

Stage 1: Preprocessing - data quality of matching fields

Number of Person ID - Pupil Census records	10,525,879	
Number of Person ID - Pupil Census pupils	1,457,198	
Number of Person ID - Attendance pupils not in Pupil Census dataset	11,863	
Number of Person ID - Exclusions pupils not in Pupil Census or Attendance datasets	3	
Number of Person ID - Total pupils	1,469,064	
Note - pupils not in Pupil Census had no postcode, DOB or gender and were not used in matching		
Number of invalid (missing or truncated) postcodes	4,357	0.0%
Number of invalid gender	-	0.0%
Number missing DOB	4,179	0.0%
Number where YOB < 1989 (oldest is 1908)	1,873	0.0%
Number where YOB > 2015 (up to 2047)	11	0.0%
Number where MOB = 1	884,639	8.4%
- January MOB expectation based on uniform distribution across calendar year	893,012	8.5%
Number where Day of Birth = 1	345,053	3.3%
- DYOB=1 expectation based on uniform distribution across months	345,682	3.3%
Academic Year		
2007/08	724,363	6.9%
2008/09	723,996	6.9%
2009/10	732,735	7.0%
2010/11	750,517	7.1%
2011/12	774,399	7.4%
2012/13	797,344	7.6%
2013/14	809,518	7.7%
2014/15	821,818	7.8%
2015/16	837,585	8.0%
2016/17	859,984	8.2%
2017/18	877,279	8.3%
2018/19	897,362	8.5%
2019/20	918,979	8.7%
Number of records input for Spine Matching	2,151,669	
(Number of unique postcode-DOB-gender combinations for all unique pupils)		

Linkage Summary Report

Stage 2: BigMatch Linkage against the Indexing Spine

BigMatch is a linkage software program developed and used in-house by the Statistical Research Division, U.S. Bureau of Census. It has been designed to undertake timely matching of very large files (e.g. linking the US census, 300 million x 300 million).

The program is strictly a linkage engine and implements traditional probabilistic record linkage methodology.

The Bigmatch program is designed to extract plausible matches from a large file using several blocking criteria without having to sort the file before each blocking run.

Further details at <https://www.census.gov/srd/papers/pdf/rcc2007-01.pdf>

In this run, probabilistic weights and match categories similar to those piloted in the pupil census linkage were generated in SAS on the plausible matches generated by BigMatch - see <http://www.isdscotland.org/Products-and-Services/eDRIS/Docs/20150421-Linking-ScotXed-Data.pdf>

The BigMatch parameters file was set up with the following hierarchical blocking criteria :

<u>Block number</u>	<u>Block description</u>
0	Exact matches on Postcode, Sex, DOB
1	Matches on Postcode & DOB
2	Matches on 1st 6 characters of Postcode, Sex, DOB
3	Matches on Postcode, Sex, Year & Month of Birth
4	Matches on Postcode, Sex, Year & Day of Month of Birth
5	Matches on Postcode, Sex, Month & Day of Birth
6	Matches on 1st 5 characters of Postcode, Sex, DOB
7	Matches on 1st 4 characters of Postcode, Sex, DOB
8	Matches on Postcode, Sex & Year of Birth

Number of pairs above threshold score output from all blocks per batch:

<u>Batch Number</u>	<u>CensusID in batch</u>	<u>Number of pairs</u>	<u>Unique CensusID/SpineID combinations above threshold(s)</u>	<u>Unique CensusID above threshold(s)</u>	<u>Unique SpineID above threshold(s)</u>	<u>Unique CensusID/SpineID combinations at best match score</u>
1 and 2 combined	2,151,669	11,671,042	4,220,661	1,450,523	2,145,037	1,492,226
TOTAL	2,151,669	11,671,042	4,220,661	1,450,523	2,145,037	1,492,226

Stage 3: DEDUPLICATION

Identify where there are duplicate CensusID across multiple SpineID

Number of CensusID/SpineID combinations at best match score (per CensusID)	1,492,226
Number of CensusID matched to single SpineID at best match score	1,383,688
Number of unique CensusID	1,450,523

An automated process is carried out in order to ensure that each CensusID can appear a maximum of only once in the final linked dataset. .

