Linking survey and administrative data to enhance understandings of risky behaviours and potential protective factors in children receiving social care: A feasibility study’

Social Care Fellowship

Helen HODGES
CASCADE, Cardiff University
1 April 2022 to 31 March 2027
Introducing the SHRN Data

• The School Health Research Network (SHRN) is a policy-practice-research partnership between Welsh Government, Public Health Wales, and Cardiff University established in 2013.

• SHRN aims to improve young people’s health and wellbeing in Wales by working with schools in both primary and secondary education to generate and use good quality evidence for health improvement.

• This includes student and school-level surveys, capturing key health and wellbeing metrics. These metrics are referenced in over 30 national policies and strategies, including:
  • The Whole School Approach to Mental Health and Wellbeing (2021)
  • Estyn’s Healthy and Happy Report (2019)

I will be using data from the 2019, 2021, 2023 and potentially the 2025 student-level survey and linking this to other datasets in SAIL Databank
So …

Just how representative is my data of the children receiving care and support cohort?
In the perfect world …

EDUW

CRCS

SHRN

LACW

4 key datasets underpin my research.
Coverage: 210 schools invited, 198 participated (94%).

Eligible: Those aged 11-17 on 31 August 2019, in mainstream secondary schools in Wales. However, not all schools with sixth forms opted to include Y12 and Y13 pupils.

Ineligible: Those aged 11-17 attending special schools or in alternative provision (EOTAS).

The Children Receiving Care and Support (CRCS) cohort includes:

- Children looked after (CLA)
- Children on the Child Protection Register (CPR)
- Children receiving care or support who are neither CLA or CPR (“Other”)

Info taken from National report & website.

In the perfect world …
In the perfect world …

Coverage: 210 schools invited, 198 participated (94%).

Eligible: Those aged 11-17 on 31 August 2019, in mainstream secondary schools in Wales. However, not all schools with sixth forms opted to include Y12 and Y13 pupils

Ineligible: Those aged 11-17 attending special school or in alternative provision (EOTAS)

Both the LACW and CRCS would be episode data and it would be possible to determine their status on any given day. Unfortunately, it is only possible to do this with the LACW.

All children looked after would be placed and educated in Wales.
In the perfect world …

We would have a whole population study where all secondary age pupils have the opportunity to respond to the Survey.

The anonymised data would be beautifully clean and each individual would have a consistent, unique identifier.

The reality is that its messy!
EDUW (PLASC + EOTAS)

- PLASC collected mid-Jan 2020, EOTAS collected mid-Feb 2020
- 188,953 uniquely identified pupils aged 11-17 on 31st August 2019
- All have an IRN for linking to the CRCS and LACW
- After boosting, **187,057 (99%) have an ALF** for linking to the SHRN data. 1,875 w/o an ALF

SHRN 2019 Responses

- Fieldwork undertaken between 16/9/2019 and 19/12/2019
- 32,558 sets of responses from those who consented to data linkage
- No IRN
- **23,233 uniquely identified pupils with an ALF (86.7%)**, aged 11-17 on 31 August 2019

CRCS

- Return reflects status on 31 March 2020. However, to be included then need to have an open case for at least 3 months prior => CRCS on 1 January 2020
- 6,878 children aged 11-17 on 31 August 2019
- 6,754 (98.2%) have an IRN (missing x124)
- All have a hybridID for linking to the LACW
- After boosting, **6,668 (96.9%) have an ALF** for linking to other datasets (missing x210).

LACW

- Episode data so able to determine if looked after when completed SHRN survey
- 3,814 children aged 11-17 on 31 August 2019 were looked after at some point during the year ending 31 March 2020.
- All have a hybridID for linking to the CRCS
- After boosting, **3,691 (96.8%) had an ALF** (missing x 123)
How messy?
How representative?

Where there is no ALF, it is not possible to link across datasets

The temporal coverage of the children’s social services data means that we do not know their status when they completed the survey (if eligible), unless they were CLA.

Between completing the Survey, some pupils may change schools or enter EOTAS.

Some CLA are place outside Wales therefore may not be educated in Wales. There are also some legal status codes which suggest that although resident in Wales, they may not be appear in PLASC or EOTAS
Piloted in 2017 with a limited number of schools, pupils have the option to consent to data linkage. If they consent, they are asked to provide unique identifiers to enable an ALF to be generated.

