	movements									
92	Radio travel news	-	No estimate possible as no signs of this type in sample data						-	-
93	No through roads	103,481	103,483	1,908	1.8%	101,575	105,391	0.0%	1,067	17,266
94	AA/RAC Telephone/service	556	No estimate possible as too few signs of this type in sample data						6	17
95	Traffic merge	736	3,376	241	7.1%	3,135	3,617	358.7%	130	500
96	Pedestrian information	-	7,065	535	7.6%	6,530	7,600		293	1,177
97	Cycle information	1,018	31,154	1,054	3.4%	30,099	32,208	2960.3%	669	5,030
98	Information signs	278	2,487	282	11.3%	2,205	2,769	794.7%	95	391
99	Controlled parking zone	18,294	68,090	2,674	3.9%	65,416	70,763	272.2%	800	11,528
100	Lane control	-	No estimate possible as too few signs of this type in sample data						14	29
101	Narrow road	5,943	15,187	681	4.5%	14,506	15,868	155.5%	486	3,162
102	Countdown markers	2,891	15,237	758	5.0%	14,479	15,995	427.1%	296	2,531
103	Crossing information	4,210	7,639	780	10.2%	6,859	8,420	81.5%	152	1,514
104	Hard shoulder signs	457	No estimate possible as no signs of this type in sample data						-	-
105	Dual carriageway	2,691	19,546	823	4.2%	18,723	20,369	626.4%	472	3,032
106	Crawler lane	0	No estimate possible as no signs of this type in sample data						-	-
89, 90 & 107	Miscellaneous - Informatory	7,567	20,511	675	3.3%	19,836	21,186	171.1%	663	3,563
108	Miscellaneous - Miscellaneous	256	11,225	801	7.1%	10,424	12,026	4284.7%	148	1,817
109	Roadworks (diagram numbers 7000s)	1,627	12,349	515	4.2%	11,834	12,864	659.0%	499	2,013

110	Parking information		637	90	14.1%	548	727		43	107
111	Police and other officer signs		1,307	131	10.0%	1,176	1,437		93	216
112	Traffic cameras		23,086	853	3.7%	22,232	23,939		610	3,832
113	Bus information		67,995	2,519	3.7%	65,476	70,514		564	10,923
114	Queues ahead		1,833	153	8.4%	1,680	1,987		103	311
115	Rural direction signs (finger posts not footpath s		No estimate possible as too few signs of this type in sample data						15	50
116	Other Misc		No estimate possible as too few signs of this type in sample data						3	5
117	Unknown		200,927	6,851	3.4%	194,076	207,778		766	34,406
120	Red Route signs		11,331	1,278	0	10,053	12,609		64	1,860
121	Congestion Charge Zone signs		No estimate possible as too few signs of this type in sample data						21	138
	Total	2,161,695	4,600,584							807,544

Note – the estimate for the total number of signs (4,571,710) was generated using all signs within each MSOA, rather than summing the estimates for the individual sign groups.

Table 1b		
Traffic signs in 1993 not included in 1a		**Estimated number of signs in 1993**
38	Bollards	131,434
54	Perm stay lane signs	49
55	Vehicle regulatory	90
110	Blank grey VMS	569
112	Pelican	11,618
113	Signals with ped control	12,671
302	Signals-no ped control	10,723
304	Rural direction	119,425
	Total	286,579
	Total from table 1a and 1b	2,448,274

Table 2 - By Sign Category										
Traffic sign category		**Estimated number of signs in 1993**	**Estimated number of signs**	**Confidence interval**	**%**	**95% lower bound**	**95% upper bound**	**% Increase or Decrease**	**Number of sample MSOAs containing sign type**	**Number in sample 1,121 MSOAs**
200	All signs	2,161,695	4,571,710	43,194	0.9%	4,528,516	4,614,904	111.5%	1,121	807,544
300	Direction		701,299	11,086	1.6%	690,213	712,385		1,104	133,144
301	Informatory		395,951	5,530	1.4%	390,421	401,480		1,119	65,201
302	Regulatory		2,346,370	28,595	1.2%	2,317,775	2,374,965		1,120	400,740
304	Warning		922,006	9,702	1.1%	912,304	931,708		1,120	167,859
400	All illuminated signs		581,724	8,757	1.5%	572,967	590,481		1,039	91,777
401	All non-illuminated signs		4,004,332	39,739	1.0%	3,964,593	4,044,071		1,120	715,767