Step 1: Where CensusID spans>1 SpineID in same block retain lowest ordered SpineID	1,450,548
Step 2: Where CensusID spans>1 SpineID in different blocks, drop higher numbered block(s)	1,450,523
Number of census records with best matches to the Spine	1,450,523
Percentage of census records with best matches to the Spine	99.5%
Number of census records with 'optimal' best matches after re-categorisation (see Step4)	1,442,430
Percentage of census records with best matches after re-categorisation (see Step4)	99.0%
Number of census records with 'optimal' best matches after backlinks added (see Step5)	1,442,820
Percentage of census records with best matches after backlinks added (see Step5)	99.0%
Total Number of Unique Seeded SpineID Numbers amongst 'optimal' best matches	1,413,018

Linkage Summary Report

Stage 4: Match categorisations

The BigMatch blocking criteria employed in this linkage and the block-specific linkage thresholds were chosen to replicate, as close as possible, the "ScotXed" linkage criteria used for linkages where only Postcode, Sex and Date of Birth are available. The BigMatch score thresholds used in this linkage were set at a value of 9.0 for Block 1 (in order to restrict pairs in this category to where gender is missing) and 5.0 for all other blocks.

The post-BigMatch processing involved re-calculating weights using the ISD Medical Record Linkage scoring system and deriving delta weights representing the distance to the closest rival Spine match.

Broader match categories (compared to the original 'ScotXed' criteria) have been assigned to each best match pair and these are defined as follows:

- 1= Exact Match (including ties).
- 2= Mis-match on last character of standardised 7-character postcode.
- 3= Mis-match on one of either Year, Month or Day of Date of Birth.
- 4= ISD MRL linkage weight >24.0.
- 5= Lower quality matches.

Further, flags were attached to best matches to create a menu of pre-defined link status as follows:-

- Unique **Exact** Matches (broad match category 1 where delta weight > 0);
- Original Scotxed **'Safe'** matches (broad match category 1-2 where delta weight >0);
- Optimal** match categories (best balance between precision and recall identified in ScotXed linkage - includes all from broad match categories 1, 2 & 4 and higher quality matches from 3).

Matches which did not reach 'Optimal' status were re-defined as non-links.

Number of initial best matches - by BigMatch Blocking Strategy :-

BestBlock	Description	Frequency	Percent
0	Exact matches on Postcode, Sex, DOB	1,421,057	98.0%
1	Matches on Postcode & DOB (missing gender)	-	0.0%
2	Matches on 1st 6 characters of Postcode, Sex, DOB	3,418	0.2%
3	Matches on Postcode, Sex, Year & Month of Birth	2,305	0.2%
4	Matches on Postcode, Sex, Year & Day of Month of Birth	925	0.1%
5	Matches on Postcode, Sex, Month & Day of Birth	2,282	0.2%
6	Matches on 1st 5 characters of Postcode, Sex, DOB	9,489	0.7%
7	Matches on 1st 4 characters of Postcode, Sex, DOB	9,666	0.7%
8	Matches on Postcode, Sex & Year of Birth	1,381	0.1%
Overall		1,450,523	100.0%

Number of initial best matches - by Broad match categories :-

- includes matches below optimal linkage threshold

Broad Category	Description	Frequency	Percent
1	Exact Match (including ties).	1,421,057	98.0%
2	Mis-match on last character of standardised 7-character postcode.	8,618	0.6%
3	Mis-match on one of either Year, Month or Day of Date of Birth.	4,318	0.3%
4	ISD MRL linkage weight >24.0.	11,120	0.8%
5	Lower quality matches.	5,410	0.4%
Overall		1,450,523	100.0%

Number of Best Matches by Linkage Criteria

	N	% of cohort	Precision* Crude
			Estimate
Default - Optimal Links	1,442,430	99.0%	98.0%
Safe Links	1,394,238	95.7%	99.8%
Unique Exact Matches	1,386,079	95.1%	99.9%

*Precision estimate based on ScotXed linkages - see <http://www.isdscotland.org/Products-and-Services/eDRIS/Docs/20150421-Linking-ScotXed-Data.pdf>

Linkage Summary Report

Stage 5: Merge with previous pupil census read-through

The 2019 pupil census dataset provided for indexing was matched to the existing pupil census read-through file (created Feb 2019) using the read-through index number provided. The existing read-through file was first matched to the latest Spine (June-2020) to verify whether the SpineIDs from that linkage still exist.

