Outline

• Background
• Quality in Data Linkage
• Study Example
• Reporting Standards
Benefits of Data Linkage

• Data is distributed over various departments, data systems and organisations

• Patient outcomes are hard to track after leaving hospital

• Many variables not captured in healthcare databases (e.g. detailed SES info, occupational surveillance data, carcinogen exposure data)

Data Linkage is Popular

• PubMed Search of “Data Linkage” returned **167679** results

• State/National centres have been established in Australia

• Health data collection is rapidly growing

• New entities are entering the market
Quality in Data Linkage

- Linked Data
  - True Links (TP)
  - False Matches (FP)
  - Total Links

- Non-links
  - Missed matches (FN)
  - True Negatives (TN)
  - Total Non-links

Actual match status

<table>
<thead>
<tr>
<th>Matches</th>
<th>Non-matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td>FN</td>
<td>TN</td>
</tr>
</tbody>
</table>

Total matches | Total Non-matches | Total Record pairs
The accuracy of data linkage has been shown to vary by:

- Participants’ age
- Gender
- Ethnicity
- Native language
- Regional location
- Education level

Quality in Data Linkage

• Different linkage operators have produced different results using the same software and data (Hoving et al, 2006)

• Different linkage software have been shown to produce different results with the same operators and data (Campbell et al, 2008)

Outline

• Background
• Quality in Data Linkage
• Study Example
• Reporting Standards
### Linkage Variables

<table>
<thead>
<tr>
<th>“Gold Standard”</th>
<th>“Practical Method”</th>
</tr>
</thead>
<tbody>
<tr>
<td>URN</td>
<td>Age at admission</td>
</tr>
<tr>
<td>Date of Birth</td>
<td>Gender</td>
</tr>
<tr>
<td>Age at admission</td>
<td>Hospital</td>
</tr>
<tr>
<td>Gender</td>
<td>Hospital LOS</td>
</tr>
<tr>
<td>Hospital</td>
<td>Hospital Admission Date</td>
</tr>
<tr>
<td>Hospital LOS</td>
<td>Hospital Discharge Date</td>
</tr>
<tr>
<td>Hospital Admission Date</td>
<td>ICU days</td>
</tr>
<tr>
<td>Hospital Discharge Date</td>
<td>ICU days</td>
</tr>
</tbody>
</table>
Results: Characteristics of Unmatched Cases

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Odds of being unmatched (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (female vs male)</td>
<td>1.23 (1.02-1.49)</td>
</tr>
<tr>
<td>Transfer (Yes vs No)</td>
<td>1.44 (1.16-1.8)</td>
</tr>
<tr>
<td>Overseas-born? (Yes vs No)</td>
<td>1.37 (1.15-1.62)</td>
</tr>
<tr>
<td>Elective admission?</td>
<td>0.66 (0.52-0.84)</td>
</tr>
<tr>
<td>Length of Hosp Stay&gt; 20 days</td>
<td>2.32 (1.73-3.12)</td>
</tr>
</tbody>
</table>
Outline

- Background
- Quality in Data Linkage
- Study Example
- Reporting Standards

Reporting Standards

- REporting of studies Conducted using Observational Routinely-collected Data (RECORD)

- Extension of the STROBE guidelines (STrengthening the Reporting of OBservational studies in Epidemiology)

- 22 items in total
### Reporting Standards

<table>
<thead>
<tr>
<th>Item</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RECORD 12.3</td>
<td>Linkage</td>
<td>State whether the study included person-level, institutional-level or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.</td>
</tr>
<tr>
<td>RECORD 13.1</td>
<td>Participants</td>
<td>Describe in detail the selection of the persons included in the study (i.e., study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.</td>
</tr>
</tbody>
</table>

---

**Linkage Process**

- Assessed for eligibility from data source 1 \( (n = ...) \)
- Assessed for eligibility from data source 2 \( (n = ...) \)
- Records pairs evaluated* \( (n = ...) \)
  - Definite matches \( (n = ...) \)
  - Possible matches \( (n = ...) \)
  - Non-matches \( (n = ...) \)
  - Clerical review required \( (n = ...) \)
- Records not linked \( (n = ...) \)
- Total records linked \( (n = ...) \)
- Analysed \( (n = ...) \)
- Excluded from analysis \( (n = ...) \) (give reasons)
- \( (n = ...) \) (for which variables)
- Excluded prior to \( (n = ...) \)
Final Recommendations

1. Increasing the use of routine data linkage will help to improve data quality

2. Systematic, quality assessment of linkage is needed

3. Ensure integrity and completeness of datasets before linkage

4. If potential linkage bias is identified, analyses should account for this

5. Data linkage studies should be reported in a clear and systematic manner

6. Unique identifiers, such as national health identifiers, would improve linkage quality