Table 3 - By Sign Category using the Population/MSOA only										
Traffic sign category		**Estimated number of signs in1993**	**Estimated number of signs**	**Confidence interval**	**%**	**95% lower bound**	**95% upper bound**	**% Increase or Decrease**	**Number of sample MSOAs containing sign type**	**Number in sample 1,121 MSOAs**
200	All signs	2,161,695	4,816,544	72,780	1.5%	4,743,764	4,889,324	122.8%	1,121	807,544
300	Direction		792,480	20,522	2.6%	771,958	813,002		1,104	133,144
301	Informatory		388,644	6,412	1.6%	382,232	395,057		1,119	65,201
302	Regulatory		2,386,459	37,373	1.6%	2,349,086	2,423,832		1,120	400,740
304	Warning		1,004,738	21,024	2.1%	983,714	1,025,762		1,120	167,859
400	All illuminated signs		547,649	10,338	1.9%	537,311	557,987		1,039	91,777
401	All non-illuminated signs		4,268,931	69,228	1.6%	4,199,704	4,338,159		1,120	715,767

Annex B: Methodology

B.1 Outline of methodology

A summary of the approach used to estimate the number of traffic signs in England is as follows:

1. GIS software was used to map geographical explanatory variables onto 1,121 MSOAs. For example Population, A road length and segments etc Data were stored in an Access database.

2. An initial analysis of the explanatory variables showed that they were highly correlated; therefore it was decided to generate the principal components for use in the regression model rather than use the original variables themselves.

3. Using the datasets from 17 Local Authorities, each sign was assigned to an MSOA.

4. To test the approach, roughly half the 1,121 were randomly selected in order to produce a regression model (based on the principal components) that was then used to estimate the signs in the other half. The results suggested the approach was applicable. We also used the model to predict 10 Local Authorities that had supplied part data sets.

5. A multivariate linear regression was then run to find the best fitting model for number of signs in 1,121 MSOAs.

6. This produced a set of regression coefficients that could be used to predict the number of signs in the unknown MSOAs.

7. The sign estimates for the unknown MSOAs were combined with the number of signs in the known MSOAs to give a national estimate.

8. Steps 5 to 7 were repeated for each sign group. To do this an automatic procedure was written in Excel/VBA to automatically cycle through the sign groups, calculate the principal components, run the regression algorithm and produce a national estimate.

More detail about each of these steps is provided below.

B.2 Collecting and preparing Local Authority/Highway Authority sign data

B.2.1 Sampling Local Authority data

It was established during the pilot study in 2012 that not all Local Authorities held up to date or comprehensive traffic sign inventories. Some Local Authorities held very good records, others had no records at all, and there was a range of coverage and quality and timeliness of records between these two extremes. For more details see Chapter 3 of the report, and Annex C.

A sampling strategy was devised for an ideal sample of Local Authority sign data. This set out to obtain a number of sign datasets in each region of England, and across different types of Local Authority, weighted by population. The target and actual number of datasets received are shown in the table below. Population proved to be one of the most significant factors in predicting the number of road signs.

	Total number of highway authorities (exc. TfL and HA)	**152**			**Target sample (exc TfL and HA)**	
Local Authority Type	**Number of…**		**Total population**	**% of population**	**From sample of 30**	**Datasets received**
	Non metropolitan counties	27	21,251,019	41.1%	12	10
	Unitary Authorities	56	12,112,835	23.4%	7	8
	Within metropolitan counties	36	10,170,395	19.7%	6	7
	Within Greater London	33	8,173,941	15.8%	5	5
Region	Number within…		**Total population**		**From sample of 30**	
	North East	12	2,596,886	5.0%	2	1
	North West	23	7,052,177	13.6%	4	2
	Yorkshire and the Humber	14	5,052,512	9.8%	3	7
	East	11	5,846,965	11.3%	3	3
	East Midlands	9	4,533,222	8.8%	3	2
	West Midlands	13	4,528,802	8.8%	3	3
	London	33	8,173,941	15.8%	5	5
	South West	16	5,288,935	10.2%	3	2
	South East	19	8,634,750	16.7%	5	5
				Total	**30**	**30**

Of the 30 Local Authority datasets collected, plus datasets from TfL and HA (32 in total), 19 were used to build the regression model. The other datasets were not suitable for either developing or testing the model.

