Applied Data Linkage: from dirty data to clean complete patient records in emergency health care

Dr Julia Crilly, John O’Dwyer, Marilla O’Dwyer, Dr James Lind, Kerri Melki, Dr Julia Peters, Assoc Prof Vivienne Tippett, Nerolie Bost, Prof Marianne Wallis, Dr Gerben Keijzers
Overview

1. Project Team
2. The Study
3. How HDI works (Efficiency)
4. Manual versus HDI* linking (Effectiveness)
5. So what? (Relevancy to health care)

*Health Data Integration….
**Robina ED Impact (REDI) Project team**

<table>
<thead>
<tr>
<th>Principal Investigators –</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dr James Lind</strong> (Director of Emergency Medicine Training, GCH)</td>
</tr>
<tr>
<td><strong>Dr Julia Crilly</strong> (ED Nurse Researcher, GCH, Griffith University)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Co-investigators</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dr Julia Peters</strong> (Registrar, GCH ED)</td>
</tr>
<tr>
<td><strong>Nerolie Bost</strong> (Research Officer, GCH ED)</td>
</tr>
<tr>
<td><strong>Dr Gerben Keijzers</strong> (Staff Specialist GCH ED, Bond University)</td>
</tr>
<tr>
<td><strong>Prof Marianne Wallis</strong> (Nursing Research Chair, GCH &amp; Griffith University)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marilla O’Dwyer, John O’Dwyer, Kerri Melki</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Australian E-Health Research Centre)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Assoc Prof Vivienne Tippett</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>(Director, Australian Centre for Pre-Hospital Research, Queensland Ambulance Service)</td>
</tr>
</tbody>
</table>

GCH: Gold Coast Hospital; ED: emergency department
Contributing Bodies

1. Queensland Health
2. Gold Coast Hospital Foundation
3. Queensland Ambulance Service
4. Australian E-Health Research Centre
5. Australian Centre for Pre-hospital Research
6. Griffith University
7. Queensland Emergency Medicine Research Foundation
The Study

1. Project Team

2. The study

3. How HDI works (Efficiency)

4. Manual vs HDI* linking (Effectiveness)

5. So what? (Relevance to health care)

*Health Data Integration
The Study

• AIM: to determine the impact of opening a new emergency department in Robina on the patient outcomes and service delivery for the Gold Coast Hospital

• It is suspected that the load balancing will improve
  • Ambulance delivery times
  • Overflow capability and
  • Patient outcomes

• Required data that can accurately trace each patient journey from Ambulance to ED to Admissions.

• Working with health data is challenging eg. guessed age, partial names (DIRTY DATA)
• Design: One month pre and post comparative study
• Data collection: Using routinely collected health data from three different health information systems (Ambulance, ED, hospital)
• Link three HIS to create one complete data set
Overview

1. Project Team
2. The Study
3. How HDI works (Efficiency)
4. Manual vs HDI linking (Effectiveness)
5. So what? (Relevance to health care)
What is Health Data Integration (HDI)?

• Facilitates linking of patient information across multiple databases, at multiple locations, while maintaining their privacy and security

  • Encrypted linking to protect identity and personal information

  • Federated approach to integrate local data repositories – i.e. many local data sources operating as one and accessed remotely and securely

  • Allows data custodians to retain control of their data - they specify how much of their locally stored data is available to a specific project and in what format

  • Allows researchers or clinicians to access linked but de-identified data sets for further analysis and reporting (as per ethics agreements)
HDI Platform Technology

De-identified linked data for analysis

Researcher queries virtual repositories

HDI links and integrates data

Custodial controlled data

Analysis and Reporting Tools
(eg SPSS/PASW, SAS, Crystal, etc)

Query

HDI Domain

Data Source

Data Source

Data Source
A different approach to matching……

- Rule based approach
- Based on easily understandable ‘human’ approach to matching (Hansen & Maeder, 2007)
- User defined and completely configurable (eg can include ‘case matching’ in the criteria)

<table>
<thead>
<tr>
<th>Name (First Last), Age, Gender, Date, Time</th>
<th>Probabilistic</th>
<th>‘rule based’</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Smith, 44, 1, Mon 3.00pm</td>
<td>58%</td>
<td>Yes*</td>
</tr>
<tr>
<td>Jon Smyth, 43, 1, Mon 3.35pm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andrea O’Conner, 23, 2, Fri 5.25am</td>
<td>91%</td>
<td>Yes*</td>
</tr>
<tr>
<td>Andrea M. O’Connor 23, 1, Fri 6.00am</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rebecca Andersen, 38, 1, Tue 10.38pm</td>
<td>99%</td>
<td>No*</td>
</tr>
<tr>
<td>Rebecca Andersen, 38, 1, Fri 4.52am</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bob Calender, 64, 1, Wed 7.49pm</td>
<td>89%</td>
<td>Yes*</td>
</tr>
<tr>
<td>Robert Calendar, 66, 1, Wed 8.15pm</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Based on linking algorithms used in REDI project
What the data looks like

