

# Linking consumer datasets to chart residential moves in private rental housing in England and Wales

Joanna Kuleszo<sup>\*1</sup>, Rory Coulter<sup>†1</sup>, Justin van Dijk<sup>‡1,2</sup> and Paul Longley<sup>§1,2</sup>

<sup>1</sup> Department of Geography, University College London

<sup>2</sup> ESRC Consumer Data Research Centre

February 15, 2021

## Summary

With sufficient preparation for purpose, consumer datasets may be used alongside conventional statistical sources to provide more granular and frequently updated estimates of housing market conditions than are otherwise available. Here we describe linkage of Zoopla rental market listings and Linked Consumer Register data to establish the pattern of residential moves in England and Wales between 2014-2019.

**KEYWORDS:** Consumer data; data linkage; housing market; private renting; residential mobility

## 1. Introduction

Residential mobility is a key outcome of local housing market activity and population dynamics. In Britain, a disproportionate share of this mobility takes place in the Private Rented Sector (PRS). In 2019-20, 53% of household moves in England took place in the PRS, despite the tenure housing only 19% of households (MHCLG, 2020).

Yet although the PRS has grown in recent decades, relatively little is known about the sector's composition and dynamics. Much of our knowledge derives from two core data resources that are purposely designed for statistically robust population analysis: (1) regular representative social surveys and (2) the decennial census. However, neither of these resources collects data both frequently and at the granular geographical scales required for local policy analysis.

In contrast, consumer data are routinely collected by customer-facing organisations active in the housing market. Such data may be highly granular, frequently updated and made available in something approaching real time (Livingston et al., 2021). However, such sources are often collected as a by-product of transactions and so are neither carefully curated nor collected to any scientific design. Furthermore, few consumer organisations have any monopoly of supply. As such, the provenance of raw consumer data sources and their suitability for research is largely unknown.

In this context, this paper contributes to the development of a consumer data infrastructure for population and housing market analysis by documenting the compilation and initial validation of a new data resource created by linking 2014-19 Zoopla/WhenFresh rental listings with the Linked Consumer Registers (LCRs) created at the Consumer Data Research Centre (CDRC; Lansley et al., 2019). Establishing the veracity of this new resource is crucial given uncertainties in the provenance, coverage

---

\* j.kuleszo@ucl.ac.uk

† r.coulter@ucl.ac.uk

‡ j.t.vandijk@ucl.ac.uk

§ p.longley@ucl.ac.uk

and curation of the commercial source datasets. In what follows we (1) undertake external validation of the Zoopla/WhenFresh listings data against 2011 census benchmarks before (2) reporting on the process of linking rental listings to the LCR.

## 2. Data sources

Zoopla/WhenFresh rental listings data are held in the Secure CDRC collections. The resource consists of a database of private rental listings in England and Wales advertised on Zoopla between 2014-2019. Zoopla is a market-leading platform for advertising properties that is used by a large, but undefined, share of UK private estate agents and landlords (Livingston et al., 2021). The creation of the LCRs database using computationally intensive methods is described in detail by Lansley and colleagues (2019), but in brief they aspire to list the names and geo-referenced addresses of a large proportion of UK adults from 1997 to 2019 (an update to 2020 is ongoing) and to link these records over time. Each record contains information about the year that each individual was first and last seen at a given address, providing an indicator of probable residential mobility history. **Table 1** provides additional information about the LCRs and the Zoopla data resources acquired for linkage.

