Analysing complex linked administrative data in health services research: issues and solutions

Associate Professor Rachael Moorin, Mr David Youens, Ms Ninh Ha
Data linkage (DL)

- “Data matching” – methods that identify records in different collections that belong to the same person
- Commonly used to enable medical and public health research (cradle to grave)
- Australia internationally recognised for DL and DL-based research
- **How possible?** Operates through exemptions in privacy laws – needs ethics approval + public benefit must exceed risk to privacy
Cradle to Grave: Health information over the lifespan

Birth Records | Hospital Records | Cancer Registries | Death Records

CRADLE | GP Records | PBS Records | Veterans’ Affairs / Ageing | GRAVE
Does continuity of primary care reduce demand on emergency department presentations and hospital admissions?

- **Ageing population**
  - High prevalence of chronic illnesses
  - Ever increasing expectations in service delivery
  - Barely manageable in the short-term
  - Projections indicate long-term significant challenge

- **Strategy to reduce unsustainable demand on services**
  - Shifting service delivery from the acute to the primary health care sector.
    - Reduce costs and demand on hospitals
    - Mechanism = Managed care
      - Prevent chronic condition reaching a level requiring acute care/intervention

- **Theory of ‘Ambulatory Care Sensitive Condition’**
  - Admission to hospital is considered potentially preventable through early diagnosis and/or earlier intervention

- **Continuity of Primary Care**
  - Move from reactionary to proactive ongoing contact
    - Managing the condition
Rationale for our study

- Potentially preventable hospitalisations (PPHs) & ED presentations indicate potential problems with access to / quality of primary care.
  - Represent an indirect measure of access to primary care and
  - The capacity to manage chronic health problems.

- Theoretical link proposed between PPHs and ongoing primary care (GP) contact BUT limited supporting evidence
  1. Patterns of accessing primary care translate into health outcomes.
  2. What aspects of primary care (eg regular care, time period, same GP) provide any/best benefit
  3. Under what circumstances (socio-demographic, clinical) this theory holds true.
Brief overview of data

Challenges & Opportunities
<table>
<thead>
<tr>
<th>Data holdings for our study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>WA Set 1</strong></td>
</tr>
<tr>
<td>Age Range</td>
</tr>
<tr>
<td>Hospitalisations</td>
</tr>
<tr>
<td>ED Presentations</td>
</tr>
<tr>
<td>Deaths</td>
</tr>
<tr>
<td>MBS (out-of-hospital medical services, Imaging &amp; Pathology)</td>
</tr>
<tr>
<td>PBS (Dispensed Medicines)</td>
</tr>
<tr>
<td>Self-reported HLth status, lifestyle behaviours, SES, Qol etc</td>
</tr>
<tr>
<td>Historical Electoral Roll</td>
</tr>
<tr>
<td>General practice data (GP contacts, prescribing &amp; imaging/pathology tests)</td>
</tr>
<tr>
<td>Population coverage</td>
</tr>
</tbody>
</table>
Defining clinical cohorts = Challenge

- Who is at risk of the outcome?
  - Markers of this in the data
    - Data NOT collected for research
    - Data represent “services delivered”
  - Some tests/medicines indicative for some conditions
  - Hospitalisations / ED - definitive diagnosis
    - But only capture those with severe disease

- If those not at risk in the cohort
  - No outcomes regardless of exposure levels

- If only those with severe disease in cohort
  - Results not/limited generalisable

Example: WA Diabetes cohort

Overall study population (no indication of diabetes risk)

At-risk of diabetes
(Diagnosis of hypertension or ischaemic heart disease)
n= 878,414

High-risk of diabetes
(Some evidence of testing for diabetes)
n=362,887

Confirmed diabetics
(Hospitalisation, or evidence of management of diabetes)
n=408,913
Longitudinal nature of the data

• Capturing person time @ risk (Exposed & Unexposed)

• Individuals move in – out- in - .... of time at risk
  - Exposure + Outcome
  - Move out of WA or Into Hospital

• Unbalanced complex panel data = Challenge

• Can incorporate time varying covariates in models
  - = Opportunity
Complex models required = Challenge

- **Outcomes**
  - Potentially preventable hospitalisations (PPH)
    - Count outcome
  - Bed Days for PPHs
    - Count outcome
  - Cost of PPHs
    - Continuous outcome

- **Other data challenges:**
  - Changes in availability of tests over time (ascertainment bias)
  - Likelihood of prior health service use influencing the dependent variable (initial conditions & simultaneity/reverse causality bias)
  - Likely correlation of observed and unobserved variables.

