

# Name Clustering on the Basis of Parental Preferences

GERRIT BLOOTHOOFT and LOEK GROOT

*Utrecht University, The Netherlands*

Parents do not choose first names for their children at random. Using two large datasets, for the UK and the Netherlands, covering the names of children born in the same family over a period of two decades, this paper seeks to identify clusters of names entirely inferred from common parental naming preferences. These name groups can be considered as coherent sets of names that have a high probability to be found in the same family. Operational measures for the statistical association between names and clusters are developed, as well as a two-stage clustering technique. The name groups are subsequently merged into a limited set of grand clusters. The results show that clusters emerge with cultural, linguistic, or ethnic parental backgrounds, but also along characteristics inherent in names, such as clusters of names after flowers and gems for girls, abbreviated names for boys, or names ending in *-y* or *-ie*.

## Introduction

The variety in personal given names has increased enormously over the past century. In the Netherlands, the top 3, top 10, and top 100 names account, respectively, for 16%, 33%, and 70% of the first names of elderly born between 1910 and 1930, while these figures are 3%, 8%, and 39% for babies born between 2000 and 2004. Comparable figures are presented by Galbi (2002, 4) for England and Wales. Along with the increase in the variety in names, the motives behind the choice of names for children by their parents have changed from a more or less prescribed naming after relatives to a free decision, a process that was facilitated in the Netherlands by the tolerant name law of 1970. This does not mean, however, that naming norms are absent in the naming practice. As Tucker (2003, xxvii) has shown, many forenames are still indicative of cultural, ethnic or linguistic (CEL) groups within a population.

In the last two decades we saw a proliferation of statistical analyses of name datasets, especially with respect to ethnicity classification. This is largely due to the availability of large digital name datasets, an increasing need for reliable data on the size and geographical distribution of ethnic minorities in countries to monitor

inequalities between ethnic groups in health status, educational achievements, and job careers, and a growing interest in international migration patterns.<sup>1</sup> Mateos (2007) provides an excellent review of the name-based ethnicity classification literature, mainly limited to studies which meet certain accuracy standards and evaluate their classification results against a non-name-based ethnicity information source. As Mateos notes, the fundamental trade-off in these classifications is between maximizing coverage and maximizing accuracy: the more disaggregated the ethnic groups are defined to maximize coverage, the more misclassifications (so-called false positives) result, compromising the accuracy of the classification.

Tucker (2003) has developed a technique to classify surnames in Cultural-Ethnic-Linguistic groups, and used the results to compile a *Dictionary of American Family Names* (DAFN). The aim of DAFN is to maximize for people the chance to find their surname as an entry in the dictionary. His method is as follows. First, using a large dataset of 89 million telephone subscribers, he showed that about 4% (slightly above 70 thousand) of all 1.75 million different surnames covers 85% of the total population. Second, a team of experts were set to the task of classifying the 70 thousand surnames into 23 pre-defined CEL groups. For a large fraction (20%), they were unable to assess with any confidence even the language of origin. Third, a statistical analysis on forename-surname correlation was performed. Specific forenames can be highly diagnostic for particular CEL groups if they are rarely used outside a particular CEL group (examples given are *Niamh* for Irish and *Giuseppe* for Italian). A limited set of 8000 highly diagnostic forenames were manually classified into CEL groups by onomastic experts. Fourth, using this diagnostic list a forename-cluster analysis was performed to merge forenames belonging to the same group. Finally, for each surname the distribution of forename CEL scores is given in the dictionary. Using this method, Tucker reduced the percentage of unidentified surnames from 20 to only 3%.

Mateos, Webber, and Longley (2007) also rely on what they call the CEL-triage technique, supplemented by other information provided by other techniques such as spatio-temporal analysis, geodemographics analysis and text mining (see sections 3.2–3.4) to subdivide the entire UK population and to classify every surname and forename with a frequency of 3 or higher in Britain in 2004 into 185 CEL-types (a subdivision of 15 CEL groups).

Harding et al. (1999, 48) use the *Nam Pelchan* South Asian Names dictionary as a reference list to estimate the size of the Sikhs, Moslem, and Hindu population in Bradford, UK. Lauerdale and Kestenbaum (2000) identify six major Asian-American ethnic groups by combining a name dataset which included country of birth and a name dataset with information on race (white, black, other) for persons 60 years or older. Each surname has a score for the proportion with the associated Asian country of birth (e.g. 80% Vietnamese) and for the proportion with race 'other'. Only surnames with a sufficiently high score on the product of both scores are considered as sufficiently predictive to identify persons with these names as belonging to one of the six categories Asian-American groups.

The limitation of the previously mentioned studies<sup>2</sup> is that some pre-defined classification is needed — largely based on language, origin and religion — in

combination with expert knowledge. Furthermore, the sizes of the classes can differ considerably. In the Mateos (2007) study, out of the 46 million British people classified, 31 million came into the CEL type England, and another 10 million in the CEL types Scotland, Wales, and Ireland. Whereas the identification of the other, much smaller CEL groups related to ethnic minorities is certainly very valuable, a further subdivision of the massive CEL groups may give additional insights, particularly in the relation between socioeconomic factors and naming. In order to make this possible, and to circumvent a lack of data that could define additional factors, we took an entirely different approach. Instead of the largely top-down methods discussed before, we adopted a bottom-up method by studying the naming preferences of parents. In this approach, the assumption is that parents tend to give names to their children on the basis of preferences that are influenced by their social group (see also Fryer and Levitt (2004) for an analysis of differences in naming patterns between blacks and whites and of the growth of distinctively black names following the Black Power movement in the early 1970s). Given a sufficient number of parents that share these preferences, we can identify names as belonging to what we label as name groups. These name groups can resemble CEL groups or CEL types, but they may also show a much finer structure. A name group could consist of, for example, Frisian names, but also of girls' names after flowers or gems, or abbreviated names.

To identify these name groups, we will exploit the statistical information in name corpora, containing the names of children born in the same family. We have access to a full sample of names for children born in between 1982 and 2005 in the Netherlands and a 40% draw of all children born in between 1982 and 2002 in the UK. By exploiting the conditional probability that a combination of names can be found for children within the same family, it will turn out that some names have a stronger association to each other than to other names. This is the basis for their clustering in name groups. A major part of this paper is devoted to the design of an appropriate statistical method for their identification. An outline of the method has been presented already in Bloothoof (2001, 2002), but is now given a new and solid mathematical foundation. The quintessence of a truly bottom-up method is that, besides some setting of parameters, no additional information is being used.

This paper is structured as follows. In section 2 we shortly describe our name databases. Sections 3 and 4 comprise the methodological part. Section 3 explains how the phenomenon that some names are strongly connected to each other can be expressed in terms of conditional probabilities, defined as the likelihood that a younger sibling of a child with name  $i$  has name  $j$ . In section 4, the conditional probabilities are used to cluster names into name groups. The purpose of the cluster process is to identify sets of names that have a high probability to be found in the same family. We will demonstrate that it is beneficial to make a distinction between a cluster level that shows fine details and a higher level of grand clusters that summarizes the major features of the fine clusters. The information processed in the derivation of grand clusters covers the whole continuum in the naming practice of parents: they can choose names for all of their children from a single fine cluster, from one or more related clusters, but also definitely *not* from certain other clusters. A paradigmatic example of the latter is that names from the Western and Arabic clusters are virtually never to be found in the same family. We conclude with an annotated

presentation of all grand and fine name clusters for both the Netherlands and the UK.

## Name databases

The Dutch Social Security Bank (SVB) made available to us the initial first name, gender and year of birth of all children born in the Netherlands between 1982 and 2005. The SVB draws these data directly from the Civil Registration. In addition to the names, we also got a code by which children in the same family could be identified. The corpus consists of the initial first name of 4.65 million children, of which 3.54 million were born in 1.46 million families with more than one child, which is a condition for our further analysis.

The same type of data was received from the HM Revenue and Customs in the UK. The sample includes the initial first name of children born in between 1982 and 2002. For privacy reasons, names with a frequency less than 60 were removed from the full sample, as were their siblings. Subsequently, a random draw of 40% was performed. This corpus includes the initial name of 4.46 million children, born in 1.80 million families.

We believe that both corpora are exquisite data sets to investigate whether, and which, name groups exists. There are however some differences between both corpora. Whereas the size of the Dutch and UK corpora are similar, the number of different names in the corpora differs considerably. In the Netherlands there are 68,230 different names for boys and 84,354 different names for girls, while for the UK corpus this is 26,253 and 35,293 respectively. This difference is a consequence of deleting names with a frequency less than sixty and does not necessarily imply that naming in the Netherlands is more varied than in the UK. Other related differences in the distributions of our corpora are that popular names in the UK cover a higher percentage of all children, while the number of low-frequent and unique names is lower than in the Netherlands.

## Name pairs

We assume that parents do not name their children in a random way. This implies that the name of the older child can be of predictive value for the choice of the names of subsequent children. We express this relationship by the conditional probability  $P(\text{name}_i | \text{name}_j)$ . For easier interpretation, we will use the example names *John* and *Mary* throughout this paper. The conditional probability  $P(\text{John} | \text{Mary})$  presents the likelihood that a younger brother of *Mary* will be named *John*. If this likelihood is high, it demonstrates a close relationship between the two names.  $P(\text{John} | \text{Mary})$  is calculated by selecting all families with a girl named *Mary*, count all occurrences of a younger brother *John* ( $N_{\text{John} | \text{Mary}}$ ) and divide this by the number of all younger brothers of *Mary* until and including a boy named *John* ( $N_{\text{YoungerBrothers} | \text{Mary}}$ ). Thus

$$P(\text{John} | \text{Mary}) = \frac{N_{\text{John} | \text{Mary}}}{N_{\text{YoungerBrothers} | \text{Mary}}} \quad (1)$$

It is not known how many younger brothers of Mary will be born after the end of our data range. However, we think that this uncertainty only has a small effect given our time span of over twenty years. In the initial years of our corpora only starting families were included.

In Table 1 we present the top ten probabilities of a name for a brother or sister of *Maria* in the Netherlands, and of *Mary* in the UK. The top ten covers about 20% of the names of all younger brothers and sisters of *Maria* and *Mary*, while this is 40% for the top ten of brothers of *Mary*. The names are of a rather traditional type and we may conclude somewhat prematurely that when parents choose the traditional name *Maria* or *Mary* for a daughter, they are likely to choose a traditional name for other children as well. It is a first indication that knowledge of names of children in a family conveys interesting information on parents' naming preferences. Note that, taking *Mary* as the English equivalent of the Dutch *Maria* and *John* of *Jan*, Table 1

TABLE 1

THE TOP TEN NAMES OF YOUNGER BROTHERS AND SISTERS OF *MARIA* IN THE NETHERLANDS (BASED ON 16,347 *MARIA*'S WITH 9201 YOUNGER BROTHERS AND 8471 YOUNGER SISTERS) AND OF *MARY* IN THE UK (BASED ON 2878 *MARY*'S WITH 1685 YOUNGER BROTHERS AND 1579 YOUNGER SISTERS)

Probabilities are expressed as percentages

<i>Maria</i> (NL)		<i>Mary</i> (UK)	
brothers	%	brothers	%
Johannes	6.17	John	7.91
Cornelis	3.21	Michael	6.12
Jan	2.51	James	5.32
Petrus	1.85	David	3.47
Willem	1.70	Patrick	2.88
Hendrik	1.65	William	2.56
Pieter	1.73	Peter	2.34
Marinus	1.49	Martin	1.85
Gerrit	1.47	Robert	1.85
Martinus	1.20	George	1.65
sisters	%	sisters	%
Johanna	5.11	Sarah	4.60
Anna	3.35	Elizabeth	2.85
Cornelia	2.52	Alice	2.49
Elisabeth	2.21	Catherine	2.18
Catharina	1.73	Anna	1.74
Adriana	1.52	Margaret	1.63
Wilhelmina	1.05	Kathleen	1.19
Petronella	0.94	Ruth	0.97
Hendrika	0.85	Ann	0.76
Jacoba	0.83	Frances	0.65

exhibits some more equivalents. Among the brothers we find (*Johannes, Jan, Willem, Petrus, Pieter, and Martinus*) for *Maria*, corresponding to (*John, William, Peter, and Martin*) for *Mary*. The same goes for the sisters, with the Dutch set (*Anna, Elisabeth, and Catharina*) corresponding to (*Ann, Anna, Elizabeth, and Catherine*).

For the full data set, the highest probabilities were found in the UK for *Mohammad* with a brother *Mohammed* (24.6%), *Tom* with a brother *Jack* (14.2%), *Shazia* with a sister *Nazia* (12.9%), and *Lowri* with a sister *Ffion* (12.1%). For the Netherlands it is *Fatima* with a brother *Mohamed* (16.6%), *Yasin* with a sister *Yasemin* (14.5%), *Björn* with a brother *Sven* (11.7%) and *Yunus* with a brother *Emre* (11.7%).