**Table 1** Summary of datasets

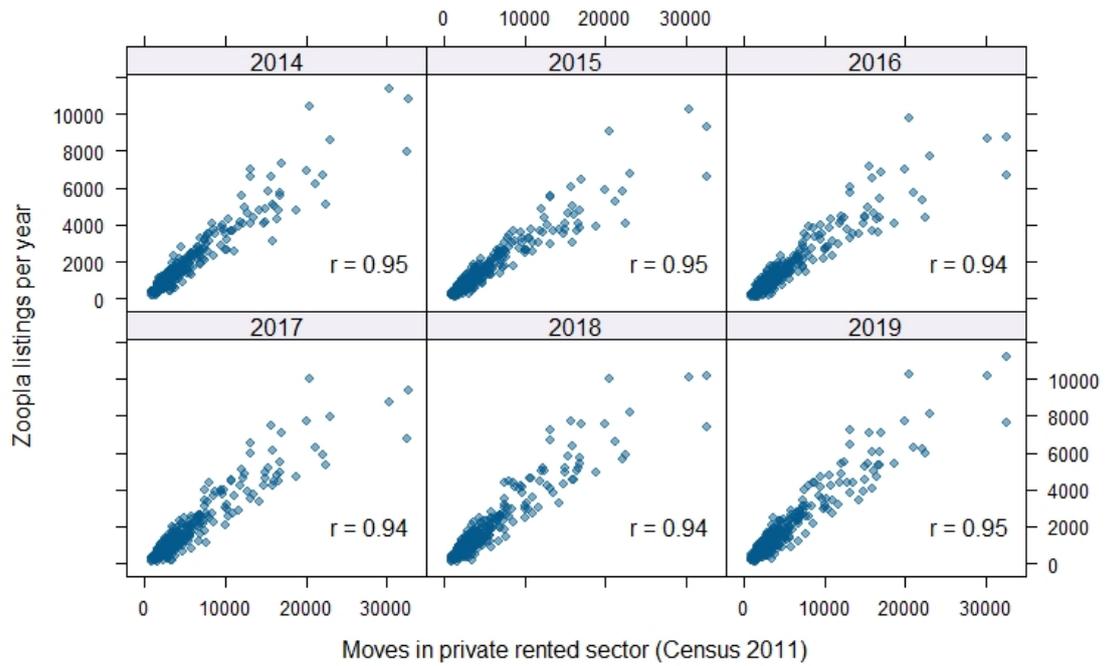
Dataset	Years	Coverage	Variables used	Records
LCR	1997-2019	UK	Geo-referenced address, forename, surname, first seen date, last seen date.	144m
Zoopla	2014-2019	England and Wales	Geo-referenced address, date property first listed on market, date (re)listing created, date last seen as listed for rent, date rented, additional property details (size, listing price etc).	3.5m ( $\approx 570,000/\text{year}$ )

Zoopla data were cleaned by discarding duplicate listings (15,460 records), calculating vacancy opening and closure dates and removing records with illogical date sequences (136,827 records). This left 3,423,091 records henceforth referred to as ‘listing events’. The occurrence of a listing event indicates a change of tenants and thus household mobility.

