resistance = know more about less
  Neglected = complexities of health & healthcare

• Solution:
  Collective insight ≠ e-plagiarism

• Deployment:
  Unpack research messages to activate pack insight
  ... descale to e-scale
3 Big Gaps in Health Knowledge

1. Prevention

2. Trials $\rightarrow$ real-world outcomes

3. Dynamics of health (care) systems
Prevention

• Healthcare affordability depends on it

• Little evidence of what works
  – Personalisation
  – Frequent reinforcement
  – Autonomy

• Most risks are interwoven: most research is not
Puddles of research around the organising principle … but policies need the big picture
Women and not men from low-income households are fatter in England.
Women from low-income households and men from high-income households are fatter in Greater Manchester.

Data from Health Survey for England.
Real-world Healthcare Evidence

• Clinical trials usually exclude
  – Women of child bearing age
  – People with multiple conditions
  – People on other medications

• More than two thirds of what will happen to patients when treated is unpredictable with current evidence
Digital Bridges Since 1990s:
Integrated Care **Pathways (Disease-specific)**
**Missing:** Patient & Community

‘**Big-picture**’ Across Diseases/Services/Pathways

- **Self Care**
  - Diabetology: Glucose control
  - Ophthalmology: Diabetic eye care
  - Nephrology: Chronic kidney disease

- **Clinical Care**

- **Primary Care**

- **Secondary Care**

- **Specialist A**

- **Specialist B**

**Future:** Realistically complex and dynamic models of individual or community care: i.e. Mr Smith’s care pathway, not diabetes + eye + kidney care pathways
Data aggregation isn’t enough!

Methods/Models/Applications ↑

Experts ↔

Data ↑↑↑↑
Typical Health Sciences Signal Path

Problem Space  Observation Space  Data Space

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

Health eScience must harness observers & thinkers not just data & methods
Dynamics of Health Systems

- Patchy early warning systems
- Report culture rather than watchfulness
- Disease & organisation specific budgets rather than tactical targeting of resources to needs
- Lack of “what if?” scenario planning tools
Early Data, Late Actions

Body Mass Index (BMI) trend in Wirral 3y-olds from 1988 to 2003

Three-monthly rolling average BMI SDS

SDS = standard deviation score from 1990 British Growth Reference charts – adjusts for age and sex of the child
An inconvenient public health truth dating back to 1066

TWO COUNTRIES: N./S. ENGLAND

<table>
<thead>
<tr>
<th>Region</th>
<th>Life Expectancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>North West</td>
<td>76.3</td>
</tr>
<tr>
<td>Yorkshire and Humber</td>
<td>77.1</td>
</tr>
<tr>
<td>South West</td>
<td>79.0</td>
</tr>
<tr>
<td>South East</td>
<td>79.2</td>
</tr>
<tr>
<td>London</td>
<td>78.2</td>
</tr>
<tr>
<td>East of England</td>
<td>78.9</td>
</tr>
<tr>
<td>West Midlands</td>
<td>77.2</td>
</tr>
<tr>
<td>East Midlands</td>
<td>77.8</td>
</tr>
<tr>
<td>Wales</td>
<td></td>
</tr>
<tr>
<td>North East</td>
<td>76.5</td>
</tr>
</tbody>
</table>

Boundaries of Government Office Regions in Northern and Southern England showing Female Life Expectancy at birth 2006-2008

<table>
<thead>
<tr>
<th>Region</th>
<th>Life Expectancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>North West</td>
<td>80.6</td>
</tr>
<tr>
<td>Yorkshire and Humber</td>
<td>81.3</td>
</tr>
<tr>
<td>South West</td>
<td>83.1</td>
</tr>
<tr>
<td>South East</td>
<td>83.0</td>
</tr>
<tr>
<td>London</td>
<td>82.7</td>
</tr>
<tr>
<td>East of England</td>
<td>82.8</td>
</tr>
<tr>
<td>West Midlands</td>
<td>81.8</td>
</tr>
<tr>
<td>Wales</td>
<td></td>
</tr>
<tr>
<td>North East</td>
<td>80.6</td>
</tr>
</tbody>
</table>

Contains Ordnance Survey data © Crown copyright and database right 2010
20% More Die Before Their Time in North v. South England

![Graph showing percentage of deaths before time for males and females from 1965 to 2007. The graph indicates a trend where males and females have a similar percentage of deaths before their time, with a slight increase in recent years.]
“38,000 people a year die sooner in North England than they would if they had life chances equivalent to those in the South”
Complex Dynamics

• Time
  – Secular
  – Life-course / age
  – Duration of treatment
  – Maturation of an intervention
  – Decay of an intervention
  – ...

Machine Learned Epidemiology

• 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

• Potential heresy: unsupervised search for patterns of sensitisation instead of hypothesis
Model: Unsupervised Clustering of Allergic Sensitisation Across Ages
From 2 to 5 Useful Classes of Atopy

![Graph showing classes of atopy and their associations with asthma, asthma exacerbation, persistent wheeze, and current wheeze.](image)
Better Prediction of Real-world Outcomes

Admitted at Any Age

First Admitted > 3 Years Old (remove early virus wheeze)
What might a usefully complex healthcare future look like?