- Names - different formats, spelling
- Age – date of birth, age, estimated age
  - Ie age difference for the same person up to ± 5 years
- Gender – different formats
- Case match – hours between Ambulance – ED and ED-Admissions
- Plus all the usual mis-spellings, data capture / entry errors and typos
How we set it up to match.....

Ambulance
PATIENT NAME: John A. Van Smith
AGE-DOB: 01/06/1956
GENDER: M
D_Triage: 02/05/2005 23:56

ED
Present Name First: Jon Anthony
Present Name Surname: Van Smith
Present Age in Years: 49
Present Gender: m
Triaged At: 03/05/2005 00:34

Admissions
LastName: Van Smythe, John Antony
Age: 50
Sex: M
Ad Date: 03/05/2005
Time: 01:02

Standardise and transform data..........................

FirstName: John
Surname1: Smith
Surname2: A. Van Smith
Surname3: Van Smith
Surname4: Van Smith
AgeRange: 44-54
Sex: 1
CaseTimeFrom: 02/05/2005 23:00
CaseTimeTo: 03/05/2005 02:00

FirstName: Jon
Surname1: Van Smith
Surname2: Van Smith
Surname3: Van Smith
Surname4: Van Smith
AgeRange: 44-54
Sex: 1
CaseTimeFrom: 03/05/2005 00:00
CaseTimeTo: 03/05/2005 03:00

FirstName: John
Surname1: Van Smythe
Surname2: Van Smythe
Surname3: Van Smythe
Surname4: Van Smythe
AgeRange: 45-55
Sex: 1
CaseTimeFrom: 03/05/2005 01:00
CaseTimeTo: 03/05/2005 04:00
How it matches……..

Ambulance
PATIENT NAME: John A. Van Smith
AGE-DOB: 01/06/1956
GENDER: M
D_Triage: 02/05/2005 23:56

ED
Present Name First:: Jon Anthony
Present Name Surname: Van Smith
Present Age in Years: 49
Present Gender: m
Triaged At: 03/05/2005 00:34

Admissions
LastName: Van Smythe, John Antony
Age: 50
Sex: M
Ad Date: 03/05/2005
Time: 01:02

Link in HDI…….. based on user-defined rules………..

FirstName: John
Surname1: Smith
Surname2: A. Van Smith
Surname3: Van Smith
Surname4: Van Smith
AgeRange: 44-54
Sex: 1
CaseTimeFrom: 02/05/2005 23:00
CaseTimeTo: 03/05/2005 02:00

FirstName: Jon
Surname1: Van Smith
Surname2: Van Smith
Surname3: Van Smith
Surname4: Van Smith
AgeRange: 44-54
Sex: 1
CaseTimeFrom: 03/05/2005 00:00
CaseTimeTo: 03/05/2005 03:00

FirstName: John
Surname1: Van Smythe
Surname2: Van Smythe
Surname3: Van Smythe
Surname4: Van Smythe
AgeRange: 45-55
Sex: 1
CaseTimeFrom: 03/05/2005 01:00
CaseTimeTo: 03/05/2005 04:00
The Challenges

• While working with the data we had to overcome some unique challenges

• Identity matching
  • Often the age of an individual must be guessed
    • This may not exactly match the age captured in the ED or admissions data
  • Names are often misheard and may only sound like the actual name
    • Like in our example “Smythe” may be captured as “Smith” in the ambulance data

• Case matching
  • Match the Ambulance data with a specific Emergency Department arrival
    • Must match person’s ID
      • AND
    • The triage time in ED must be within a few hours of the Ambulance triage time
  • Similar measures must be taken with the matching from ED to Hospital Admissions
Overview

1. Project Team
2. The Study
3. How HDI works (Efficiency)
4. Manual vs HDI linking (Effectiveness)
5. So what? (Relevance to health care)
## Linking Results: Manual vs HDI

### Gold Standard Results (Manual)

<table>
<thead>
<tr>
<th>Case</th>
<th>EDIS</th>
<th>Rows</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>QAS</td>
<td>3,469</td>
<td>3,192</td>
<td>(92.0%)</td>
</tr>
<tr>
<td>HBCIS</td>
<td>3,431</td>
<td>3,244</td>
<td>(94.5%)</td>
</tr>
</tbody>
</table>