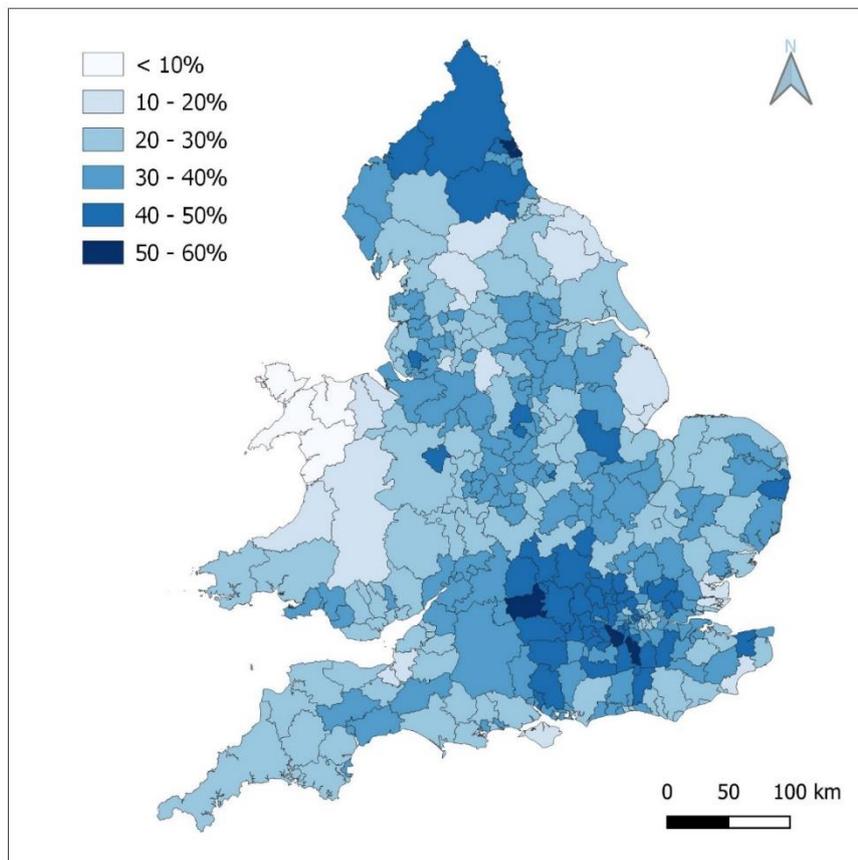
## 3. External validation

Our first objective was to assess whether Zoopla listing events provide a proxy for residential moves in the PRS. To assess this, we correlated the annual number of Zoopla listings in each local authority against the number of PRS households moving in the year preceding the 2011 census. **Figure 1** shows strong positive correlations with  $r=95$  for Zoopla listings from 2014, 2015 and 2019, and  $r=94$  for listings from 2016, 2017 and 2018.

**Figure 2** maps the ratio of Zoopla listings to 2011 census PRS mobility rates. This ratio varies regionally and is generally higher (0.4-0.6) for authorities in the South, particularly those located in the triangle bounding London, Reading and Oxford. There is also good coverage in other urban centres, and in Northumberland and Durham area. Parts of the rural North and Wales on the other hand have a lower ratio, indicating poorer Zoopla coverage of PRS moves.



**Figure 1** Comparison of Zoopla listings per local authority with 2011 census PRS mobility rates.



**Figure 2** Zoopla listings as percentage of census moves across England and Wales.

#### 4. Linking Zoopla listings to LCR records

We first excluded LCR records outside the geographical and temporal scope of the Zoopla dataset (**Table 1**) to leave 26 million records in England and Wales for which either the first or last seen year fell between 2014-2019. Zoopla listings longer than a year were then excluded as agents do not always de-list let properties (50,824). We also discarded records from addresses re-listed >4 times a year on average (13,024). This left 3,359,244 records for potential linking.

Zoopla listing events were then linked to LCR records by simultaneously:

- 1) matching Zoopla listing addresses to the LCRs using a common reference file compiled from AddressBase Premium 2018 and the Postcode Address File 2018;
- 2) matching Zoopla listing dates to both first seen and last seen dates in the LCRs.