The mirror image of the likelihood that *Mary* gets a younger brother *John*, is the likelihood that *John* will get a younger sister *Mary*. This probability  $P(\text{Mary}|\text{John})$  is usually not the same as  $P(\text{John}|\text{Mary})$ . Actual data for *John* and *Mary* can illustrate this. It turns out that  $P(\text{John}|\text{Mary}) = 7.91\%$  and  $P(\text{Mary}|\text{John}) = 0.91\%$ , a difference that originates in the much higher popularity of *John* (21,740 *John*'s) compared to *Mary* (2878 *Mary*'s). The theoretical relationship between both conditional probabilities is expressed as

$$P(\text{John}|\text{Mary}) = \frac{P(\text{Mary}|\text{John}) \cdot P(\text{John})}{P(\text{Mary})} \quad (2)$$

Since we wish to express the attraction between two names by a single measure, the dependency of conditional probabilities on the popularity of a name poses a problem. However, from (2) it immediately follows that

$$\frac{P(\text{John}|\text{Mary})}{P(\text{John})} = \frac{P(\text{Mary}|\text{John})}{P(\text{Mary})} \quad (3)$$

and we could use this value as an expression of the attraction  $A(\text{John}, \text{Mary})$  between two names, that is independent of their order and individual popularity. The measure tells us how more often *John* is chosen as the name of a brother of *Mary* than as a name for a boy in general, which equals how more often *Mary* is chosen as a name for a sister of *John* than for a girl in general.

Although theoretically attractive, this approach still does not work well in practice, which can be illustrated by an example. Suppose a population has two distinct religious groups, Christians and Muslims, of which there are nine times as many Christians as Muslims. Members of both groups only choose typical names from their own religion. And although *Mary* and *Fatima*, and *John* and *Mohammed* may be equally popular within their groups, for the whole population the frequencies of *John* and *Mary* will be by nine times higher than those for *Mohammed* and *Fatima*, respectively. This implies that  $P(\text{John}|\text{Mary}) = P(\text{Mohammed}|\text{Fatima})$  and  $P(\text{Mary}|\text{John}) = P(\text{Fatima}|\text{Mohammed})$ , but  $P(\text{Mohammed}|\text{Fatima}) / P(\text{Mohammed}) = 9 * P(\text{John}|\text{Mary}) / P(\text{John})$ . In other words, the attraction between two names is dependent on the size of the subgroup in which they are popular. Normally, we do not know this subgroup nor its size beforehand. Otherwise we could define a corrected attraction measure  $A_c$  such that

$$A_c(\text{John}, \text{Mary}) = \frac{P(\text{John}|\text{Mary})}{P'(\text{John})} = \frac{P(\text{Mary}|\text{John})}{P'(\text{Mary})} \quad (4)$$

where  $P'(John)$  and  $P'(Mary)$  are the probabilities of *John* and *Mary* relative to the size of their subgroup. However, using equation (3), we can estimate these probabilities by considering that for all names related to some subgroup the following conditional relationships hold

$$\begin{aligned} P'(Mary) &= P'(John) P(Mary|John) / P(John|Mary) \\ P'(William) &= P'(John) P(William|John) / P(John|William) \\ P'(Elizabeth) &= P'(John) P(Elizabeth|John) / P(John|Elizabeth) \\ &\dots \\ P'(John) &= P'(John) \end{aligned}$$

The sum of the probabilities on the left-hand side is 1 for all boys and 1 for all girls since we apply gender specific probabilities. By taking the sum at both sides we arrive at

$$P'(John) = \frac{2}{1 + \sum \frac{P(name_i | John)}{P(John | name_i)}} \quad (5)$$

where the sum is taken over all known brothers and sisters of *John*. The relative size  $R$  of the subgroup is  $P(John) / P'(John)$ . In our example this value would be 0.9 for Christian names and 0.1 for the Muslim names. In that case,  $A_c(John, Mary)$  equals  $A_c(Mohamed, Fatima)$ .

Although theoretically  $A_c(John, Mary)$  should be exactly the same as  $A_c(Mary, John)$ , in practice this is not the case because of inaccuracies in the estimation of the relative size of their subgroup. Therefore we take the average of both values as the final estimate of the attraction between the names *John* and *Mary*. If the estimated relative size of the subgroup for the two names differs more than a factor 3, possibly because the names reside in two different subgroups, we exclude the pair.

Since children with names of low popularity have few brothers and sisters, we will arrive at poor estimates of the conditional and individual probabilities of their names, which would seriously hamper further analysis. To avoid this, we required that for any name there should be at least in total 100 younger brothers to consider for a pair with a male name, and at least 100 younger sisters for a pair with a female name. One more brother or sister with some name then roughly increases the conditional probability by maximally 0.5%, which we considered acceptable given observed probabilities up to 25%.

Finally, we neglect name pairs for which  $A_c$  is less than one, which implies that a name would have a lower likelihood to be found in that pair than in general. These quite severe restrictions result for the Dutch data in 24,435 name pairs from 1409 names, and for the UK data in 30,815 name pairs from 912 names. This is much less than the maximum number of 2 million and 0.8 million pairs<sup>3</sup> that could be formed from those names in the Netherlands and UK respectively, which suggest severe limitations on the possible pairs. It suggests that clustering of names is viable. The number of 1409 and 912 names that fulfill our conditions is only around 1% of the total number of different names. However, because these names are highly frequent, their coverage of the total number of children is 75% (NL) and 87% (UK). For the

UK, the share of children covered in the full population would probably be lower than 87%, since our sample does not include names with a frequency less than 60.

Using equation (4), we can now list the name pairs that have the highest attractions, to see whether our results are plausible and already show some typical features. The top twenty name pairs with the highest attraction scores (limited to highly popular names with a frequency over 10,000) is presented in Table 2.

For the Dutch top twenty, it can be seen that *Lars* scores highest with *Niels*. The attraction tells us that the likelihood of finding a brother *Niels* with *Lars* is 4.59 times higher than the probability of finding the name *Lars* in general (within the group of parents that could consider Nordic names). There are also combinations of *Niels* or *Lars* with other Nordic names like *Sven*, *Jesper*, *Bjorn*, and *Jorn*, but these combinations fall outside the top twenty range. Apparently these parents prefer Nordic names for their children. From Table 2 we observe in the Dutch data already some likely clusters of three or more names, such as (*Martijn*, *Jeroen*, *Sander*, *Jasper*), (*Bas*, *Tom*, *Tim*, *Bart*, *Daan*, *Koen*), (*Mike*, *Nick*, *Roy*, *Kim*), and (*Maria*, *Johannes*, *Johanna*).

TABLE 2  
THE 20 PAIRS OF POPULAR NAMES WITH HIGHEST ATTRACTION, BOTH IN THE NETHERLANDS  
(1982–2005) AND THE UK (1982–2002)

The frequency of each name is higher than 10,000			
NL		UK	
name pair	attraction	name pair	attraction
Lars, Niels	4.59	Ben, Sam	8.12
Martijn, Jeroen	4.52	Edward, William	5.56
Bas, Tom	4.38	George, Harry	5.04
Maarten, Wouter	3.95	Elizabeth, Katherine	4.58
Martijn, Sander	3.46	Ross, Scott	4.64
Bas, Tim	3.25	Elizabeth, Catherine	4.60
Mike, Roy	3.22	Samuel, Benjamin	3.61
Daan, Koen	3.14	Samuel, Joseph	3.55
Mike, Nick	3.03	Mark, Paul	3.46
David, Ruben	3.01	Elizabeth, Victoria	3.34
Bram, Daan	3.00	Edward, George	3.28
Martijn, Jasper	2.91	Eleanor, George	3.23
Mark, Linda	2.88	Dean, Lee	3.19
Johannes, Maria	2.81	Alice, Emily	3.09
Roy, Kim	2.78	Alice, Edward	2.97
Bart, Koen	2.75	Jennifer, Katherine	2.95
Bart, Tom	2.74	Elizabeth, Edward	2.93
Jeffrey, Wesley	2.72	Craig, Scott	2.93
Patrick, Chantal	2.68	George, William	2.88
Johanna, Maria	2.67	Callum, Connor	2.80



No cluster-crossing combination like (*Mike, Lars*) is seen. This suggests that it may be possible to cluster names into groups based on parental preferences. The sets also immediately evoke associations to original language, length of the names (notably in very short names like *Bas* and *Tom*), and the time they were most popular (the traditional names as we have already seen with *Maria*, but also in the set with *Martijn* which names have passed their peak several years ago).

For the UK top twenty highly popular name pairs, the same type of observations can be made. Possible sets are (*Edward, William, George, Harry, Victoria, Elizabeth, Eleanor, Alice, Emily, Katherine, Catherine*), including quite a few royal names, the Scottish names (*Scott, Ross, Craig*) and the Hebrew names (*Samuel, Benjamin, Joseph*). Note that the attraction values for the UK do not differ considerably from those for the Netherlands.

If we put no limitation on the frequencies of names (other than set by our analysis method) the top twenty of name pairs is different and shown in Table 3.

The names in Table 3 are less common, but form very plausible pairs. For the Dutch top twenty, only *Marjolein, Evelien, Annemiek*, and *Carolien* are a set, but for

TABLE 3  
THE 20 PAIRS OF NAMES WITH HIGHEST ATTRACTION, BOTH IN THE NETHERLANDS (1983–2005) AND THE UK (1982–2002)

NL		UK	
name pair	attraction	name pair	attraction
Oscar, Victor	9.50	Ffion, Lowri	34.25
Gijs, Teun	7.86	Nia, Aled	32.18
Allard, Ewoud	7.69	Aoife, Eoin	29.63
Noud, Ward	7.37	Ffion, Nia	29.00
Jill, Lynn	7.18	Bethan, Rhian	25.76
Jildou, Marrit	7.15	Jimmy, Tommy	24.97
Evelien, Marjolein	7.06	Lowri, Nia	24.77
Auke, Sietse	6.93	Aisling, Roisin	23.93
Carolien, Marjolein	6.91	Lowri, Tomos	23.11
Caitlin, Megan	6.91	Niamh, Orla	22.78
Joram, Tamar	6.86	Lowri, Aled	21.87
Björn, Sven	6.79	Dafydd, Sion	20.86
Eric, Marc	6.68	Cerys, Rhian	20.69
Jet, Pien	6.68	Ffion, Tomos	20.67
Esther, Judith	6.67	Roisin, Sinead	20.10
Lynn, Tess	6.60	Ceri, Nia	20.01
Jip, Puck	6.59	Albert, Arthur	19.90
Annemiek, Evelien	6.48	Eoin, Niall	19.84
Dave, Mike	6.25	Ciara, Orla	19.80
Gideon, Jonathan	6.25	Aine, Aiofe	19.45

the UK top twenty there are Welsh names in (*Ffion, Lowri, Nia, Aled, Tomos, Ceri*) and (*Rhian, Bethan, Cerys*), and Irish names in (*Aoife, Eoin, Niall, Aine*) and (*Aisling, Roisin, Sinead*). While the attraction of the Dutch pairs is only slightly higher than that for the popular name pairs presented in Table 2, for the UK name pairs the attraction is more than three times higher (also in comparison to the highest Dutch attraction scores). This may originate in an underestimation of the size of the subgroup of parents that may choose for such a name in the UK, as we did using formula (5). Alternatively, it might be that in the UK there are more indigenous names. For the Netherlands, highest values for the subgroup size were obtained for *Laura* and *Mark* with 54% and 51% of all parents. For the UK these were *James* and *Emma* with even 87% and 84%. The popular names with highest attraction (Table 2) typically relate to 30–50% of all parents for the Netherlands and 45–65% for the UK. The names in Table 3 belong to a smaller subgroup of parents, typically between 10–20% in both countries. However, this may be still a too high estimate in the UK case.

## Clustering of names

The aim of this section is to identify name groups. Obviously, to combine names into name groups based on the values of their mutual attraction cannot be done by hand. The purpose of our clustering method is to separate groups of comparable names from others, so that names within one cluster are more similar than names of different clusters. Cluster analysis covers a wide array of statistical techniques used to group objects in homogeneous sub-groups on the basis of similarity (see Everitt et al., 2001). In principle, by using a clustering technique choices have to be made. The similarity between objects can be measured by the distance such as the squared Euclidean distance, by the correlation between objects or still another (dis)similarity criterion. In addition, the cluster method can be hierarchical or non-hierarchical, where a hierarchical method makes combinations in successive rounds (objects combined in the first rounds are more closely related than objects combined in subsequent rounds), whereas a non-hierarchical method is mostly an iterative technique, revising divisions until an optimum is reached. Finally, a clustering algorithm has to be chosen. A cluster algorithm can be seen as an amalgamation rule which determines when two clusters are sufficiently close to be combined. At the start, each object is considered its own cluster, but in successive rounds, clusters are formed. One such algorithm that we will use in the first step of our analysis (see below) is known as Single-linkage or Nearest neighbour, which combines two clusters when any two objects in the two clusters are closer together than to any other object not in these two clusters. To make the picture complete, a fundamental and unresolved problem in cluster analysis is that there are no rules for the choice of the optimal number of clusters. In principle, the number of clusters can vary from 1 to the total number of objects.

In what follows, we will explain our preferred cluster technique, motivated by both practical and conceptual considerations. At the practical level, we have to cope with the problem that a large number of objects (1409 names for the Netherlands, 912 for

the UK) have to be entered in the clustering process, which rules out any standard cluster method due to computational constraints. As said, by concentrating on the highest observed attractions, in the first step we use a variant of the Single-linkage algorithm to arrive at an initial clustering, where the number of clusters is determined endogenously. After this first step, the availability of the attraction scores between two names also allow us to compute the attraction between a name and an initial cluster, or the attraction between two clusters.

