

The **Networked Data Lab**: Statistical analysis plan  
for Topic 3 studying unpaid carers' and looked  
after children's health

Satellite analysis for North West London

Evgeny Galimov, Roberto Fernandez Crespo, Sandeep Prashar, Emma  
Sharpe-Jones, Clare McCrudden, Rachel McCarthy, Matthew Chisambi

17.06.2022

Version 6.0

## Contents

1. Research rationale and objectives .....	3
Network Data Lab .....	3
Unpaid carers .....	3
Looked after children .....	3
Aims, research questions and objectives .....	4
2. Data and data linkages .....	7
3. Statistical methods .....	7
3.1 Study design .....	7
3.2 Study period .....	7
3.3 Study population .....	7
3.4 Definitions of outcomes and exposures .....	8
3.6 Methods for addressing missing data .....	11
3.7 Known limitations .....	11
4. Governance .....	11
5. Impact, dissemination and engagement .....	11
6. Appendix .....	12
Local Stakeholders .....	12
References .....	12

# 1. Research rationale and objectives

## Network Data Lab

The Network Data Lab is a collaborative network of 5 analytical teams across the UK working together on shared challenges and promoting the use of analytics in improving health and social care.

North West London Networked Data Lab (NWL NDL) is one of the 5 analytical teams funded by The Health Foundation and includes specialists from Imperial College Health Partners (IHP), North West London Health and Care Partnership, and the Institute of Global Health Innovation (IGHI).

The overall aim of the NDL is to improve health and care services and reduce health inequalities in the UK, with the current project specifically aiming to understand the needs, health issues and pathways to services of unpaid carers and looked after children.

## Unpaid carers

Social care is provided by professional carers working in care homes, local services, etc. (estimated to be 1.5 million adult workers) and unpaid carers - children and adults who look after a family member, partner or friend who needs help because of their illness, frailty, disability, a mental health problem. The number of unpaid carers is estimated between 7 and 13 million in England (1). The value of the care that unpaid carers provide to the state has been estimated to be £132 billion a year – almost the equivalent of a second NHS.

With access to formal care services disrupted by the pandemic, unpaid carers are now under increasing pressure. 81% of carers are providing more care since the COVID-19 pandemic began, partly due to increased needs of the person they care for (40% of carers), partly due to reduced local services (38% of carers) (2).

Many carers suffer from loneliness and isolation or face financial difficulties if they are unable to work alongside their caring role. Isolation, stress and performing caring are associated with poorer health. 36 out of 37 long-term conditions recorded in primary care were more prevalent among unpaid carers than non-carers, with largest differences seen for cancer, depression and irritable bowel syndrome (3). The COVID-19 pandemic worsened unpaid carers health even more. 64% of them reported a decline in mental health and 56% in physical health as a result of the pandemic (2).

Early stages of PPIE research (semi-structured qualitative interviews) suggests that unpaid carers may suffer, among others, from depression, anxiety, cardiovascular issues and high stress levels. Their caring responsibilities also make it more difficult for them to access healthcare services especially when waiting times to be seen are long. PPIE research findings also suggest that the COVID-19 pandemic may have led to an increased hours spent caring after others, isolation and a deterioration of their mental health.

## Looked after children

Looked after children or children in care are the children living with foster parents, ill parents, in a residential children's home or children who have experienced abuse at home.

There are approximately 100,000 looked after children in the UK, and their number has increased by 10% in the last 5 years. Looked after children are much more likely than other children to have experienced abuse and neglect, family dysfunction or disability, which are associated with the development of mental health disorders. It has been shown that looked-after children are four times more likely to suffer from mental health problems than their peers (4). Looked after children may also have higher prevalence of physical disorders, particularly of cerebral palsy, epilepsy and cystic fibrosis (5).

### Aims, research questions and objectives

The team will aim to explore questions from at least one of the three subtopics under the 2 main topics. The final scope of the project will depend on the feasibility of the specific question due to external factors e.g. data availability. For both cohorts we will seek to understand:

- Demographic profiles
- Geographical distribution
- Care service utilisation
- Health condition prevalence
- Health outcomes
- Wider determinants of health

Unpaid carers		
Type of analysis	Question	Outcome of interest
Descriptive analysis of unpaid carers cohort	1. What are the demographic characteristics of the unpaid carers' population?	% of population providing unpaid care split by age, IMD, gender and ethnicity (further split by ethnicity, gender and age, where possible)
	2. What are the geographical variations in the amount of unpaid caring at the local authority level?	% of population providing unpaid care per borough (further split by ethnicity, sex and age, where possible).
	3. What health impact (measured in A&E/primary care/outpatient visits) did the COVID-19 pandemic have on the unpaid carers' population?	A&E/primary care/outpatient visits per carer to describe the effect of COVID-19.
Analysis of health conditions more prevalent in unpaid carers and the effect of COVID-19	1. Is there an association between being an unpaid carer and higher healthcare risk (assessed through rising risk score, PAM score, BMI and/or frailty score)? BMI (as well as frailty index) is a proxy of health. Obese people are at increased risk of diabetes, hypertension, heart disease, stroke and total mortality.	PAM score, BMI, eFI to find associations with being a carer adjusted for age, gender, IMD and ethnicity.