ENVISIONING USEFUL COMPLEXITY
Large scale inference

Unified Graphical Model

Open Unifying Modelling: Across mechanisms and contexts

Health Records & Knowledge Silos

Health Avatars & Dynamic Models

Electronic Health Records (eHR)

Data-intensive Paradigm shift

Expertise

Expertise

Expertise

Unified Graphical Model

\[ \bigcup \text{models} = \text{Avatar} \]

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

Data-intensive Paradigm shift

Large scale inference

Model refinement

- e.g. Coronary heart disease
- e.g. Chronic obstructive pulmonary disease
- e.g. Lung cancer
Digital Leap

• From
  – Silos of reward for knowing more about less
  – Silos of healthcare evidence & practice
  – Public health complexity hidden

• To
  – Reward for “big picture contribution”
  – Complex models linked to practical decision making
  – Palpable public health
What might the future of healthcare intelligence look like?

FROM MANAGING KNOWLEDGE TO INTELLIGENT HEALTHCARE SYSTEMS
Toward Digital Health Economies

• Natural experiment capture

• Collective insight capture
tacit knowledge $\rightarrow$ explicit metadata

• Open unified modelling
(not simply multi-scale)

...i.e. not top-down or bottom-up: middle-out
Building a ‘sense-making layer’ on top of NHS care records...

REAL WORLD
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.
NHS e-Lab
Turning “tombs” of data into useful intelligence

Usual suppliers

Local Community Integrated Health Record

NHS no.

Optometrist
Community nurse
Podiatrist

Biobanks
Local surveys
Individual research

ONS vital statistics
Local authority socio-economic
Public health

Commissioning
Audit
Public Health
Research

“unified sense-making”
Data queries: From this...
...to this
Prevalence of high HBA1C values for diabetics

Investigation into diabetes in Salford

Data
- High HBA1C Prevalence
- Normalised HBA1C by Patient
- Repository

Documents
- Diabetes ruleset 14.0 as PDF

Data Exploration
- HBA1C
- example Cube

Notes
- Info about data
- Prevalence of high HBA1C values for diabetics
- This is the QOF ruleset

Snapshots
- HBA1C by ward Mean HBA1C values by ward Map
- Normalised HBA1C by Patient High HBA1C Map

People
- gary
- dammers
- gmoultont
eLab basics

- Web browser delivered
- Natural language queries, not battling with clinical codes
- Role based access
- Search for worked examples, not separate data and tools
- Just-in-time training
- Peer-to-peer support and visibility
- Audited (security & credit for contribution)
e-Lab

Research/Work Object

Find
Share
Reuse

Research protocol
Data-sources
Data-preparation scripts
Working datasets
References

Statistical analysis scripts
Analysis-logs & notes
Figures/Graphics
Manuscripts
Slides

Data - sources
Research/Work
Object

Socially-stimulating science & service, in-silico
Any data sharing is context-specific

Localities only share the data items relevant to the work, packaged into a work object that is checked by a local officer before being shared – all subsequent work on the object is audited and visible to the originator.
“Borrowing Strength” along Service Buses

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

Strength is borrowed and costs reduced by pooling expertise.
Exploit Heterogeneity

• Incorporating more information about differences between centres increases the estimated effects of interventions from meta-analysis of observational studies

• Federation of e-Labs could generate extremely useful healthcare intelligence at scale

Future transparency of decision support

Usual suppliers

Local Community Integrated Health Record

NHS no.

Optometrist
Community nurse
Podiatrist

Biobanks
Local surveys
Individual research

ONS vital statistics
Local authority socio-economic
Public health

Decision support models

Model (re)validation alongside clinical audit

EU Directive 2007/47/EC
Software = device
Disseminating New Models: e.g. Policy Simulator: Probabilistic Graphical Models into Practice

Export “Method Object” from research → into NHS practice via e-Lab

Ask “what if” scenario planning questions

e.g. “what is the likely public health impact of 500k spend on statins vs. smoking cessation”

www.healthimpact.org.uk
E-Lab approach to open linked data and the instrumented web

CO-THINKING ACROSS DISCIPLINES
Linked data ≠
Linked data, methods & investigators

Social Research:
Data, methods & investigators

Biomedical Research:
Data, methods & investigators

MIND THE GAP

Co-Reason
From data **archives** to data **playgrounds**

Browsing for **data extracts** made by a **social network** of ‘archive’ users...
Shopping for variables from across different years of a survey collection...
Instant access to relevant parts of survey documentation...