### ***Initial clustering***

The first step in our procedure is to organize an initial, self-organizing clustering on the basis of descending attractions. For this, we order all name pairs according to their attraction from high to low, while for each pair — starting with the pair with highest attraction — we take the following decisions:

1. If both names are not yet assigned to a cluster, they constitute a new cluster
2. If one of the names has been assigned already to one cluster and the other name is new, the latter is assigned to the same cluster
3. If both names already were assigned to some cluster, either the same or different ones, no further action is taken.

This procedure results in a moderate clustering of names. For the Dutch data the 1409 names are combined in 302 clusters, while for the UK data the 912 names are combined in 160 clusters. These initial results are not yet optimal, however, since the start of a new cluster is very much dependent on the (accidental) order of attractions. That is, if two pairs of names (*John, Mary*) and (*William, Albert*) exist, each with high attraction, while the bridging attractions (*John, William*), (*John, Albert*), (*Mary, William*) or (*Mary, Albert*) do not reach that high, initially two clusters will be generated, one around *John* and *Mary* and one around *William* and *Albert*. But if all information could be taken together, it may be that a single cluster of these names would provide a better description of the data.

### ***Initial clusters reconsidered***

In the second phase we reconsider the initial clusters. For this, we focus on the attraction between a cluster and a target name. This attraction tells us how more likely it is to find for some target name the names for brothers and sisters in a cluster, than to find them in general in the population. With this knowledge we can find the cluster that has the highest attraction to some target name. That may be the cluster the name is already in, but it may be another cluster as well. If the latter is the case, we reassign the name to that cluster.

The attraction between a cluster and some target name is simply the sum of the attractions of all names in the cluster and the target name. As an example, if the name under consideration is *Mary*, and *John* and *William* are in one cluster already, we sum  $A_c(\textit{Mary}, \textit{William})$  and  $A_c(\textit{Mary}, \textit{John})$  as the attraction of the cluster (*John, William*) to *Mary*. Or, in general terms, the attraction  $A$  of cluster  $K$  to the name *Mary* is

$$A(\textit{Mary}, \textit{cluster}_K) = \sum_{i \in K} A_c(\textit{Mary}, \textit{name}_i) \quad (6)$$

After having reassigned the names to the clusters which exert the highest attraction — when necessary — their distribution over clusters may have changed, and the procedure has to be repeated. In this iterative process, the number of clusters gradually decreases since some clusters lose all their names, but the process converges after several iterations.

Unfortunately, it shows that a straightforward summation of attractions does not work well. For the UK, all Western names gather in one big cluster, while names from other ethnic and religious origin join in some much smaller clusters. For the Dutch data, the result is more diverse, but still unbalanced with a few big and many small clusters. The reason is that the attraction between a limited number of typical or idiosyncratic names is not high enough to stand the collective (individually much lower) attraction of many other atypical names in big clusters. The optimization problem of the second phase consists of making the best use of strong connections of name pairs, while neutralizing the aggregate effect of many weak connections of popular names with other names in the cluster outcomes. To achieve this, we put much more emphasis on closely related names by applying an exponential weighting of the attraction by a so-called Minkowski coefficient  $m$ . The attraction by cluster  $K$  on some name then is

$$A(\text{name}, \text{cluster}_K) = \left\{ \sum_{i \in K} A_c(\text{name}, \text{name}_i)^m \right\}^{1/m} \quad (7)$$

For both the Dutch set, with initially 302 clusters, and the UK set, with initially 160 clusters, the final number of clusters as a function of the Minkowski coefficient  $m$  is shown in Figure 1.

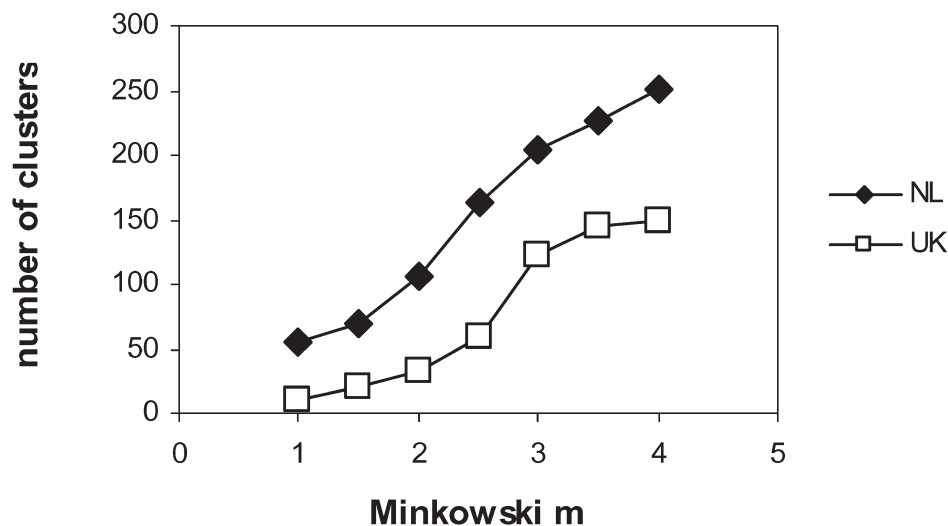


FIGURE 1 Number of name clusters as a function of weighting with Minkowski coefficient  $m$ , for both Dutch and UK data

For high values of the Minkowski coefficient  $m$ , the number of clusters approaches the number of clusters in the initial phase (302 for the Dutch data, 160 for the UK data), because then only the nearest neighbours of a name will have an influence on the result. Recall that in the initial clustering phase only the single nearest neighbour was in play. As mentioned before, a straightforward summation of all attractions (equivalent with  $m=1$ ) will reduce the number of clusters to a minimum, but with an unbalanced outcome.

Obviously there is some optimum value of  $m$  for which clusters convey a maximum amount of information while not generating too much irrelevant detail. For this value of  $m$  there is no recipe, however, because it depends on specific properties of the data. The best  $m$  gives enough granularity in clusters (no big cluster that takes all), while expected clusters like regional names or Hebrew names, show up. For the Dutch data this was the case for  $m=2.5$  (164 clusters), for the UK data for  $m=3.5$  (144 clusters).

The resulting clusters are shown in Tables 4 and 5. But before discussing these, we still have one step to make, because the number of clusters is still quite high.

### **Grand clusters**

So far, our analysis was concerned with the relationship between a name and another name, or between a name and a cluster of names. This approach is based on the assumption that parents choose names for all their children from a single cluster of names. However, parents that have affinity to some typical group of names may also like related names from neighbouring clusters. If certain Dutch parents prefer some cluster of Frisian names for their children, they may also like other clusters of Frisian names or certain types of traditional Dutch names, while disliking English names and not even thinking of Arabic names. The result is that, among this group of parents, some will choose names for all their children from that specific Frisian name set, but other parents may choose only one name from it and another name out of other Frisian names or traditional Dutch names. Therefore, after having looked for parents that choose names for all their children from a single name cluster, we may widen our view and look for parents that share a similar pattern of preferences among the entire spectrum of name clusters we have distinguished already.

The relationship between name clusters can be quantified by the computation of the attraction that clusters exerts on each other. Analogous to formula (7) but with an extension, formula (8) gives the attraction of clusters vis-à-vis each other. For the computation of the attraction of cluster  $G$  on cluster  $K$ , we compute the attraction of cluster  $K$  on each individual name in  $G$ , take the sum over all names in  $G$ , and normalize the result for the number of names involved. It shows that optimal results are obtained if we use the maximum of the number of names in cluster  $K$  or  $G$  for the latter. The attraction between clusters  $G$  and  $K$  is then (symmetrically) defined as:

$$A(\text{cluster}_G, \text{cluster}_K) = \left[ \frac{\sum_{j \in G} \sum_{i \in K} A_c(\text{name}_i, \text{name}_j)^m}{\max(N_G, N_K)} \right]^{1/m} \quad (8)$$

On this basis, we can compute the attraction that all name clusters exert on each other. The result is a cluster attraction matrix of  $164 * 164$  values for the Dutch data and  $144 * 144$  values for the UK data in which we can search for common attraction patterns. We use factor analysis for this (see e.g. Gorsuch, 1983), and apply varimax rotation to find patterns on which the loading (L) of individual name clusters is maximized. For both the Dutch and UK data we set a limit to 25 independent patterns, which together explained 48.7% and 63.4% of total variance in the attraction values, respectively. Each of the 164 (NL) and 144 (UK) name clusters has the highest loading on a particular pattern, and they are grouped accordingly in grand clusters. Maximum loadings are typically higher than  $L=0.5$ . In a few cases the loading was scattered over many patterns. The name cluster was then still associated to the best fitting grand cluster, or a number of name clusters were taken together in a separate, weakly classified grand cluster. In all, we distinguished 34 grand clusters for the Netherlands and 28 for the United Kingdom.

The final results are presented in Table 4 for the Netherlands and in Table 5 for the United Kingdom. The tables are enriched with general information on the grand name group, and include a short description that helps to identify the properties. This description can be based on the dominant language, the overall length of the names, or another striking feature, possibly including the dominant gender. Besides the total number of children in a name group, we also distinguished two periods of birth, 1985–1989 and 2000–2005 (NL) or 2000–2002 (UK), for which we give the percentage of coverage. This makes it possible to identify trends in the general popularity of a name in a name group, although some individual names may deviate from the change in popularity of the name group to which it belongs. The order of presentation is governed by these trends. Per country, we start with grand name groups that are in decline, followed by those which represent upcoming names. We conclude with non-Western name groups.

As we look to the first grand cluster for the Netherlands in Table 4 as an example, we observe traditional names in Latin form, as usually used by the Catholic part of the population. This group consists of 69 names, and declined between about 1987 and 2003 from 5.6 to 2.2% of all children. Names are equally distributed among boys and girls. The grand cluster is composed of four finer name groups, JOHANNES, MARCUS, REGINA, and JOACHIM, of which JOHANNES stands out. Their loading (L) on the grand cluster is strongest for the groups JOHANNES and MARCUS (0.7 and 0.8). The estimate of their total subgroup size (R), i.e. the share of the parents that might consider a name from this group, is 31% for the group JOHANNES, and drops to only 4% for the group JOACHIM. Subsequently, Table 4 lists the important names of this grand cluster, with name, gender, and frequency (N) in the whole period 1982–2005. The order is by fine name group and frequency. In the same way, all grand clusters are presented, in Table 4 for the Netherlands and in Table 5 for the UK.

Note that the emerging groups and clusters are entirely defined by a statistical analysis, and therefore can be heterogeneous due to accidental properties of the data. This may especially be relevant for the names with lower frequencies. Nevertheless, the result seems surprisingly plausible for both countries, which adds to the general validity of the approach.



TABLE 4 (Continued)

Theresia	f	558	Jacoba	f	2908	
<b>Joachim</b>	m	528	Geert	m	2904	
<b>TRADITIONAL NAMES</b>			Janna	f	2836	
<b>Dutch form</b>			Leendert	m	2650	
128 names - 248,803 children			Berend	m	2639	
69% male - 31% female			Frederik	m	2517	
6.7% 1987 > 3.5% 2003			Roelof	m	2423	
Names that were traditionally predominantly given by the non-Catholic population. Serious in decline for fifty years. It is of interest that more boys than girls are still given a name from this group which is possibly due to more naming after the grandfather(s) in original spelling than for girls.			Evert	m	2414	
	L	R(%)	N	Gerard	m	2407
JAN	0.4	24	241,869	Gerben	m	2077
JANTINE	0.6	6	2140	Aaltje	f	2034
BAREND	0.7	6	1566	Abraham	m	2009
OTTO	0.7	4	1217	Geertje	f	2008
WILLEMPIE	0.6	3	1061	Neeltje	f	2003
CARINA	0.4	7	950	Teunis	m	1902
<b>Jan</b>	m	19,975	Arend	m	1821	
Willem	m	13,232	Gijsbert	m	1767	
Hendrik	m	12,322	Herman	m	1767	
Cornelis	m	12,299	Grietje	f	1704	
Pieter	m	12,032	Geertruida	f	1697	
Gerrit	m	8342	Jannetje	f	1627	
Dirk	m	7382	Marina	f	1619	
Cornelia	f	7115	Anne	m	1571	
Jacob	m	6545	Frans	m	1562	
Johan	m	5714	Dirkje	f	1464	
Marinus	m	5585	Harmen	m	1459	
Adriana	f	4922	Antje	f	1361	
Hendrika	f	4148	Alida	f	1358	
Albert	m	3657	Martina	f	1353	
Elizabeth	f	3547	Aart	m	1346	
Arie	m	3399	Hendrikje	f	1344	
Klaas	m	3106	Willemina	f	1339	
Harm	m	3070	Derk	m	1263	
Adriaan	m	3063	Andries	m	1232	
			Pietermella	f	1228	
			Lena	f	1189	
			Dina	f	1172	
			Trijntje	f	1165	
			Karel	m	1116	



TABLE 4 (Continued)