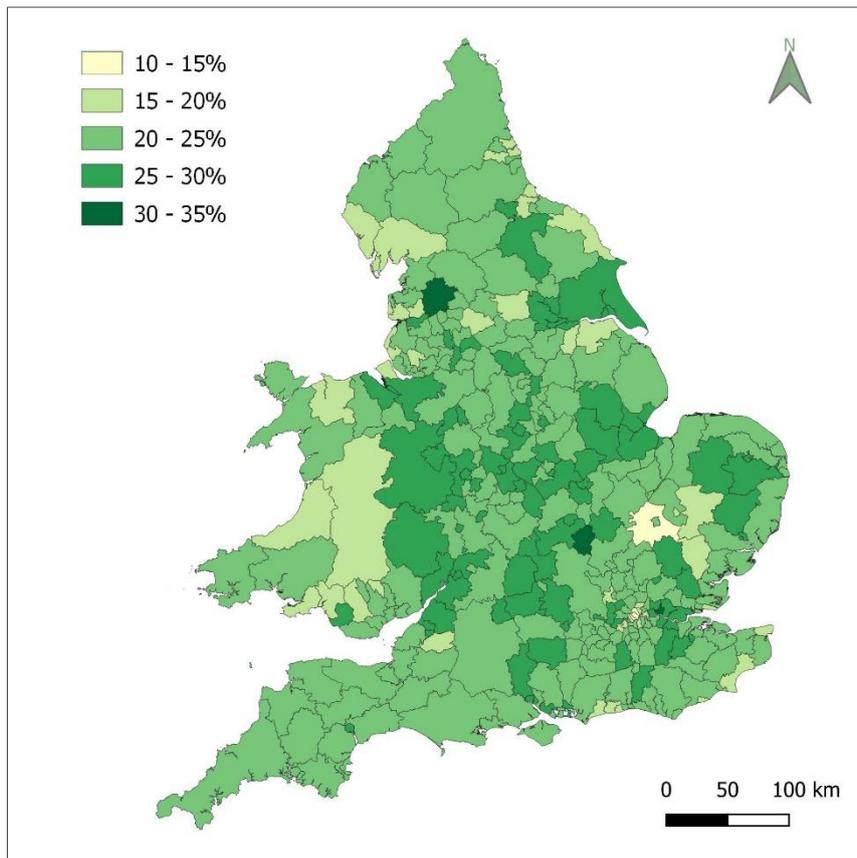
23% of Zoopla listings were attached to 1+ address records in the LCRs. Further analysis showed that this proportion is uncorrelated with the number of Zoopla listings per authority. **Table 2** shows that there is a temporal trend of diminishing proportions of Zoopla listings attached to LCR records. This is most likely related to LCR coverage diminishing with time because of increased opt-out from the public electoral roll components of the Registers (Lansley et al., 2019).

**Table 2** Summary of matching efficacy

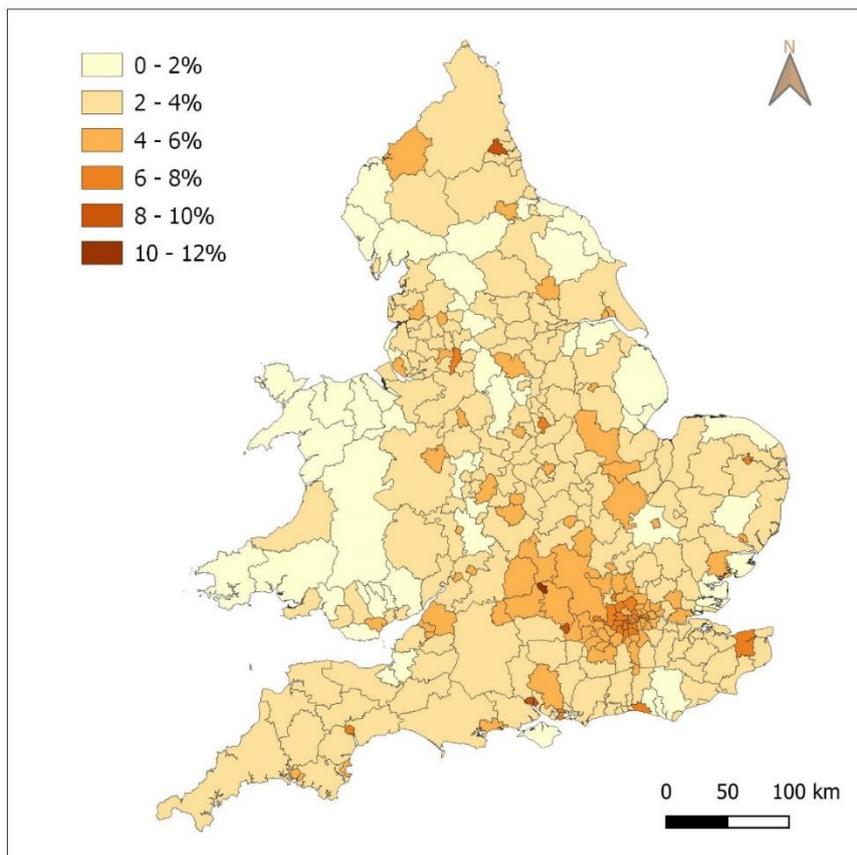
Year	N Zoopla listings	N listings matched to LCR	% listings matched to LCRs
2014	597,602	201,507	33.72
2015	513,862	118,252	23.01
2016	522,325	185,259	35.47
2017	548,697	76,878	14.01
2018	581,894	109,081	18.75
2019	594,864	73,556	12.37
<b>Total</b>	<b>3,359,244</b>	<b>764,533</b>	<b>22.76</b>

**Figure 3** shows how matching efficacy (defined as percentage of listings matched to the LCRs) varies between local authorities. No clear rural-urban patterns are evident and higher match rates are achieved in a zone around London, parts of East Anglia, Humberside and the East/West Midlands. The lowest rates are achieved in Inner London.

