



FAQ 7: Why Origins totals and percentages differs from ONS country of birth statistics

7 December 2016

Purpose of Information Note

When the numbers and percentages of names by Origins are compared with the numbers and percentages of UK names by country of birth, it is evident that there are some categories where they are quite similar, others where they are quite different.

The purpose of this note is to set out the principal reasons why these numbers and percentages differ.

Neither figure is necessarily the more accurate. There are some instances when it is more relevant to understand the size of particular communities irrespective of whether their members are first or second generation. On other occasions it is more relevant to adhere to the actual number of people born abroad.

Comparing ONS and Origins at the national level

Origins was conceived originally to enable users to infer the cultural background of people whose names appear on customer and client files and, more particularly, to compare the percentage of names on a file that are associated with particular cultural backgrounds with the overall percentage of names in the adult population as a whole.

Subsequently the classification system has been used to examine the population mix of geographical areas, whether postcodes, wards or local authority areas and how this has changed since the time of the 2011 census.

The main difference between the census and Origins is that the census, among other things, measures where people were born. Origins, by contrast, tends to identify the origins of a person's parents or grandparents. Origins therefore tends to give much larger counts for minorities which have been settled in Britain for more than one generation.

Table one, below, uses ONS estimates for 2016 to compare the distribution of the population by country of birth with distribution of adults by Origins.

Comparison of ONS estimates of country of birth and Origins codes				
Country of birth	2016 ONS estimates (% of UK population)	2016 Origins (% of UK adults)	Census estimate as % of Origins estimate	Likely reason for variance
Republic of Ireland	0.59	6.15	9.7	Immigration mostly pre-1945
India + Sri Lanka + Mauritius + Nepal	1.60	1.99	80.6	Second generation often born in Britain; Origins includes East African Asians
Pakistan + Afghanistan	0.90	1.60	56.3	Second generation often born in Britain
Poland	1.29	1.22	106.0	
Italy + Malta	0.30	1.22	24.3	Immigration mostly pre-1945
Germany	0.45	0.97	45.9	Immigration mostly pre-1945
Spain + Philippines	0.40	0.66	60.6	
France	0.24	0.63	37.8	Immigration mostly pre-1945
Nigeria	0.31	0.59	52.5	Origins may have included some other Black Africans
Bangladesh	0.34	0.47	71.8	Second generation often born in Britain
Jamaica	0.21	0.45	47.4	Origins figure upweighted; Origins includes rest of West Indies
Portugal + Brazil	0.30	0.44	68.2	
China + Taiwan + Singapore	0.45	0.43	103.5	
Netherlands + Flemish Belgium	0.15	0.41	35.7	Second generation often born in Britain
Turkey	0.11	0.39	28.7	Many born in Germany or UK
Russia	0.07	0.35	20.5	Immigration mostly pre-1945
Cyprus + Greece	0.18	0.32	56.9	Early immigrants
Lithuania + Latvia	0.38	0.29	131.4	Mostly adult population
Czech Republic + Slovakia	0.21	0.27	77.0	
Ghana	0.17	0.27	62.2	
Romania	0.34	0.26	131.7	Mostly adult population
Iran	0.13	0.26	51.5	Early immigrants
South Africa	0.31	0.23	135.3	
Hungary	0.12	0.21	59.3	Early immigrants
Somalia	0.18	0.13	136.5	Origins may include some other Muslim names
Vietnam + Thailand	0.11	0.09	118.9	
Bulgaria	0.11	0.07	153.4	Mostly adult population
Albania	0.04	0.07	62.2	Origins figure includes ethnic Albanians born elsewhere in the Balkans
Japan	0.07	0.05	133.8	Mostly adult population

Table one: comparison of 2016 country of birth with 2016 Origins codes

To clarify the reasons for some of these differences the following tables (which use census statistics from 2011 and for England and Wales only) highlight some of the circumstance in which the two sources of information come up with different distributions.

Reason 1: 2011 vs 2016

The Origins data relates to May 2016, the census statistics to April 2011. During the intervening five years there has been an increase in the number and percentage of non white-British names and, within that group, particularly high increases (up to 30%) of names from Romania and Bulgaria, from the Baltic States and from Black Southern Africa.

Reason 2: Adults vs Total Population

The Origins data collects information on adults only, the census on all persons present. Because the ratio of children to adults is higher among non-white populations than among the population of white British and white European origin, the Origins figures, being based on adults only, tend to underestimate the percentage of the total population that is of non-white British origin.

