International Journal of Population Data Science (2019) 4:3:119

International Journal of Population Data Science

Journal Website: www.ijpds.org

Allocating Unique Property Reference Numbers (UPRNs) to general practitioner-recorded patient addresses using a deterministic address-matching algorithm: evaluation of representativeness and bias in an ethnically-diverse inner city population

Harper, G[1*], Stables, D[2], Boomla, K[1], Robson, J[1], Fry, R[3], Ahmed, Z[1], and Dezateux, C[1]

[1]Queen Mary University of London

[2]Endeavour Health

[3]Swansea University

Background

Pseudonymised UPRNs based on patient addresses can be used to link environmental information to electronic health records (EHRs), however the representativeness and potential demographic or health-related biases in linkage using existing address-matching algorithms have not been evaluated using patient addresses.

Main aim

To evaluate representativeness and bias in assigning UPRNs using an address-matching algorithm based on general practitioner (GP)-recorded patient addresses for a geographically-defined multi-ethnic inner city population.

Methods

We evaluated the Discovery Programme deterministic address-matching algorithm, comprising 213 rules applied, in rank order of minimising false positives, to the GP-recorded address of 879,286 (48% female) patients currently registered with all GP practices in four boroughs in inner east London.

We used logistic regression to estimate the adjusted odds (aOR) of an address not being linked to a UPRN by: age band (reference group: <1 year), sex (female), ethnic group (White British), Index of Multiple Deprivation (IMD) quintile (most deprived), number of long-term conditions (none); and timing of GP registration (most recent quartile). We evaluated the linkage and algorithm error rates in an independent validated NHS address dataset using best practice linkage reporting standards.

Results

99% of patients had a UPRN assigned. Men (aOR;95%CI:0.87;0.8,0.91), and patients aged 15-19 (0.51;0.39,0.68), 20-24 (0.67;0.51,0.89), or ≥90 years (0.35;0.83,0.91), of Chinese ethnic background (95% CI; 0.50; 0.45,0.56), or living in the least deprived IMD quintile (0.24; 0.20,0.30) were less likely, and those with a GP-registration preceding mid-2016 (p-value 0.00) more likely, to have a UPRN assigned. The sensitivity, specificity, positive and negative predictive-values and F-measure of the algorithm were, respectively: 0.993, 0.019, 0.914, 0.204, and 0.9516.

Conclusion

We have demonstrated, for the first time, a high GP-address UPRN match rate and quantified error rates and biases for users. Further work is needed to investigate addresses in patients with more complex address histories.

*Corresponding Author:
Email Address: g.harper@qmul.ac.uk (G Harper)

https://doi.org/10.23889/ijpds.v4i3.1282