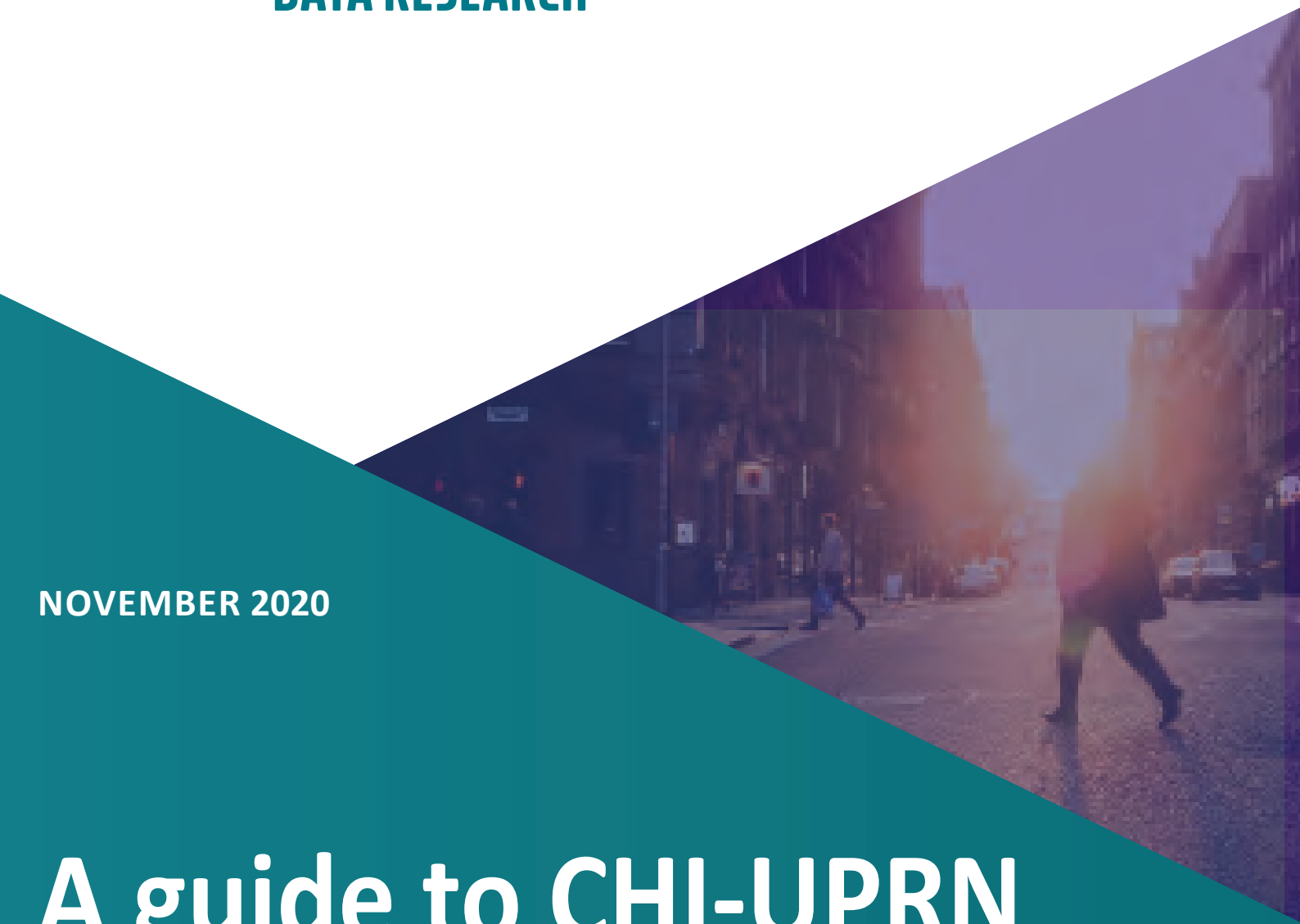
SCOTTISH CENTRE FOR ADMINISTRATIVE DATA RESEARCH

Public Health Scotland



NOVEMBER 2020

A guide to CHI-UPRN Residential Linkage (CURL) File

Authors: David Clark, Public Health Scotland
Professor Chris Dibben, University of Edinburgh