B.2.2 Data cleaning and preparation

All datasets required the data to be cleaned to make them consistent. Geographical location codes (Eastings and Northings in the majority of cases) needed to be checked (e.g. did the sign fall into the Local Authority concerned?), multiple signs recorded in one record split out, and sign diagram numbers checked against the TRSGD list of signs, and then classified into one of approximately 100 sign groups, which were the same sign groups as used in the 1993 study. For some Local Authorities, this process was lengthy, as the quality of the sign diagram number data was poor, and there was a need to interpret text field descriptors of signs. Traffic Signs Policy provided guidance and advice

on how to classify some signs. If signs were located outside of the boundary of the Local Authority (usually due to data errors), they were discarded.

Once the data was cleaned, it was imported into ArcGIS and each sign allocated to an MSOA.

TfL and HA datasets for signs were overlaid on the Local Authority sign data, and where they intersected an MSOA, added to the sign data for that MSOA. In addition, BPRN (B and Primary Road Network) data were added to London Borough sign data where it was not included in the data provided by the Local Authorities.

A total of 807,544 signs were assigned across 1,121 MSOAs.

B.3 Preparing the explanatory variables

B.3.1 Explanatory variables

For each MSOA, the following data was collected using ArcGIS and map data on roads, population and geographical features, and attributed to the relevant MSOA:

- Total road length and number of road segments (as counted from the ArcGIS map data) for 6 different road categories: A roads, B roads, Minor roads, Local roads, Motorways and Pedestrianised streets.
- Total length of each of canals, main rivers, secondary rivers and minor rivers.
- Total area of each of woodland, lakes and foreshore
- Total population. This was calculated by locating the centroid of each LSOA (Lower layer Super Output Area) census data point, and attributing the population in that LSOA to the MSOA containing it. LSOAs have populations of around 1,200 – 1,400, and the area of an LSOA can vary depending on the population density of the area.
- Area of MSOA.

This set of explanatory variables was stored in an Access database.

B.4 Multivariate regression models

B.4.1 Total signs estimate

In estimating the total number of traffic signs across the sample Local Authorities, multivariate linear regression provided good fitting models. The Population variable tended to be most closely related to the number of signs in a MSOA. The other variables that were included were based on road length and number of segments. However, all these variables tended to highly correlated with each other – which might be expected - and suggested that there may be problems in using the variables in a regression model due to multicollinearity. It was therefore decided to use the principal components in all of the regression modelling.

A regression model using principal components, gave an estimate for all signs of 4,571,710 +/- 43,194 (95% confidence interval). We believe that the confidence interval is understated – partly because of the quality of the data and partly because we had no control over which MSOAs that were used – i.e. we were not able to take a random sample but were limited to data that were available.

In order to validate the approach, the 1,121 MSOAs where the number of signs was known, a subset of roughly half the MSOAs were randomly selected. The selected MSOAs were used to produce a regression model based on the principal components of the explanatory variables. The model was then used to make predictions of the number of signs in the MSOA that had not been used to generate the regression model. The errors were examined to see that they were normally distributed.

Models for lit and unlit signs were then generated and used to estimate 8 Local Authorities that had partial data. While only 4 of these were within the predicted 95% confidence interval, the estimate of the total number of signs for all 8 was within the confidence 95% interval.

This analysis suggested that at MSOA and local Authority levels, predictions were likely to be unreliable but larger consolidations were more likely to provide reasonable estimates.

As a comparative benchmark for the approach, we also ran a simple ordinary linear regression model based on Population (per MSOA). This gave an estimate of 4,816,544 signs – higher than the estimate using principal components; however, this may be due to some sign groups not exclusively driven by population.

B.4.2 Estimating sign groups

The same approach was taken to find a best fitting model for estimating the number of traffic signs in individual sign groups – although the models produced for sign groups were not as good as for all signs. It was decided to not to estimate a sign group unless there were at least 30 MSOAs that had at least one of that sign group.