*Large propn of zero’s*

Overdispersion in count outcomes

Models require separate components for zero + non-zero outcomes
Opportunity: Defining the exposure: GP contact

• Aim = To evaluate & separate out three components of primary care contact as provided by GPs

1. i) ‘Regularity’ (As previously defined + some enhancements)
2. ii) ‘Cover’ (a time-limited protective effect of primary care)
3. iii) Provider specific continuity (Same GP and/or Same practice)
Chronic Disease Management

- Encourage contact with GP at regular intervals

**Defining the EXPOSURE:**
Metric built from individual level linked data

1. **Regularity**
Measures how well distributed the service utilisation is, not how often.

David will present on an updated metric
2. Cover (new metric)

Ninh will present on the new metric

\[ C = \frac{\sum \text{Days under warranty}}{\text{Ascertainment period (days)}} \]

Value between 0 & 1
3. Continuity of provider (Usual provider index)

= provider A
= provider B

Time

This measure does not incorporate regularity

Continuity = 0.75

Paper in development: Continuity of provider-adjusted regularity
Regular General Practitioner Contact –
Methods of measurement using administrative data

David Youens, Mark Harris, Suzanne Robinson, David Preen, Rachael Moorin
Background
Provider continuity

Jim:
- • = provider A
- x = provider B

Time

Continuity = 0.75

Frequency = 4

Continuity scores are the same yet patterns look quite different

Joan:

Time

Continuity = 0.75

Frequency = 4
Regularity of General Practitioner (GP) contact

Previous Australian research:

• Regular GP contacts were associated with increased time to first hospitalization or death among patients with certain chronic conditions (1)

• Where GPs claimed certain Medicare care planning incentive items in relation to a patient, GP contact was more regular in the following year with no change in the frequency (amount) of contacts (2)

Previous American research

• Regular primary care contacts were associated with early detection of breast cancer (3)

3. The Journal of Rural Health 31:269-281
Aims

Compare two existing and one new measure of regularity of GP contact

Assess whether existing regularity indices are associated with frequency of contacts

Assess whether any association with frequency biases the estimation of hospitalization outcomes
Methods
Methods – regularity indices

- Indicates date of GP contact

\[
r = \frac{1}{1 + \text{var}(\text{days})}
\]

<table>
<thead>
<tr>
<th>Year start</th>
<th>Year end</th>
</tr>
</thead>
<tbody>
<tr>
<td>(days_1)</td>
<td>(days_2)</td>
</tr>
</tbody>
</table>
Methods – regularity indices

![Diagram of regularity indices]

- Indicates date of GP contact

\[ r = \frac{1}{1 + var(days)} \]

Levels:
- Annual
- Biannual
- Quarterly
- Bimonthly
Methods – regularity indices

- Indicates date of GP contact

\[ r = \frac{1}{1 + var(days)} \]

Levels:
- Annual
- Biannual
- Quarterly
- Bimonthly
Methods – regularity indices

- Indicates date of GP contact

\[ r = \frac{1}{1 + \text{var}(\text{days})} \]

Levels:
- Annual
- Biannual
- Quarterly
- Bimonthly
Methods – regularity indices

- Indicates date of GP contact

\[ r = \frac{1}{1 + \text{var}(\text{days})} \]

Levels:
- Annual
- Biannual
- Quarterly
- Bimonthly
Methods – regularity indices

• Indicates date of GP contact

\[
\begin{align*}
\text{Year start} & \quad \text{days}_1 \quad \text{days}_2 \quad \text{days}_3 \quad \text{Year end} \\
\text{Indicates date of GP contact} & \\
\end{align*}
\]

\[
r = \frac{1}{1 + \text{var} (\text{days})}
\]

\[
r = 1/(1 + \left( \frac{\text{sd} (\text{days})}{\text{mean} (\text{days})} * 100 \right))
\]
Methods – additional notes

