Healthcare e-Labs:
Opening & Integrating Models of Health

Microsoft eScience, Pittsburgh, 17 Oct 2009

Iain Buchan
University of Manchester
This Talk

• Problem: Late & fragmented healthcare

• Target: Usefully complex models of health

• Vehicle: e-Lab (open ∪ data, models, people)
Do you have healthcare?

• Clinical model:
  – Rescue the ill
  – Resource \( \propto \) illness
  – Specialise to optimise

• Health model:
  – Rescue the at-risk
  – Resource \( \propto \) disease/risk
  – Generalise to optimise
After 5,000 years of the clinical model
A life-course view of personal health

Birth

Instability

Hip fracture

Personal cost: healthy years of life lost

Societal cost: large proportion of healthcare resource used

Death

+ structured exercise programme
False dichotomy: prevention vs. care

Multiple Budgets/Insurance + Citizens’ Disposable Income

Public Health Prevention Wellbeing

Unified Model

Healthcare Budgets/Insurance

Clinical Care Biomedical Science

Life-course view (barely visible investment)
Life of a person

Sickness view (fire-fighting investments)
Life of a parliament
eScience can help

Start with data-flooded healthcare...
Problem: Ravines of Healthcare Communication & Organisation

Self Care

Primary Care

Hospital A

Clinical Care

Secondary Care

Hospital B
Disease Knowledge Bridges: Integrated Care Pathways

Self Care
Primary Care
Clinical Care
Secondary Care
Hospital A
Hospital B
**Missing**: Patient & Community

‘Big-picture’ Across Pathways

Self Care

Primary Care

Hospital A

Clinical Care

Secondary Care

Hospital B

∪ = Avatar
Pulling evidence together into one, realistically-complex model: e.g. MRC IMPACT II

Outputs: Population-based incidence, prevalence; Deaths prevented; Life-Years; Life expectancy; Costs; Cost-effectiveness ratios
Recombine evidence on graphical models of disease plus care-services.

Ask ‘what if’ invest in statins vs. smoking cessation etc.
Sufficient input for unified models?

Natural Experiments of Healthcare

Healthcare Evidence Base
Digital Dust (data deposit > use)
Cloud of millions of care messages in the local health economy

- Organise
- Transform & Examine

Structured Data

Structured Data & Metadata
Data-flood vs. -intensity

Methods/Models ↑

Contextual expertise ↔

Data ↑↑↑↑
NHS e-Lab: Salford Pilot

Real-time Data Repository in PCT

Person-identifiable and sensitive information removed

Anonymised Data Repository in PCT

24-hourly updates

FIREWALL

Trusted person poses question(s)

Sense-making software & support

Outputs

Optometrist
Eye screening
Community nurses
Podiatry

Deaths, Demographics etc.

Biomics Data

Hosp.

Link on NHS number

GP

GP

GP

GP

Optometrist
Eye screening
Community nurses
Podiatry
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
Linked Health Records
Audit; Research; Intelligence

Clear Public Good
Work Objects
e-Lab for a defined community

Unclear Public Good
Linked Health Records

De-identify
Local Ownership
Asset Enrichment
Trust & Benefit in Research across Health Records

Now
Database-centred

Clinical Information Governance

Research Governance

Ethical Oversight

Future?
e-Lab: Community-centred

Anonymised e-Lab

Integrated Health

x Health Agencies

Research

x Research Agencies

x Health Communities
Salford Integrated Record (SIR)

Trial Recruitment Tool

Salford e-Lab

Bulk ETL with Anonymisation

Trial Protocol Designer

Clinical Care Boundary

Salford Health Community

Clinician

Query Objects

HTTPS

email

HTTPS

HTTPS

HTTPS

SQL

HTTPS

HTTPS

HTTPS

HTTPS

Researcher/Manager

Clinical Care Boundary
Federation: More local use → better quality data

Community development

Health Community 1
Health Community 2
Health Community 3

e-Lab

Research Networks

Research

10 developers

NWH
North West e-Health
Work Object in Science: Research Object

Services

Workflows

Packages of Outputs

Reproduce
Share
Morph
Inherit

Research Object

Data/Workflow
Analysis Scripts
Annotations
Slides
Manuscripts
References

Encapsulated $\rightarrow$ (DAG) discovery?
Typical Health Sciences Signal Path

Problem Space  Observation Space  Data Space

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

ey = b1x1 + b2x2 + b3x3 + c

e-Lab must harness observers & thinkers not just data & methods
Additional signal paths

• Genome-wide genotyping
• Proliferation of biomarkers
• Systems biology

‘omics

• More complete reflection of observation space
• Image-like
• Variably more reflection of the problem space
Orthodoxy: Sieving Associations

<table>
<thead>
<tr>
<th>Association</th>
<th>Bias</th>
<th>Type</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C→M</td>
<td>Cause-effect</td>
<td>Real</td>
<td>Cause-effect</td>
</tr>
<tr>
<td>MI→C</td>
<td>Reverse</td>
<td>Real</td>
<td>Effect-cause</td>
</tr>
<tr>
<td>C←?→MI</td>
<td>Confounding</td>
<td>Real</td>
<td>Effect-effect</td>
</tr>
<tr>
<td>C &lt;&gt; MI</td>
<td>Random error</td>
<td>Spurious</td>
<td>Chance</td>
</tr>
<tr>
<td>C &lt;&gt; MI</td>
<td>Systematic error</td>
<td>Spurious</td>
<td>Bias</td>
</tr>
</tbody>
</table>

