Using administrative data for social work research

Rick Hood and Allie Goldacre
Department of Social Work and Social Care, Kingston University
12 November 2021
Aims of the session

Hopefully by the end of this morning you will be able to:

• Understand what is meant by administrative data as well as the difference between aggregated and individual-level data

• Describe some research questions that could be answered using administrative data and the appropriate methods to answer such questions

• Refer to examples of research that have used administrative data to examine aspects of social care provision

• Appraise the advantages and limitations of doing research with this kind of data.
Information created when people interact with public services, such as schools, the NHS, the courts or the benefits system, and collated by government (ADR-UK, 2021)

Examples: educational attainment (DfE), tax and incomes (HMRC), hospital records (NHS), National Census (ONS)

By-product of government services, i.e. not created for research purposes

May include personal and/or sensitive data
Benefits of using AD for research

• Resource implications – purpose-built datasets often time-consuming and difficult to complete.

• Access to population-level data – greater statistical power, ability to detect modest but meaningful relationships, detect rare events, and study heterogenous effects

• Feasibility of data linkage, e.g. to Census data about same population

• Real time nature of data – allows topical questions, e.g. effects of recent events / new policies

• Possibility of ‘natural experiments’

• Data quality and comprehensiveness
Aggregated vs individual AD

**Aggregated data**
- Averaged or combined by year, area, organisation, etc.
- Often publicly available
- Not personal or sensitive

**Individual-level data**
- Refers to individuals
- Often personal and/or sensitive
- Restrictions around access and use
Questions you could answer with aggregated data

• How many referrals were made to children’s social care last year?
• Has demand for children’s social care services gone up in recent years?
• Do rates of children in need vary between Local Authorities?
• Is there a relationship between local authority rates of children in care and levels of deprivation?
• Which Local Authorities have seen the biggest cuts to expenditure in children’s social care?
• Are changes in expenditure on Early Help associated with subsequent changes in rates of statutory provision?
• What is the impact of an Ofsted inspection on rates of intervention?
• Is it possible to predict a local authority’s Ofsted rating using the administrative data?

Explore the data: https://webb.shinyapps.io/cwip-app-v2/
Correlations

Hood et al., 2019

Notes:
1. Children in need during the year
2. CP plans at 31 March
3. Percentage of CIN plans ceased
4. Percentage of CP plans ceased
5. Percentage of referrals
6. Rate per 10,000 LA child population
7. Average IMD score

Key:
- - Positive correlation
— — Negative correlation
Limitations of aggregated data

- Difficult to perform subgroup analysis, e.g. whether an intervention works better for some people than for others
- Risk of ‘ecological fallacy’ – correlations between two variables measured at aggregate level cannot substitute for correlations at the individual level as they are not the same (Robinson, 1950)
- Not possible to do some types of analysis, e.g. not possible to follow individuals longitudinally
- Variation in reporting practices and quality of data
Questions you would (usually) need individual-level data to answer

- Can we identify and profile the underlying types of demand for CSC services in England?
- How do children’s needs evolve over time for those receiving multiple episodes of intervention?
- Which children are more likely to experience a re-referral to children’s social care, or re-enter care after they return home?
- What are the gaps in educational attainment of children who receive statutory services compared with those who don’t?
- Do children who receive statutory services experience poorer health outcomes later in life?
### Event data for individuals

#### EVENTS

<table>
<thead>
<tr>
<th>Event</th>
<th>pre-birth</th>
<th>infancy</th>
<th>early childhood</th>
<th>childhood</th>
<th>teenage years</th>
<th>adulthood</th>
<th>older age</th>
<th>death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSC episode</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLA episode</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KS2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KS4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital admission</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GP consultation</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Death</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Children in need census
- Intervention pathways (e.g. CIN, S47, CPP)
- Factors at assessment
- CP Plan category of abuse
- Re-referrals / repeat CP Plans
- Sources of referral

#### Children looked after returns
- Reason for new episode
- Reason for episode cease
- Placement type
- Legal status
- Category of need

#### Schools census
- KS1
- KS2
- KS4
- SEND
- Free School meals

#### Hospital episode statistics (HES)
- Diagnosis - 3 characters (DIAG_3_NN)
- Method of admission (ADMIMETH)
- Source of admission (ADMISORC)
- Treatment specialty (TRETSPEF)

#### GP clinical records
- Diagnosis
- Treatment details

#### Office for National Statistics (ONS) mortality data
- Causes of death
Survival analysis

Figure 1: Kaplan-Meier plot showing re-entry to care following reunification
Figure 2: Characteristics of the reunified cohort and factors associated with re-entry to care

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Hazard ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (Ref: Males):</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Females</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Age (Ref: Under 1):</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>1-4 Years</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>5-9 Years</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>10-15 Years</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>16-17 Years</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Ethnicity (Ref: White):</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Asian</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Black</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Mixed</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Other</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Prior POC (Ref: No):</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Yes</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>POC length (Ref: &lt; 12 months):</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>12+ months</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Avg. placement length (Ref: &lt; 3 months):</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>3 to 9 months</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>9+ months</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Placement moves within 1 year (Ref: 0 to 1):</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>2+</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Section 20 (exit) (Ref: No):</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Yes</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Placement (exit) (Ref: Foster care):</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Foster care (kin)</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Childrens homes</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Other residential</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Independent living</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Secure unit / young offender</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Other</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Placement provider (exit) (Ref: LA):</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Private</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Voluntary/third sector</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Distance from home (exit) (Ref: &lt;5 miles):</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>6 to 20 miles</td>
<td>1x (the same rate)</td>
</tr>
<tr>
<td>Over 20 miles</td>
<td>1x (the same rate)</td>
</tr>
</tbody>
</table>

0.5x lower 1x (the same rate) 1.5x higher 2x higher

Hazard ratio
Limitations of using individual-level data

• Access, ethics and data protection
• Statistical disclosure control (SDC)
• Correlations vs explanations
• Population comparisons
• Barriers to data linkage
To find out more…

**Useful websites:**

[https://www.adruk.org/](https://www.adruk.org/)
[https://saildatabank.com/](https://saildatabank.com/)

**Projects on children’s social care:**

[https://www.nuffieldfoundation.org/project/outcomes-of-different-types-of-demand-for-childrens-social-care](https://www.nuffieldfoundation.org/project/outcomes-of-different-types-of-demand-for-childrens-social-care)

**Work by other UK researchers:**


**Email:** rick.hood@sgul.Kingston.ac.uk
References


