

S3. GEOCODING PROCESS, ACCURACY AND PRECISION

For all data except the education records, geocoding of addresses takes place within the NIA database which involves addresses being matched against a regularly updated address database and allocated Easting and Northing coordinates which were extracted as part of the datasets. Of a random sample of 400 records (100 from each dataset) with coordinates checked for geocoding accuracy, 98% were geocoded to the correct address (95% CI: 96%-99%), 99% to the correct SA1 (95% CI: 98%-99.8%). 'Street records' based on street name rather than specific addresses (constituting 24%/10%/7%/2% of the address/offense/incident/family address samples respectively, 89% overall) were deemed accurate when the coordinates fell on an address on the correct street and at least part of the street fell within the same SA1/SA2 as the coordinates.

The potential imprecision of street records was measured as the maximum distance between the record's map point and any point along the street (or street segment within the postcode or suburb specified in the address string). The median potential imprecision was 805.5 meters (95% CI 309 – 1083, n=31), thus even if street record points were maximally imprecise, the actual location of the offense/incident/address would typically fall within 1km of the SA1 into which the record fell.

The school/education data was geocoded in R using a custom function that queried the name of the school/institution in Google Maps using an API and returned the top matching address and its latitude and longitude. The coordinates were then transformed into Easting and Northing using ArcGIS. Checks of a random sample of school/education records with coordinates (n=372) showed 95% were geocoded to the correct address and SA1 (95% CI: 93%-97%).