Using data linkage to understand the impact of social and ethnic inequalities in health

24th Sept 2022
Professor Ronan Lyons
Secure Anonymised Information Linkage (SAIL) Databank

- Built from the ground up as a safe, legal, social acceptable solution to data linkage and data sharing
- Underpinned by world-leading technical infrastructure (SeRP)
- “No data leaves” model, keeping data under control at all times
- Highly automated, highly efficient, highly secure, completely trustworthy
- Creating unsurpassed research-ready linked data resources for both research community and public sector analytical teams
- SAIL is not limited to Welsh data but supports many UK and international projects
1) Data sharing agreements and secure transfer
2) Reliable matching process
3) Anonymisation and encryption
4) Disclosure control and risk mitigation
5) Data access controls and agreements
6) Scrutiny of data use proposals and results
7) External verification of compliance with IG

Safe Projects
Safe People
Safe Data
Safe Settings
Safe Outputs

The ‘Five Safes’: a framework for planning, designing and evaluating data access solutions (Felix Ritchie)
Stakeholder & Public Engagement

- The SAIL Consumer Panel

- SAIL Databank has a long standing Consumer Panel, established in 2011

- It actively recruits and is involved in all aspects of SAIL, from bids to IGRP to dissemination and impact
Total population data linkage

- Tracking 3.2M pop with multiple datasets in privacy protecting environment with public engagement
- Population demographic spine, including 2011 Census data and health, social care and education workforce
- Infection: antigen and antibody
- Viral subtypes
- Emergency department attendances
- Hospital outpatient and admissions
- Critical care dataset
- General practice
- All laboratory tests
- Deaths
- Longer term outcomes
- Vaccination
- Embedded trials and cohorts
Evaluating the effect of COVID-19 on dispensing patterns: a national cohort analysis

Fatemeh Torabi, Ashley Akbari, Laura North, Daniel Harris, Gareth Davies, Mike Gravenor, Rowena Griffiths, Jane Lyons, Neil Jenkins, Andrew Morris, Julian Holcik, Ronan A. Lyons
doi: https://doi.org/10.1101/2021.02.15.21251552
This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.


Joe Hollinghurst, Jane Lyons, Richard Fry, Ashley Akbari, Mike Gravenor, Alan Watkins, Fiona Verity, Ronan A Lyons
Age and Ageing, Volume 50, Issue 1, January 2021, Pages 25–31,
https://doi.org/10.1093/ageing/afa207
Published: 19 September 2020

Surgical activity in England and Wales during the COVID-19 pandemic: a nationwide observational cohort study

doi: https://doi.org/10.1101/2021.02.27.21251593
This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

Staff–pupil SARS-CoV-2 infection pathways in schools in Wales: a population-level linked data approach

Daniel A Thompson, Hoda Abbassianjani, Richard Fry, Emily Marchant, Lucy Griffiths, Ashley Akbari, Joe Hollinghurst, Laura North, Jane Lyons, Fatemeh Torabi, Gareth Davies, Mike B Gravenor, Ronan A Lyons

What is known about the subject?
- Evidence of the role school play in the transmission of SARS-CoV-2 is limited.
- Higher positivity rates are observed in school staff compared to pupils.
- Lack of evidence on transmission pathways transmissi on into and within schools.
PHIRI: EU Project Use Case A

• Has the COVID19 pandemic changed existing patterns of non-COVID-19 health care utilisation for (vulnerable) populations within and between countries?

  - Heart attack and strokes (Cohort 1)
  - Hip and knee replacements (Cohort 2)
  - Serious trauma admissions (Cohort 3)

• Method: Compare age-standardised utilization rates for each month of 2020 (and possibly 2021) compared with pre-existing trends during 2017-2019, supplemented by ecological analyses and comparisons using data on infections and hospitalizations from ECDC.

Objectives of the study:

✓ Demonstrate how a broad variety of secondary data (e.g. administrative and survey data) can be pooled and/or reused in a distributed way across Europe
✓ Gain insights into the situation of socially (and potentially clinically) vulnerable groups during the COVID-19 pandemic
✓ Understand gaps in health system performance during crisis
✓ Develop learnings on system resilience and inclusive pandemic preparedness
EASR for myocardial infarction and inequality ratio in Wales, UK

The European age standardised rate of strokes (A) and heart attacks (C) occurring in Wales with the corresponding ratio of individuals with the highest deprivation score to those with the lowest (B and D)
The European age standardised rate of elective knee replacement surgeries (A) and elective hip replacement surgeries (C) occurring in Wales with the corresponding ratio of individuals with the highest deprivation score to those with the lowest (B and D).
Use Case B: Breast cancer during COVID-19: detection and treatment

Courtesy of Sarah Aldridge – PHIRI project www.phiri.eu
Wales Multimorbidity Cohort: Survival by Deprivation
Fifth in those in their 40s or 60s at inception
Developing and comparing a population-scale linked data ethnicity-spine in Wales to assess COVID-19 outcomes

Ashley Akbari, Fatemeh Torabi, Stuart Bedston, Emily Lowthian, Hoda Abbasizanjani, Rich Fry, Jane Lyons, Rhiannon Owen, Kamlesh Khunti, Ronan A. Lyons

• 26 data sources within SAIL

• Derive a set of harmonised ethnic group spine for the population of Wales.
• Longitudinal, varied ethnicity values into a harmonised, de-duplicated lean research ready data asset (RRDA).