Acknowledgements

The CHI-UPRN Residential Linkage (CURL) has been developed by David Clark, eDRIS, and Professor Chris Dibben, Director of Scottish Centre for Administrative Data Research (SCADR), in collaboration with the Data Hub at the Improvement Service. With thanks to Dr Jenni Burton, University of Glasgow, who co-developed the care home flag.

What is it?

Background

The CHI-UPRN Residential Linkage (CURL) - a Reference or Lookup File - has been derived in Public Health Scotland (PHS) from the Community Health Index (CHI) database of all General Practice registrations in Scotland. PHS have been using CHI for record linkage purposes for many years, but thus far most linkage developments have focussed on the ability to link at the person level with the CHI number as the unique patient identifier used in NHS Scotland.

Although the CHI database also holds patient addresses, to date there has not been a consistent method for recording these, and the same address can appear multiple times but in a variety of formats, making intra-property linkage problematic. With the help of the Data Hub service provided by the Improvement Service (custodians of the One Scotland Gazetteer), Public Health Scotland have been seeding the addresses recorded in CHI with the Unique Property Reference Number (UPRN) from the Ordnance Survey Address Base. The UPRN can then be used in CURL to bring CHI numbers for different people living in the same property together.

Contents

The initial version of CURL is a lookup between the CHI Unique Patient Identifier and the UPRN for current GP registered patients in Scotland but also includes any person who has died since January 2020. The CURL file held in PHS contains a number of processing fields including the full address, but the fields that can be made available to researchers, subject to requirement and approval, are as follows:

Field Name	Description
PersonIndexNumber	A study-specific index number that will replace the CHI number in linked datasets for a research project hosted in the National Safe Haven.
CurrentStatus	An indicator with the status of the CHI address record. In the initial version this will just be an indicator stating the person is currently alive and registered with a GP in Scotland or the person is dead.
DateOfDeath	The date of death as recorded on CHI.
DateAddressChanged	The date recorded on CHI for a change in the current address fields.
GPPracticeNumber	A unique number identifying the GP Practice.
PropertyIndexNumber	A study-specific index number that will replace the UPRN in linked datasets for a research project hosted in the National Safe Haven.

UPRN_LinkFlag	A flag variable that will indicate how the CHI record address string was matched to the PropertyIndexNumber (e.g. quality indicator of automated DataHub matches, non-UPRN internal links, manual match etc.)
In_CH	A flag variable that will indicate if the PropertyIndexNumber relates to a CareHome UPRN. UPRNs associated with these flagged records can potentially be linked to a separate table of information about the care home facility from Care Inspectorate. The Care Inspectorate address records have been seeded with UPRN by Ordnance Survey.

Methods

Address fields (3 address lines plus postcode) were extracted from the PHS CHI monthly download (dated 03-Aug-2020) from current records of persons still alive in Scotland or who died since 1st January 2020. This yielded 3,208,951 unique address strings relating to 5,828,951 people.

Some minimal formatting and pre-processing was carried out in order to extract a Town entity from the address lines, trim leading spaces, format postcodes, and apply a series of rules to expand abbreviations at end of address lines (e.g. "RD" to "ROAD", "DR" to "DRIVE" etc.)

Files of unique address strings were processed at DataHub - https://datahub.scot/home/ - using the Non-residential UPRN seeding templates and programs. These returned 5 categories of output file with best matching UPRN attached to the input strings where applicable. The categories of output were EXCELLENT/GOOD/FAIR/MULTIPLE/NO_MATCH.

