

Home (<https://ijpds.org/index>) / Archives (<https://ijpds.org/issue/archive>)

/ Vol. 5 No. 5 (2020): Conference Proceedings for International Population Data Linkage Conference 2020 (<https://ijpds.org/issue/view/19>)

/ Conference Proceedings

# Allocating Unique Property Reference Numbers to Patient Addresses Using A Deterministic Address-Matching Algorithm: Evaluation of Accuracy, Match Rate and Bias

**Gill Harper**

Clinical Effectiveness Group, Institute of Population Health Sciences, Queen Mary University of London

**Kambiz Boomla**

Clinical Effectiveness Group, Institute of Population Health Sciences, Queen Mary University of London

**John Robson**

Clinical Effectiveness Group, Institute of Population Health Sciences, Queen Mary University of London

**David Stables**

Endeavour Health Charitable Trust

**Zaheer Ahmed**

Clinical Effectiveness Group, Institute of Population Health Sciences, Queen Mary University of London

**Richard Fry**

Biomedical Sciences, Swansea University

**Carol Dezateux**

Clinical Effectiveness Group, Institute of Population Health Sciences, Queen Mary University of London

## Abstract

## Introduction

Representing patient-registered addresses as pseudonymised Unique Property Reference Numbers (UPRNs) enables linkage of environmental and household information to electronic health records (EHRs). However, the accuracy and potential biases in address-matching algorithm results applied to patient addresses is unknown.

## Objectives and Approach

To investigate accuracy, match rate, and biases in assigning UPRNs to general practitioner (GP)-registered patient addresses for a geographically-defined UK population, using a bespoke deterministic address-matching algorithm comprising 213 rules applied in rank order of minimising false-positives, developed for the Discovery Data Service.

We ran this algorithm to match 906,220 adult patient GP-registered addresses (48% female, 47% non-White, 89% 20-64) sampled in mid-2018 from 159 GP practices in four London boroughs to Ordnance Survey's AddressBase Premium database.

We evaluated the error rates using a gold-standard dataset. We used binary logistic regression to estimate the likelihood (Odds Ratio [OR]; 95% Confidence Intervals [CI]) of no UPRN match according to and adjusting for patient age, sex, ethnic background, deprivation, residential mobility and multiple GP registrations.

## Results

96% of patient addresses were successfully assigned a UPRN. Algorithm sensitivity, specificity, positive and negative predictive-values and F-measure were, respectively: 0.993, 0.019, 0.914, 0.204, and 0.9516.

After mutual adjustment, UPRN assignment was less likely for: men (OR: 0.87; 95%CI: 0.83,0.91); adolescents and the elderly (15-19 years: 0.57;0.43,0.77;  $\geq 90$  years: 0.39;0.18,0.84); those from Chinese ethnic backgrounds (0.87;0.8,0.91), living in the least deprived areas (0.25;0.21,0.31), or with two or more distinct UPRNs across multiple registrations (0.37;0.28,0.49); and more likely for: those from Bangladeshi ethnic backgrounds (1.79;1.61,2.00), registered before 2018 (5.10;4.42,5.87), or with multiple GP registrations (2.36;1.82,3.05).

## Conclusion / Implications

The Discovery open-source algorithm achieves a high accurate match rate and quantifies the demographic groups that may be under-represented among those successfully matched. This is the first time that bias in matching rates for an address-matching algorithm has been evaluated using patient-registered addresses.