Need to determine amongst these groups whether they would have been eligible to participate or not:

⇒ Eligible but did not consent / participate
⇒ Ineligible
### Key Tasks

<table>
<thead>
<tr>
<th>To what extent has consent introduced biased? [Using SHRN + EDUW]</th>
<th>How representative are the SHRN responses in SAIL of the CRCS cohort? [Using CRCS &amp; LACW, and comparing this to SHRN + EDUW]</th>
</tr>
</thead>
<tbody>
<tr>
<td>What proportion were eligible to participate on basis of type of establishment?</td>
<td>In addition to the first set of questions ... How do the profiles compare relative to peers?</td>
</tr>
<tr>
<td>How does the profile of those <strong>eligible</strong> compare to those that were <strong>ineligible</strong>?</td>
<td>Is the CRCS cohort under-/over- represented relative to their peers? - there is scope here to look at differences by CRCS status, legal status and placement type / location as well as pupil characteristics</td>
</tr>
<tr>
<td>Of those who were eligible, what proportion gave consent for data linkage ie in SAIL?</td>
<td>How does the approached used outside SAIL to determine if care experienced compared to the ‘official’ status from the LACW? Is there scope for improvement?</td>
</tr>
<tr>
<td>How does the profile of those <strong>eligible and completed</strong> compare to those <strong>eligible but no responses in SAIL</strong>?</td>
<td></td>
</tr>
<tr>
<td>Of those <strong>eligible and completed</strong>, how many provided adequate information for an ALF to be generated?</td>
<td></td>
</tr>
<tr>
<td>How does the profile of those <strong>with an ALF</strong> compare to those <strong>without and ALF</strong>?</td>
<td></td>
</tr>
<tr>
<td>To what extent has consent introduced biased? [Using SHRN + EDUW]</td>
<td>How representative are the SHRN responses in SAIL of the CRCS cohort? [Using CRCS &amp; LACW, and comparing this to SHRN + EDUW]</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>What proportion were eligible to participate on basis of type of establishment?</td>
<td>In addition to the first set of questions ... How do the profiles compare relative to peers?</td>
</tr>
<tr>
<td>How does the profile of those eligible compare to those that were ineligible?</td>
<td>Is the CRCS cohort under-/over-represented relative to their peers? - there is scope here to look at differences by CRCS status, legal status and placement type / location as well as pupil characteristics</td>
</tr>
<tr>
<td>Of those who were eligible, what proportion gave consent for data linkage ie in SAIL?</td>
<td>How does the approached used outside SAIL to determine if care experienced compared to the ‘official’ status from the LACW? Is there scope for improvement?</td>
</tr>
<tr>
<td>How does the profile of those eligible and completed compare to those eligible but no responses in SAIL?</td>
<td></td>
</tr>
<tr>
<td>Of those eligible and completed, how many provided adequate information for an ALF to be generated?</td>
<td></td>
</tr>
<tr>
<td>How does the profile of those with an ALF compare to those without and ALF?</td>
<td></td>
</tr>
</tbody>
</table>
Next Steps

• Comparisons of the profiles to establish if there are key demographic characteristics and/or status which are under-/over-represented in the SHRN data held in SAIL Databank

• Repeat the exercise with the SHRN 2021 responses

• Address the question:

  Is consent a source of bias when linking survey and administrative data to enhance understandings of risky behaviours and potential protective factors in children receiving social care?

•Ascertain whether the measure being used outside SAIL to identify those who are looked after can be improved upon -> recommendations

• Feedback to the SHRN Advisory Group
Future Plans

• Drill down to undertake cross-sectional analysis with respect to each of the risky behaviours – **cross sectional analysis**

• Use datasets such as EDUW (for truanting, exclusions, bullying), SMDS (for substance misuse) and health datasets to supplement what is known about the pupil’s engagement in the different behaviours

• Track individual responses over time to see if there are different trajectories for different sub-groups – **longitudinal analysis**

• At each stage, link in with CASCADE Voices and/or ALPHA to check interpretation, language used and for their views on implications