	2. What health conditions are more prevalent among unpaid carers compared with non-carers?	Age-sex standardised prevalence rate for a set of long-term conditions to describe the effect of being a carer.
	3. What is the effect of COVID-19 on health conditions in unpaid carers?	Presence of long-term conditions from a predefined set predicted by being a carer, age, sex, IMD, ethnicity, COVID_19 pandemic period, interaction between being a carer and COVID-19 pandemic period.
	4. How does support, employment status (if available) and the amount of care influence unpaid carers' health?	Presence of long-term conditions from a predefined set predicted by being a carer, age, sex, IMD, ethnicity, employment status, carers' support and the amount of provided care
Analysis of healthcare services used by unpaid carers and the effect of COVID-19	1. Is there any difference in healthcare access between unpaid carers and non-carers?	GP visits/outpatients/ urgent care/referrals/prescriptions (related to long-term conditions from a predefined set) predicted by being a carer, age, sex, IMD, ethnicity.
	2. What is the effect of COVID-19 on healthcare access in unpaid carers?	Referral by specialty rates (to long-term conditions from a predefined set) predicted by being a carer, COVID-19 pandemic period, interaction between carer's Cohort and COVID-19 pandemic period.

Looked after children		
Type of analysis	Question	Outcome of interest
Descriptive analysis of looked after children cohort	1. What are the demographic characteristics of the looked after children population?	% of looked after children split by age, IMD, gender and ethnicity (further split by ethnicity, sex and age, where possible).
	2. What are the geographical variations in the number of looked after children?	% of looked after children per borough (further split by ethnicity, sex and age, where possible).
Analysis of health conditions more prevalent in looked after children and the effect of COVID-19	1. Is there an association between being a looked after child and healthcare risk (assessed through rising risk score, PAM score and/or BMI)?	PAM score/BMI to find associations with being a looked after child adjusted for age, gender, IMD and ethnicity.
	2. What health conditions are more prevalent among looked after children compared with other children?	Age-sex standardised prevalence rate a set of long-term conditions to describe the effect of being a looked after child.
	3. What is the effect of COVID-19 on health conditions of looked after children?	Presence of long-term conditions from a predefined set predicted by being a looked after child, age, sex, IMD, ethnicity, COVID_19 pandemic period, interaction between being a looked after child and COVID-19 pandemic period.
	4. How do health conditions vary dependent on the category on the type of looked after children	Presence of long-term conditions from a predefined set predicted by being a looked after child, age, sex, IMD, ethnicity, type of looked after children (fostered children, disabled children, children in residential children's homes).
Analysis of healthcare services used by looked after children and the effect of COVID-19	1. Is there any difference in healthcare access between looked after children and other children?	GP visits/outpatients/ urgent care/referrals/prescriptions (related to long-term conditions from a predefined set) predicted by being a looked after child, age, sex, IMD, ethnicity.
	2. What is the effect of COVID-19 on healthcare access in the case of looked after children?	Referral by specialty rates (to long-term conditions from a predefined set) predicted by being a looked after child, COVID-19 pandemic period, interaction between looked after children's Cohort and COVID-19 pandemic period.

## 2. Data and data linkages

In this study we use the longitudinal Discover dataset (the deidentified version of the Whole Systems Integrated Care (WSIC) database). This dataset provides linked coded primary care, secondary care, acute, mental health, community health and social care record for over 2.7 million patients who live and are registered with a GP in North West London. This dataset consists of data from over 400 provider organisations including 360 GP practices, 2 mental health and 2 community trusts and all acute providers attended by North West London patients (in the form of Secondary Uses Service (SUS) data). We will use demographic variables (age, gender, geographic location, ethnicity, IMD), various health metrics (BMI, PAM score, eFI<sup>1</sup>), data about GP visits, diagnosis, prescriptions and referrals.

Social care data from local authorities about unpaid carers and looked after children are in the process of integration into the Discover dataset. If the data are successfully integrated, we will use them to identify unpaid carers, type of care, employment status, the amount of provided care, type of looked after children.

The Discover dataset will be accessed through a trusted research environment provided by NHS North West London Commissioning Support Unit.