In a number of cases, especially where the number of signs in a group was small, multivariate linear regression did not produce a satisfactory model, and so both Poisson, Negative Binomial regression was considered. In general, neither of these variations tended to improve the quality of the model. In general the more unusual/scarce the sign group the more difficult it was to find a good fitting model. If an attempt to use this approach is made in future, we suggest that a wider range of explanatory variables are collected.

Some sign groups only appear in parts of the country – e.g. Red Routes only appear in London and Birmingham. In those cases the analysis only included relevant MSOAs in both fitting the regression model and making the prediction.

Annex C: Data summary

The following table provides a summary of the data provided by each Local Authority or Highways Authority, and includes details of any data limitations specific to that data provider.

	Local Authority or Highway Authority	Notes on data quality/content	Total number of signs	% "unknown" sign type	Total area of Local Authority (sq kms)
Highway Authority	**Highways Agency**	Reasonable - some diagram numbers missing. "The data are maintained by the Service Providers, we cannot guarantee its accuracy or completeness".	172,194	11%	
	TfL	Good. Does not include bus stops nor wayfinding signs.	57,608	none	
	London BPRN	Good - all London primary route signs. Duplicated with some other London Local Authority data.	99,204	3%	
Data to use in Model (Local Authority data plus HA/TfL/LondonBPRN) - Full datasets	**Birmingham**	Good. TSRGD signs only and no street furniture i.e. way markers, way finders, street name plates or bus stops etc.	64,062	0.4%	376
	Blackpool	Good. “Hazard Illumination” signs classified as illuminated.	9,073	0.3%	40.39
	Cambridgeshire	Good	54,834	3%	3043.69
	Camden	Good	23,450	2%	21.78
	Derbyshire	Good	73,384	1%	2526.24
	Hammersmith and Fulham	Good	21,116	2%	17.15
	Hertfordshire	Good	121,870	7%	1642.09
	Lincolnshire	Good	85,518	0.6%	6008.9
	Medway	Reasonable - some diagram numbers missing, and not possible to classify these further. “Regarding the sign data for Medway. Sign number 0 or NA or * is what we know as a sign with no reference. This is due to our internal processes in Medway. If a sign cannot be fitted into a particular sign ref number then it is given a null value. We will then have a photograph attached to the sign which allows officers see what exactly is there. An example of this is non primary directional signs.	12,908	46%	250.86

	Local Authority or Highway Authority	Notes on data quality/content	Total number of signs	% "unknown" sign type	Total area of Local Authority (sq kms)
		Regarding bus stop flags. We do not managed these in Highways so they are not included in the register which was provided to you. Footpath signs will be in the list but they are not categorised as "Footpath Directional Signs". Within our asset management system I have started sub categorising Footpath signs as "PROW directional" however this is not complete so not all footpath signs are categorised in this way.			
	Newcastle	Good	22,876	5.4%	115.05
	Norfolk	Good, but no data on illumination	85,832	8%	5358.3
	North Yorkshire	Good	93,339	5%	7830.56
	Redbridge	85% of signs covered. Good data. No evidence received that the missing signs are biased to any one type of road or area. No adjustment was made for the missing data in the data feeding the model.	14,372	4%	56.43
	Surrey (East)	Good - although many text descriptions rather than diagram numbers, and coordinates difficult to place on map of Surrey. Data had to be transposed to sit on map of Surrey, so approximations introduced.	39,628	0.2%	764.69
	Westminster	Good, except no data on whether or not a sign was illuminated.	16,539	1.0%	22.04
	West Sussex	Good	55,071	6%	2441.68
	York	Old data, and may not be 100% complete coverage. Unknown signs were all direction signs - allocated proportionately across different direction sign groups	13,722	0%	264.13
	Staffordshire	Good data, but for illuminated signs only. Used for validation.			