The information from these merges was used to identify or establish:

- 1) Number of existing Spine-matched read-through numbers whose matched SpineID no longer exists on the latest Spine (June-2020).
- 2) Number of existing Spine-matched read-through numbers who now match to a different SpineID.
- 3) Number of cases from the 2019 pupil census which are new.
- 4) Update the Match Categories and best matching SpineIDs, using linkage information from both this indexing and the previous read-through indexing for pupils with information in both sources.

Number of read-through records from previous existing indexing (Feb-2019) in this indexing	1,392,394	
Number of read-through records with 'optimal' matches from existing indexing (Feb-2019)	1,378,410	99.0%
Number of deleted SpineIDs since existing read-through indexing (Feb-2019)	742	0.05%
Total number of merged pupil IDs across both this and previous indexing	1,457,198	
Number of new pupil IDs in this indexing (not in previous) - including non-matches	64,804	4.4%
Number of new SpineID links amongst previous existing non-links	4,570	0.33%
Number of amended SpineID links amongst previous existing links	1,700	0.12%
Number of deleted SpineIDs amongst previous existing links (excl. amended)	203	0.01%
Number of new SpineIDs amongst 2019 Pupil Census back-linked from previous read-through	459	0.03%
Number of amended SpineIDs amongst 2019 Pupil Census back-linked from previous read-through	1,700	0.12%

Number of final best matches - by Source and Linkage Criteria :-

Denominator	Description	OPTIMAL LINKS		SAFE LINKS		UNIQUE EXACT	
		Frequency	Percent	Frequency	Percent	Frequency	Percent
1,392,394	Pupil Census 2019 - in previous read-thru	1,381,313	99.2%	1,333,499	95.8%	1,325,892	95.2%
64,804	Pupil Census 2019 - new pupils	61,507	94.9%	58,774	90.7%	58,223	89.8%
1,457,198	Total read-thru - as at Aug 2020	1,442,820	99.0%	1,392,273	95.5%	1,384,115	95.0%

Dataset Name: Pupil Census
Dataset ID: 2020-A001

Stage 6: Pupil Census 2019 - Linkage Rates by Demography

SEX			% Optimal Links			% Safe Links			% Unique Exact Matches		
Sex	Frequency		Sex	Total	Row %	Sex	Total	Row %	Sex	Total	Row %
1	Male	742,154	1	735,224	99.1%	1	709,843	95.6%	1	705,461	95.1%
2	Female	714,540	2	707,108	99.0%	2	681,960	95.4%	2	678,191	94.9%
3	Unknown	504	3	488	96.8%	3	470	93.3%	3	463	91.9%
Total		1,457,198	Total	1,442,820	99.0%	Total	1,392,273	95.5%	Total	1,384,115	95.0%