## 3. Statistical methods

### 3.1 Study design

A retrospective observational analysis of the Discover dataset which includes case and control matched cohorts.

### 3.2 Study period

Timeline: February 2015 – February 2020 (Pre-COVID) and March 2020 to May 2022 (During COVID-19)

### 3.3 Study population

#### **Cohort identification**

#### **Target cohort 1 – an unpaid carers cohort**

#### **Inclusion Criteria**

We will use two independent approaches to define the unpaid carers population using:

1. Flags of unpaid carer received through data integration from local authorities
2. Any mention in regard to unpaid carer status in SNOMED codes in GP records (see appendix).

The control cohort for cohort 1 will include randomly selected patients who are not in the unpaid carers cohort but have similar characteristics (age, sex, IMD/GP practice). We will do an initial exploratory analysis of the carer population and then identify a comparable control cohort.

The data integration from local authorities is subject to approval from Local Authority partners and North West London ICS's information governance procedures. Therefore, this cohort identification process cannot be guaranteed at this stage.

## Target cohort 2 - a looked after children cohort

### Inclusion Criteria

Given the data from local authorities are integrated into Discover, we will use the flags for looked after children to generate the cohort and include all patients aged under 18.

The control cohort 2 will include randomly selected patients with age under 18 who are not in the looked after children cohort but have similar characteristics (age, sex, IMD/GP practice).

**3.4 Definitions of outcomes and exposures** All the variables will be defined using the Discover dataset.

Outcome/Exposure	Definition
<b>Outcomes</b>	
% of population of the target cohort	% of the people in the target cohort among the general population
A&E/primary care/outpatient visits for the target cohort	A&E/primary care/outpatient visits registered in the Discover dataset, the optimal period of time for obtaining counts will be defined after the data extraction.
PAM score	This 13-item instrument measures confidence in self-management and knowledge of health condition, latest value available in the Discover dataset will be taken
BMI	Body mass index, latest value available in the Discover dataset will be taken
eFI	Electronic Frailty Index, latest value available in the Discover dataset will be taken
Age-sex standardised prevalence rate for a set of long-term conditions	The presence of a condition from a predefined set will be defined based on flags prebuilt WSIC or using SNOMED codes approved by a trained clinician. Then the age-sex standardised prevalence rate will be calculated using latest Clinical commissioning group population estimates
Presence of long-term conditions from a predefined set	The presence of a condition from a predefined set will be defined based on flags prebuilt WSIC or using SNOMED codes approved by a trained clinician



GP visits/referrals/prescriptions (related to long-term conditions from a predefined set)	GP visits/referrals/prescriptions registered in the Discover dataset, the optimal period of time for obtaining counts will be defined after the data extraction
Referral by specialty rates (to long-term conditions from a predefined set)	Referral by specialty rates obtained from Discover dataset, the optimal period of time for obtaining counts will be defined after the data extraction.
<b>Exposure</b>	
Age	Age at date of data extraction
IMD Decile	Index of Multiple Deprivation is a measure of relative deprivation for small areas
Gender	Male, Female or other gender
Ethnicity	White, Asian, Black, Mixed, Other ethnic groups
eFI	The electronic frailty index
Borough	Ealing, Hillingdon, Brent, Harrow, Hounslow, Westminster, Kensington and Chelsea, Hammersmith and Fulham
COVID-19 pandemic period	March 2020 to May 2022
Being in a target cohort	A flag defining if a patient belongs to the target cohort (unpaid carers or looked after children).
Interaction between being a target cohort and COVID-19 pandemic period	See individual definitions above
Type of care	Provided care for children, adults or elderly people, physical illness or mental illness/dementia (exact definitions will be provided after data integration).
Employment status	The status will be obtained from local authorities data (exact definitions will be provided after data integration)
Carers' support	The exact definitions will be provided after data integration
The amount of provided care	The exact definitions will be provided after data integration
Type of looked after children	Fostered children, disabled children, children in residential children's homes

### 3.5 Statistical approaches

The methods described here will be applied to analyse both cohorts.

#### a) Descriptive analysis of the target cohort

*Core analysis:*

General demographic summary of the population for each year of the period defined above (including pre and COVID-19 periods), split by:

- i. Specific age groups based on guidance from PPIE
- ii. Indices of Multiple Deprivation (IMD)
- iii. Sex
- iv. Ethnicity (further split by sex and age, where possible)
- v. Borough

## **b) Analysis of health conditions more prevalent in the target cohort and the effect of COVID-19**

### *Core analysis:*

Calculate age-sex standardised prevalence rates for a predefined set of long-term conditions (6) and compare them between the target and control cohorts (also split by type of unpaid carer (target cohort 1) or type of looked after children (target cohort 2)). We will use the total North West London population, obtained from the ONS CCG population estimates, to standardise our analysis.