### HDI Results

<table>
<thead>
<tr>
<th>Case</th>
<th>EDIS</th>
<th>Rows</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>QAS</td>
<td>3,469</td>
<td>3,049</td>
<td>(87.9%)</td>
</tr>
<tr>
<td>HBCIS</td>
<td>3,431</td>
<td>3,240</td>
<td>(94.4%)</td>
</tr>
</tbody>
</table>
Overview

1. Project Team
2. The Study
3. How HDI works (Efficiency)
4. Manual vs HDI linking (Effectiveness)
5. So what? (Relevance to health care)
Study Findings

• 10,835 patient presentations were made to the ED during two months

• Offload time decreased sig (11 mins to 10 mins, p < 0.001)

• Proportion offload delay (>15 mins) decreased sig (38% to 27%, p < 0.001)

• Hosp LOS decreased sig

• ED LOS > 8 hrs about 19% in both groups (NS)
The opening of a new ED had a positive effect on ambulance access to the ED, time to unloading of patients and hospital length of stay but had no effect on general ED functioning

<p>| 2 | . | . | . | . | . | . | . | . | . | 456222 | 55.0 | 2 | 3111 | T41.3 | W | 3 | 2007- |
| 3 | . | . | . | . | . | . | . | . | . | 688234 | 45.0 | 2 | 4203 | K59.0 | AR | 3 | 2007- |
| 4 | 30 | 2 | DRUG INTOXICATION | 2006-05-17 01:29 | 2006-05-17 01:46 | 2006-05-17 02:13 | 0 | 2006-05-17 02:23 | 2006-05-17 02:38 | 786436 | 30.0 | 2 | 4218 | AR | 3 | 2006- |
| 5 | . | . | . | . | . | . | . | . | . | 234788 | 67.0 | 2 | 4211 | O72.1 | AR | 3 | 2006- |
| 6 | . | . | . | . | . | . | . | . | . | 674586 | 57.0 | 2 | 2462 | K59.9 | W | 3 | 2008- |
| 7 | . | . | . | . | . | . | . | . | . | 612945 | 43.0 | 1 | 4662 | Z53.2 | W | 4 | 2008- |
| 8 | . | . | . | . | . | . | . | . | . | 674533 | 32.0 | 1 | 4573 | R11 | W | 3 | 2008- |
| 9 | . | . | . | . | . | . | . | . | . | 670945 | 47.0 | 2 | 4216 | W | 4 | 2007- |
| 10 | . | . | . | . | . | . | . | . | . | 777834 | 4.0 | 1 | 4227 | L02.8 | W | 3 | 2007- |
| 11 | . | . | . | . | . | . | . | . | . | 56123 | . | 2 | 4214 | J06.9 | W | 3 | 2008- |
| 12 | . | . | . | . | . | . | . | . | . | 396127 | 26.0 | 2 | 4211 | J16.1 | W | 3 | 2007- |</p>
<table>
<thead>
<tr>
<th>diaglnsis_id</th>
<th>diagnosis_desc</th>
<th>e_arrival_date</th>
<th>e_traged_at</th>
<th>e_treated_doctor_seen_at</th>
<th>e_departure_ready_at</th>
<th>e_departure_actual_at</th>
<th>e_departure_status_desc</th>
</tr>
</thead>
<tbody>
<tr>
<td>539.0</td>
<td>AR</td>
<td>2007-01-22 12:24:00.000</td>
<td>2007-01-22 12:44:00.000</td>
<td>2007-01-22 12:55:00.000</td>
<td>2007-01-22 13:00:00.000</td>
<td>2007-01-22 18:24:00.000</td>
<td>H ED SERVICE EVENT COMPLETED - DISCHARGED</td>
</tr>
<tr>
<td>741.3</td>
<td>W</td>
<td>2007-09-02 10:22:00.000</td>