YEAR OF BIRTH			% Optimal Links			% Safe Links			% Unique Exact Matches		
YOB	Frequency		YOB	Total	Row %	YOB	Total	Row %	YOB	Total	Row %
<=1988	1,134		<=1988	0	0.0%	<=1988	0	0.0%	<=1988	0	0.0%
1989	740		1989	727	98.2%	1989	679	91.8%	1989	666	90.0%
1990	24,481		1990	24,236	99.0%	1990	23,277	95.1%	1990	23,014	94.0%
1991	46,325		1991	45,865	99.0%	1991	44,174	95.4%	1991	43,527	94.0%
1992	62,923		1992	62,311	99.0%	1992	59,916	95.2%	1992	59,066	93.9%
1993	61,282		1993	60,847	99.3%	1993	58,745	95.9%	1993	58,154	94.9%
1994	60,902		1994	60,546	99.4%	1994	58,542	96.1%	1994	58,070	95.3%
1995	60,085		1995	59,773	99.5%	1995	57,789	96.2%	1995	57,442	95.6%
1996	60,104		1996	59,762	99.4%	1996	57,804	96.2%	1996	57,517	95.7%
1997	61,343		1997	60,992	99.4%	1997	58,955	96.1%	1997	58,680	95.7%
1998	60,535		1998	60,194	99.4%	1998	58,177	96.1%	1998	57,933	95.7%
1999	59,110		1999	58,687	99.3%	1999	56,913	96.3%	1999	56,705	95.9%
2000	58,256		2000	57,876	99.3%	2000	56,087	96.3%	2000	55,906	96.0%
2001	58,064		2001	57,652	99.3%	2001	55,654	95.8%	2001	55,466	95.5%
2002	57,617		2002	57,148	99.2%	2002	55,142	95.7%	2002	54,904	95.3%
2003	58,498		2003	58,023	99.2%	2003	55,996	95.7%	2003	55,763	95.3%
2004	59,967		2004	59,542	99.3%	2004	57,533	95.9%	2004	57,311	95.6%
2005	59,744		2005	59,351	99.3%	2005	57,306	95.9%	2005	57,081	95.5%
2006	60,750		2006	60,360	99.4%	2006	58,222	95.8%	2006	57,962	95.4%
2007	62,159		2007	61,776	99.4%	2007	59,617	95.9%	2007	59,375	95.5%
2008	63,191		2008	62,743	99.3%	2008	60,502	95.7%	2008	60,252	95.3%
2009	62,148		2009	61,671	99.2%	2009	59,425	95.6%	2009	59,167	95.2%
2010	61,325		2010	60,795	99.1%	2010	58,571	95.5%	2010	58,290	95.1%
2011	60,670		2011	59,966	98.8%	2011	57,756	95.2%	2011	57,451	94.7%
2012	59,581		2012	58,716	98.5%	2012	56,581	95.0%	2012	56,281	94.5%
2013	56,764		2013	55,541	97.8%	2013	53,469	94.2%	2013	53,106	93.6%
2014	54,922		2014	53,285	97.0%	2014	51,175	93.2%	2014	50,797	92.5%
>=2015	4,540		>=2015	4,413	97.2%	>=2015	4,244	93.5%	>=2015	4,207	92.7%
99	missing	38	99	22	57.9%	99	22	57.9%	99	22	57.9%
Total		1,457,198	Total	1,442,820	99.0%	Total	1,392,273	95.5%	Total	1,384,115	95.0%

SIMD 2020 DECILE			% Optimal Links			% Safe Links			% Unique Exact Matches		
SIMD 2020 DECILE	Frequency		SIMD 2020	Total	Row %	SIMD 2020	Total	Row %	SIMD 2020	Total	Row %
1 Most deprived	159,586		1	158,112	99.1%	1	152,067	95.3%	1	150,707	94.4%
2	156,798		2	155,381	99.1%	2	149,824	95.6%	2	148,861	94.9%
3	144,392		3	143,030	99.1%	3	138,114	95.7%	3	137,364	95.1%
4	139,453		4	138,221	99.2%	4	133,607	95.8%	4	132,880	95.3%
5	138,753		5	137,462	99.1%	5	132,916	95.8%	5	132,187	95.3%
6	135,932		6	134,536	99.1%	6	130,021	95.7%	6	129,196	95.0%
7	140,439		7	139,176	99.1%	7	134,368	95.7%	7	133,639	95.2%
8	149,752		8	148,425	99.2%	8	143,403	95.8%	8	142,744	95.3%
9	148,556		9	147,257	99.2%	9	142,173	95.7%	9	141,652	95.4%
10 Least deprived	136,928		10	135,728	99.2%	10	131,102	95.7%	10	130,747	95.5%
99 Missing	6,609		99	5,492	82.5%	99	4,678	70.8%	99	4,138	62.6%
Total		1,457,198	Total	1,442,820	99.0%	Total	1,392,273	95.5%	Total	1,384,115	95.0%