Overall just over 1m LCR records (4%) were assigned 1+ Zoopla listing. **Figure 4** shows the geographical distribution of these matches. In general, urban centres have higher match levels, for example 12% in Newcastle and 10% in Oxford. The lowest levels are found in Wales with Anglesey, Conwy and Gwynedd the only three authorities where fewer than 1% of LCR records matched to Zoopla. The overall distribution of matches resembles the size of the PRS recorded in the 2011 census and the distribution of PRS moves shown in **Figure 2**.



**Figure 3** Percentage of Zoopla records matched to LCR



**Figure 4** Percentage of LCR records matched to Zoopla record(s)

## 5. Conclusions

The procedures of external validation suggest sufficient correspondence with 2011 census estimates to justify using Zoopla and LCRs to produce estimates of PRS activity in more recent non-census years. However, these sources do not, as yet, amount to a substitute for census estimates. Linked data are in other respects more than a substitute, in that precise georeferencing makes it possible to ascertain the exact distances and also social gradients over which household moves take place.

We also anticipate the development of internal validation procedures of LCR and Zoopla data. In methodological terms, the advent of consumer data resources entails a shift from a conception of data sources as being correct to prespecified levels of precision with known margins of error, to an environment in which unknown errors in data collection are managed by triangulation of different sources as best endeavours to produce reliable estimates. In a broader sense, this marks a shift from reliance upon scientific inference to procedures for merging diverse data sources with different structures. We propose to advance this research agenda to develop timely, granular and data-rich representations of residential mobility and its social outcomes.

## 6. Acknowledgements

This work is funded by the UK ESRC Consumer Data Research Centre (CDRC) grant reference ES/L011840/1, ESRC grant ES/S016422/1 ('Modelling Housing Career Trajectories in Great Britain') and EPSRC grant EP/M023583/1 ('UK Regions Digital Research Facility').

The data for this research have been provided by the Consumer Data Research Centre, an ESRC Data Investment, under project ID CDRC229, ES/L011840/1; ES/L011891/1.

## References

- Lansley G, Li W, Longley P A (2019). Creating a linked consumer register for granular demographic analysis. *Journal of the Royal Statistical Society: Series A*, 182(4), 1587-1605.
- Livingston M, Pannullo F, Bowman A, Scott E, Bailey N (2021). Exploiting new forms of data to study the private rented sector: Strengths and limitations of a database of rental listings. *Journal of the Royal Statistical Society: Series A*. DOI:10.1111/rssa.12643.
- Ministry of Housing, Communities and Local Government (MHCLG) (2020). *English Housing Survey 2019-2020: Headline Report*. MHCLG: London.

## Biographies

Rory Coulter is Lecturer in Quantitative Human Geography at UCL. His work focuses on housing and life course dynamics, neighbourhoods and population mobility processes.

Joanna Kuleszo is a Research Associate in the Department of Geography at University College London. Her research focuses on spatial analysis of built and natural environment systems.

Justin van Dijk is a Research Associate and Teaching Fellow in the Department of Geography at University College London. His primary research interests are grouped around the analysis and visualisation of large-scale spatial data, urban mobility, socio-spatial inequalities, and geospatial data in general.

Paul A. Longley is Professor of Geographic Information Science at UCL and Director of the UK Consumer Data Research Centre at UCL. His research focuses on the application of geographic information science, with a strong emphasis on the development and deployment of geo-temporal data infrastructures developed from Big or Open Data.