Seeding Results:	Address Rows		People	
Excellent	2,378,655	74.1%	4595099	78.8%
Good	294,294	9.2%	482532	8.3%
Fair	89,096	2.8%	129758	2.2%
Multiple	5,406	0.2%	9129	0.2%
No Match	441,500	13.8%	612433	10.5%
Total	3,208,951	100.0%	5828951	100.0%
Total Seeded: Fair-Excellent	2,762,045	86.1%	5,207,389	89.3%
Not seeded: No match-Multiple	446,906	13.9%	621562	10.7%

Although, overall >89% of CHI records were matched automatically to a UPRN, there are geographical variations in match rate. Also, when looking only at the CHI records with an InstitutionFlag denoting a resident in a Care Home, only around 60% of these records were returned with a UPRN.

Care Home Flag

As only around 60% of the current CHI records that were already identified as people living in a care home (by the CHI Institution Code) could be automatically matched by DataHub, a manual exercise was carried out in order to assign a UPRN to unmatched records. Instances in CHI of care homes registered by Care Inspectorate were combed, and the Ordnance Survey assigned UPRN in the Care Inspectorate list of addresses was manually entered against the CHI record where expert knowledge determined it was a care home address. This was very challenging due to a number of reasons. A small example of these issues include:

- Care Home names change over time, but the CHI address may not change when the care home is renamed.
- Differences in the way the same care home name is referenced in the CHI address.
- Reference to sub-unit or suite names in the CHI address file.
- Multiple UPRNs assigned to the same care home.

The outcome of this exercise was that 42,334 people in adult care homes could be flagged, compared with 40,196 people in any care home, using the traditional CHI Institution variable. Not only did this allow identification of more residents, it also enabled allocation to specific care home locations which can be the basis of more meaningful research.

Linkage Quality and Biases

Researchers who wish to use the initial version of the CURL file should be aware that, as yet, a thorough clerical review of the automatic matches returned by DataHub has not been undertaken. Preliminary small random samples of the non-exact address matches categorised by DataHub as Excellent/Good/Fair showed very few potential false positives, indicating generally very high precision, but as yet unquantified. eDRIS intend to carry out more extensive sample-based clerical review that will target some of the areas or types of property that are most problematic for matching. This will give a better indication of the precision of the UPRN matches overall and in different areas.

In terms of sensitivity of matches, the overall automatch rate of just under 90% masks some differences by area and property type. For example, only 81% of properties in very remote rural areas (category 8) as defined by the Scottish Government 8-fold Urban Rural Classification 2016 were matched this way, whereas 94% of properties away from large urban areas but in accessible small towns or other urban areas (categories 2&3) automatically linked to a UPRN. As a further example of these differences, 96% of people living in North Ayrshire have an address that was automatched, whilst in remote Argyll & Bute the match rate is only 77% and in built-up City of Glasgow, only 81%.

There are also some UPRNs with a very large number of people apparently living in the same property. Some of these occurrences are valid, such as children at boarding school, but in many of these instances it appears to be the address of the GP Practice that is recorded in CHI, presumably in registrations where the patient has been unable to provide a fixed permanent address.

The initial lookup file has been made available on the understanding that it is experimental and that there may be other issues that have still to be discovered. eDRIS would welcome any feedback from

researchers, if they identify any further quality issues when using the file. eDRIS would like to develop the linkage further and this may involve:

- Additional bespoke matching of the residual addresses unmatched by the DataHub, developed in-house and/or in collaboration with University-based data scientists.
- Incorporate flags for other types of institution in addition to adult care homes.
- UPRN-seed historic CHI addresses.
- Productionise a process for keeping the UPRN matches up to date.

Why is this tool helpful?

The CURL File has been created in response to the COVID-19 pandemic in order to facilitate understanding of the impact on people living in the same property as another person infected with the disease. The potential utility of this resource, however, goes much further than just COVID-19 related research, and is seen as a major development to the Community Health Index (CHI) as a data linkage tool in Scotland. CHI has long been used as a population spine for patient-level data linkage in health based research projects, but the addition of a property level dimension through CURL opens up brand new avenues for researchers.