Gerritje	f	1058	<b>Friso</b>	m	795		
Jantje	f	1053	Ewout	m	670		
Henk	m	1049	<b>ELITE NAMES 2</b>				
Egbert	m	1027	11 names - 33,122 children				
<b>Jantine</b>	f	700	78% male - 22% female				
Henri	m	535	0.9% 1987 > 0.4% 2003				
Aline	f	534	Another group of international names in decline with an elite connotation.				
<b>Barend</b>	m	795		L	R(%)		
Annigje	f	511			N		
<b>Carina</b>	f	548	ALEXANDER	0.5	39	22,716	
<b>ELITE NAMES 1</b>			ALEXANDRA	0.3	14	3052	
19 names - 13,662 children			BARBARA	0.7	18	2585	
54% male - 46% female			NORA	0.6	7	2504	
0.3% 1987 > 0.2% 2003			RUDOLF	0.2	12	2265	
The names in this group associate to the elite. They are quite long and include some typical Dutch and French names. They are somewhat in decline.			<b>Alexander</b>		m	8352	
	L	R(%)	N	Sebastiaan	m	8198	
RODERICK	0.5	7	4934	Christiaan	m	6166	
LISELOTTE	0.6	9	2982	<b>Alexandra</b>	f	2235	
EMILIE	0.6	4	2260	Victoria	f	817	
MAXIMILIAAN	0.6	7	2021	<b>Barbara</b>	f	1859	
FRISO	0.6	7	1465	Caspar	m	726	
<b>Roderick</b>		m	1012	<b>Nora</b>	f	1685	
Ferdinand		m	971	Sofia	f	819	
Boudewijn		m	916	<b>Rudolf</b>	m	1288	
Magdalena		f	706	Eduard	m	977	
Bernadette		f	525	<b>MIXED NAMES 1</b>			
<b>Liselotte</b>		f	1171	23 names - 17,046 children			
Rozemarijn		f	1044	64% male - 36% female			
Annemijn		f	767	0.4% 1987 > 0.3% 2003			
<b>Emilie</b>		f	655	A small group of names, perhaps with some elite flavour. The cluster INGEBORG includes Nordic names.			
Frédérique		f	572		L	R(%)	
Philippe		m	518			N	
Etienne		m	515	LUKAS	0.5	15	4013
<b>Maximiliaan</b>		m	939	INGEBORG	0.4	9	3985
Justus		m	560	LENNARD	0.7	10	3353
Constantijn		m	522	ANTONIE	0.6	8	2445
				JOANNE	0.5	10	2373

TABLE 4 (Continued)

JURIAN	0.3	5	877	Ilse	f	9193
<b>Lukas</b>		m	1759	Frank	m	9124
Kasper		m	1537	Peter	m	9117
Jurriaan		m	717	Erik	m	8214
<b>Ingeborg</b>		f	1116	Paul	m	7316
Sigrid		f	727	Inge	f	6482
Arne		m	701	Rob	m	6118
Kristian		m	571	Sandra	f	5861
Roald		m	501	Saskia	f	5156
<b>Lennard</b>		m	1035	Ellen	f	5083
Arnoud		m	838	Yvonne	f	4317
Madeleine		f	531	Martin	m	4159
<b>Antonie</b>		m	872	René	m	4110
Machiel		m	590	Irene	f	3579
Aleida		f	566	Karin	f	3455
<b>Joanne</b>		f	1306	Ruud	m	3229
Corine		f	607	Alex	m	3193
<b>INTERNATIONAL &amp; DUTCH NAMES</b>				Susan	f	2615
<b>premodern</b>				Petra	f	2485
92 names - 250,732 children				Astrid	f	2444
41% male - 59% female				Ingrid	f	2308
8.0% 1987 > 2.4% 2003				John	m	2301
The clusters in this group are in serious decline and have had their most popular years in the seventies or earlier. The big cluster LAURA shows many internationally used names ( <i>Laura</i> was the most popular girls name in 1991, 1994–1997), while the other clusters predominantly include Dutch names.				Anke	f	2238
				Jolanda	f	2185
				Elisa	f	2119
				Anita	f	1999
				Jos	m	1990
				Sylvia	f	1823
				Suzan	f	1761
				Moniek	f	1708
				Ben	m	1570
				Paula	f	1561
				Kristel	f	1535
				Nico	m	1501
				Ron	m	1407
				Tanja	f	1356
				Karen	f	1285
				Paulien	f	1222
				André	m	1160
	L	R(%)	N			
LAURA	0.6	34	218,569			
MARIEKE	0.7	31	14,463			
MARIJKE	0.7	16	8970			
HANS	0.7	16	4397			
HILDE	0.6	17	2460			
ELSKE	0.6	5	1873			
<b>Laura</b>		f	20,225			
Mark		m	18,413			
Linda		f	13,414			
Robert		m	11,003			

TABLE 4 (Continued)

José	f	1113	Wouter	m	13,869
Mario	m	1083	Jasper	m	12,574
Anja	f	1010	Maarten	m	11,988
<b>Marieke</b>	f	7750	Suzanne	f	9442
Janneke	f	4440	Eline	f	9353
Hanneke	f	1568	Lisanne	f	7986
Marjan	f	705	Matthijs	m	7106
<b>Marijke</b>	f	2862	Simone	f	6897
Anneke	f	1588	Joris	m	6653
Tineke	f	820	Marloes	f	6383
Perry	m	721	Steven	m	6068
Ineke	f	688	Marjolein	f	6058
Gert-Jan	m	684	Michiel	m	5975
Annette	f	622	Rianne	f	5677
Devin	m	620	Bastiaan	m	5067
<b>Hans</b>	m	2197	Leonie	f	4784
Bert	m	782	Marleen	f	4415
Wim	m	598	Wessel	m	4366
<b>Hilde</b>	f	1876	Mathijs	m	4205
Else	f	584	Elise	f	4140
<b>Elske</b>	f	619	Evelien	f	4077
Theo	m	517	Rutger	m	4071
<b>DUTCH NAMES</b>			Robbert	m	3975
<b>premodern - THOMAS</b>			Menno	m	3955
87 names - 343,881 children			Marije	f	3928
58% male - 42% female			Lisette	f	3774
8.3% 1987 > 5.4% 2003			Jochem	m	3644
The big cluster THOMAS is presented separately but has spread loadings to both premodern names and Dutch modern names. The name <i>Thomas</i> itself was the most popular boys name in 1995–1998, 2000–2003 and formed a kind of a bridge between premodern and modern names. Most other names had their peak period (much) earlier.			Susanne	f	3576
			Lianne	f	3548
			Marijn	m	3507
			Karlijn	f	3029
			Rosanne	f	2846
			Eveline	f	2767
			Mariëlle	f	2686
			Martine	f	2608
			Carlijn	f	2568
			Emiel	m	2448
			Janine	f	2339
			Caroline	f	2289
	L	R(%)	N		
THOMAS	0.5	34	343,881		
<b>Thomas</b>		m	25,933		
Jeroen		m	21,125		
Sander		m	16,205		
Martijn		m	16,192		

TABLE 4 (Continued)

Tijmen	m	2272	<b>Ronnie</b>	m	954
Marianne	f	2268	Dennie	m	541
Thijmen	m	2266	<b>ENGLISH NAMES</b>		
Aniek	f	2123	<b>premodern</b>		
Willemijn	f	2115	231 names - 759,960 children		
Lennart	m	2068	51% male - 49% female		
Pauline	f	2008	18.0% 1987 > 11.5% 2003		
Annemarie	f	1953	The big cluster KEVIN largely exists of English names but also includes some Roman names. <i>Kevin</i> itself was the most popular name from 1990-1994, but many other names from this cluster attained a high frequency as well. Currently most names are in decline.		
Rogier	m	1927		L	R(%)
Marissa	f	1925			N
Mirthe	f	1835	KEVIN	0.5	27
Maureen	f	1795	<b>Kevin</b>	m	22,586
Marlies	f	1758	Dennis	m	18,965
Jolien	f	1740	Robin	m	16,904
Marnix	m	1664	Michael	m	14,699
Merijn	m	1570	Stefan	m	14,565
Jurgen	m	1562	Jeffrey	m	13,281
Annemiek	f	1505	Michelle	f	13,174
Annemieke	f	1495	Patrick	m	12,976
Anton	m	1377	Danny	m	11,334
Annelies	f	1326	Wesley	m	11,226
Anniek	f	1321	Melissa	f	11,196
Dorien	f	1307	Chantal	f	11,000
Margot	f	1298	Daniëlle	f	9948
Roland	m	1291	Naomi	f	9915
Carolien	f	1255	Denise	f	9546
Jeanine	f	1231	Vincent	m	9451
Heleen	f	1224	Jordy	m	9375
Annabel	f	1217	Romy	f	9116
<b>DUTCH NAMES</b>			Joey	m	8919
<b>unclassified premodern</b>			Daphne	f	8363
4 names - 9,953 children			Sharon	f	8132
100% male - 0% female			Samantha	f	8029
0.3% 1987 > 0.1% 2003			Jessica	f	8021
		R(%)	Wendy	f	7988
		N			
MARC		31			8458
RONNIE		9			1495
<b>Marc</b>	m				5667
Eric	m				2791

TABLE 4 (Continued)

Richard	m	7878	Ramon	m	3834
Remco	m	7833	Bryan	m	3828
Demi	f	7782	Christian	m	3757
Nicole	f	7343	Damian	m	3692
Dylan	m	7314	Esmée	f	3683
Justin	m	7077	Ryan	m	3679
Melanie	f	6957	Youri	m	3630
Stephanie	f	6837	Cynthia	f	3622
Marco	m	6794	Ashley	f	3582
Ricardo	m	6551	Edwin	m	3531
Michel	m	6397	Amanda	f	3531
Jennifer	f	6290	Larissa	f	3511
Nathalie	f	6290	Miranda	f	3468
Tamara	f	6197	Nadine	f	3358
Kimberley	f	6126	Quinten	m	3321
Brian	m	5400	Sabrina	f	3297
Danique	f	5305	Arjan	m	3280
Kimberly	f	5268	Yannick	m	3275
Priscilla	f	5206	Stefanie	f	3093
Julian	m	5185	Jeremy	m	2972
Sabine	f	5080	Raymond	m	2955
Mariska	f	5004	Guido	m	2890
Marcel	m	4968	Tristan	m	2859
Stephan	m	4702	Leroy	m	2856
Claudia	f	4657	Mitchel	m	2835
Bianca	f	4641	Marvin	m	2827
Ronald	m	4632	Shirley	f	2737
Melvin	m	4603	Angelique	f	2659
Pascal	m	4595	Natasja	f	2658
Patricia	f	4594	Brenda	f	2601
Mitchell	m	4559	Kyra	f	2507
Angela	f	4545	Jason	m	2488
Erwin	m	4544	Celine	f	2480
Leon	m	4220	Renate	f	2340
Maikel	m	4204	Joëlle	f	2338
Maurice	m	4106	Jamie	f	2333
Ilona	f	4075	Roxanne	f	2305
Monique	f	4021	Kayleigh	f	2291
Carmen	f	3861	Davey	m	2245

TABLE 4 (Continued)

Ralph	m	2225	Christel	f	1376
Nikita	f	2224	Monica	f	1375
Yvette	f	2193	Desiree	f	1365
Kaylee	f	2166	Jenny	f	1363
Jacqueline	f	2099	Ferry	m	1352
Tara	f	2063	Jaimy	f	1316
Anthony	m	2015	Dionne	f	1306
Natascha	f	2001	Kenneth	m	1304
Randy	m	1961	Carlo	m	1274
Nigel	m	1911	Selina	f	1263
Andrea	f	1881	Marcella	f	1248
Diana	f	1872	Rowena	f	1237
Rowan	m	1866	Jerry	m	1228
Quinty	f	1811	Calvin	m	1225
Debbie	f	1800	Manuela	f	1224
Shannon	f	1737	Carola	f	1217
Sylvana	f	1734	Sascha	f	1177
Vivian	f	1708	Céline	f	1176
Tycho	m	1699	Arno	m	1168
Madelon	f	1642	Wilco	m	1119
Xander	m	1625	Jack	m	1115
Remy	m	1624	Veronique	f	1113
Jamie	m	1620	Alyssa	f	1107
Esmeralda	f	1619	Brigitte	f	1089
Raoul	m	1581	Lesley	m	1079
Brandon	m	1526	Kenny	m	1078
Lindsey	f	1522	Kelsey	f	1073
Sharona	f	1508	Danielle	f	1063
Jaimy	m	1497	Desirée	f	1049
Kelvin	m	1486	Shanna	f	1031
Charissa	f	1478	Elvira	f	1028
Ramona	f	1463	Lauren	f	1018
Robbin	m	1455	Arnold	m	1004
Manouk	f	1454			
Lindsay	f	1439			
Dominique	m	1421			
Sebastian	m	1410			
Babette	f	1397			
<b>ENGLISH NAMES</b>					
<b>royal names</b>					
3 names - 2966 children					
100% male - 0% female					
0.1% 1987 > 0.0% 2003					
				R(%)	N
WILLIAM				13	2966

TABLE 4 (Continued)

