Data Linkage for Pharmacovigilance

Evidence Synthesis

Brad Kirby

Supervisors: Dr James McLay & Dr Colin Simpson
Advisers: Prof Peter Helms, Prof Marion Bennie, Mary Sweetland, Dr Rachael Wood
Overview

- Background
- Aims
- Methods
- Results
- Future Directions
- Questions
Background

- ADRs
- Burden on NHS
- Better ways of looking for ADRs
- Pharmacovigilance
- Current PV systems are limited
- Routine NHS data
Child Medical Records for Safer Medicines (CHIMES)

- Acceptability and validity of datasets derived from linked routinely acquired NHS data for post marketing surveillance of medicines in children

- Work Package 1: User Communities
- Work Package 2: Evidence Synthesis
- Work Package 3: Pharmacovigilance
Child Medical Records for Safer Medicines (CHIMES)

- Accuracy and validity of routinely acquired linked NHS data to support a routine mechanism for post marketing surveillance of medicines in children

- Work Package 1
  - User Communities

- Work Package 2
  - Evidence Synthesis

- Work Package 3
  - Pharmacovigilance

✓ ACCURACY
✓ VALIDITY
Research Questions

- What routine data are there?
- What's the best way to link it?
- What are the limitations of these data?
- How accurate are the data?
- What are the technical barriers to linking the data?
Methods

**PRIMARY CARE**
- Prescribing
- General Practice

**SECONDARY CARE**
- Hospital Admissions
- Cancer Registrations
- Maternity
- Outpatient Attendances
- Deaths
## Routine Datasets (Fields for linkage)

<table>
<thead>
<tr>
<th></th>
<th>SMR01</th>
<th>SMR06</th>
<th>GRO</th>
<th>SMR4</th>
<th>SMR00</th>
<th>SMR02</th>
<th>A&amp;E</th>
<th>PCCIU</th>
<th>PTI</th>
<th>PIS</th>
<th>CMS/AMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hospital discharges</td>
<td>Cancer Reg</td>
<td>Deaths</td>
<td>Mental Health</td>
<td>OP</td>
<td>Maternity</td>
<td>A&amp;E</td>
<td>PCCIU GP Data</td>
<td>PTI GP Data</td>
<td>Prescr.</td>
<td>E-Pharm</td>
</tr>
<tr>
<td>Seeded Chi Number</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHI Number</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surname</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Forename</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Date of Birth</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Sex</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Marital status</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postcode</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
# Routine Datasets (Fields for analysis)

<table>
<thead>
<tr>
<th></th>
<th>SMR01</th>
<th>SMR06</th>
<th>GRO</th>
<th>SMR4</th>
<th>SMR00</th>
<th>SMR02</th>
<th>A&amp;E</th>
<th>PCCIU Data</th>
<th>PTI Data</th>
<th>PIS</th>
<th>CMS/AMS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Condition</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Procedure</strong></td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Drug</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Demographics</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Methods...

- **Accuracy**

- **Validity:**
  - Is a record true for a specific person?

- **Completeness:**
  - If a patient has a disease, what is the likelihood that this will be recorded in the database?

- **External Patient Registries**

- **Record Linkage**: validation; case-control; case-cohort studies

WOSCOPS. *J Clin Epidemiol* 1995;48(12):1441
Methods...

- Accuracy
- Record Linkage:
  - Bringing two records together
  - Database linkage
- Duplicates
  - John Smith
  - John Smith
  - John Smith
Methods...

- Accuracy
- Record Linkage:
  - **Deterministic**
  - Probabilistic

- Exact agreement
- Community Health Index (CHI) as a unique identifier

<table>
<thead>
<tr>
<th>CHI</th>
<th>Fname</th>
<th>Sname</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>Bradley</td>
<td>Kirby</td>
</tr>
</tbody>
</table>

Table A:

<table>
<thead>
<tr>
<th>CHI</th>
<th>Fname</th>
<th>Sname</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>Brad</td>
<td>Kirby</td>
</tr>
</tbody>
</table>

Table B:
Methods...

- Accuracy
- Record Linkage
  - Deterministic
  - Probabilistic
- Unique ID errors
- Multiple Variables
- Discrepancies

Table A:

<table>
<thead>
<tr>
<th>CHI</th>
<th>Fname</th>
<th>Sname</th>
<th>D.o.B</th>
<th>Postcode</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL</td>
<td>Bradley</td>
<td>Kirby</td>
<td>24/11/76</td>
<td>EH1 1LB</td>
</tr>
</tbody>
</table>

Table B:

<table>
<thead>
<tr>
<th>CHI</th>
<th>Fname</th>
<th>Sname</th>
<th>D.o.B</th>
<th>Postcode</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL</td>
<td>Brad</td>
<td>Kirby</td>
<td>24/11/46</td>
<td>EH1 1LB</td>
</tr>
</tbody>
</table>
Methods...

- Accuracy
- Record Linkage
  - Deterministic
  - Probabilistic

- Probability Weights
  - Data Quality (M prob)
  - Commonness (U prob)

\[ M = \Pr(\text{the values agree} \mid \text{the records are a match}) \]
\[ U = \Pr(\text{the values agree} \mid \text{the records are not a match}) \]
Methods...