C = caffeine, MI = myocardial infarction (heart attack)

Disciplined approach to causal inference, Bradford-Hill: Criteria (temporality, strength, dose-response, consistency, plausibility, consideration of alternatives, open to experiment, specificity, coherence)
Exhausted Orthodox Platform

Problem 1: Dwindling hits from tools to detect independent “causes”

Problem 2: Knowledge can’t be managed by reading papers any more

The big public health problems e.g. Type 2 Diabetes have “complex webs of causes”

The “data-set” and structure extend beyond the study’s observations
• Population-based birth cohort
• Subjects recruited in utero
  – parents skin tested + questionnaire
• 1085 children born into the study
• Reviews at age 1, 3, 5 and 8 years
  – 980 children reviewed at age 8 years
    • 90.3% follow-up rate
• 11 year follow up to be completed in 2009
Outcomes

• Subjective outcomes, age 1, 3, 5 and 8
  – Validated questionnaires,
    • symptoms of asthma, eczema and rhinitis
    • medication use

• Atopic status, age 1, 3, 5 and 8
  – Skin tests to inhalant and food allergens
  – Total and specific IgE

• Lung function
  – Specific airways resistance (sRaw), Spirometry
  – Airway responsiveness
  – Exhaled Nitric Oxide (eNO), age 8
Lung Function in Pre-School Children

Mask with integrated mouthpiece

Lowe at al, Lancet 2002; 359: 1904-8
Environmental Exposures

- Allergen levels (mite, cat and dog)
- Endotoxin
- Pet ownership and contact
- Sibship
- Tobacco smoke exposure
- Childcare arrangements
- Vaccination uptake
- Duration of breastfeeding
- Dietary intake (Diet-Q), age 5 and 8
- Antibiotic and other medication usage (from GP records)
Genetics

• DNA collected during age 5, 8 and 11 year follow-ups in virtually all study participants (n=1003)

• All DNA quantified and normalised

• Recent 660k SNP chip processing
The Atopy Model

Unsupervised Clustering

Sensitization Class

Switch between Sens Classes

Prob of Sensitization at Year 1
Prob of Gaining Sensitization
Prob of Losing Sensitization

3 Intervals (1-3, 3-5, 5-8)
States of Sens Class (2-4)

Acquired Sensitization at age 1
Acquired Sensitization at age 3
Acquired Sensitization at age 5
Acquired Sensitization at age 8

Skin Test at age 1
Skin Test at age 3
Skin Test at age 5
Skin Test at age 8

IgE Test at age 1
IgE Test at age 3
IgE Test at age 5
IgE Test at age 8

Prob Pos Skin Test given sensitized
Prob Pos Skin Test given NOT sensitized
Prob Pos IgE Test given sensitized
Prob Pos IgE Test given NOT sensitized

Children (1186)

To Infer

Allergens = (mite, cat, dog, pollen, egg, milk, mold, peanut)

infer.net
2 Sensitization Classes

The children are clustered into 2 groups.
The first group has many sensitizations and high probability of having asthma.
1. The first cluster has many sensitizations.
2. The multiple early onset sensitization and multiple late onset sensitization children are also separated.
3. Cluster of children with sensitization to predominantly dust mite.
4. Clusters of children with sensitization to predominantly pollen.
Atopy Classes and Asthma

Multiple Early Sensitisations
odds ratio (OR) [95% CI]: 30.5 [12.4-75.3]
c.f. Ever atopic group:

% with asthma
(symptomatic on challenge)

No atopy  "Atopy 4"  "Atopy 3"  "Atopy 2"  "Atopy 1"
Usefully-complex model building needs...

• Privacy-preserved access to healthcare and clinical study data

• Invisible HPC

• Methods/sub-models that find people describing problems
  → statistical ‘borrowing of strength’
Open Unifying e-Lab

\[ \bigcup (\text{open sub-models}) \]
\[ \bigcup \]

Easy computation =

\[ \bigcup \]

Abstract reasoning and motivation of domain experts

- More complete insights
- More reusable evidence
- Better management across diseases
- Earlier intervention
- Greater citizen involvement
Large scale inference

Unified Graphical Model

Health Records & Knowledge Silos

Open Unifying Modelling: Across mechanisms and contexts

Data-intensive Paradigm -shift

Health Avatars & Dynamic Models

∪ models = Avatar

e.g. Chronic obstructive pulmonary disease

Electronic Health Records (eHR)

Expertise

Unified Graphical Model

Expertise

Expertise

Data-intensive Paradigm -shift

Large scale inference

Model refinement

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

e.g. Lung cancer
Both contexts need:

- Open
- Unified & unifying
- Usefully-complex

Models

Jim Gray’s legacy may have wider implications