Reason 3: First generation vs second and subsequent generation migrants

Origins tends to give higher numbers and percentages for countries from which migrants came to Britain many generations ago. Ireland is a good example. Origins gives a percentage of (England and Wales) adults with Irish family names nine times as large as the census figure for the percentage of people born in Ireland. Less dramatic differences occur for similar reasons for most western European countries, for Black Caribbeans, Jews, Turks and Greek Cypriots, in other words the minorities that by now are relatively well established in Britain.

Total: Country of birth	2011 census: % by country of birth	2016 % by Origins
Europe: Ireland	0.73	6.60
Europe: Italy	0.24	0.94
Europe: Turkey	0.16	0.39
Europe: Cyprus, Greece	0.20	0.32
Middle East and Asia: India and Sri Lanka	1.47	2.74

Reason 4: Country of birth vs religion

Though most names are indicative of country of origin, there are many which are indicative of religion. This is why some of the Origins categories are described on the basis of religious rather than national identity. Neither Sikhs and Jews can be unambiguously associated with a single country of origin. The 2011 census would assign most of the population identified by Origins as Sikh to “India” or “Pakistan” as their country of birth.

Total: Country of birth	2011 census: % by country of birth	2016 % by Origins
Middle East and Asia: India and Sri Lanka	1.47	1.99
Sikh	0.00	0.75

Reason 5: Geographical vs cultural groupings

There are good reasons why the census and Origins should group countries into larger regions in slightly different ways. In Africa for instance, the Origins classification groups Somalia, Morocco, Algeria, Tunisia, Libya and Egypt as part of the “Muslim World” for the reason that their inhabitants tend to have been given names from Arab or Islamic culture. The rest of Africa is grouped together into a sub-Saharan “Black Africa” group that shares very different names. The census, for good reason, avoids what may appear to be a contestable definition by equating its definition of “Africa” with its continental land mass.

Likewise whilst Origins groups Iberian, South American and Central American names into a larger “Hispanic” grouping, the census makes separate provision for the populations born in Europe (Spain and Portugal) and South and Central America.

In South Asia, Origins groups Tamil names irrespective of whether they originate from India or Sri Lanka. The census partitions the same population by country of birth.

These are just some of the examples of the difference between a grouping based on geography and one based on culture.

Reason 6: Adoption of British names

The Jewish and the Black Caribbean populations are two that Origins tends to underestimate. In the case of the Jewish population, many have adopted British family and personal names. Those that tend not to are mostly found among the most Orthodox members of the Jewish community. Origins tends therefore to underestimate the more assimilated members of the Jewish community. The under-recording of the Black Caribbean community is consistent and has been addressed by upweighting the size of that community by a factor of four.

Reason 7: The Anglo-Saxon diaspora

Not all immigrants from the United States, Canada, Australia and New Zealand bear white British names but the majority do. Origins does not distinguish between these immigrants and the host white British population.

Total: Country of birth	2011 census: % by country of birth	2016 % by Origins
The Americas and the Caribbean: United States, Canada	0.44	Not applicable
Antarctica and Oceania: Australia, New Zealand	0.31	Not applicable

Reason 8: Multiple migrations

East African Asians are just one example of a group which has experienced more than one set of migrations. A very significant proportion of Hindu Asians living in Leicester will, as a result, be recorded by the 2011 census as having been born in Africa despite their culture being Indian.

Summary of inconsistencies due to inability to associate names with Origins codes correctly

Overall Origins identifies nearly 250 different codes which are grouped into around 40 “sub-groups”. Taking these seven reasons into account, we believe that at the sub-group level there is a relatively good correspondence between the number and percentage of adults in a categories and the number and percentage of adults with particular backgrounds.

One of the exceptions relates to the Jewish and Black Caribbean populations which, as we have explained in reason six, are underestimated by virtue of having adopted British names.

There are some other categories which are subject to some error as follows:

- Philippines: Origins under-records the number of Filipino names, quite a few of which are recorded as Hispanic.
- Black Africa: Other than Ghana, Nigeria and South Africa, Origins tends to under-record the number of Black African names. Most not found in the Origins master reference files tend to be from other areas of Black Africa.
- Somalia: Origins somewhat under-estimates the number of Somali names due to overlap with other Muslim names.
- Pakistan: Due to the overlap between Pakistani and other Muslim names, Origins tends to under-estimate the number of Pakistani names.
- Sri Lanka: Reason five explains that many immigrants from Sri Lanka are categorised as Tamil rather than Sri Lankan. Britain’s non-Tamil Sri Lankan population is probably under-recorded.

As a general rule users should not expect numbers and percentages to be accurate below the level of the sub-group. For example though Origins contains codes for Laos, Iraq and Fiji, Origins does not claim to have had an exhaustive list of Lao, Iraqi or Fiji names. Where these finer categories are used they should be used for profiling, where distributions can be compared with a national base, rather than for the estimation of size.