<td>2007-09-02 10:28:00.000</td>
<td>2007-09-02 15:20:00.000</td>
<td>2007-09-02 16:20:00.000</td>
<td>2007-09-02 16:20:00.000</td>
<td>H ED SERVICE EVENT COMPLETED - DISCHARGED</td>
</tr>
<tr>
<td>599.0</td>
<td>AR</td>
<td>2006-09-09 03:20:00.000</td>
<td>2006-09-09 03:24:00.000</td>
<td>2006-09-09 03:43:00.000</td>
<td>2006-09-09 03:43:00.000</td>
<td>2006-09-09 20:43:00.000</td>
<td>H ED SERVICE EVENT COMPLETED - DISCHARGED</td>
</tr>
<tr>
<td>4</td>
<td>AR</td>
<td>2006-09-17 02:24:00.000</td>
<td>2006-09-17 02:24:00.000</td>
<td>2006-09-17 02:55:00.000</td>
<td>2006-09-17 02:55:00.000</td>
<td>2006-09-17 02:55:00.000</td>
<td>W DID NOT WAIT</td>
</tr>
<tr>
<td>527.1</td>
<td>AR</td>
<td>2006-11-17 01:07:00.000</td>
<td>2006-11-17 01:28:00.000</td>
<td>2006-11-17 03:19:00.000</td>
<td>2006-11-17 08:02:00.000</td>
<td>2006-11-17 08:02:00.000</td>
<td>A ADMITTED (EXCL ED BED)</td>
</tr>
<tr>
<td>550.9</td>
<td>W</td>
<td>2006-09-29 14:48:00.000</td>
<td>2006-09-29 15:56:00.000</td>
<td>2006-09-29 18:38:00.000</td>
<td>2006-09-29 20:12:00.000</td>
<td>2006-09-29 20:12:00.000</td>
<td>A ADMITTED (EXCL ED BED)</td>
</tr>
<tr>
<td>753.2</td>
<td>W</td>
<td>2006-03-10 18:34:00.000</td>
<td>2006-03-10 18:38:00.000</td>
<td>2006-03-10 23:25:00.000</td>
<td>2006-03-10 23:25:00.000</td>
<td>2006-03-10 23:25:00.000</td>
<td>W DID NOT WAIT</td>
</tr>
<tr>
<td>811</td>
<td>W</td>
<td>2006-04-12 0:24:00.000</td>
<td>2006-04-14 12:05:00.000</td>
<td>2006-04-14 13:46:00.000</td>
<td>2006-04-14 17:13:00.000</td>
<td>2006-04-14 17:13:00.000</td>
<td>H ED SERVICE EVENT COMPLETED - DISCHARGED</td>
</tr>
<tr>
<td>9</td>
<td>W</td>
<td>2007-04-10 21:36:00.000</td>
<td>2007-04-10 21:36:00.000</td>
<td>2007-04-10 21:48:00.000</td>
<td>2007-04-10 21:48:00.000</td>
<td>2007-04-10 21:48:00.000</td>
<td>W DID NOT WAIT</td>
</tr>
<tr>
<td>11</td>
<td>W</td>
<td>2003-04-11 10:00:00.000</td>
<td>2003-04-11 10:00:00.000</td>
<td>2003-04-11 10:23:00.000</td>
<td>2003-04-11 10:23:00.000</td>
<td>2003-04-11 11:00:00.000</td>
<td>H ED SERVICE EVENT COMPLETED - DISCHARGED</td>
</tr>
<tr>
<td>12</td>
<td>W</td>
<td>2007-08-08 22:11:00.000</td>
<td>2007-08-08 22:21:00.000</td>
<td>2007-08-08 02:45:00.000</td>
<td>2007-08-09 05:10:00.000</td>
<td>2007-08-09 05:10:00.000</td>
<td>A ADMITTED (EXCL ED BED)</td>
</tr>
</tbody>
</table>
| e_departure_actual_at | e_departure_status_desc | e_occurrances | h_medical_record_number | h_enc_no | h_Sex | h_AgeAtDiagnosis | h_post_code | h_ad_source | h_ad_datetime | h_tro_category | h_care_type | h_ad_unit | h_ad_ward | Visible: 44 of 44 Variables