Understanding the household

Unlike traditional surveys, which typically ask about the household of residence, administrative data derived research datasets will typically offer little or no information on the household within which a person lives. This means an important level of understanding will be missing from any analysis. The household is arguably the most important sphere of social, psychological and economic life for an individual. With the UPRN linked to the CHI, it is now possible to explore the many issues where the household is important, as individual records can be grouped into units with the same UPRN. This, for example, will allow an understanding of informal care and its wider impact: who might be available to give that care and what is the economic and psychological impact of the care giving. It can also help with the understanding of educational outcomes of children by allowing an examination of who else is in the household who might provide 'human capital' or indeed other support that ultimately leads to social mobility or the transfer of social advantage.

Examples of use

Understanding Hospital Discharges to Care Homes

Public Health Scotland and academics from the Universities of Glasgow and Edinburgh were commissioned by the Cabinet Secretary for Health and Sport to identify discharges from NHS Scotland Hospitals to Care Homes during March to May 2020. This work required identifying who was discharged to a care home, by linking multiple national datasets and then identifying care home outbreaks of COVID-19 and analysing factors related to those outbreaks. The CURL file was used to help in identification of tests performed on care home residents and in allocating tests to specific care home locations.

The full report and methodology documents are available from: https://beta.isdscotland.org/find-publications-and-data/population-health/covid-19/discharges-from-nhsscotland-hospitals-to-care-homes/.

Case-Control Study

PHS maintains a case-control study, regularly updated, which has been used to inform policy on the control of the COVID-19 pandemic. Care home residence is a very important predictor of the risk of severe COVID-19, but it is based on the CHI-institution variable which is known to have limited sensitivity. The UPRN project, by systematically identifying all care home addresses, will address this deficiency, improving the accuracy of results using the case control study.

Care Homes Study

PHS has longitudinal data on care home residence status going back to 2007. We have used this data to estimate the effect of COVID-19 on mortality among care home residents, both overall and compared to seasonal respiratory infections. We have also used these data to compare mortality in care home residents versus older people who are not resident in care homes. However, this work is based on the CHI-institution variable which is known to have limited sensitivity. Nor, for large parts of Scotland, is the specificity known. The UPRN project will allow us to compare differences in mortality rates based on the original CHI-derived care home variable to that based on the UPRN-derived care home variable. If the differences are similar, this will provide some reassurance that the historical comparisons are valid.

How do I access it?

Access to CURL, for inclusion of a residential linkage field in research studies, will follow the standard approvals process required for any study requiring national health datasets or linkage involving CHI numbers.

The full product isn't expected to be available to researchers. It is a mechanism to index households/properties where people in research studies are residing. This hasn't previously been possible, due to the multitude of ways addresses could be recorded on CHI.

Access to data analysts within PHS and eDRIS will be via the standard Access to Data process rules. Researchers who wish to add a residential marker or the new care home marker to their research datasets assembled from health and/or other administrative data sources will need to apply to the Public Benefit and Privacy Panel (PBPP) and any other data controller approval processes. Researchers would need to request inclusion of the CHI-UPRN Residential Linkage File as a data source in an application to PBPP (https://www.informationgovernance.scot.nhs.uk/pbpphsc/) and access will be via the National Safe Haven.

In the first instance, please download and complete an enquiry form from the eDRIS website (https://www.isdscotland.org/Products-and-Services/eDRIS/) and send it to the eDRIS team at phs.edris@phs.scot who can provide further advice.

eDRIS
www.isdscotland.org/edris

improvement service

Data Hub

ADR Scotland Publication:

A Guide to CHI-UPRN Residential Linkage (CURL) File

ADR SCOTLAND

UKRI Economic and Social Research Council

scadr@ed.ac.uk | www.scadr.ac.uk | @scadr_data