	Local Authority or Highway Authority	Notes on data quality/content	Total number of signs	% "unknown" sign type	Total area of Local Authority (sq kms)
	Portsmouth	Good data for non-illuminated. It includes both illuminated and non-illuminated signs. No diagram numbers for illuminated signs. Used for validation.			
	IoW	No diagram numbers available - high level sign descriptions provided. Rights of Way signs were not included in the surveys (survey related to Highway Assets only) and bus stop signs are the responsibility of Southern Vectis. Used for validation.	14,250		
	Lancashire	Summary data only. Not used for model or validation.			
	Plymouth	Total number of signs by category. Only covers A and B roads. Not used for model or validation.			
	Sheffield	No diagram numbers (or descriptors). Used for validation.	34,274		
	Cornwall	Summary data only - not used for model or validation. Includes pedestrian direction signs but not Bus Stops			
	Portsmouth	No diagram numbers (or descriptors) for illuminated signs - used for validation.			
	Lambeth	Illuminated signs only. No Eastings and Northings. Inventory covers all but 30km of Lambeth's roads (they are still completing) – equates to missing 10% of Lambeth's roads. Not used for model or validation.			
	Coventry	Full list of illuminated signs, but only some diagram numbers. Validation only.			
	East Riding	Illuminated signs only. No Eastings and Northings. Seemingly low number of signs for size of Local Authority. Not used for model or validation.			
	Barnsley	Location info only, for illuminated signs only. No diagram numbers. Validation only.			

	Local Authority or Highway Authority	Notes on data quality/content	Total number of signs	% "unknown" sign type	Total area of Local Authority (sq kms)
	Calderdale	Reasonable data only (text descriptors of signs rather than diagram number), but for illuminated signs only, and not all had Eastings and Northings. Query over what has been included as total number is high – also whether it include non traffic signs. Not used for model or validation.			
	Bradford	Location info only, for illuminated signs only. No diagram numbers. Validation only.			

Annex D: Middle Super Output Areas

D.1.1 Introduction

Middle Super Output Areas (MSOAs) were chosen as they possessed several useful features:

- They are well defined in geographical terms with easting/northing, boundaries and a centre.
- They are supported by the Office for National Statistics (ONS) with population projections.
- They have similar populations – although this means they vary in shape and area.
- They respect local authority boundaries.
- They were small enough to define distinct areas without oversimplifying the problem.

The remainder of this section provides a more detailed description of MSOAs and how they relate to Super Output Areas (SOAs) and Lower Layer Super Output Areas (LSOAs).

D.1.2 Super Output Areas (SOAs)

Super Output Areas (SOAs) are a geography designed for the collection and publication of small area statistics. They are used on the Neighbourhood Statistics site and across National Statistics.

There are currently two layers of SOA, with areas intermediate in size between census Output Areas (OAs) and local authorities and each layer nesting inside the layer above. This offers a choice of scale for the collection and publication of data, and allows for the release of local data that could be disclosive if published for OAs.

SOAs give an improved basis for comparison across the country because the units are more similar in size of population than, for example, electoral wards. They are also intended to be stable, enabling the improved comparison and monitoring of policy over time. In addition, figures for user defined geographies are aggregated and best fitted from data held for OAs and SOAs.

The link below gives more information on the SOAs.[1]

D.1.3 Lower Layer Super Output Areas (LSOAs)

The 33,323 Lower Layer SOAs (LSOAs) in England were generated automatically and released to the public in February 2004.

[1]http://neighbourhood.statistics.gov.uk/dissemination/Info.do?page=/aboutneighbourhood/geography/geography.htm

The LSOAs were built using 2001 Census data from groups of Output Areas (typically four to six) and were constrained by the Standard Table wards used for 2001 Census outputs. They had a minimum size of 1,000 residents and 400 households, but average 1,500 residents. Measures of proximity (to give a reasonably compact shape) and social homogeneity (to encourage areas of similar social background) were also included.

(Note: the specific homogeneity criteria used related to type of dwelling - e.g., detached/semi-detached etc. - and nature of tenure - e.g., owner-occupied, private rented etc.).

D.1.4 Middle Super Output Areas (MSOAs)

The Middle Layer SOAs (MSOAs) were defined in a two-stage process: an initial set was generated automatically but the boundaries were then modified in consultation with local authorities and other local bodies. The final boundaries were released to the public in August 2004.