Outcomes
Diabetes potentially preventable hospitalisations

Covariates
Sex, age, indigenous status, socio-economic status, remoteness, comorbid condition count, count of specialist physician contacts, risk level

<table>
<thead>
<tr>
<th>Interval index</th>
<th>Variance index</th>
<th>Relative variance index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 = annual contact</td>
<td>1 = least regular contact</td>
<td>1 = least regular contact</td>
</tr>
<tr>
<td>2 = biannual</td>
<td>2 = 2nd least regular</td>
<td>2 = 2nd least regular</td>
</tr>
<tr>
<td>3 = quarterly</td>
<td>3 = 2nd most regular</td>
<td>3 = 2nd most regular</td>
</tr>
<tr>
<td>4 = bimonthly</td>
<td>4 = most regular</td>
<td>4 = most regular</td>
</tr>
</tbody>
</table>

Panel data: years within individuals
Methods – analysis

1. Is regularity associated with frequency for each of these indices?
   Regress frequency (count) on each regularity index separately - negative binomial models, with least regular group as reference category
   Compare coefficients at each “level” of regularity
   Compare BIC values

2. Does any association identified in (1) confound the relationship between regularity and an outcome of interest?
   Regress diabetes-related hospitalisation in year $t$ on regularity in year $t-1$ separately for each regularity index
   Repeat with frequency included as a covariate
   Interpret changes in coefficients for models with / without frequency included
Diagrammatically

Regularity of contact

Frequency of contact

A
Diagrammatically

Regularity of contact → B → Hospital admission

Frequency of contact

A → B

C → B
Diagrammatically

Regularity of contact

Hospital admission

Frequency of contact

A

B

C
Results
## Cohort characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>76,642</td>
<td>49.96</td>
</tr>
<tr>
<td>Male</td>
<td>76,772</td>
<td>50.04</td>
</tr>
<tr>
<td><strong>SEIFA quintile</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest disadvantage</td>
<td>27,967</td>
<td>18.44</td>
</tr>
<tr>
<td>High disadvantage</td>
<td>39,990</td>
<td>26.36</td>
</tr>
<tr>
<td>Moderate disadvantage</td>
<td>22,280</td>
<td>14.69</td>
</tr>
<tr>
<td>Less disadvantage</td>
<td>23,688</td>
<td>15.62</td>
</tr>
<tr>
<td>Least disadvantage</td>
<td>37,763</td>
<td>24.90</td>
</tr>
<tr>
<td><strong>ARIA category</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very remote</td>
<td>4,418</td>
<td>2.91</td>
</tr>
<tr>
<td>Remote</td>
<td>2,490</td>
<td>1.64</td>
</tr>
<tr>
<td>Moderately Accessible</td>
<td>7,470</td>
<td>4.92</td>
</tr>
<tr>
<td>Accessible</td>
<td>8,153</td>
<td>5.37</td>
</tr>
<tr>
<td>Highly Accessible</td>
<td>129,184</td>
<td>85.15</td>
</tr>
<tr>
<td><strong>Indigenous status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indigenous</td>
<td>5,689</td>
<td>4.00</td>
</tr>
<tr>
<td>Non-indigenous</td>
<td>136,670</td>
<td>96.00</td>
</tr>
<tr>
<td><strong>Risk status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confirmed diabetes</td>
<td>44,998</td>
<td>29.33</td>
</tr>
<tr>
<td>High risk</td>
<td>108,416</td>
<td>70.67</td>
</tr>
<tr>
<td><strong>Variable</strong></td>
<td><strong>mean</strong></td>
<td><strong>SD</strong></td>
</tr>
<tr>
<td>Age</td>
<td>56.24</td>
<td>14.56</td>
</tr>
<tr>
<td>Frequency</td>
<td>10.15</td>
<td>8.77</td>
</tr>
<tr>
<td>Diabetes-related hospital separations in year</td>
<td>0.19</td>
<td>3.45</td>
</tr>
<tr>
<td>Comorbid condition count</td>
<td>0.94</td>
<td>2.10</td>
</tr>
<tr>
<td>Bed days associated with diabetes-related separations in year</td>
<td>0.89</td>
<td>9.26</td>
</tr>
<tr>
<td>Total</td>
<td><strong>153,414</strong></td>
<td><strong>100.00</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regularity level</th>
<th>% (variance / relative variance index)</th>
<th>% (interval index)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least / annual</td>
<td>25.00%</td>
<td>12.76%</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; least / biannual</td>
<td>25.00%</td>
<td>40.32%</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; most / quarterly</td>
<td>25.00%</td>
<td>23.75%</td>
</tr>
<tr>
<td>Most / bimonthly</td>
<td>25.00%</td>
<td>23.18%</td>
</tr>
</tbody>
</table>
Results 1 – associations between regularity and frequency

