An Introduction to Complex Data for Professionals in Rheumatology

Kelvin Jordan
Presenter Disclosure Information:
This speaker has no conflicts of interest
The context

• Data collected at multiple time points on the same individuals (longitudinal)

• Ideas / opportunities for design and analysis
  – Whilst thinking about missing data
  – And sources of data

• IDEAS GENERATION RATHER THAN “HOW TO”
Short term change may be stronger predictor than baseline measurement e.g. for predictors of outcome following consultation.
Course of pain intensity

Month of follow-up

Pain intensity
Course of pain intensity

Pain intensity

Month of follow-up
Course of pain intensity

Pain intensity

Month of follow-up

-2 -1 0 1 2 . . . 34 35 36 37 38
The Challenge

• Longitudinal studies often use single measure taken at baseline with one follow-up point
  – “one-off” estimates cannot accurately reflect the recurrent and fluctuating nature of many conditions
  – they may also miss key transitions in the long-term prognosis of conditions
Issues with follow-up in a trial

Mean pain intensity over time for intervention and control groups.

Month of follow-up:
- Baseline
- 3 months
- 6 months
- 12 months

- Intervention (blue)
- Control (red)
Analysis issues

• Analysing data at each time point separately, or aggregating data, does not make full use of information

• Can get complicated with many time points!

• Ignoring association between repeated measures on same people is unwise
  
  – Stats speak! “correlated residuals, biased parameter estimates, underestimated standard errors”
2 potential approaches

Many approaches, we are going to consider 2:
1. Multilevel modelling
2. Latent class analysis (LCA)
Multilevel Modelling

- Extension of linear / logistic regression
- Modelling of hierarchical data ("clustering")
  - patients within clinician within hospital (3 level)
  - repeated measurements within participants (2 level)
Example - 1

Vasseljen et al., Natural course of acute neck and low back pain in the general population: The HUNT study, *Pain* 2013; 154:1237-44

- 219 people with new episode of neck or low back pain
- Pain intensity measured 1m, 2m, 3m, 6m, 12m
- Compare course of pain by initial site of pain (neck, back, both)
Analysis

• Multilevel linear regression with 2 levels: repeated measures within patients

• Relationship of pain intensity with initial site (neck, back, or both neck & back)
  – Adjusting for age, gender, previous history of pain, work status (“confounders”)
  – Interaction with time (“effect modification” by time)

• Results presented as predicted levels of pain, holding other variables constant at mean values

Vasseljen et al., *Pain* 2013; 154:1237-44
# Predicted mean pain intensity score by initial site of pain

<table>
<thead>
<tr>
<th>Baseline pain</th>
<th>Month 1*</th>
<th>Month 2</th>
<th>Month 3</th>
<th>Month 6</th>
<th>Month 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neck only</td>
<td>3.6 (3.3, 4.0)</td>
<td>2.7</td>
<td>2.5</td>
<td>2.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Back only</td>
<td>3.7 (3.3, 4.1)</td>
<td>2.2</td>
<td>2.5</td>
<td>2.0</td>
<td>2.1</td>
</tr>
<tr>
<td>Neck &amp; Back</td>
<td>3.6 (2.9, 4.4)</td>
<td>3.7</td>
<td>3.3</td>
<td>3.9</td>
<td>3.4</td>
</tr>
</tbody>
</table>

Adjusted for age, sex, work status, previous pain history

* With 95% CI

Vasseljen et al., *Pain* 2013; 154:1237-44
Why use multilevel model?

• Analyse all patients and all time points together
• More efficient and statistically robust than analysing time points separately

• Stats Speak! Include growth terms (linear, quadratic)
• Can assess how effect of something (e.g. treatment) changes over time
• Can be adapted to all standard regression approaches (linear, logistic,...)

• Can include patients with missing data!
• Time points can vary between patients
• Applicable to interventional or observational studies
2 potential approaches

Many approaches, we are going to consider 2:

1. Multilevel modelling
2. Latent class analysis (LCA)
Example - 2

Dunn et al., Characterizing the course of low back pain: a latent class analysis, *Am J Epidemiol* 2006; 163:754-61.

- 342 back pain consulters - monthly pain intensity for following 6 months
- Identify different trajectories (pathways) of back pain
- Analysed using latent class analysis (LCA)
Characteristics of LCA

• We can group (cluster) people together based on something we don’t already know ("latent") — Trajectories (pathways) of pain over 6 months
• Use statistical measures and interpretation of clusters to decide number of clusters
• Each participant is allocated to 1 cluster
• Participants should match the cluster-specific trajectory
Back pain trajectories

Mean pain intensity scores

Baseline 1 2 3 4 5 6

Month

Dunn et al., *Am J Epidemiol* 2006; 163:754-761
Persistent mild self-reported back pain trajectories (n=122, 36%)

Dunn et al., Am J Epidemiol 2006; 163:754-761
Fluctuating self-reported back pain trajectories (n=45, 13%)

Dunn et al., *Am J Epidemiol* 2006; 163:754-761
Characteristics of LCA

• Probabilities used to describe the characteristics of each cluster
  – Probability of severe pain given cluster membership

• Probabilities used to allocate people to a cluster
  – Probability of belonging to each cluster
  – Allocated to cluster with highest probability

• Missing data allowed!
Short term trajectories as predictors of long term outcomes

• At 12m, those with ‘severe’ short term trajectory had more:
  – Disability
  – Time off work
  – Anxiety
  – Depression...

  ➢ And still had most severe pain 7 years later
Issues with longitudinal self-reported / clinic measured data

• Missing data
• Non-response
• Time consuming
• Expensive
• “Black holes”
  – Apps, diaries, ...?
Routine medical record data

• Increased availability (national datasets)
• Large populations
• Good quality (primary care)
• Continuous measurement over long periods of time (restricted information)

➢ Course, management, and outcome of morbidity
➢ Link to self-report?
Summary

1. Introduction to study design
   – Interventional vs observational
   – Case-control vs cohort

2. Statistical tests and interpretation
   – Build up analysis – descriptive first
   – Assess confounders vs effect modifiers

3. Ideas for longitudinal data analysis
   – Several follow-up points
   – Analyses that allow for missing data