

Supplementary Appendices

Supplementary Appendix 1. The allocation of an Anonymised Linkage Field (ALF)

Demographic data comprising first name, surname, gender, date of birth, and postcode, are separated from content data such as clinical diagnostic testing and results data, or placement details of children looked after. Data is pseudonymised and allocated an ALF in the place of demographic data (by a trusted third party) where a successful match is made. The ALF is then transferred and joined to the personal data (i.e. placement details). Datasets provided to

the SAIL Databank are already split at the source organisation into demographic and content data. An individual is allocated an ALF either using their unique 10-digit NHS number (deterministic record linkage, DRL), or a combination of other quasi-identifiers (probabilistic record linkage, PRL) held and maintained by the NHS Administrative Register (name, address (including historical), postcode, gender, date of birth, general practice of registration and NHS number). Within the SAIL Databank, records of individuals with an ALF can be linked to other data sources such as demographic, health, social and education datasets.



Supplementary Appendix 2: Cohort characteristics of individuals and match rate

		Total number of individuals	Routine (pre-algorithm) match rate (n,%)	Boosted (post-algorithm) match rate (n,%)	Percentage change between routine (pre-) and boosted (post-match) rate (%)
Total		38,260 (100)	14,024 (100)	23,409 (100)	24.5
Sex	Male	20,390 (53.3)	7375 (52.6)	12,405 (53)	24.7
	Female	17,870 (46.7)	6649 (47.4)	11,004 (47)	24.4
Age	<1	7842 (20.5)	1839 (13.1)	3797 (16.2)	25.0
	1-2	4568 (11.9)	1501 (10.7)	2348 (10.0)	18.5
	3-5	5357 (14.0)	2635 (18.8)	3976 (17.0)	25.0
	6-8	4752 (12.4)	2471 (17.6)	3755 (16.0)	27.0
	9-11	4457 (11.6)	1988 (14.2)	3147 (13.4)	26.0
	12-14	6186 (16.2)	2112 (15.1)	3684 (15.7)	25.4
	15-17	5032 (13.2)	1467 (10.5)	2687 (11.5)	24.2
	18+	66 (0.2)	11 (0.1)	15 (0.1)	6.1
Year of birth	pre-1988	1569 (4.1)	*	10 (0.0)	*
	1988	699 (1.8)	23* (0.2)	35 (0.1)	1.7
	1989	766 (2.0)	54 (0.4)	151 (0.6)	12.7
	1990	861 (2.3)	95 (0.7)	244 (1.0)	17.3
	1991	946 (2.5)	107 (0.8)	307 (1.3)	21.1
	1992	961 (2.5)	137 (1.0)	444 (1.9)	31.9
	1993	1042 (2.7)	134 (1.0)	590 (2.5)	43.8
	1994	1154 (3.0)	166 (1.2)	724 (3.1)	48.4
	1995	1133 (3.0)	173 (1.2)	767 (3.3)	52.4
	1996	1285 (3.4)	216 (1.5)	897 (3.8)	53.0
	1997	1323 (3.5)	248 (1.8)	941 (4.0)	52.4
	1998	1270 (3.3)	454 (3.2)	877 (3.7)	33.3
	1999	1341 (3.5)	610 (4.3)	927 (4.0)	23.6
	2000	1385 (3.6)	654 (4.7)	944 (4.0)	20.9
	2001	1431 (3.7)	670 (4.8)	933 (4.0)	18.4
	2002	1459 (3.8)	709 (5.1)	970 (4.1)	17.9
	2003	1493 (3.9)	768 (5.5)	1043 (4.5)	18.4
	2004	1447 (3.8)	772 (5.5)	1017 (4.3)	16.9
	2005	1403 (3.7)	734 (5.2)	1056 (4.5)	23.0
	2006	1366 (3.6)	743 (5.3)	1049 (4.5)	22.4
	2007	1300 (3.4)	672 (4.8)	1030 (4.4)	27.5
	2008	1298 (3.4)	691 (4.9)	1008 (4.3)	24.4
	2009	1321 (3.5)	678 (4.8)	998 (4.3)	24.2
	2010	1232 (3.2)	627 (4.5)	902 (3.9)	22.3
	2011	1193 (3.1)	646 (4.6)	894 (3.8)	20.8
	2012	1237 (3.2)	665 (4.7)	878 (3.8)	17.2
	2013	1117 (2.9)	597 (4.3)	783 (3.3)	16.7
	2014	990 (2.6)	519 (3.7)	677 (2.9)	16.0
	2015	892 (2.3)	462 (3.3)	627 (2.7)	18.5
	2016	884 (2.3)	383 (2.7)	559 (2.4)	19.9
	2017	771 (2.0)	175 (1.2)	325 (1.4)	19.5
	2018	716 (1.9)	176 (1.3)	324 (1.4)	20.7
	2019	534 (1.4)	151 (1.1)	252 (1.1)	18.9
	2020	389 (1.0)	101 (0.7)	197 (0.8)	24.7
	2021	52 (0.1)	14 (0.1)	29 (0.1)	28.8
WIMD 2019	1 (most deprived)	11,517 (30.1)	5554 (39.6)	8674 (37.1)	27.1
	2	5827 (15.2)	2856 (20.4)	4385 (18.7)	26.2
	3	3646 (9.5)	1787 (12.7)	2700 (11.5)	25.0
	4	2310 (6.0)	1160 (8.3)	1717 (7.3)	24.1
	5 (least deprived)	1301 (3.4)	626 (4.5)	966 (4.1)	26.1
	Missing	13,659 (35.7)	2041 (14.6)	4967 (21.2)	21.4

Continued

Supplementary Appendix 2: Continued

		Total number of individuals	Routine (pre-algorithm) match rate (n,%)	Boosted (post-algorithm) match rate (n,%)	Percentage change between routine (pre-) and boosted (post-match) rate (%)
Ethnicity	Asian or Asian British	750 (2.0)	291 (2.1)	423 (1.8)	17.6
	Black; African; Caribbean or Black	585 (1.5)	237 (1.7)	336 (1.4)	16.9
	British				
	Mixed ethnic groups	1095 (2.9)	476 (3.4)	738 (3.2)	23.9
	Information not obtained	2067 (5.4)	443 (3.2)	894 (3.8)	21.8
	Other ethnic group	706 (1.8)	240 (1.7)	368 (1.6)	18.1
	White	33,057 (86.4)	12,337 (88.0)	20650 (88.2)	25.1
Asylum seeking status	No	26,713 (69.8)	12,507 (89.2)	194 (0.8)	26.5
	Yes	539 (1.4)	124 (0.9)	19577 (83.6)	13.0
	Missing	11,008 (28.8)	1393 (9.9)	3638 (15.5)	20.4

*numbers combined to prevent disclosure of small numbers (<5).

Demographics are at the time of first CLA episode.

WIMD: Welsh Index of Multiple Deprivation.