Figure: Associations between regularity and frequency according to the three indices, derived from negative binomial models.
Results 1 – associations between regularity and frequency

Figure: Associations between regularity and frequency according to the three indices, derived from negative binomial models
Results 1 – associations between regularity and frequency

Figure: Associations between regularity and frequency according to the three indices, derived from negative binomial models

<table>
<thead>
<tr>
<th>Coefficients from regressions of frequency on levels of regularity (least regular group forming reference category)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variance index</strong></td>
</tr>
<tr>
<td>-0.6  0.0  1.2</td>
</tr>
<tr>
<td><strong>Relative variance index</strong></td>
</tr>
<tr>
<td>-0.6  0.0  0.8</td>
</tr>
<tr>
<td><strong>Interval index</strong></td>
</tr>
<tr>
<td>-0.6  0.0  1.6</td>
</tr>
</tbody>
</table>

- **Regularity level (baseline least regular)**
  - Biannual
  - Quarterly
  - Bimonthly

- **Regularity level (baseline least regular)**
  - Biannual
  - Quarterly
  - Bimonthly

- **Regularity level (baseline annual contact)**
  - Biannual
  - Quarterly
  - Bimonthly
Results 1 – associations between regularity and frequency

Figure: BIC values from negative binomial models predicting frequency of GP contact. Four separate models with frequency regressed on set of covariates without regularity, then with each index separately.
Results

- Regularity of contact
- Frequency of contact
- Hospital admission

Connections:
- A connects Regularity of contact to Frequency of contact
- B connects Frequency of contact to Hospital admission
- C connects Hospital admission to Regularity of contact
Results 2 – associations between regularity and hospitalization outcomes

Figure: Associations between regularity and diabetes-related hospitalisation rate according to the three indices. Six models are estimated; one with and without frequency of GP visits included as a covariate for each index
Results 2 – associations between regularity and hospitalization outcomes

Figure: Associations between regularity and diabetes-related hospitalisation rate according to the three indices. Six models are estimated; one with and without frequency of GP visits included as a covariate for each index.
Results 2 – associations between regularity and hospitalization outcomes

Figure: Associations between regularity and diabetes-related hospitalisation rate according to the three indices. Six models are estimated; one with and without frequency of GP visits included as a covariate for each index.
Results 3 – other considerations

Other considerations

• Two of these indices require exact dates of GP contacts while one doesn’t – data availability
• Levels of the interval index are more meaningful which might make interpretation of results easier
• The indices respond differently to changes in the underlying pattern of contacts
Conclusion

1. Regularity of contacts could be considered to capture additional information on continuity of care, alongside continuity of provider

2. The choice of index can matter, depending on research aims

3. Controlling for frequency, as in some previous literature, resolves this issue

4. Limitations – lack of contextual information on contacts, lack of objective information on health status
Development of a time-duration measure of continuity of primary care: A threshold effects approach to identify optimal primary health care use for diabetes

Ninh Ha, Mark Harris, David Preen, Suzanne Robinson, Rachael Moorin

CIC seminar 2018
Background

• Regularity is suitable as a target for health policy intervention

• Lack of time component in regularity index

• Hypothesis: Interaction with a GP can have a protective effect from adverse events like hospitalisations within a particular optimal time interval
Aim

Our study aimed to develop a methodology for determining “cover” of primary care using individual-level linked administrative data by