• Four different approaches: mode, most recent (similar to population estimates), weighted mode, composite (if >1 then code to mixed)

• Compared the coverage and frequency of records for the entire population of Wales using an existing national-scale e-cohort (C20) across sex, age groups and major COVID-19 outcomes: infection, vaccination, hospitalisation and death.
Datasets used
Classifications of ethnicity – from COVID19 NERVTAG and ONS
Ethnicity distribution – by NERVTAG groupings and method

• 46 million ethnicity codes found between 2000 and 2021
• across 26 data sources for 4,297,694 individuals
• 3,345,694 individuals who were alive and residing in Wales as identified in the C20 national-scale e-cohort for Wales on the 1st January 2020
• 4.4% had not had any associated ethnicity records in their health records

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>3,080,277 (89.0%)</td>
<td>3,093,817 (89.0%)</td>
<td>2,987,420 (86.0%)</td>
<td>2,987,245 (86.0%)</td>
</tr>
<tr>
<td>Bangladeshi</td>
<td>25,780 (0.7%)</td>
<td>16,243 (0.5%)</td>
<td>55,675 (1.6%)</td>
<td>9,204 (0.3%)</td>
</tr>
<tr>
<td>Chinese</td>
<td>17,670 (0.5%)</td>
<td>17,851 (0.5%)</td>
<td>20,302 (0.6%)</td>
<td>13,703 (0.4%)</td>
</tr>
<tr>
<td>Indian</td>
<td>25,058 (0.7%)</td>
<td>25,190 (0.7%)</td>
<td>27,078 (0.8%)</td>
<td>17,075 (0.5%)</td>
</tr>
<tr>
<td>Pakistani</td>
<td>17,035 (0.5%)</td>
<td>17,043 (0.5%)</td>
<td>20,287 (0.6%)</td>
<td>11,030 (0.3%)</td>
</tr>
<tr>
<td>Black African</td>
<td>21,578 (0.6%)</td>
<td>22,003 (0.6%)</td>
<td>32,671 (0.9%)</td>
<td>9,904 (0.3%)</td>
</tr>
<tr>
<td>Black Caribbean</td>
<td>4,730 (0.1%)</td>
<td>4,083 (0.1%)</td>
<td>7,145 (0.2%)</td>
<td>1,135 (&lt;0.1%)</td>
</tr>
<tr>
<td>Mixed</td>
<td>43,672 (1.3%)</td>
<td>43,572 (1.3%)</td>
<td>66,895 (1.9%)</td>
<td>217,059 (6.3%)</td>
</tr>
<tr>
<td>Other</td>
<td>68,262 (2.0%)</td>
<td>64,560 (1.9%)</td>
<td>86,889 (2.5%)</td>
<td>38,007 (1.1%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>153,332 (4.4%)</td>
<td>153,332 (4.4%)</td>
<td>153,332 (4.4%)</td>
<td>153,332 (4.4%)</td>
</tr>
</tbody>
</table>

Ronan Lyons and Rhiannon Owen, Swansea University Medical School
on behalf of the One Wales Team
All-cause mortality Wales: 2016-Jan 2022

- Expected (95% Prediction Interval)
- Observed

Cause of death
- All-cause
- COVID

Dominant Variant
- Alpha
- Delta
- No variant information
- Omicron

Royal Statistical Society
Data | Evidence | Decisions
Place of Death
All-cause and COVID mortality by ethnicity (a) White population (b) Asian population, by dominant variant variant
Mortality by ICD10 chapter: (a) respiratory (b) circulatory
Mortality by ICD10 chapter: (a) neoplasms, (b) digestive disorders
Mortality by ICD10 chapter: (a) mental health, (b) trauma
Ethnic differences in COVID-19 mortality during the vaccine roll-out in England

Matthew Bosworth, Vahé Nafilyan
Health Analysis and Life Events Division
Office for National Statistics
Background

- Ethnic differences in Covid-19 mortality well documented
- What are the factors explaining these differences?
- What is the role of differences in COVID-19 vaccination uptake?
- Use new linked data based on Census

Source: Public Health England CHIME tool
Obtaining NHS numbers for Census respondents