- Accuracy
- Record Linkage
  - Deterministic
  - Probabilistic

- Probability Weights
  - Data Quality (M prob)
  - Commonness (U prob)

Agreement field weight = \( \log \left( \frac{m}{u} \right) \)

Disagreement field weight = \( \log \left( \frac{(1-m)}{(1-u)} \right) \)
Methods...

- **Accuracy**
- **Record Linkage**
  - Deterministic
  - **Probabilistic**

**Probability Weights**
- Data Quality (M prob)
- Commonness (U prob)

<table>
<thead>
<tr>
<th>Field</th>
<th>M prob</th>
<th>U prob</th>
<th>Agree Log [m/u]</th>
<th>Disagree Log [(1-m)/(1-u)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sname</td>
<td>0.95</td>
<td>0.01</td>
<td>6.57</td>
<td>-4.31</td>
</tr>
<tr>
<td>D.o.B</td>
<td>0.9</td>
<td>0.01</td>
<td>6.49</td>
<td>-3.31</td>
</tr>
<tr>
<td>P.C</td>
<td>0.7</td>
<td>0.01</td>
<td>6.13</td>
<td>-1.72</td>
</tr>
</tbody>
</table>
Methods...

- Accuracy
- Record Linkage
  - Deterministic
  - Probabilistic

- Probability Weights
  - Data Quality (M prob)
  - Commonness (U prob)

<table>
<thead>
<tr>
<th>CHI</th>
<th>Fname</th>
<th>Sname</th>
<th>D.o.B</th>
<th>Postcode</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL</td>
<td>Bradley</td>
<td>Kirby</td>
<td>24/11/76</td>
<td>EH1 1LB</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

+ 6.57

- 3.31

+ 6.13

TOTAL WEIGHT SCORE: = 9.39

<table>
<thead>
<tr>
<th>CHI</th>
<th>Fname</th>
<th>Sname</th>
<th>D.o.B</th>
<th>Postcode</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL</td>
<td>Brad</td>
<td>Kirby</td>
<td>24/11/46</td>
<td>EH1 1LB</td>
</tr>
</tbody>
</table>

LINK?
Methods...

All record pair comparisons:

1,000 x 1,000 = 1,000,000

Total Comparisons = 1 million!

NON-LINKS = 999,000 | LINKS = 1,000
Methods...

Blakely, T. *Int. J. Epidemiol.* 2002;31:1246
Methods...

Grannis, S.J. *AIMA Annu Symp Proc.* 2003: 259
Methods...

<table>
<thead>
<tr>
<th>CHI</th>
<th>Fname</th>
<th>Sname</th>
<th>Married</th>
<th>Postcode</th>
</tr>
</thead>
<tbody>
<tr>
<td>1010</td>
<td>M</td>
<td>Smith</td>
<td>Yes</td>
<td>EH1 1LB</td>
</tr>
</tbody>
</table>

![](images)

- 0
- 4.31
- 1
+ 6.13

= 0.82
Methods...

Single Threshold

METHODS

- Record Linkage
  - Deterministic
  - Probabilistic

- Linkage accuracy

- Match rates
- False +
- False –

Acceptability depends on purpose:
- Statistical
- Administrative
Methods...

- Accuracy
- Record Linkage
  - Deterministic
  - Probabilistic
- Data Mining

On Electronic Health Records for Pharmacovigilance

“Application of statistical techniques to databases”

Data mining v traditional statistical techniques

Methods...

- Accuracy
- Record Linkage
  - Deterministic
  - Probabilistic
- Data Mining
- Prediction:
  - Classification
  - Regression
  - Decision Trees
- Description:
  - Clustering
  - Association Rules
Results – Database Validation

The chart shows the rate per 100,000 for Prescribing, External Patient Registry, and Hospital Admissions, with the Incidence in 0-14 year olds represented by blue bars.
Results...

Pundziute-Lycka. *Diabetologia* 2002;45:783
Future Directions

- Asthma
- Cancer
- etc.

- Longitudinal
- Retrospective / Prospective Cohort
- Data Mining
- Medicine & Disease

DIABETES

VIRTUAL Linkage

Prescribing

Hospital Admissions

PHARMACOVIGILANCE
Summary

- Adverse drug reactions are a problem
- We need better ways of looking for them
- Record linkage & data mining using Routine health data offers one way
Questions?

Brad Kirby – Child Medical Records for Safer Medicines

CHIMES acknowledges the financial support of NHS Research Scotland (NRS), through NHS Grampian
CHI Completeness – Insulin

CHI “Missingness”:
Patient Based v Random
Burden of Adverse Drug Reactions (ADRs)

1 in 16 Admissions ➔ £466m / yr

ADR -> Off-label (use of medicine outside licence)
35-92% Children receive off-label medicine
11-16% in Primary Care

Pharmacovigilance

“Science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other possible drug related problem” (World Health Organization, 2002)

Spontaneous Reporting

Specific Medicines
Limitations

- Under reporting
- Lack of a denominator
- Reporting bias due to media attention
- Quality of received data is variable
- Low % of reports relating to children

- England Only
- Low response rate
- Specific drugs of interest (currently approx 70 new drugs)

METHODS

- **Accuracy**
- **Record Linkage**
  - Deterministic
  - **Probabilistic**

- Bringing pairs of records together for comparison
- **Probability Weights**
  - M prob – Quality & Reliability of variable
  - U prob – Randomness of the value
- **Linkage Decision**
  - Thresholds