• (i) estimating the optimal maximum time interval over which primary care affords an increased protection from PPH using threshold effects models; and

• (ii) using the derived optimal time period to operationalize “cover” at the individual level.
The proportion of time that an individual is under the potentially protective effect from GP visit over a pre-specified ascertainment period.
Methods

Estimating the optimal maximum time interval
### Study population

People with diabetes living in a period between 1998/99 and 2003/04

### Inclusion

- Aged 45+ years
- Being resident in WA at least 2 consecutive years

### Measures

- Number of diabetes related potentially preventable hospitalisation (PPH)
- GP utilisation
- Demographic characteristics

### Data sources

- Hospital records
- Medicare claim records
- WA Electoral roll records
- Mortality data
- Hospital records
- Medicare claim records
- WA Electoral roll records
How the studied cohorts were constructed?


July 1998

0 complication

1-2 complications

3+ complications

Couplet years

Main measures

Individual’s characteristics
- Age
- Gender
- SEIFA
- Comorbidity
- Complications
- Hospital & GP usage

Main measures
- PPH
- GP usage

June 2004
**Measure GP usage**

<table>
<thead>
<tr>
<th>svc_type</th>
<th>date_care</th>
</tr>
</thead>
<tbody>
<tr>
<td>hospital admission</td>
<td>15feb1999</td>
</tr>
<tr>
<td>GP visit</td>
<td>15mar1999</td>
</tr>
<tr>
<td>GP visit</td>
<td>01apr1999</td>
</tr>
<tr>
<td>GP visit</td>
<td>04apr1999</td>
</tr>
<tr>
<td>GP visit</td>
<td>21may1999</td>
</tr>
<tr>
<td>GP visit</td>
<td>27aug1999</td>
</tr>
<tr>
<td>hospital admission</td>
<td>05nov1999</td>
</tr>
<tr>
<td>GP visit</td>
<td>29nov1999</td>
</tr>
<tr>
<td>GP visit</td>
<td>03mar2000</td>
</tr>
<tr>
<td>GP visit</td>
<td>02jun2000</td>
</tr>
<tr>
<td>GP visit</td>
<td>02aug2000</td>
</tr>
<tr>
<td>hospital admission</td>
<td>04oct2000</td>
</tr>
<tr>
<td>GP visit</td>
<td>25oct2000</td>
</tr>
</tbody>
</table>

**Time intervals to GP visits**

**GP measures in a financial year:**
- The maximum time interval to GP (MTI)
- Number of GP visit
- Regularity
Analysis

- All analyses were conducted separately for each complication cohort
- Data: panel structure, complex and unbalanced
- Outcome variables: count number of PPH, over-dispersed count outcome variables
- Main predictor: the maximum time interval to GP visit (MTI) (months)
- Control variables: demographic characteristics, comorbidities, and history of PPH and GP usage
- The optimal maximum time interval to a GP visit was estimated using threshold effects model (Gannon, Harris, and Harris 2014).
Empirical model

The standard negative binomial model for relationship between PPH and MTI is as follow:

\[ PPH_{it}^* = \beta MTI_{it} + \alpha X_{it} + U_i \]

The model is extended to threshold effects to allow for differential effect of the MTI on PPH with respect to an individual’s position with regard to MTI

\[ PPH_{it} = \sum_{m=1}^{M} Y_m(R_{m,i} \ast MTI_{i,t}) + \bar{R}_{i,t} + U_i \]

\[ R_{m,i} = 1 \text{ if } \{\tau_{m-1} < MTI_{i,t} \leq \tau_m\}, \text{ and 0 otherwise; where } m \text{ is the number of subpopulation and } \tau \text{ is the threshold parameters} \]

The preferred model was the one which minimized the appropriate information criteria (AIC and BIC)
Cover metric calculation

- Measuring total days under GP cover using the optimal maximum time interval found in the threshold effects model

- Measuring total days in the ascertainment period, excluding Days in hospital or temporary residence outside of WA

\[
\text{Cover index} = \frac{\sum \text{(time under cover)}}{\text{Ascertainment period}}
\]