ACADEMIC YEAR*			% Optimal Links			% Safe Links			% Unique Exact Matches		
ACYEAR	Frequency		ACYEAR	Total	Row %	ACYEAR	Total	Row %	ACYEAR	Total	Row %
2007/08	71,356		2007/08	69,913	98.0%	2007/08	66,583	93.3%	2007/08	65,273	91.5%
2008/09	65,310		2008/09	64,175	98.3%	2008/09	61,627	94.4%	2008/09	60,665	92.9%
2009/10	64,203		2009/10	63,567	99.0%	2009/10	61,257	95.4%	2009/10	60,569	94.3%
2010/11	64,994		2010/11	64,318	99.0%	2010/11	62,126	95.6%	2010/11	61,650	94.9%
2011/12	63,239		2011/12	62,684	99.1%	2011/12	60,565	95.8%	2011/12	60,161	95.1%
2012/13	62,957		2012/13	62,464	99.2%	2012/13	60,463	96.0%	2012/13	60,126	95.5%
2013/14	62,207		2013/14	61,674	99.1%	2013/14	59,561	95.7%	2013/14	59,286	95.3%
2014/15	62,014		2014/15	61,493	99.2%	2014/15	59,529	96.0%	2014/15	59,315	95.6%
2015/16	61,714		2015/16	61,270	99.3%	2015/16	59,313	96.1%	2015/16	59,130	95.8%
2016/17	60,821		2016/17	60,359	99.2%	2016/17	58,507	96.2%	2016/17	58,334	95.9%
2017/18	59,299		2017/18	58,834	99.2%	2017/18	56,839	95.9%	2017/18	56,685	95.6%
2018/19	60,195		2018/19	59,691	99.2%	2018/19	57,659	95.8%	2018/19	57,492	95.5%
2019/20	698,889		2019/20	692,378	99.1%	2019/20	668,244	95.6%	2019/20	665,429	95.2%
Total		1,457,198	Total	1,442,820	99.0%	Total	1,392,273	95.5%	Total	1,384,115	95.0%

*Note: Most recent academic year taken for each individual in the table

Dataset Name: Attendance
Dataset ID: 2020-A002

Stage 6: Attendance 2019 - Linkage Rates by Demography

SEX			% Optimal Links			% Safe Links			% Unique Exact Matches		
Sex		Frequency	Sex	Total	Row %	Sex	Total	Row %	Sex	Total	Row %
1	Male	691,218	1	686,386	99.3%	1	663,292	96.0%	1	659,505	95.4%
2	Female	666,194	2	661,603	99.3%	2	638,611	95.9%	2	635,300	95.4%
3	Unknown	43,826	3	31,142	71.1%	3	29,764	67.9%	3	29,369	67.0%
Total		1,401,238	Total	1,379,131	98.4%	Total	1,331,667	95.0%	Total	1,324,174	94.5%

YEAR OF BIRTH			% Optimal Links			% Safe Links			% Unique Exact Matches		
YOB		Frequency	YOB	Total	Row %	YOB	Total	Row %	YOB	Total	Row %
<=1989		610	<=1989	573	93.9%	<=1989	537	88.0%	<=1989	527	86.4%
1990		24,129	1990	23,894	99.0%	1990	22,957	95.1%	1990	22,702	94.1%
1991		46,025	1991	45,569	99.0%	1991	43,902	95.4%	1991	43,263	94.0%
1992		62,693	1992	62,090	99.0%	1992	59,706	95.2%	1992	58,860	93.9%
1993		61,101	1993	60,681	99.3%	1993	58,587	95.9%	1993	58,004	94.9%
1994		60,503	1994	60,194	99.5%	1994	58,230	96.2%	1994	57,774	95.5%
1995		59,741	1995	59,484	99.6%	1995	57,543	96.3%	1995	57,201	95.7%
1996		59,668	1996	59,412	99.6%	1996	57,514	96.4%	1996	57,238	95.9%
1997		60,962	1997	60,670	99.5%	1997	58,678	96.3%	1997	58,415	95.8%
1998		60,092	1998	59,826	99.6%	1998	57,875	96.3%	1998	57,642	95.9%
1999		58,719	1999	58,377	99.4%	1999	56,638	96.5%	1999	56,438	96.1%
2000		57,704	2000	57,421	99.5%	2000	55,700	96.5%	2000	55,535	96.2%
2001		57,607	2001	57,285	99.4%	2001	55,333	96.1%	2001	55,157	95.7%
2002		56,697	2002	56,387	99.5%	2002	54,459	96.1%	2002	54,237	95.7%
2003		57,612	2003	57,270	99.4%	2003	55,330	96.0%	2003	55,120	95.7%
2004		59,051	2004	58,740	99.5%	2004	56,810	96.2%	2004	56,599	95.8%
2005		58,745	2005	58,470	99.5%	2005	56,487	96.2%	2005	56,289	95.8%
2006		59,183	2006	58,922	99.6%	2006	56,907	96.2%	2006	56,688	95.8%
2007		61,041	2007	60,766	99.5%	2007	58,679	96.1%	2007	58,466	95.8%
2008		61,644	2008	61,301	99.4%	2008	59,166	96.0%	2008	58,954	95.6%
2009		61,020	2009	60,629	99.4%	2009	58,455	95.8%	2009	58,216	95.4%
2010		59,750	2010	59,322	99.3%	2010	57,204	95.7%	2010	56,952	95.3%
2011		59,569	2011	58,947	99.0%	2011	56,801	95.4%	2011	56,523	94.9%
2012		57,926	2012	57,143	98.6%	2012	55,104	95.1%	2012	54,822	94.6%
2013		54,617	2013	53,496	97.9%	2013	51,530	94.3%	2013	51,201	93.7%
>=2014		4,841	>=2014	4,746	98.0%	>=2014	4,574	94.5%	>=2014	4,544	93.9%
99	missing	19,988	99	7,516	37.6%	99	6,961	34.8%	99	6,807	34.1%
Total		1,401,238	Total	1,379,131	98.4%	Total	1,331,667	95.0%	Total	1,324,174	94.5%