<b>William</b>	m	1705	<b>Johnny</b>	m	1331	
James	m	696	Benny	m	514	
Harry	m	565	<b>Sidney</b>	m	753	
<b>ENGLISH NAMES</b>			<b>MIXED NAMES 2</b>			
<b>y-suffix</b>			5 names - 2408 children			
33 names - 61,863 children			51% male - 49% female			
43% male - 57% female			0.1% 1987 > 0.0% 2003			
1.8% 1987 > 0.6% 2003						
These clusters include English names with the -y suffix. Just like the cluster KEVIN they are in decline.						
	L	R(%)	N			
KELLY	0.5	23	47,066	REMKO	0.4 6 1422	
JIMMY	0.6	15	9365	REMI	0.3 7 986	
QUINCY	0.6	4	2396	<b>Remko</b>	m 620	
JOHNNY	0.4	11	1845	<b>Remi</b>	m 598	
SIDNEY	0.6	3	1191			
<b>Kelly</b>	f	9359	<b>UPCOMING NAMES</b>			
Mandy	f	8189	<b>FRISIAN NAMES</b>			
Nicky	m	4666	67 names - 100,871 children			
Daisy	f	4177	40% male - 60% female			
Cindy	f	3745	1.8% 1987 < 2.5% 2003			
Davy	m	2310	This group including names that originate from the province of Friesland. They are highly traditional, but especially names in the cluster FEMKE have gained nationwide popularity and are responsible for the growth of this group.			
Nancy	f	2185		L	R(%)	N
Ricky	m	1678	FEMKE	0.7	34	40,518
Andy	m	1510	JELMER	0.7	13	24,087
Barry	m	1441	DOUWE	0.8	8	10,824
Donny	m	1424	HIDDE	0.7	9	8202
Debby	f	1289	NOORTJE	0.5	14	6432
Patty	f	1097	FAMKE	0.8	7	3674
<b>Jimmy</b>	m	3332	WIETSKE	0.7	7	2536
Tommy	m	1578	SIETSKE	0.6	9	1921
Lizzy	f	1234	BAUKJE	0.6	4	1670
Gaby	f	990	JOUKE		2	1007
Bobby	m	817	<b>Femke</b>	f	10,535	
Sonny	m	783	Jelle	m	9940	
Francis	f	631	Maaikje	f	8804	
<b>Quincy</b>	m	1389	Nienke	f	7574	
Shanice	f	569				

TABLE 4 (Continued)

Renske	f	3665	<b>Sietske</b>	f	1005	
<b>Jelmer</b>	m	3401	Froukje	f	916	
Nynke	f	2515	<b>Baukje</b>	f	547	
Fenna	f	2052	<b>Jouke</b>	m	524	
Arjen	m	1921				
Jorrit	m	1883	<b>HEBREW NAMES</b>			
Marten	m	1610	60 names - 152,291 children			
Hester	f	1418	66% male - 34% female			
Hessel	m	998	2.9% 1987 < 3.4% 2003			
Jurjen	m	963	The clusters DANIËL, ESTHER and RUTH are gender specific. The cluster JOSHUA deviates as it includes typical English names.			
Jelte	m	888				
<b>Douwe</b>	m	1679				
Sietse	m	1093				
Auke	m	1061	DANIËL	0.7	37	67,942
Ella	f	1027	ESTHER	0.8	33	27,196
Tjeerd	m	1002	NOAH	0.6	10	16,474
Wiebe	m	889	JOSHUA	0.6	13	16,432
Aukje	f	880	RACHEL	0.8	23	11,778
Lieve	f	782	JOËL	0.6	13	5232
Bauke	m	662	BOAZ	0.2	4	3403
Bouke	m	594	HANNAH	0.7	17	2968
<b>Hidde</b>	m	2683	ARIANNE	0.3	5	866
Jurre	m	1625	<b>Daniël</b>	m	14,083	
Melle	m	987	Ruben	m	13,152	
Benthe	f	897	David	m	10,716	
Siebe	m	804	Lucas	m	8582	
Tjerk	m	674	Simon	m	5333	
Kars	m	532	Benjamin	m	4615	
<b>Noortje</b>	f	2130	Jonathan	m	3946	
Lonneke	f	1876	Daniel	m	2992	
Eefje	f	1017	Samuel	m	2535	
Dieuwertje	f	621	Joseph	m	1003	
<b>Famke</b>	f	980	<b>Esther</b>	f	10,846	
Bregje	f	797	Judith	f	6526	
Wietse	m	739	Mirjam	f	3548	
Lobke	f	700	Hanna	f	1688	
<b>Wietske</b>	f	887	Miriam	f	1659	
Sjoukje	f	604	Lydia	f	1603	
Rinske	f	593	Ruth	f	1326	



TABLE 4 (Continued)

				L	R(%)	N
<b>Noah</b>	m	3240	the cluster CHARLOTTE has female French names, the cluster FLORIS has old noble Dutch male names, while the cluster CASPER has these with a more international flavor.			
Levi	m	2614				
Nathan	m	2046				
Aron	m	1726				
Aaron	m	1307	AMBER	0.6	24	51,499
Jonas	m	1101	EMMA	0.7	31	41,088
Ayla	f	814	CHARLOTTE	0.5	23	27,980
Tamar	f	777	FLORIS	0.7	21	19,536
Chloë	f	762	CASPER	0.5	22	14,851
Joram	m	589	OLIVIER	0.6	8	9227
<b>Joshua</b>	m	3335	FABIAN	0.6	15	7608
Timothy	m	3107	ROSALIE	0.7	20	7593
Matthias	m	1575	ROELAND	0.3	9	5646
Christopher	m	1513	MADELIEF	0.4	7	2043
Matthew	m	1231				
Talitha	f	926	<b>Amber</b>		f	10,271
Andrew	m	816	Fleur		f	9205
Elena	f	813	Merel		f	8454
Samuël	m	674	Myrthe		f	4857
Gregory	m	673	Esmee		f	4128
<b>Rachel</b>	f	4077	Jasmijn		f	3559
Deborah	f	3451	Sterre		f	3045
Rebecca	f	3408	Veerle		f	2762
Debora	f	842	Claire		f	2073
<b>Joël</b>	m	2937	Jade		f	1604
Thirza	f	979	Linde		f	1541
Micha	m	809	<b>Emma</b>		f	9992
Esmé	f	507	Julia		f	9475
<b>Boaz</b>	m	846	Sophie		f	8893
Tirza	f	688	Sarah		f	7102
Yoran	m	661	Rosa		f	3370
Ezra	m	651	Sophia		f	2256
Noah	f	557	<b>Charlotte</b>		f	9874
<b>Hannah</b>	f	2968	Isabelle		f	3103
<b>ELITE NAMES 3</b>			Louise		f	2446
70 names - 187,071 children			Valerie		f	2077
30% male - 70% female			Emily		f	1880
2.4% 1987 < 5.8% 2003			Josephine		f	1333
Names in this group have a typical elite flavor. The cluster AMBER has female names originating in nature,			Juliette		f	1258
			Christine		f	1209

TABLE 4 (Continued)

			<b>DUTCH NAMES</b>			
			<b>modern</b>			
Fabienne	f	1114				
Frederique	f	928				
<b>Floris</b>	m	5828	158 names - 575,780 children			
Laurens	m	5066	60% male - 40% female			
Pepijn	m	2990	8.0% 1987 < 16.2% 2003			
Maurits	m	2375	Dutch modern names are short (abbreviated), seldom more than five letters, for boys in many cases just one syllable. <i>Lisa</i> was the most popular name for girls in 1992, <i>Iris</i> in 1993, and <i>Sanne</i> from 1998–2006, <i>Tim</i> for boys in 1996 and 1999. Many high-ranked names are still in the top twenty. The cluster has weak relations to the English names in the cluster KEVIN, which were most popular in the years before.			
Philip	m	1657				
Reinier	m	1620				
<b>Casper</b>	m	4589				
Hugo	m	3105				
Victor	m	2796				
Arthur	m	1841		L	R(%)	N
Oscar	m	1839	TIM	0.4	29	575,780
Edward	m	681	Tim		m	22,915
<b>Olivier</b>	m	2458	Sanne		f	21,600
Julius	m	1319	Anne		f	19,853
Diederik	m	1182	Lisa		f	16,874
Valentijn	m	994	Tom		m	16,630
Roosmarijn	f	660	Rick		m	16,541
Leander	m	572	Iris		f	16,373
Lodewijk	m	553	Bart		m	15,996
Olga	f	544	Daan		m	14,672
Nicolas	m	515	Eva		f	13,228
<b>Fabian</b>	m	3217	Bas		m	12,690
Tobias	m	2582	Max		m	12,485
Florian	m	1334	Bram		m	12,455
<b>Rosalie</b>	f	3081	Lotte		f	12,129
Isabel	f	2782	Koen		m	11,709
Isabella	f	1730	Thijs		m	11,511
<b>Roeland</b>	m	1043	Tessa		f	10,861
Jacco	m	955	Jesse		m	10,239
Ernst	m	704	Joost		m	8665
Allard	m	688	Luuk		m	8573
Folkert	m	598	Stijn		m	8512
Eleonora	f	588	Nina		f	7881
Alissa	f	568	Lieke		f	7291
Ewoud	m	502	Sjoerd		m	6897
<b>Madelief</b>	f	1069	Vera		f	6777
Merlijn	m	974	Gijs		m	6615

TABLE 4 (Continued)

Sam	m	6136	Boris	m	1980
Maud	f	5924	Imke	f	1948
Pim	m	5834	Sam	f	1874
Stan	m	5808	Pien	f	1873
Rik	m	5672	Jaap	m	1868
Floor	f	5618	Silke	f	1867
Sem	m	5434	Ties	m	1693
Niek	m	5427	Dana	f	1632
Roos	f	5278	Jessie	f	1629
Sara	f	5186	Wout	m	1605
Maartje	f	5140	Noor	f	1583
Roel	m	4688	Evi	f	1552
Isa	f	4636	Liza	f	1504
Bob	m	4462	Elke	f	1475
Rens	m	4359	Jort	m	1457
Luc	m	4321	Floortje	f	1443
Joep	m	4236	Nils	m	1435
Teun	m	4196	Lex	m	1399
Job	m	4156	Mieke	f	1399
Loes	f	3810	Mara	f	1367
Bo	f	3493	Tomas	m	1356
Stef	m	3375	Janne	f	1342
Jens	m	3091	Kees	m	1333
Chris	m	3085	Jip	m	1329
Cas	m	3054	Chiel	m	1323
Ivo	m	3008	Malou	f	1317
Twan	m	2965	Marlou	f	1268
Mees	m	2902	Jop	m	1233
Meike	f	2900	Ruby	f	1222
Thom	m	2838	Jet	f	1221
Guus	m	2594	Sil	m	1192
Jorn	m	2585	Pleun	f	1182
Coen	m	2511	Felix	m	1157
Sjors	m	2477	Marijn	f	1150
Sofie	f	2359	Maik	m	1108
Tijn	m	2172	Tijs	m	1107
Mart	m	2104	Kaj	m	1062
Julie	f	2083	Sacha	f	1062
Freek	m	2082	Brent	m	1048

TABLE 4 (Continued)

Lindy	f	1039	Finn	m	1973	
Jolijn	f	1037	Joy	f	1702	
Renee	f	1037	Sean	m	1243	
Evy	f	1030	Liam	m	1125	
<b>ENGLISH NAMES</b>						
<b>short</b>						
36 names - 133,924 children						
57% male - 43% female						
2.2% 1987 < 3.3% 2003						
The current trend to short and abbreviated names also extends to the English names in this group. Many have just one syllable.						
	L	R(%)	N			
KIM	0.6	41	75,092	Owen	m	1089
BRITT	0.7	15	24,869	Collin	m	1078
AMY	0.8	13	18,122	Yentl	f	734
JIM	0.5	18	7231	<b>Jim</b>	m	2866
MEGAN	0.6	7	4319	Mick	m	2494
GLENN	0.7	15	4291	Luke	m	1871
<b>Kim</b>		f	18,509	<b>Megan</b>	f	2011
Nick		m	16,170	Duncan	m	1196
Mike		m	13,846	Caitlin	f	1112
Roy		m	13,225	<b>Glenn</b>	m	3387
Joyce		f	9317	Scott	m	904
Dave		m	4025	<b>ENGLISH NAMES</b>		
<b>Britt</b>		f	7690	<b>unclassified</b>		
Jill		f	3238	5 names - 5372 children		
Lynn		f	3229	23% male - 77% female		
Tess		f	2840	0.0% 1987 < 0.2% 2003		
Kay		m	2035		R(%)	N
Mitch		m	1451	CHEYENNE	5	3812
Kai		m	1346	CHAYENNE	4	1560
Gwen		f	1056	<b>Cheyenne</b>	f	2198
Lois		f	856	Chelsea	f	991
Quint		m	591	Damiën	m	623
<b>Amy</b>		f	4198	<b>Chayenne</b>	f	939
Colin		m	2354	Jermaine	m	621
Ian		m	2098	<b>FRENCH NAMES</b>		
17 names - 19,459 children						
23% male - 77% female						
0.3% 1987 < 0.4% 2003						
A group of small clusters with French names, with highest frequencies for girls names.						
	L	R(%)	N			
MAXIME	0.7	11	5038			
DOMINIQUE	0.8	22	4967			

TABLE 4 (Continued)