The proportion of time that an individual is under the GP cover over a pre-specified ascertainment period.
Characteristics of studied cohorts

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>No complication</th>
<th>One or two complications</th>
<th>Three complications or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>20039 (46.0)</td>
<td>15866 (41.6)</td>
<td>1969 (4.6)</td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-59</td>
<td>8323 (46.0)</td>
<td>3863 (25.9)</td>
<td>1,863 (11.6)</td>
</tr>
<tr>
<td>60-74</td>
<td>8050 (43.2)</td>
<td>7578 (48.3)</td>
<td>4,706 (29.4)</td>
</tr>
<tr>
<td>≥75</td>
<td>2166 (10.8)</td>
<td>3619 (22.9)</td>
<td>2,151 (13.2)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>9741 (48.6)</td>
<td>7263 (48.8)</td>
<td>5080 (46.6)</td>
</tr>
<tr>
<td>Male</td>
<td>10298 (51.4)</td>
<td>7603 (51.2)</td>
<td>5,730 (53.4)</td>
</tr>
<tr>
<td>Indigenous status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>17911 (95.9)</td>
<td>13,937 (93.7)</td>
<td>9,880 (92.1)</td>
</tr>
<tr>
<td>Yes</td>
<td>771 (4.1)</td>
<td>929 (6.2)</td>
<td>850 (7.9)</td>
</tr>
<tr>
<td>SEIFA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest Disadvantage</td>
<td>3951 (19.8)</td>
<td>3,232 (21.9)</td>
<td>2,445 (22.9)</td>
</tr>
<tr>
<td>High disadvantaged</td>
<td>5540 (27.8)</td>
<td>4,120 (26.2)</td>
<td>3,238 (29.4)</td>
</tr>
<tr>
<td>Moderate disadvantage</td>
<td>2702 (14.0)</td>
<td>2,100 (13.4)</td>
<td>1,495 (13.4)</td>
</tr>
<tr>
<td>Less disadvantaged</td>
<td>3225 (16.5)</td>
<td>2,238 (15.1)</td>
<td>1,582 (14.5)</td>
</tr>
<tr>
<td>Least disadvantaged</td>
<td>4412 (22.2)</td>
<td>2,878 (19.5)</td>
<td>2,085 (19.0)</td>
</tr>
<tr>
<td>Accessibility</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very remote</td>
<td>553 (2.8)</td>
<td>686 (4.2)</td>
<td>525 (4.9)</td>
</tr>
<tr>
<td>Remote</td>
<td>359 (1.8)</td>
<td>277 (1.8)</td>
<td>186 (1.7)</td>
</tr>
<tr>
<td>Moderate</td>
<td>945 (4.7)</td>
<td>770 (5.2)</td>
<td>602 (5.6)</td>
</tr>
<tr>
<td>Accessible</td>
<td>1339 (6.5)</td>
<td>1,041 (6.7)</td>
<td>620 (5.6)</td>
</tr>
<tr>
<td>Highly accessible</td>
<td>17004 (85.6)</td>
<td>12265 (83.0)</td>
<td>8,716 (81.9)</td>
</tr>
<tr>
<td>Number of comorbidity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>3.0 (2.8)</td>
<td>5.7 (3.2)</td>
<td>8.3 (3.3)</td>
</tr>
<tr>
<td>Duration of diabetes (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>5.4 (3.8)</td>
<td>7.3 (4.2)</td>
<td>9.2 (4.2)</td>
</tr>
<tr>
<td>Regularity quantiles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No regularity</td>
<td>4708 (23.8)</td>
<td>3,225 (22.2)</td>
<td>2,382 (22.2)</td>
</tr>
</tbody>
</table>

No complication cohort:
- Younger age
- Lower number of comorbidities
- Lower number of years in diabetes

Higher complication cohorts:
- Older age
- Higher number of comorbidities
- Higher number of years in diabetes
Estimation of the optimal time intervals for each diabetes cohort: threshold effects model

