

S2. DATA FILTERS AND SAMPLE ATTRITION

Data filtering steps

The following steps were taken to filter out inaccurate or imprecise data. First, offenses/incident codes which do not reliably establish the offender's presence at the offense/incident address were excluded, because they could not be assumed to represent the offender's activity space. This step is described in greater detail in the following section. Second, offenses/incidents recorded with start and end timestamps (within which the exact time of the offense/incident could not be established) covering more than 30 days were excluded. This threshold minimized data loss (under 2% of records). Third, offenses/incidents with no coordinates, or whose coordinates were not located in a census unit ('SA1') due to data entry errors were excluded. Addresses recorded against generic locations such as suburbs or cities as indicated by keywords such as 'NFA'/'no fixed abode' and 'unknown' in the address string were also excluded, as their coordinates were insufficiently precise to link the offender to a specific location. Finally, offenses/incidents relating to offenders who no longer appeared in the reference offense dataset (having been excluded in any of the above steps) were removed from the activity node datasets. Table S2.1 shows the number and proportion of records excluded by each filter. Sample sizes are expressed as the number of unique person-locations, where a location is either a discrete event in space and time (offenses/incidents) or an address without a specific related event record (offender, family and education addresses).

Table S2.1. Sample size per dataset following exclusions.

Exclusions	Reference offenses ^a	Offenses ^a	Incidents	Addresses	Family addresses	Education
Initial N	144,839	842,582*	647,548	1,501,873	2,462,611	35,470
			New N (-%)			
Offender not present	-	802,614 (-4.7%)	499,449 (-22.9%)	-	-	-
Long timespan	142,267 (-1.8%)	793,224 (-1.2%)	493,282 (-1.2%)	-	-	-
No coordinates	140,765 (-1.1%)	773,120 (-2.5%)	484,957 (-1.7%)	1,375,571 (-8.4%)	2,162,276 (-12.5%)	33,931 (-4.3%)
No SA1	140,707 (-0.04%)	772,629 (-0.06%)	484,662 (-0.06%)	1,372,004 (-0.3%)	2,156,828 (-0.3%)	33,931 (-0%)
Imprecise address	140,670 (-0.02%)	772,474 (-0.02%)	484,516 (-0.03%)	1,352,327 (-1.4%)	2,135,972 (-1.0%)	-
Offender not in ref. offense data	-	749,428 (-3.1%)	454,462 (-6.6%)	1,282,080 (-5.2%)	2,014,328 (-5.7%)	32,831 (-3.2%)

^a The 'reference offense' dataset included all *possible* reference offenses i.e., all burglaries, robberies and extra-familial sex offenses committed between 2009 and 2018, from which the most recent per offender and offense type were identified. The 'offenses' dataset included the 'reference offense' dataset plus all other offenses committed between 2004 and 2018. For each individual offender, there is no more than one reference offense per offense type (burglary, robbery, or extra-familial sexual offense). All prior offenses were committed before the reference offense. For offenders who committed multiple types of offenses (burglary and/or robbery and/or extra-familial sexual offense) a particular offense (say a burglary on January 16, 2016) can be included as a burglary reference offense and also as a prior offense of a reference robbery or a reference sexual crime committed subsequently (e.g., on March 2, 2017).

Exclusion of codes not reliably indicating the offender's presence

Small samples of all codes in each of the categories shown in Table S2.2 were checked against NIA records to determine whether the offender was present at the offense/incident address. These categories were then excluded because:

1. For offenses (committed by the offender), less than 75% involved the offender being present at the recorded address.
2. For incidents (offenses involving the offender in a non-offending role or non-crime incidents), less than 75% involved the offender being present at the recorded address.

Including these offenses would have meant overestimating activity space. But excluding them does not result in a large underestimation of activity space: of the offenses in these categories which do involve the person's presence, many occur at their home address which is already captured in the address dataset. Further, over 80% of the records excluded from the prior incidents dataset for this reason were breaches of bail, which are predominantly recorded against the person's home address or against the location of the offense for which they are on bail. Both locations are already in the address and prior offense datasets respectively. Table S2.2 lists the codes excluded with this filter.

Table S2.2. Codes excluded from offenses (O) or incidents (I) offender not present.

Category	Code excluded	From
Make/publish/supply objectionable material	2792, 2793, 2794, 2795, 2946, 2961, 2965, 2966, 2967, 6563	O, I
Possess/make/publish intimate visual recording	2991, 2992, 2993	O, I
Other publication offense (e.g., breach name suppression)	3811, 3816, 7444	I
Harmful digital communication	1765	O
Telephone based offense	6551, 6552, 6554, 6555, 6559	O
Post based offense	6569	O
Unauthorized computer access	4631, 4632, 4634, 4636	O, I
Harassment	1841, 1849, 6553	O
Receiving stolen goods	4413, 4417, 4418, 4419, 4422, 4423, 4429	O, I
Fraud involving loss by deception	4553, 4554, 4555, 4556, 4557, 4558, 4567	O
Forgery/altering documents	4511, 4512, 4531, 4532, 4538, 4541, 4542, 4575, 4576, 4581	I
Unspecified "other" incident	1Z	I
Complaint	6A	I
Breach of police/court bail	7193, 7194, 6B, 6D, 6E	O, I
Fail to attend court or program	3746, 3832, 3835, 3855, 3855, 7191, 7192, 7955, 7956	O, I