SIMD 2020 DECILE			% Optimal Links			% Safe Links			% Unique Exact Matches		
SIMD 2020 DECILE		Frequency	SIMD 2020	Total	Row %	SIMD 2020	Total	Row %	SIMD 2020	Total	Row %
1	Most deprived	150,829	1	149,843	99.3%	1	144,244	95.6%	1	143,027	94.8%
2		148,981	2	148,026	99.4%	2	142,840	95.9%	2	141,958	95.3%
3		137,410	3	136,479	99.3%	3	131,899	96.0%	3	131,221	95.5%
4		132,224	4	131,471	99.4%	4	127,237	96.2%	4	126,587	95.7%
5		131,996	5	131,181	99.4%	5	126,978	96.2%	5	126,300	95.7%
6		129,211	6	128,302	99.3%	6	124,094	96.0%	6	123,341	95.5%
7		133,056	7	132,199	99.4%	7	127,699	96.0%	7	127,056	95.5%
8		140,762	8	139,912	99.4%	8	135,253	96.1%	8	134,668	95.7%
9		140,582	9	139,723	99.4%	9	135,034	96.1%	9	134,575	95.7%
10	Least deprived	129,599	10	128,827	99.4%	10	124,568	96.1%	10	124,283	95.9%
99	Missing	26,588	99	13,168	49.5%	99	11,821	44.5%	99	11,158	42.0%
Total		1,401,238	Total	1,379,131	98.4%	Total	1,331,667	95.0%	Total	1,324,174	94.5%

ACADEMIC YEAR*			% Optimal Links			% Safe Links			% Unique Exact Matches		
ACYEAR		Frequency	ACYEAR	Total	Row %	ACYEAR	Total	Row %	ACYEAR	Total	Row %
2007/08		72,170	2007/08	65,713	91.1%	2007/08	62,592	86.7%	2007/08	61,390	85.1%
2008/09		66,466	2008/09	64,561	97.1%	2008/09	62,020	93.3%	2008/09	61,053	91.9%
2009/10		65,626	2009/10	64,056	97.6%	2009/10	61,736	94.1%	2009/10	61,017	93.0%
2010/11		127,191	2010/11	125,386	98.6%	2010/11	121,236	95.3%	2010/11	120,405	94.7%
2011/12		0	2011/12	0		2011/12	0		2011/12	0	
2012/13		124,534	2012/13	122,735	98.6%	2012/13	118,734	95.3%	2012/13	118,172	94.9%
2013/14		0	2013/14	0		2013/14	0		2013/14	0	
2014/15		122,473	2014/15	121,038	98.8%	2014/15	117,272	95.8%	2014/15	116,892	95.4%
2015/16		0	2015/16	0		2015/16	0		2015/16	0	
2016/17		119,887	2016/17	118,426	98.8%	2016/17	114,667	95.6%	2016/17	114,363	95.4%
2017/18		0	2017/18	0		2017/18	0		2017/18	0	
2018/19		702,891	2018/19	697,216	99.2%	2018/19	673,410	95.8%	2018/19	670,882	95.4%
2019/20		0	2019/20	0		2019/20	0		2019/20	0	
Total		1,401,238	Total	1,379,131	98.4%	Total	1,331,667	95.0%	Total	1,324,174	94.5%