FABIËNNE	0.8	10	4527	<b>Beau</b>	f	1349	
STÉPHANIE	0.6	8	3154	Emile	m	766	
DIMITRI	0.3	8	1773	Alain	m	722	
<b>Maxime</b>		f	2720	Mathieu	m	624	
Aimée		f	846	Guy	m	610	
Thierry		m	781	Julien	m	590	
Noël		m	691	Yves	m	572	
<b>Dominique</b>		f	4194	Valérie	f	556	
Pascalie		f	773	<b>Jules</b>	m	1455	
<b>Fabiënne</b>		f	1527	Louis	m	953	
Rachelle		f	1410	Camiel	m	778	
Juliette		f	1031	Inez	f	681	
Florine		f	559	<b>Maxim</b>	m	1087	
<b>Stéphanie</b>		f	942	Beau	m	966	
Geoffrey		m	632	Roman	m	523	
Xavier		m	630	<b>Marie</b>	f	1409	
Mylène		f	512	Jean	m	731	
<b>Dimitri</b>		m	1043	<b>Anique</b>	f	618	
Dominic		m	730				
<b>FRENCH NAMES</b>				<b>MIXED NAMES</b>			
<b>short</b>				<b>short</b>			
26 names - 33,234 children				18 names - 17,557 children			
36% male - 64% female				50% male - 50% female			
0.6% 1987 < 0.8% 2003				0.2% 1987 < 0.6% 2003			
Another group of clusters with French names, including the somewhat shorter names.				The clusters in this group include Nordic names in the cluster BENTE, and international names in the others. Names have no more than five letters. Fewer than 10% of the parents consider the names.			
	L	R(%)	N		L	R(%)	N
ROBIN	0.4	26	16,368	BENTE	0.5	10	8549
BEAU	0.6	6	6765	IVAN	0.5	8	4642
JULES	0.8	10	3867	ABEL	0.5	6	2379
MAXIM	0.5	5	2576	LISE	0.6	9	1987
MARIE	0.5	12	2140	<b>Bente</b>		f	2980
ANIQUE	0.6	6	1518	Mats		m	1998
<b>Robin</b>		f	5511	Mirte		f	1419
Nikki		f	5385	Merle		f	1297
Renée		f	3227	Sten		m	855
Nicky		f	2245	<b>Ivan</b>		m	1046

TABLE 4 (Continued)

Rolf	m	827	Ivar	m	1237
Rudy	m	713	Dagmar	f	1237
Carla	f	622	Lilian	f	1094
Frits	m	523	Birgit	f	895
Kitty	f	505	<b>Joran</b>	m	1197
<b>Abel</b>	m	849	Leanne	f	816
Ward	m	813	Alwin	m	735
Noud	m	717	Duco	m	589
<b>Lise</b>	f	930	<b>Yorick</b>	m	1203
Hanne	f	580	Ingmar	m	698

**MIXED NAMES****Nordic & French**

25 names - 108,484 children

60% male - 40% female

1.7% 1987 < 2.7% 2003

This group is dominated by increasingly popular Nordic names in the clusters NIELS and OLAF, and some French popular names in the cluster ANOUK.

	L	R(%)	N
NIELS	0.8	33	60,901
ANOUK	0.6	34	34,870
OLAF	0.7	12	7475
JORAN	0.7	8	3337
YORICK	0.3	7	1901
<b>Niels</b>		m	17,602
Lars		m	13,477
Sven		m	9313
Kirsten		f	5742
Marit		f	5352
Jesper		m	3599
Bjorn		m	3281
Björn		m	2535
<b>Anouk</b>		f	13,986
Manon		f	8861
Timo		m	5140
Milou		f	4275
Joeri		m	2608
<b>Olaf</b>		m	1552
Annika		f	1460

**MODERN NAMES 1**

11 names - 14,964 children

23% male - 77% female

0.1% 1987 < 0.6% 2003

A mixture of modern short names, predominantly female, are found in this group. *Jip* is typical Dutch (and unisex).

	L	R(%)	N
PUCK	0.5	11	5161
MIKA	0.4	7	3190
DONNA	0.8	10	2668
ADAM	0.5	5	2035
SELMA	0.1	7	1910
<b>Puck</b>		f	2273
Kiki		f	2190
Jip		f	698
<b>Mika</b>		m	1643
Senna		f	1547
<b>Donna</b>		f	1499
Gina		f	1169
<b>Adam</b>		m	1166
Lina		f	869
<b>Selma</b>		f	1233
Ferdi		m	677

**MODERN NAMES 2**

29 names - 58,573 children

53% male - 47% female

0.4% 1987 < 2.6% 2003

TABLE 4 (Continued)

This group also includes modern names, with a variety of backgrounds.

	L	R(%)	N
MILAN	0.8	14	13,610
LARA	0.7	13	9400
JARNO	0.5	15	8890
ZOË	0.7	16	8772
LOÏS	0.6	8	6747
DION	0.5	11	5277
LUCA	0.5	8	3925
NOËLLE	0.5	10	1952
<b>Milan</b>		m	4659
Jordi		m	4224
Luna		f	2507
Dani		m	1501
Stella		f	719
<b>Lara</b>		f	3199
Indy		f	1982
Yara		f	1741
Mila		f	1445
Isis		f	1033
<b>Jarno</b>		m	2956
Dewi		f	1865
Jari		m	1588
Rico		m	1561
Renzo		m	920
<b>Zoë</b>		f	4590
Noa		f	4182
<b>Loïs</b>		f	1591
Boy		m	1531
Jay		m	1503
Vince		m	1077
Dean		m	1045
<b>Dion</b>		m	3219
Roan		m	1179
Rowan		f	879
<b>Luca</b>		m	2478
Nino		m	1447
<b>Noëlle</b>		f	1262
Romée		f	690

**ITALIAN & SPANISH NAMES**

30 names - 30,345 children

64% male - 36% female

0.4% 1987 < 0.8% 2003

Italian names dominate this group, but it shows also *Juan* and *Diego*.

	L	R(%)	N
LORENZO	0.7	7	9183
SORAYA	0.5	6	4938
GINO	0.7	8	4238
ALICIA	0.5	8	4071
DIEGO	0.6	3	2582
SERENA	0.2	7	2309
GABRIËLLA	0.1	7	1846
GIANNI	0.5	2	1178
<b>Lorenzo</b>		m	2999
Giovanni		m	2454
Delano		m	1298
Romano		m	1154
Celina		f	698
Marciano		m	580
<b>Soraya</b>		f	1947
Felicia		f	1202
Stefano		m	896
<b>Gino</b>		m	1804
Angelo		m	1616
Sergio		m	818
<b>Alicia</b>		f	1543
Roberto		m	863
Miguel		m	652
Selena		f	597
<b>Diego</b>		m	1103
Juan		m	550
<b>Serena</b>		f	1192
Chiara		f	1117
<b>Gabriëlla</b>		f	686
Daniëlla		f	606
Gabriël		m	554

TABLE 4 (Continued)

<b>Gianni</b>	m	611	<b>Mohamed</b>	m	9964		
Fabio	m	567	Fatima	f	3684		
<b>ITALIAN NAMES</b>			Youssef	m	2007		
<b>unclassified</b>			Khadija	f	1299		
3 names - 1830 children			Brahim	m	787		
67% male - 33% female			Zahra	f	578		
0.0% 1987 < 0.0% 2003			Halima	f	574		
	R(%)	N	Mustapha	m	526		
LEONARD	7	1830	<b>Khalid</b>	m	1584		
<b>Leonard</b>	m	877	Rachid	m	1434		
Louisa	f	597	Said	m	1057		
<b>SLAVIC NAMES</b>			Jamal	m	958		
3 names - 1,794 children			Laila	f	519		
35% male - 65% female			<b>Ahmed</b>	m	2511		
0.0% 1987 < 0.0% 2003			Hassan	m	832		
	R(%)	N	Saida	f	525		
IVANA	3	1794	<b>Karima</b>	f	1226		
<b>Ivana</b>	f	722	Latifa	f	730		
Igor	m	633	Salima	f	603		
<b>NON-WESTERN NAMES</b>			<b>Rachida</b>	f	697		
<b>ARABIC NAMES 1</b>			Hayat	f	623		
<b>in decline</b>			Fatiha	f	544		
36 names - 39,980 children			Najat	f	501		
60% male - 40% female			<b>Asma</b>	f	951		
1.2% 1987 > 0.5% 2003			Maryam	f	909		
This group with well-known Arabic names seems to represent names of declining popularity.			<b>ARABIC NAMES 2</b>				
	L	R(%)	N	14 names - 17,447 children			
MOHAMED	0.7	3	20,969	36% male - 64% female			
KHALID	0.7	2	6874	0.4% 1987 > 0.3% 2003			
AHMED	0.9	3	4183	Whereas Arabic names are considered by no more than 3% of all parents, the cluster NADIA is an exception with 13%, probably because <i>Nadia</i> itself also is a well-known Slavic name.			
KARIMA	0.7	2	2934		L	R(%)	N
RACHIDA	0.8	1	2662	NADIA	0.5	13	6128
ASMA		2	2358	SIHAM	0.7	2	4024
				ACHRAF	0.7	1	3853
				SAMIR	0.9	2	3442
				<b>Nadia</b>	f		3741
				Samira	f		2387



TABLE 4 (Continued)

<b>Siham</b>	f	1317	FADOUA	0.7	1	1814
Amal	f	1221	ASMAE	0.7	1	1502
Naoual	f	787	<b>Hicham</b>		m	1285
Ilham	f	699	Yasmina		f	1279
<b>Achraf</b>	m	995	Yassin		m	735
Anouar	m	907	<b>Hanane</b>		f	1121
Soufiane	m	818	<b>Fadoua</b>		f	602
Mounir	m	690	Imad		m	599
<b>Samir</b>	m	1445	<b>Asmae</b>		f	568
Karim	m	1358				
Saloua	f	639				

**ARABIC NAMES 3**

16 names - 9,740 children  
 65% male - 35% female  
 0.2% 1987 > 0.1% 2003

	L	R(%)	N
TARIK	0.8	2	4287
BOUCHRA	0.6	2	1682
IKRAM	0.8	2	1449
REDOUAN	0.8	2	1314
FOUAD	0.6	1	1008
<b>Tarik</b>		m	1156
Adil		m	915
Nabil		m	823
Jaouad		m	652
<b>Bouchra</b>		f	663
<b>Ikram</b>		f	994
<b>Redouan</b>		m	949
<b>Fouad</b>		m	797

**ARABIC NAMES 4**

15 names - 8785 children  
 34% male - 66% female  
 0.2% 1987 > 0.2% 2003

	L	R(%)	N
HICHAM	0.8	2	3641
HANANE	0.6	2	1828

**ARABIC NAMES 5****upcoming**

31 names - 30,203 children  
 50% male - 50% female  
 0.3% 1987 < 1.0% 2003

The Arabic names in this group are increasingly popular, but *Mohamed* is not in this group.

	L	R(%)	N
YASSINE	0.7	2	10,027
OMAR	0.9	2	8214
HAMZA	0.6	2	3594
IMANE	0.7	2	3489
YASMINE	0.9	2	3340
OUMAIMA	0.7	1	1539
<b>Yassine</b>		m	2120
Youssra		f	1410
Younes		m	1228
Kaoutar		f	938
Anissa		f	751
Loubna		f	747
Dounia		f	721
Marouane		m	585
Sana		f	578
Sanae		f	564
<b>Omar</b>		m	1985
Zakaria		m	1609
Ayoub		m	1368

TABLE 4 (Continued)

			Though the loadings for the clusters in this group are low, they definitely join Turkish names.		
			R(%)	N	
Hajar	f	1127			
Mariam	f	951			
Chaima	f	726			
<b>Hamza</b>	m	1803	IBRAHIM	3 3905	
Soumaya	f	639	MERVE	2 2656	
<b>Imane</b>	f	1482	HAKAN	1 2402	
Ilias	m	1318	YASIN	1 1951	
Amine	m	689	DENIZ	1 1916	
<b>Yasmine</b>	f	1399	HASAN	1 1781	
Anass	m	1116	SERKAN	1 1469	
Oussama	m	825	WALID	2 1032	
<b>Oumaima</b>	f	922	KÜBRA	1 801	
Chaimae	f	617	DUYGU	1 783	
<b>TURKISH NAMES</b>			AZIZ	1 718	
<b>unclassified 1</b>			ZAINAB	1 693	
6 names - 7127 children			<b>Ibrahim</b>	m 2240	
63% male - 37% female			Ismail	m 1665	
0.1% 1987 < 0.2% 2003			<b>Merve</b>	f 1538	
	L	R(%)	N		
EMRE		1	4488	Meryem	f 1118
ESRA		2	2639	<b>Hakan</b>	m 833
<b>Emre</b>	m		1729	Gökhan	m 556
Yusuf	m		1238	Volkan	m 545
Yunus	m		816	<b>Yasin</b>	m 1069
Enes	m		705	Yasemin	f 882
<b>Esra</b>	f		1911	<b>Deniz</b>	m 804
Esmā	f		728	Derya	f 735
<b>TURKISH NAMES</b>			<b>Hasan</b>	m 1032	
<b>unclassified 2</b>			Hüseyin	m 749	
28 names - 20,107 children			<b>Serkan</b>	m 683	
63% male - 37% female			<b>Walid</b>	m 542	
0.5% 1987 > 0.3% 2003			<b>Kübra</b>	f 801	