Of the 53,483,502 Census records:
- 50,019,451 were linked deterministically.
- 555,291 additional matches were obtained using probabilistic matching.
- **Total linkage rate: 94.6%**
## Linkage failure by ethnic group

<table>
<thead>
<tr>
<th>Ethnic Group</th>
<th>N at 2011 Census</th>
<th>N linked to Patient Register</th>
<th>Rate of linkage failure (%)</th>
<th>Unadjusted OR (95% CI)</th>
<th>OR (95% CI) adjusted for age and sex</th>
<th>OR (95% CI) adjusted for age, sex, ethnicity, region and IMD decile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bangladeshi</td>
<td>245,164</td>
<td>226,755</td>
<td>7.5</td>
<td>1.86 (1.83-1.89)</td>
<td>1.28 (1.26-1.30)</td>
<td>0.92 (0.91-0.94)</td>
</tr>
<tr>
<td>Black African</td>
<td>599,118</td>
<td>527,772</td>
<td>11.9</td>
<td>3.10 (3.07-3.12)</td>
<td>2.36 (2.34-2.38)</td>
<td>1.76 (1.74-1.77)</td>
</tr>
<tr>
<td>Black Caribbean</td>
<td>445,503</td>
<td>413,411</td>
<td>7.2</td>
<td>1.78 (1.76-1.80)</td>
<td>1.72 (1.70-1.74)</td>
<td>1.30 (1.28-1.31)</td>
</tr>
<tr>
<td>Chinese</td>
<td>286,419</td>
<td>242,791</td>
<td>15.2</td>
<td>4.11 (4.07-4.16)</td>
<td>3.00 (2.97-3.03)</td>
<td>2.69 (2.66-2.72)</td>
</tr>
<tr>
<td>Indian</td>
<td>1,036,678</td>
<td>979,485</td>
<td>5.5</td>
<td>1.34 (1.33-1.35)</td>
<td>1.05 (1.04-1.06)</td>
<td>0.92 (0.92-0.93)</td>
</tr>
<tr>
<td>Mixed</td>
<td>524,235</td>
<td>439,168</td>
<td>16.2</td>
<td>4.44 (4.40-4.47)</td>
<td>3.20 (3.18-3.23)</td>
<td>2.76 (2.73-2.78)</td>
</tr>
<tr>
<td>Pakistani</td>
<td>650,738</td>
<td>610,439</td>
<td>6.2</td>
<td>1.51 (1.50-1.53)</td>
<td>1.09 (1.08-1.10)</td>
<td>0.84 (0.83-0.85)</td>
</tr>
<tr>
<td>White British</td>
<td>31,923,883</td>
<td>30,587,948</td>
<td>4.2</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>White other</td>
<td>2,440,994</td>
<td>2,220,570</td>
<td>9.0</td>
<td>2.27 (2.26-2.28)</td>
<td>1.78 (1.77-1.79)</td>
<td>1.56 (1.55-1.57)</td>
</tr>
<tr>
<td>Other</td>
<td>1,095,623</td>
<td>873,949</td>
<td>20.2</td>
<td>5.81 (5.78-5.84)</td>
<td>4.54 (4.51-4.56)</td>
<td>3.72 (3.70-3.74)</td>
</tr>
</tbody>
</table>
Data on 40.0 million people aged 9+, alive on 24th Jan 2020

Demographics:
sex, age, ethnicity, religion

Census

Socio-economics:
Household deprivation, education, SES, tenure, housing, household composition

Census

Geographical factors:
Place of residence, Care home/private household, Population density, rural/urban, area deprivation

2019 Patient register

Occupational exposure:
Occupation; O*NET exposure measures

Census

Pre-existing conditions:
As per QCOVID risk model: BMI, learning disability, cancer and immunosuppression, chronic kidney disease, diabetes, COPD, other pulmonary diseases, coronary heart disease, stroke, dementia, severe mental illness, etc

GPES/HES

Outcomes:
COVID-19 mortality & hospitalisation
Hospitalisation & diagnosis in primary care

Office for National Statistics

<table>
<thead>
<tr>
<th>Age</th>
<th>Coverage (% of total pop)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9+</td>
<td>78.6%</td>
</tr>
<tr>
<td>40+</td>
<td>85.3%</td>
</tr>
<tr>
<td>65+</td>
<td>89.1%</td>
</tr>
</tbody>
</table>
Vaccination coverage by ethnic groups (1 dec 2021)
HRs for death involving COVID-19 by ethnic group during the third wave of the pandemic (13 June 2021 to 1 December 2021)
Summary

• SAIL has a lot to offer for future research on inequalities
• Updating the ethnicity spine with 2021 Census will enable much better analyses to be undertaken
• Funds needed for data analysts to answer questions
• Social Justice theme of the Administrative Data Research Wales plan analyses on many marginalised groups, including ethnic minority groups
• Need for studies of interventions to reduce inequalities