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Category</th>
<th>Survey</th>
<th>Dataset</th>
<th>Year</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>jntinc</td>
<td>Joint income</td>
<td>Income</td>
<td>HSE</td>
<td>hse05ai</td>
<td>2005</td>
<td>0</td>
</tr>
<tr>
<td>jntinc</td>
<td>Joint income</td>
<td>Income</td>
<td>HSE</td>
<td>hse07ai</td>
<td>2007</td>
<td>3</td>
</tr>
<tr>
<td>srcin08b</td>
<td>Income: Income Support</td>
<td>Income</td>
<td>HSE</td>
<td>hse05ai</td>
<td>2005</td>
<td>0</td>
</tr>
<tr>
<td>srcin08b</td>
<td>Income: Income Support</td>
<td>Income</td>
<td>HSE</td>
<td>hse07ai</td>
<td>2007</td>
<td>3</td>
</tr>
<tr>
<td>srcin15b</td>
<td>Income: No source of income</td>
<td>Income</td>
<td>HSE</td>
<td>hse05ai</td>
<td>2005</td>
<td>0</td>
</tr>
<tr>
<td>srcin15b</td>
<td>Income: No source of income</td>
<td>Income</td>
<td>HSE</td>
<td>hse07ai</td>
<td>2007</td>
<td>1</td>
</tr>
<tr>
<td>srcin03b</td>
<td>Income: Child Benefit</td>
<td>Income</td>
<td>HSE</td>
<td>hse05ai</td>
<td>2005</td>
<td>0</td>
</tr>
<tr>
<td>srcin07b</td>
<td>Income: Pension Credit</td>
<td>Income</td>
<td>HSE</td>
<td>hse05ai</td>
<td>2005</td>
<td>0</td>
</tr>
<tr>
<td>hhinc</td>
<td>Total household Income</td>
<td>Income</td>
<td>HSE</td>
<td>hse05ai</td>
<td>2005</td>
<td>0</td>
</tr>
</tbody>
</table>

Category: Income

Derivation type: Hhold

Derivation method:

SHOW CARD E
Thinking of the income of your household as a whole, which of the groups on this card represents the total income of the whole household before deductions for income tax, National Insurance, etc.
ENTER BAND NUMBER. DON'T KNOW = 96, REFUSED = 97.
Range: 1..31, 96, 97

Value information:

This variable is numeric, the SPSS measurement level is scale.
SPSS user missing values = -9 thru -1
Value label information for hhinc
Value = -9 Label = No answer/refused
Value = -8 Label = Don't know
Value = -7 Label = Refused/not obtained
Value = -6 Label = Schedule not obtained
Value = -5 Label = Schedule not applicable
Value = -4 Label = - Item not applicable

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Category</th>
<th>Survey</th>
<th>Dataset</th>
<th>Year</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>srcin05b</td>
<td>Income: Child Benefit</td>
<td>Income</td>
<td>HSE</td>
<td>hse07ai</td>
<td>2007</td>
<td>1</td>
</tr>
<tr>
<td>srcin07b</td>
<td>Income: Pension Credit</td>
<td>Income</td>
<td>HSE</td>
<td>hse07ai</td>
<td>2007</td>
<td>1</td>
</tr>
<tr>
<td>hhinc</td>
<td>Total household income</td>
<td>Income</td>
<td>HSE</td>
<td>hse07ai</td>
<td>2007</td>
<td>1</td>
</tr>
</tbody>
</table>
Sharing and visibility

Linking a data extract with a script for deriving variables...

Making the data extract visible...
Enabling **user-visibility** for data extraction or derivation contributions...
Social vs. Technical Scalability

... assemble a digital ecosystem of **data, models** and **experience**
across molecular, clinical and population worlds

- People with relevant expertise and authorisation
- State-of-the-art algorithms
- Quality assured integrated data
- Intelligence
Lessons from e-Science

• Visibility motivates contribution

• Sharing & reusing is not e-plagiarism

• Open resource ≠ open for business
A Tale of Two Clouds

Me + Stuff = my view

Us + Stuff = our view

electronic
digital

Me + Stuff = my view
Re-pack to:
Update
Learn
Expand
Combine
Repurpose
Translate
Mobilise

Percentage excess deaths in North vs South England - all ages

From the mid-1970s the Northern-excess mortality became greater in men than women, whereas in the late 1960s it had been higher in women. From 1975 to 2008 the male Northern-excess mortality has been 15.0% (14.8% to 15.1%), whereas this gap was 12.5% (12.4% to 12.7%) in women.
N:S Divide Example: Sobering Facts

- Publicly available data
- Very simple analysis
- Many PH workers unfamiliar with methods
- Reusable in clinical audit etc.
- Me + stuff = credit for a paper
- Us + stuff =, potentially, equitable healthcare resource allocation
Paradigms?

E-Science

(Electronic → Enabling: Established Brand)

Science^k  Co-Science

‘Abstractio’ ad absurdum  At e-scale
Conclusion: descale to e-scale

• Public health needs usefully complex & timely health sciences & systems

• Curiosity-led research will drive technical scale

• E-Science / Co-Science could drive social scale