TABLE 5  
FIRST NAMES FROM THE UK (1982–2002)

See Table 4 for an explanation

<b>NAMES IN DECLINE</b>						
				Stephen	m	19,005
				Steven	m	17,957
				Jennifer	f	17,910
				Nicola	f	17,590
				Kirsty	f	17,426
				Louise	f	17,138
				Stephanie	f	16,920
				Kelly	f	15,184
				Lisa	f	14,417
				Peter	m	14,070
				Simon	m	13,245
				Zoe	f	13,008
				Hayley	f	13,006
				Anthony	m	12,759
				Stacey	f	11,739
				Leanne	f	10,779
				Michelle	f	10,659
				Martin	m	10,637
				Dean	m	10,350
				Stuart	m	10,217
				Anna	f	10,076
				Catherine	f	9896
				Darren	m	9439
				Gary	m	9265
				Shaun	m	8791
				Helen	f	8647
				Philip	m	8629
				Rachael	f	8570
				Kimberley	f	8330
				Joanne	f	8176
				Kevin	m	7904
				Kerry	f	7578
				Carl	m	6611
				Kathryn	f	6477
				Amanda	f	6444
				Clare	f	5958
				Alan	m	5042
<b>TRADITIONAL NAMES</b>						
146 names - 1,470,464 children						
55% male - 45% female						
46.7% 1987 > 13.0% 2001						
Until the 1990s, naming in the UK was very much dominated by this group of traditional names, without subclusters. Popularity has dramatically dropped since then. The group is strongly related to the Scottish and Gaelic names and has a loading of 0.628 to those. Because of its size and distinct character, the group is presented separately.						
	L	R(%)	N			
DANIEL	0.6	71	1,470,464			
<b>Daniel</b>		m	76,508			
James		m	75,337			
Matthew		m	58,591			
Christopher		m	54,417			
Rebecca		f	48,739			
Emma		f	46,644			
Michael		m	45,720			
Sarah		f	44,304			
Laura		f	44,278			
David		m	43,946			
Adam		m	39,186			
Andrew		m	39,017			
Robert		m	29,832			
Samantha		f	26,832			
Mark		m	25,756			
Rachel		f	25,660			
Paul		m	22,920			
Richard		m	22,367			
Gemma		f	22,182			
Jonathan		m	21,564			
Natalie		f	20,070			
Craig		m	19,334			
Claire		f	19,321			
Lee		m	19,201			



TABLE 5 (Continued)

<b>Patricia</b>	f	983	A group with a couple of typical name pairs of declining popularity.				
Bernadette	f	572					
Theresa	f	548					
<b>Gerard</b>	m	915		L	R(%)	N	
			RUTH	0.46	46	4851	
			JANE	0.63	45	3254	
<b>TRADITIONAL NAMES 3</b>			<b>Ruth</b>		f	3848	
18 names - 50,684 children			Judith		f	656	
4% male - 96% female			<b>Jane</b>		f	1992	
1.3% 1987 > 0.9% 2001			Anne		f	1262	
A group of mainly female names with a traditional flavor.			<b>SCOTTISH &amp; GAELIC NAMES</b>				
	L	R(%)	N	52 names - 122,368 children			
ELIZABETH	0.7	66	26,453	75% male - 25% female			
JOANNA	0.8	54	9978	3.1% 1987 > 2.1% 2001			
ANNABEL	0.5	35	6151	The group of Scottish and Gaelic names is in decline, just like the other traditional English names. The cluster HEATHER seems a bit of an outlier with some names linked to nature.			
ALICIA	0.6	37	5861		L	R(%)	N
JULIA	0.6	39	2241	SCOTT	0.8	54	54,098
<b>Elizabeth</b>		f	15,095	FIONA	0.8	42	25,284
Katherine		f	9993	IAN	0.5	66	12,384
Katharine		f	1365	HEATHER	0.5	55	12,149
<b>Joanna</b>		f	5370	ALISTAIR	0.7	37	9485
Philippa		f	2207	GREGORY	0.7	41	8968
Robin		m	1732	<b>Scott</b>		m	20,107
Susannah		f	669	Ross		m	11,266
<b>Annabel</b>		f	1905	Cameron		m	8560
Louisa		f	1753	Grant		m	3570
Lucinda		f	1109	Fraser		m	2458
Camilla		f	1005	Stewart		m	1871
<b>Alicia</b>		f	2574	Kirstie		f	1680
Felicity		f	1311	Greg		m	1605
Annabelle		f	1014	Blair		m	717
Verity		f	962	Murray		m	565
<b>Julia</b>		f	1728	<b>Fiona</b>		f	5358
Rosalind		f	513	Calum		m	3550
<b>JANE &amp; RUTH</b>			Lorna		f	2551	
5 names - 8105 children			Iain		m	2444	
0% male - 100% female							
0.3% 1987 > 0.1% 2001							

TABLE 5 (Continued)

Ewan	m	1871	<b>Jade</b>	f	17,620
Kirsten	f	1746	Amber	f	5705
Alastair	m	1646	Jasmine	f	4807
Euan	m	1558	Jasmin	f	1382
Catriona	f	1184	Keeley	f	1215
Eilidh	f	1038	Crystal	f	707
<b>Ian</b>	m	7361	Coral	f	616
Neil	m	5023	Kirk	m	585
<b>Heather</b>	f	6136	<b>Danny</b>	m	4190
Frances	f	2154	Ricky	m	3508
Rosemary	f	1628	Tony	m	2561
Hazel	f	1488	Terry	m	1710
Tessa	f	743	Nicky	m	861
<b>Alistair</b>	m	2393	Deanna	f	515
Duncan	m	1655	<b>Shane</b>	m	6277
Kenneth	m	1585	Charlene	f	3423
Finlay	m	923	Sadie	f	1475
Alasdair	m	832	Tammy	f	1310
Donald	m	573	<b>Tara</b>	f	3347
Bonnie	f	562	Tanya	f	2732
Malcolm	m	543	<b>Russell</b>	m	2206
<b>Gregory</b>	m	3248	Nigel	m	758
Douglas	m	1619	Tania	f	739
Gordon	m	1285	Kelvin	m	620
Allan	m	1254	<b>Lindse</b>	f	2168
Alec	m	734			

**GEMS & NAMES IN -y**

29 names - 72,719 children

33% male - 67% female

1.8% 1987 > 1.3% 2001

Gems like jade, amber, crystal and coral inspire parents, who also seem to have affinity to names in -y.

	L	R(%)	N
JADE	0.4	44	32,637
DANNY	0.5	46	13,815
SHANE	0.5	47	13,287
TARA	0.7	53	6079
RUSSELL	0.6	42	4323
LINDSEY	0.6	46	2578

**DIANE**

17 names - 13,642 children

9% male - 91% female

0.5% 1987 > 0.1% 2001

Also this group has two clusters with female names in -y.

	L	R(%)	N
PAULA	0.4	33	3844
VICKY	0.6	32	3207
TRACY	0.7	31	3016
DIANE	0.7	28	1854
RAYMOND	0.5	36	1721
<b>Paula</b>		f	1747
Denise		f	796

TABLE 5 (Continued)

Sandra	f	583	<b>Selina</b>	f	994
<b>Vicky</b>	f	1609	Anita	f	721
Becky	f	1043	<b>TONI</b>		
Carley	f	555	11 names - 13,464 children		
<b>Tracy</b>	f	1212	16% male - 84% female		
Wendy	f	714	0.5% 1987 > 0.1% 2001		
Mandy	f	698	A small group of mainly female names, consisting of typical pairs of names with quite a few ending with an i sound.		
<b>Diane</b>	f	790		L	R(%)
Carol	f	722			N
<b>Raymond</b>	m	1279	TONI	0.7	45
<b>FEMALE NAMES IN -a</b>			DAMIEN	0.5	41
21 names - 35,157 children			STACY	0.4	31
4% male - 96% female			KAY	0.4	28
0.9% 1987 > 0.5% 2001			<b>Toni</b>	f	3538
Many names in this group of female names are ending on -a, with the exception of the clusters GABRIELLE and TERENCE.			Terri	f	1971
	L	R(%)	<b>Damien</b>	m	1810
NATASHA	0.5	65	Carrie	f	1412
KATRINA	0.7	42	<b>Stacy</b>	f	1337
GABRIELLE	0.4	31	Ami	f	638
CARLA	0.5	45	<b>Kay</b>	f	863
TERENE	0.4	35	Jody	f	836
SELINA	0.7	27	<b>FRANK</b>		
<b>Natasha</b>	f	13,900	6 names - 4540 children		
Nikita	f	1398	67% male - 33% female		
<b>Katrina</b>	f	2745	0.1% 1987 > 0.0% 2001		
Sabrina	f	1581		L	R(%)
Sonia	f	1205	MICHEAL		40
Monica	f	609	FRANK	0.7	26
<b>Gabrielle</b>	f	1972	TERESA	0.5	23
Dominique	f	801	<b>Micheal</b>	m	1786
Dionne	f	785	<b>Frank</b>	m	668
<b>Carla</b>	f	2896	Roy	m	590
Justine	f	819	<b>Teresa</b>	f	673
Gina	f	657	<b>DARRYL</b>		
<b>Terence</b>	m	1070	16 names - 16,452 children		
Sarah-Jane	f	638	49% male - 51% female		
Anne-Marie	f	535	0.5% 1987 > 0.2% 2001		

TABLE 5 (Continued)

	L	R(%)	N		L	R(%)	N
MARC	0.6	56	5898	THOMAS	0.7	58	171,897
NIKKI	0.5	33	2976	SAMUEL	0.5	60	126,442
CASSIE	0.5	29	2322	CHARLOTTE	0.5	66	88,596
DARRYL	0.7	29	2242	ALEXANDER	0.5	67	68,840
ROSANNA	0.5	25	1927	OLIVER	0.8	40	30,112
FERN	0.3	17	1087	ALEX	0.7	37	22,140
<b>Marc</b>		m	4601	JOEL	0.6	27	19,286
Jon		m	1297	MOLLY	0.7	23	17,594
<b>Nikki</b>		f	1797	DOMINIC	0.6	41	16,016
Trevor		m	635	JACOB	0.6	32	10,798
Kris		m	544	LYDIA	0.8	27	10,599
<b>Cassie</b>		f	886	ELLA	0.5	31	9844
Kellie		f	723	IMOGEN	0.8	17	8860
Christie		f	713	LOUIS	0.5	28	7879
<b>Darryl</b>		m	1051	LOIS	0.7	19	6839
Maxine		f	821	ELLIOTT	0.6	25	6626
<b>Rosanna</b>		f	1066	FLORENCE	0.5	13	5810
<b>Fern</b>		f	769	LILY	0.6	17	5584
<b>JEFFREY</b>				NATHANIEL	0.8	20	4215
2 names - 808 children				MILLIE	0.5	13	4127
70% male - 30% female				FREYA	0.6	17	3795
0.0% 1987 > 0.0% 2001				HARRISON	0.5	25	3540
				ISOBEL	0.5	30	2909
	L	R(%)	N	<b>Thomas</b>	m		66,014
JEFFREY	0.6	23	808	William	m		27,049
<b>Jeffrey</b>		m	563	George	m		17,861
<b>UPCOMING NAMES</b>				Harry	m		16,335
<b>POPULAR CLASSIC NAMES</b>				Edward	m		11,247
150 names - 652,348 children				Eleanor	f		10,703
72% male - 28% female				Charles	m		9161
10.9% 1987 < 20.0% 2001				Harriet	f		5776
Where the traditional group DANIEL collapsed, this group of popular names doubled in size. The clusters, which mostly have a classic flavour, are varied with male royal names in the cluster THOMAS, Hebrew names in the clusters SAMUEL, JOEL, and JACOB, French names in the clusters CHARLOTTE, DOMINIC, and LOUIS, abbreviated names in the cluster ALEX, names in -y in the cluster MOLLY, names on -a in the cluster ELLA, and so on.				Henry	m		5426
				Frederick	m		2045
				<b>Samuel</b>	m		32,627
				Joshua	m		32,103
				Benjamin	m		31,821
				Joseph	m		29,891
				<b>Charlotte</b>	f		35,957



TABLE 5 (Continued)

Victoria	f	21,627	Serena	f	852
Alexandra	f	9000	Reuben	m	750
Georgina	f	8375	Caleb	m	605
Francesca	f	5634	<b>Molly</b>	f	6653
Claudia	f	1346	Rosie	f	4121
Lucas	m	1342	Daisy	f	3339
Gabriella	f	1325	Poppy	f	1881
Daniella	f	1286	Polly	f	966
Antonia	f	1248	Nancy	f	634
<b>Alexander</b>	m	27,809	<b>Dominic</b>	m	6968
Nicholas	m	15,962	Christian	m	3288
Jason	m	12,716	Sebastian	m	2061
Timothy	m	7093	Tristan	m	1225
Justin	m	1826	Benedict	m	930
Laurence	m	1607	Nicolas	m	522
Jeremy	m	974	Emilia	f	518
Julian	m	853	Beatrice	f	504
<b>Oliver</b>	m	19,109	<b>Jacob</b>	m	8895
Toby	m	3477	Zachary	m	1903
Lawrence	m	1632	<b>Lydia</b>	f	4409
Helena	f	1558	Phoebe	f	3336
Oscar	m	1551	Mia	f	2202
Tobias	m	1400	Esme	f	652
Felix	m	824	<b>Ella</b>	f	4549
Barnaby	m	561	Lara	f	1716
<b>Alex</b>	m	7491	Nina	f	1316
Max	m	5341	Maya	f	762
Elliot	m	3958	Anyia	f	599
Guy	m	513	Jak	m	505
Miles	m	1222	<b>Imogen</b>	f	2857
Leo	m	1134	India	f	986
Hugh	m	1034	Madeline	f	866
<b>Joel</b>	m	4120	Hugo	m	741
Ethan	m	4101	Meghan	f	687
Deborah	f	2812	Maximilian	m	588
Isaac	m	1982	Miranda	f	580
Esther	f	1317	<b>Louis</b>	m	5502
Martha	f	1104	Gabriel	m	889
Miriam	f	895	<b>Lois</b>	f	1754