1. 2007-04-22 18:24:00.000   ED SERVICE EVENT COMPLETED - DISCHARGED   1   345698   345698-1   2   45.0   4225   2   2007-04-22 18:10   2   1   MED   ED   4.0   2007-01-2
2. 2007-09-02 18:20:00.000   ED SERVICE EVENT COMPLETED - DISCHARGED   1   .   .   .   .   .   .   .   .   .   .   .   .   .
3. 2008-09-08 08:43:00.000   ED SERVICE EVENT COMPLETED - DISCHARGED   1   .   .   .   .   .   .   .   .   .   .   .   .   .
4. 2008-09-17 02:55:00.000   YY   DID NOT WAIT   1   .   .   .   .   .   .   .   .   .   .   .   .   .
5. 2008-11-17 08:02:00.000   A   ADMITTED (EXCL ED BED)   1   .   .   .   .   .   .   .   .   .   .   .   .   .
6. 2008-05-29 20:12:00.000   A   ADMITTED (EXCL ED BED)   1   674508   674508-4   2   57.0   2452   3   2008-05-29 18:33   2   1   GYN ED   2.0   2008-05-3
7. 2008-03-10 23:25:00.000   YY   DID NOT WAIT   1   .   .   .   .   .   .   .   .   .   .   .   .   .
8. 2008-05-04 17:13:00.000   H   ED SERVICE EVENT COMPLETED - DISCHARGED   1   .   .   .   .   .   .   .   .   .   .   .   .   .
9. 2007-04-11 00:58:00.000   YY   DID NOT WAIT   1   .   .   .   .   .   .   .   .   .   .   .   .   .
10. 2007-01-23 01:46:00.000   A   ADMITTED (EXCL ED BED)   1   777834   777834-2   1   4.0   4213   3   2007-01-22 21:48   2   1   PAE ED   4.0   2007-01-2
11. 2008-04-11 11:00:00.000   H   ED SERVICE EVENT COMPLETED - DISCHARGED   1   .   .   .   .   .   .   .   .   .   .   .   .   .
12. 2007-06-09 05:10:00.000   A   ADMITTED (EXCL ED BED)   1   386127   386127-3   2   26.0   4211   3   2007-06-09 02:40   2   1   MD1   GMAU   1.0   2007-06-0
The REDI Result: a linked health care journey

John Van Smith, approx 45, Male

Linked data for John allows us to answer research questions:

How long did it take the ambulance to get to John, from the time he called the ambulance?

How long did John wait to see a Dr in the ED?

How long did he wait in the ED?

Did he get admitted to hospital? Did he die in hospital?
So what?
(Relevance to health care RESEARCHERS)

1. Time saving

2. Allows researchers to access linked but de-identified data sets for further analysis and reporting (as per ethics agreements)

3. Inform policy and practice development
So what?
(Relevance to health care PROVIDERS)

- Potential to have an integrated health care record
- Utilizing existing systems and information
- More ‘real time’
- Multiple hospitals - benchmark
So what?
(Relevance to health care PROVIDERS)

ED activity – now

The table below provides an up-to-date view of the status of the Emergency Department in each of the metropolitan hospitals that contains an Emergency Department.

**Preview of Emergency Department (ED) figures at Tuesday, 11 August 2009 02:24 PM**

<table>
<thead>
<tr>
<th>Hospital</th>
<th>Ambulance diverted</th>
<th>Triage 4 patients average waiting time (minutes)</th>
<th>Number of patients waiting to be seen in ED</th>
<th>Total patients in ED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Armadale/Kalamunda District Memorial Hospital</td>
<td>No</td>
<td>45</td>
<td>12</td>
<td>38</td>
</tr>
<tr>
<td>Fremantle Hospital</td>
<td>No</td>
<td>71</td>
<td>3</td>
<td>40</td>
</tr>
<tr>
<td>Joondalup Health Campus</td>
<td>No</td>
<td>67</td>
<td>15</td>
<td>53</td>
</tr>
<tr>
<td>King Edward Memorial Hospital for Women</td>
<td>No</td>
<td>35</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Princess Margaret Hospital for Children</td>
<td>No</td>
<td>25</td>
<td>7</td>
<td>31</td>
</tr>
<tr>
<td>Rockingham General Hospital</td>
<td>No</td>
<td>66</td>
<td>15</td>
<td>31</td>
</tr>
<tr>
<td>Royal Perth Hospital</td>
<td>No</td>
<td>63</td>
<td>10</td>
<td>53</td>
</tr>
<tr>
<td>Sir Charles Gairdner Hospital</td>
<td>No</td>
<td>145</td>
<td>13</td>
<td>56</td>
</tr>
<tr>
<td>Swan District Hospital</td>
<td>No</td>
<td>27</td>
<td>4</td>
<td>17</td>
</tr>
</tbody>
</table>

Print the whole page
Improved outcomes and safety
Benefit from efficiencies in process improvement
Now What…

• Stage 1 - Completion of pilot
  (2 months data, 1 hospital)
  • Create one dataset showing all patient presentations to emergency via ambulance, then admitted into hospital.
  • Analyse data
• Stage 2 – Complete research project
  (2 years data, 3 hospitals)
  • Setup and configure a production ready HDI Server
  • Link and merge a larger dataset showing all patient presentations to emergency via ambulance, then admitted into hospital.
  • Analyse data
Questions?