<table>
<thead>
<tr>
<th>Complication cohorts</th>
<th>No complication (a)</th>
<th>One or two complications (b)</th>
<th>Three or more complications (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subpopulations</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>AIC</td>
<td>45169.0</td>
<td>45091.2</td>
<td>45049.5</td>
</tr>
<tr>
<td>BIC</td>
<td>45398.5</td>
<td>45329.9</td>
<td>45297.3</td>
</tr>
<tr>
<td>Threshold parameters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_1$</td>
<td>8</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>$t_2$</td>
<td>13</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>$t_3$</td>
<td>13</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>$t_4$</td>
<td>13</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>$t_5$</td>
<td>13</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Coefficients</td>
<td>$g_1$</td>
<td>-0.229***</td>
<td>-0.249***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.03)</td>
</tr>
<tr>
<td></td>
<td>$g_2$</td>
<td>-0.083***</td>
<td>-0.126***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td></td>
<td>$g_3$</td>
<td>-0.013</td>
<td>-0.101*</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td></td>
<td>$g_4$</td>
<td>0.004</td>
<td>0.072***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td></td>
<td>$g_5$</td>
<td>0.024</td>
<td>0.057***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td></td>
<td>$g_6$</td>
<td>0.022</td>
<td>-</td>
</tr>
</tbody>
</table>
Estimation of the optimal time intervals for each diabetes cohort: spline function

No complication

One or two complications

Three or more complications
Cover score by each optimal period boundary over the studied period