Develop mixed-effects multiple logistic regressions (to account for confounders: gender, age, IMD, ethnicity, borough) for predicting each of the long-term conditions from the predefined list (6) and assess the effect and significance of COVID-19 pandemic period as well as of being an unpaid carer (Target cohort 1) or a child in care (Target cohort 2). We will model variation with a random intercept for each borough and random slope (effect of Covid) for each patient. We will also account for the interaction between being a carer and COVID-19 pandemic period. Data about employment status, carers' support and the amount of provided care will be included in the analysis if they are provided by local authorities and sufficiently populated.

We will develop a linear model to predict BMI, PAM score, or eFI from being an unpaid carer (Target cohort 1) or a child in care (Target cohort 2), age, sex, IMD and ethnicity.

## **c) Analysis of healthcare services used by the target cohort and the effect of COVID-19**

### *Core analysis:*

Conduct multiple negative binomial regression models (to account for confounders) to predict GP visits/referrals/prescriptions related to long-term conditions from the predefined list and assess the effect and significance of being an unpaid carer (Target cohort 1) or a child in care (Target cohort 2). Data about employment status, carers' support and the amount of provided care will be included in the analysis if they are provided by local authorities and sufficiently populated.

Assess the effect of COVID-19 on the referral by specialty rates (for long-term conditions found in the previous analysis) in the control and target cohorts using interrupted time-series (including the effect of interaction between Cohort type and COVID-19 era).

Split by:

- i. Specific age groups

- ii. IMD
- iii. Sex
- iv. Ethnicity (further split by sex and age, where possible)
- v. Geography

### 3.6 Methods for addressing missing data

Where needed the mechanism of missingness will be assessed and where the data are missing at random, multiple random imputation method will be applied.

### 3.7 Known limitations

Identifying carers using SNOMED codes – there is a challenge in identifying unpaid carers through SNOMED code usage. These codes are often ambiguous, and there are multiple codes which could be used. Also, there is no guarantee that clinicians use the codes consistently and appropriately. It is highly likely that significant numbers of the true unmet carer population will be excluded.

Accessing MH services over the years – the data lacks information about points of access to health services.

## 4. Governance

### ***Availability of data and materials***

The project has been presented to the data access committee at Information Governance Services Limited and the access to the Discover dataset has been approved until September 2022 at the moment. An additional submission for extension will be made.

### ***Ethics approval and consent to participate***

Given the retrospective nature of the study there will be no ethics approval required beyond the Data Access Committee mentioned above.

## 5. Impact, dissemination and engagement

We plan to disseminate our findings through the Health Foundation, which publishes a report based on the overall findings from each NDL partner. We also plan to communicate our findings to local commissioners through our local system Boards and professional reference group and local authorities. Our lab collaborates with our North West London Integrated Care Partnership, which comprises of NHS providers of healthcare services & clinical commissioning colleagues. Beyond the ICS, we will also share our findings with Imperial College London's research community. We will also publish the findings in public-facing formats (e.g. blogs and social media channels) and through our youth and charity networks to reach our the voluntary sector and the public. Where possible, findings may also be translated into an academic publication(s), to share insights with the broader academic community.

## 6. Appendix

### Local Stakeholders

- North West London ICS
- Data Access Committee for North West London
- Professional Reference Groups

### SNOMED codes used to define unpaid carers (we are still working on this):

224484003	Patient themselves providing care (finding)
513901000000109	Informal carer needing support plan (finding)
224512006	Informal caregiver unsupported (finding)
224493002	Caregiver willing to share care with professionals (finding)

## References

1. Paddison, C. Vaccinating unpaid carers: the need for a clear plan. Nuffield Trust. 2021;
  2. Caring behind closed doors: six months on. Carers UK. 2020;
  3. Huang F, Song J, Davies AR. The creation of an e-cohort to understand long-term health conditions amongst unpaid carers in Wales. :33.
  4. O'Donnell C, Sandford R, Parker A. Physical education, school sport and looked-after-children: health, wellbeing and educational engagement. Sport Educ Soc. 2020 Jul 23;25(6):605–17.
  5. Croft GA. Meeting the physical health needs of our looked after children. Arch Dis Child. 2014 Feb 1;99(2):99–100.
  6. Cassell A, Edwards D, Harshfield A, Rhodes K, Brimicombe J, Payne R, et al. The epidemiology of multimorbidity in primary care: a retrospective cohort study. Br J Gen Pract. 2018;68(669):e245–51.
-