The MSOAs used in this analysis were formulated with the 2001 census. Since the 2011 census, certain boundaries have changed, resulting in eight new MSOAs. However, this data was unavailable when the analysis was performed.

As with the LSOAs, initial Middle Layer SOAs were generated automatically by zone-design software. They were built using 2001 Census data from groups of Lower Layer SOAs and had a minimum size of 5,000 residents and 2,000 households. They also fitted within the boundaries of local authorities as at the end of 2002 (corresponding with the geography of the Census).

A nationwide consultation exercise gave local authorities the opportunity to amend the initial Middle Layer SOAs to define areas more suited to local requirements. The consultation resulted in 6,781 MSOAs in England (and 413 in Wales) with an average population size of 7,200.

Annex E: 1993 results

ROAD SIGNS: ENGLAND-WIDE ESTIMATES

SIGN CATEGORY	SIGN TYPE	SIGNS ON ALL ROADS	SUPPLEMENTARY SIGN %	% VARIABLILTY
WARNING	S1 Dist. stop/give way	11655	88	3
	S2 Junction warning	62427	4	3
	S3 Roundabout warning	12779	29	7
	S4 Bend warning	65275	13	2
	S5 Road narrows	22557	16	4
	S6 Two-way traffic	13324	1	6
	S7 Steep hill	4613	24	7
	S8 Bridge/tunnel warn.	2833	12	8
	S9 Misc. danger warning	12007	42	10
	S10 Cycle warning	809	18	17
	S11 Vehicle warning	488	6	25
	S12 Sharp deviation	36940	2	4
	S13 Roundabt. deviation	22961	1	8
	S14 Road merge	1743	15	38
	S15 Speed humps	4675	91	9
	S16 Slow lorries	103	76	208
	S17 Height warning	11486	39	7
	S18 Train warning	3143	51	10
	S19 Traff. signals ahead	5836	19	11
	S20 Pedestrian warning	73352	81	2
	S21 Animal warning	12478	10	5
	S22 Water warning	1568	19	11
REGULATORY	S23 Stop	6298	2	4
	S24 Give way	92449	5	2
	S25 Turning regulation	40739	35	5
	S26 Turning restriction	21028	5	12
	S27 Vehicle prohibition	11114	66	5
	S28 Restriction ends	4541	1	7
	S29 Speed limit	120730	0	3
	S30 No speed restriction	67626	0	3
	S31 Speed limit repeaters	104155	0	5
	S32 Minimum speed limit	37	0	760
	S33 Mini roundabout	8534	2	13
	S34 No entry	53354	2	5
	S35 No entry/except buses etc	978	0	12
	S36 No stopping	34289	61	10
	S37 No stopping repeater	9112	0	19
	S38 Bollards	131434	0	5
	S39 Waiting/loading rest	421948	0	4
	S40 Waiting limit	21842	0	7
	S41 Urban clearway	855	0	20
	S42 Vehicle regulations	2026	0	13
	S43 Clearway regulation	3444	0	14
	S44 Veh. weight limits	25083	80	7
	S45 Pedests/cycle route	5690	3	17
	S46 Veh. size prohib.	12654	58	6
	S47 Vehicle priority	1572	21	10

TABLE D1

Annex F: Analytical Assurance Statement

The level of assurance for the results of the analysis undertaken is low.

The In House Analytical Consultancy (IHAC) has undertaken an analysis of a subset of Local Authorities (Local Authorities), Transport for London TFL and Highways sign inventory data to estimate the number of traffic signs in England – both in total and for a number of specified groups. In areas where the number of signs were known (about 25% by area) these were related to explanatory variables such as population, road lengths/type etc within those specific areas. This enabled estimates to be made for those areas where sign data were not available.

The data had a number of limitations in terms of coverage and accuracy and these are fully identified in the report. Significant improvements in the quality and depth of coverage of the data would be needed, the responsibility of which lies with Local Authorities, TFL and Highways, in order to obtain more reliable estimates. A number of slightly different approaches were used and these gave similar answers. - that the total number of signs had roughly doubled from just under 2.2 million in 1993 to just over 4.5 million in 2013.

The work was undertaken by experienced and skilled staff with sufficient time and resources to analyse the data. Given the quality of the data however, the results should be treated as “ball park” figures.