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Low bound cover</th>
<th>Middle bound cover</th>
<th>Upper bound cover</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cover index</td>
<td>95%CI</td>
<td>Cover index</td>
</tr>
</tbody>
</table>
| Overall                         | 0.88 (0.87;0.89)| 0.91 (0.90;0.92)   | 0.93 (0.92;0.94)  | 90% of ascertainment period is under GP cover for people with diabetes.  
The lowest average cover index scores:  
- 75 years or older,  
- Male,  
- Indigenous,  
- Highest disadvantage  
- Living very remote areas
| Age group (years)               |                 |                    |                   |                    |             |                    |
| 45-59                           |                 |                    |                   |                    |             |                    |
| 60-74                           |                 |                    |                   |                    |             |                    |
| ≥75                             | 0.85 (0.85;0.86)| 0.89 (0.88;0.89)   | 0.91 (0.90;0.92)  | 90% of ascertainment period is under GP cover for people with diabetes.  
The lowest average cover index scores:  
- 75 years or older,  
- Male,  
- Indigenous,  
- Highest disadvantage  
- Living very remote areas
| Gender                          |                 |                    |                   |                    |             |                    |
| Female                          | 0.91 (0.91;0.93)| 0.93 (0.92;0.94)   | 0.95 (0.94;0.96)  | 90% of ascertainment period is under GP cover for people with diabetes.  
The lowest average cover index scores:  
- 75 years or older,  
- Male,  
- Indigenous,  
- Highest disadvantage  
- Living very remote areas
| Male                            | 0.87 (0.87;0.89)| 0.90 (0.90;0.92)   | 0.92 (0.91;0.93)  | 90% of ascertainment period is under GP cover for people with diabetes.  
The lowest average cover index scores:  
- 75 years or older,  
- Male,  
- Indigenous,  
- Highest disadvantage  
- Living very remote areas
| Indigenous status               |                 |                    |                   |                    |             |                    |
| No                              | 0.90 (0.90;0.92)| 0.93 (0.92;0.94)   | 0.95 (0.94;0.96)  | 90% of ascertainment period is under GP cover for people with diabetes.  
The lowest average cover index scores:  
- 75 years or older,  
- Male,  
- Indigenous,  
- Highest disadvantage  
- Living very remote areas
| Yes                             | 0.72 (0.71;0.74)| 0.78 (0.77;0.79)   | 0.82 (0.81;0.83)  | 90% of ascertainment period is under GP cover for people with diabetes.  
The lowest average cover index scores:  
- 75 years or older,  
- Male,  
- Indigenous,  
- Highest disadvantage  
- Living very remote areas
| SEIFA                           |                 |                    |                   |                    |             |                    |
| Highest Disadvantage            | 0.87 (0.87;0.89)| 0.90 (0.90;0.92)   | 0.92 (0.91;0.93)  | 90% of ascertainment period is under GP cover for people with diabetes.  
The lowest average cover index scores:  
- 75 years or older,  
- Male,  
- Indigenous,  
- Highest disadvantage  
- Living very remote areas
| High disadvantaged              | 0.89 (0.89;0.90)| 0.92 (0.92;0.93)   | 0.94 (0.93;0.95)  | 90% of ascertainment period is under GP cover for people with diabetes.  
The lowest average cover index scores:  
- 75 years or older,  
- Male,  
- Indigenous,  
- Highest disadvantage  
- Living very remote areas
| Moderate disadvantage           | 0.89 (0.89;0.90)| 0.92 (0.91;0.93)   | 0.94 (0.93;0.95)  | 90% of ascertainment period is under GP cover for people with diabetes.  
The lowest average cover index scores:  
- 75 years or older,  
- Male,  
- Indigenous,  
- Highest disadvantage  
- Living very remote areas
| Less disadvantage               | 0.90 (0.89;0.90)| 0.92 (0.92;0.93)   | 0.94 (0.93;0.95)  | 90% of ascertainment period is under GP cover for people with diabetes.  
The lowest average cover index scores:  
- 75 years or older,  
- Male,  
- Indigenous,  
- Highest disadvantage  
- Living very remote areas
| Least disadvantage              | 0.90 (0.90;0.91)| 0.93 (0.92;0.94)   | 0.94 (0.93;0.95)  | 90% of ascertainment period is under GP cover for people with diabetes.  
The lowest average cover index scores:  
- 75 years or older,  
- Male,  
- Indigenous,  
- Highest disadvantage  
- Living very remote areas
| Accessibility                   |                 |                    |                   |                    |             |                    |
| Very remote                     | 0.68 (0.66;0.70)| 0.75 (0.73;0.77)   | 0.80 (0.78;0.82)  | 90% of ascertainment period is under GP cover for people with diabetes.  
The lowest average cover index scores:  
- 75 years or older,  
- Male,  
- Indigenous,  
- Highest disadvantage  
- Living very remote areas
| Remote                          | 0.83 (0.81;0.85)| 0.87 (0.85;0.89)   | 0.91 (0.89;0.93)  | 90% of ascertainment period is under GP cover for people with diabetes.  
The lowest average cover index scores:  
- 75 years or older,  
- Male,  
- Indigenous,  
- Highest disadvantage  
- Living very remote areas
| Moderate                        | 0.86 (0.85;0.87)| 0.90 (0.89;0.91)   | 0.92 (0.91;0.93)  | 90% of ascertainment period is under GP cover for people with diabetes.  
The lowest average cover index scores:  
- 75 years or older,  
- Male,  
- Indigenous,  
- Highest disadvantage  
- Living very remote areas
| Accessible                      | 0.88 (0.87;0.89)| 0.91 (0.90;0.92)   | 0.93 (0.92;0.94)  | 90% of ascertainment period is under GP cover for people with diabetes.  
The lowest average cover index scores:  
- 75 years or older,  
- Male,  
- Indigenous,  
- Highest disadvantage  
- Living very remote areas
| Highly accessible               | 0.90 (0.90;0.91)| 0.93 (0.92;0.94)   | 0.94 (0.93;0.95)  | 90% of ascertainment period is under GP cover for people with diabetes.  
The lowest average cover index scores:  
- 75 years or older,  
- Male,  
- Indigenous,  
- Highest disadvantage  
- Living very remote areas

90% of ascertainment period is under GP cover for people with diabetes.
The lowest average cover index scores:
- 75 years or older,
- Male,
- Indigenous,
- Highest disadvantage
- Living very remote areas
Discussion

- This new measure of COC adds a time parameter to capturing longitudinal continuity
- Cover has the potential to better capture underuse of PC and will significantly contribute to the sparsely available methods for analysis of linked administrative data in evaluating COC

Limitations:
- Using data from 1990 to 2004, hence it cannot provide evidence regarding current utilization of GP services
- Administrative data are not collected for research purposes, hence, they do not include some details about severity of disease
Acknowledgements:

National Health and Medical Research Council

WA Data Linkage Branch, data custodians and all those whose data were used

Thank you

Make tomorrow better.