*Note: Most recent academic year taken for each individual in the table

Dataset Name: Exclusions
Dataset ID: 2020-A003

Stage 6: Exclusions 2019 - Linkage Rates by Demography

SEX			% Optimal Links			% Safe Links			% Unique Exact Matches		
Sex	Frequency		Total	Row %		Total	Row %	Total	Row %		Row %
1 Male	58,837	1	58,570	99.5%	1	56,487	96.0%	1	56,049	95.3%	
2 Female	20,078	2	20,007	99.6%	2	19,254	95.9%	2	19,090	95.1%	
3 Unknown	457	3	354	77.5%	3	339	74.2%	3	329	72.0%	
Total	79,372		78,931	99.4%	Total	76,080	95.9%	Total	75,468	95.1%	

YEAR OF BIRTH			% Optimal Links			% Safe Links			% Unique Exact Matches		
YOB	Frequency		Total	Row %	YOB	Total	Row %	YOB	Total	Row %	
<=1990	133	<=1990	133	100.0%	<=1990	126	94.7%	<=1990	125	94.0%	
1991	772	1991	763	98.8%	1991	719	93.1%	1991	705	91.3%	
1992	4,600	1992	4,518	98.2%	1992	4,300	93.5%	1992	4,195	91.2%	
1993	7,307	1993	7,239	99.1%	1993	6,897	94.4%	1993	6,764	92.6%	
1994	8,115	1994	8,069	99.4%	1994	7,757	95.6%	1994	7,655	94.3%	
1995	8,033	1995	8,007	99.7%	1995	7,748	96.5%	1995	7,685	95.7%	
1996	7,019	1996	7,002	99.8%	1996	6,778	96.6%	1996	6,728	95.9%	
1997	6,070	1997	6,062	99.9%	1997	5,847	96.3%	1997	5,811	95.7%	
1998	5,078	1998	5,064	99.7%	1998	4,883	96.2%	1998	4,866	95.8%	
1999	4,604	1999	4,595	99.8%	1999	4,465	97.0%	1999	4,443	96.5%	
2000	4,358	2000	4,353	99.9%	2000	4,235	97.2%	2000	4,232	97.1%	
2001	3,882	2001	3,873	99.8%	2001	3,751	96.6%	2001	3,747	96.5%	
2002	3,796	2002	3,788	99.8%	2002	3,636	95.8%	2002	3,627	95.5%	
2003	3,804	2003	3,791	99.7%	2003	3,676	96.6%	2003	3,667	96.4%	
2004	3,565	2004	3,558	99.8%	2004	3,451	96.8%	2004	3,438	96.4%	
2005	2,719	2005	2,715	99.9%	2005	2,616	96.2%	2005	2,612	96.1%	
2006	1,811	2006	1,807	99.8%	2006	1,738	96.0%	2006	1,730	95.5%	
2007	1,042	2007	1,041	99.9%	2007	1,014	97.3%	2007	1,010	96.9%	
2008	784	2008	780	99.5%	2008	743	94.8%	2008	742	94.6%	
2009	589	2009	588	99.8%	2009	560	95.1%	2009	557	94.6%	
2010	415	2010	415	100.0%	2010	400	96.4%	2010	399	96.1%	
2011	272	2011	269	98.9%	2011	259	95.2%	2011	259	95.2%	
2012	154	2012	154	100.0%	2012	151	98.1%	2012	150	97.4%	
>=2013	67	>=2013	66	98.5%	>=2013	62	92.5%	>=2013	61	91.0%	
99 missing	383	99	281	73.4%	99	268	70.0%	99	260	67.9%	
Total	79,372	Total	78,931	99.4%	Total	76,080	95.9%	Total	75,468	95.1%	

SIMD 2020 DECILE			% Optimal Links			% Safe Links			% Unique Exact Matches		
SIMD 2020 DECILE	Frequency		Total	Row %	SIMD 2020	Total	Row %	SIMD 2020	Total	Row %	
1 Most deprived	16,074	1	16,020	99.7%	1	15,413	95.9%	1	15,247	94.9%	
2	13,799	2	13,759	99.7%	2	13,232	95.9%	2	13,128	95.1%	
3	10,947	3	10,920	99.8%	3	10,532	96.2%	3	10,464	95.6%	
4	8,865	4	8,838	99.7%	4	8,557	96.5%	4	8,493	95.8%	
5	7,062	5	7,051	99.8%	5	6,832	96.7%	5	6,793	96.2%	
6	5,776	6	5,741	99.4%	6	5,532	95.8%	6	5,495	95.1%	
7	5,168	7	5,147	99.6%	7	4,980	96.4%	7	4,946	95.7%	
8	4,464	8	4,445	99.6%	8	4,299	96.3%	8	4,275	95.8%	
9	3,526	9	3,505	99.4%	9	3,379	95.8%	9	3,368	95.5%	
10 Least deprived	2,318	10	2,308	99.6%	10	2,233	96.3%	10	2,228	96.1%	
99 Missing	1,373	99	1,197	87.2%	99	1,091	79.5%	99	1,031	75.1%	
Total	79,372	Total	78,931	99.4%	Total	76,080	95.9%	Total	75,468	95.1%	

ACADEMIC YEAR*			% Optimal Links			% Safe Links			% Unique Exact Matches		
ACYEAR	Frequency		Total	Row %	ACYEAR	Total	Row %	ACYEAR	Total	Row %	
2007/08	11,369	2007/08	11,201	98.5%	2007/08	10,694	94.1%	2007/08	10,474	92.1%	
2008/09	10,350	2008/09	10,275	99.3%	2008/09	9,830	95.0%	2008/09	9,697	93.7%	
2009/10	9,804	2009/10	9,758	99.5%	2009/10	9,398	95.9%	2009/10	9,319	95.1%	
2010/11	11,503	2010/11	11,462	99.6%	2010/11	11,080	96.3%	2010/11	11,002	95.6%	
2011/12	0	2011/12	0		2011/12	0		2011/12	0		
2012/13	9,296	2012/13	9,267	99.7%	2012/13	8,992	96.7%	2012/13	8,958	96.4%	
2013/14	0	2013/14	0		2013/14	0		2013/14	0		
2014/15	8,856	2014/15	8,837	99.8%	2014/15	8,584	96.9%	2014/15	8,564	96.7%	
2015/16	0	2015/16	0		2015/16	0		2015/16	0		
2016/17	9,014	2016/17	8,982	99.6%	2016/17	8,654	96.0%	2016/17	8,632	95.8%	
2017/18	0	2017/18	0		2017/18	0		2017/18	0		
2018/19	9,180	2018/19	9,149	99.7%	2018/19	8,848	96.4%	2018/19	8,822	96.1%	
2019/20	0	2019/20	0		2019/20	0		2019/20	0		
Total	79,372	Total	78,931	99.4%	Total	76,080	95.9%	Total	75,468	95.1%	

*Note: Most recent academic year taken for each individual in the table

Dataset Name: Pupil Census

Dataset ID: 2020-A001

Linkage Summary Report

Stage 1: Preprocessing

Total records in Pupil Census dataset	10,525,879
Total people in Pupil Census dataset	1,457,198
Total number of unique Demographic keys	8,907,450
Total number of unique Storage keys	8,908,491

Stage 2: Match to Spine

%Match Rate

Number of records matched to the Spine in Pupil Census dataset	10,484,661	99.6%
Number of people matched to the Spine in Pupil Census dataset	1,442,820	99.0%
Number of unique seeded SpineID numbers	1,413,018	

Dataset Name: Attendance

Dataset ID: 2020-A002

Linkage Summary Report

Stage 1: Preprocessing

Total records in Attendance dataset	86,888,306
Total people in Attendance dataset	1,401,238
Total number of unique Demographic keys	86,888,306
Total number of unique Storage keys	86,888,306

Stage 2: Match to Spine

%Match Rate

Number of records matched to the Spine in Attendance dataset	86,440,543	99.5%
Number of people matched to the Spine in Attendance dataset	1,379,131	98.4%
Number of unique seeded SpineID numbers	1,352,682	

Dataset Name: Exclusions

Dataset ID: 2020-A003

Linkage Summary Report

Stage 1: Preprocessing

Total records in Exclusions dataset	208,915
Total people in Exclusions dataset	79,372
Total number of unique Demographic keys	206,591
Total number of unique Storage keys	207,782

Stage 2: Match to Spine

%Match Rate

Number of records matched to the Spine in Exclusions dataset	207,881	99.5%
Number of people matched to the Spine in Exclusions dataset	78,931	99.4%
Number of unique seeded SpineID numbers	78,372	