TABLE 5 (Continued)

				L	R(%)	N
Matilda	f	747				
Eva	f	742	HANNAH	0.5	56	172,678
Jemima	f	721	SOPHIE	0.5	54	118,246
Tegan	f	690	KYLE	0.6	45	56,725
Tabitha	f	612	LUKE	0.4	50	54,359
Fergus	m	570	GEORGIA	0.5	43	35,946
Gregor	m	568	CHELSEY	0.6	34	20,270
<b>Elliott</b>	m	2597	MITCHELL	0.8	19	16,393
Harvey	m	1784	LEAH	0.7	4	13,850
Shelby	f	828	COURTNEY	0.7	28	8224
Spencer	m	726	BRANDON	0.7	27	6730
Frazer	m	691	KANE	0.6	17	6555
<b>Florence</b>	f	1063	LEIGH	0.4	35	6437
Arthur	m	752	DEMI	0.6	24	5726
Alfred	m	715	CURTIS	0.6	29	3750
Theo	m	669	BRYONY	0.4	26	3112
Eliza	f	551	KIERON	0.5	32	2702
Edmund	m	507	HOPE	0.5	15	1973
<b>Lily</b>	f	2896				
Ruby	f	1840	<b>Hannah</b>		f	35,958
Scarlett	f	848	Jessica		f	32,167
<b>Nathaniel</b>	m	1610	Danielle		f	19,397
Myles	m	972	Nathan		m	17,895
Theodore	m	640	Bethany		f	13,716
Flora	f	516	Abigail		f	11,633
<b>Millie</b>	f	1799	Nicole		f	9528
Maisie	f	1539	Naomi		f	6158
Madison	f	789	Marcus		m	3816
<b>Freya</b>	f	1890	Chantelle		f	3604
Rowan	m	875	Rebekah		f	2882
Saskia	f	671	Alisha		f	1935
<b>Harrison</b>	m	2608	Roxanne		f	1754
Maxwell	m	932	Nadine		f	1481
<b>Isobel</b>	f	2439	Simone		f	1436
			Elisha		f	1151
<b>POPULAR TRENDY NAMES</b>			<b>Sophie</b>		f	32,326
141 names - 533,676 children			Katie		f	27,490
34% male - 66% female			Chloe		f	27,141
8.6% 1987 < 15.2% 2001			Ashley		m	14,018
This group almost doubled in size. The clusters seem rather varied, and unorthodox in several cases.			Bradley		m	9337

TABLE 5 (Continued)

Hollie	f	4721	Chelsie	f	732
Wesley	m	1309	<b>Mitchell</b>	m	3415
Lucie	f	1028	Tyler	m	3385
<b>Kyle</b>	m	11,190	Ellis	m	1564
Kayleigh	f	10,346	Taylor	m	1512
Melissa	f	8998	Mason	m	1211
Reece	m	6148	Jordon	m	1107
Karl	m	4976	Charley	f	964
Kelsey	f	1977	Bailey	m	798
Brett	m	1892	Harley	m	723
Vanessa	f	1880	Ebony	f	714
Arron	m	1714	<b>Leah</b>	f	7334
Kimberly	f	1329	Jay	m	3362
Vincent	m	1038	Corey	m	1725
<b>Luke</b>	m	32,391	Casey	f	1429
Jake	m	15,165	<b>Courtney</b>	f	5902
Zara	f	2469	Brooke	f	1332
Zak	m	1579	Chelsey	f	990
Kai	m	1258	<b>Brandon</b>	m	4633
Kira	f	926	Drew	m	959
Jed	m	571	Brad	m	571
<b>Georgia</b>	f	11,069	Chad	m	567
Jodie	f	9425	<b>Kane</b>	m	2030
Jemma	f	4581	Tia	f	1503
Robyn	f	3825	Troy	m	809
Jamie	f	1778	Summer	f	683
Abby	f	1688	Paris	f	646
Damian	m	1203	<b>Leigh</b>	m	1443
Billie	f	866	Warren	m	1374
Todd	m	766	Glenn	m	1355
Stevie	f	745	Hayden	m	1271
<b>Chelsea</b>	f	6885	Keely	f	668
Leon	m	3199	<b>Demi</b>	f	2186
Jordan	f	2437	Morgan	f	1637
Charlie	f	1676	Taylor	f	1373
Leigh	f	1383	Alexandria	f	530
Alex	f	1154			
Levi	m	863			
Cory	m	741			

TABLE 5 (Continued)

<b>Curtis</b>	m	2616	<b>Garry</b>	m	1092	
Candice	f	703	<b>Kaylee</b>	f	508	
<b>Bryony</b>	f	1760	<hr/>			
Brogan	f	689	<b>JOSH</b>			
Jared	m	663	8 names - 23,991 children			
<b>Kieron</b>	m	1945	15% male - 85% female			
Kurtis	m	757	0.3% 1987 < 1.2% 2001			
<b>Hope</b>	f	842	Another few names of trendy character, with a small cluster of female names an i sound.			
Jesse	m	577	<hr/>			
Casey	m	554		L	R(%)	N
<b>NAMES IN -ie</b>			ELLIE	0.5	38	19,494
31 names - 66,474 children			JOSH	0.6	23	3629
79% male - 21% female			KERI	0.3	16	868
1.3% 1987 < 2.0% 2001			<b>Ellie</b>	f	7005	
This group is largely dominated by the cluster JAMIE, whose name ending in -ie is typical. Popularity is growing.			Aimee	f	5904	
	L	R(%)	N	Abbie	f	4538
JAMIE	0.4	46	61,370	Amie	f	2047
APRIL	0.5	37	1988	<b>Josh</b>	m	3051
GARRY	0.4	25	1712	Zack	m	578
KAYLEE	0.5	17	1404	<hr/>		
<b>Jamie</b>	m	24,413	<b>ABBREVIATED NAMES</b>			
Charlie	m	6869	26 names - 238,599 children			
Dale	m	4465	33% male - 67% female			
Billy	m	4030	4.0% 1987 < 7.2% 2001			
Alfie	m	2135	Abbreviated, short names are increasingly popular as well, with clusters that are gender specific.			
Robbie	m	1950		L	R(%)	N
Mollie	f	1911	JACK	0.6	43	77,331
Annie	f	1801	AMY	0.8	63	65,049
Kylie	f	1617	EMILY	0.5	45	57,786
Glen	m	1375	HOLLY	0.5	48	16,617
Tommy	m	1328	KATE	0.8	55	10,394
Kerrie	f	1274	FAYE	0.5	42	6728
Josie	f	1087	SALLY	0.7	50	4694
<b>April</b>	f	1673	<b>Jack</b>	m	41,801	
<hr/>			Ben	m	14,426	
			Sam	m	10,616	
			Joe	m	7226	

TABLE 5 (Continued)

Tom	m	3262	<b>ELISE</b>			
<b>Amy</b>	f	36,408	9 names - 6,230 children			
Lucy	f	24,806	27% male - 73% female			
Katy	f	3835	0.1% 1987 < 0.2% 2001			
<b>Emily</b>	f	30,624		L	R(%)	N
Alice	f	12,028	ELISE	0.8	28	3166
Grace	f	7356	DANE	0.5	22	1915
Paige	f	5389	BRITTANY	0.4	17	1149
Rose	f	1768	<b>Elise</b>		f	1230
Faith	f	621	Tamsin		f	719
<b>Holly</b>	f	13,390	Briony		f	662
Abbey	f	1146	Byron		m	555
Ria	f	805	<b>Dane</b>		m	724
Louie	m	706	Stacie		f	668
Penny	f	570	Stefanie		f	523
<b>Kate</b>	f	7331	<b>Brittany</b>		f	732
Beth	f	3063	<b>WELSH NAMES</b>			
<b>Faye</b>	f	3694	39 names - 99,265 children			
Eve	f	2117	55% male - 45% female			
Fay	f	917	1.4% 1987 < 3.6% 2001			
<b>Sally</b>	f	2715	The clusters with Welsh names in this group enjoy an increasing popularity.			
Jenny	f	1979		L	R(%)	N
<b>FRENCH FEMALE NAMES</b>			LEWIS 0.8 37 52,764			
8 names - 14,983 children			MEGAN 0.8 47 41,820			
0% male - 100% female			RHIANNON 0.9 30 4681			
0.2% 1987 < 0.8% 2001			<b>Lewis</b> m 19,734			
	L	R(%)	N	Rhys	m	5739
AMELIA	0.7	30	14,983	Owen	m	5279
<b>Amelia</b>		f	4479	Dylan	m	4913
Madeleine		f	2313	Morgan	m	2069
Isabel		f	2046	Lloyd	m	1989
Isabelle		f	1810	Kerri	f	1427
Josephine		f	1740	Ceri	f	974
Eloise		f	1583	Carys	f	925
Clara		f	617	Evan	m	919

TABLE 5 (Continued)

<b>Megan</b>	f	20,311	Ciaran	m	2307	
Gareth	m	7913	Sinead	f	2098	
Sian	f	5120	Brendan	m	1773	
Bethan	f	3683	Aiden	m	1668	
Rhian	f	1288	Francis	m	1480	
Cerys	f	1046	Lianne	f	1105	
Cassandra	f	980	Roisin	f	1101	
Kayley	f	926	<b>Conor</b>	m	5462	
Dillon	m	553	Erin	f	3891	
<b>Rhiannon</b>	f	2964	Michaela	f	3170	
Angharad	f	678	Niamh	f	2942	
Haydn	m	614	Ciara	f	2090	
<b>IRISH NAMES</b>			Shauna	f	1702	
45 names - 263,171 children			Ronan	m	867	
69% male - 31% female			Orla	f	698	
3.5% 1987 < 7.3% 2001			<b>Niall</b>	m	2765	
Also the clusters with Irish names more than doubled in size over the last fifteen years.			Aoife	f	668	
	L	R(%)	N	Eoin	m	506
RYAN	0.7	51	236,517	<b>Keira</b>	f	563
CONOR	0.9	35	21,194	<b>FEMALE NAMES IN -a</b>		
NIALL	0.8	26	4409	24 names - 35,497 children		
KEIRA	0.4	20	1051	13% male - 87% female		
<b>Ryan</b>	m	35,587	0.5% 1987 < 1.3% 2001			
Lauren	f	32,162	Names in this group, ending in -a have a Romanic origin. The few male names in this group are Germanic.			
Liam	m	25,151		L	R(%)	N
Jordan	m	20,732	OLIVIA	0.7	31	18,581
Callum	m	16,565	ADELE	0.4	35	11,463
Aaron	m	16,490	NADIA	0.6	27	2584
Sean	m	14,918	ELENA	0.8	19	1,564
Connor	m	13,540	BIANCA	0.5	24	1305
Kieran	m	11,480	<b>Olivia</b>	f	11881	
Shannon	f	10,512	Sophia	f	2723	
Caitlin	f	5794	Isabella	f	1726	
Ashleigh	f	5488	Lucia	f	518	
Siobhan	f	3516	Natalia	f	504	
Aidan	m	3496	<b>Adele</b>	f	1942	
Declan	m	3496				

TABLE 5 (Continued)

Tiffany	f	1873	Heidi	f	1522		
Stefan	m	1870	Aisha	f	1497		
Alana	f	1433	Ayesha	f	1022		
Kristian	m	1326	Hassan	m	981		
Tamara	f	1202	Sana	f	613		
Shanice	f	746	Leila	f	582		
Anton	m	545	Farah	f	504		
Chantel	f	526	<b>Antony</b>	m	2464		
<b>Nadia</b>	f	1813	Janine	f	1068		
Layla	f	771	Kristopher	m	879		
<b>Elena</b>	f	857	Kristina	f	803		
<b>Bianca</b>	f	958	Ellis	f	728		
<b>KARA</b>			Karina	f	578		
6 names - 3,524 children			Corinne	f	524		
0% male - 100% female			<b>Blake</b>	m	769		
0.1% 1987 < 0.1% 2001			Hanna	f	674		
	L	R(%)	N	Leona	f	580	
KARA	0.7	10	1936	<b>Sasha</b>	f	1705	
SAFFRON	0.6	4	1588	Laurie	f	628	
<b>Kara</b>		f	1163	<b>NON-WESTERN NAMES</b>			
<b>Saffron</b>		f	726	<b>ARABIC NAMES 1</b>			
<b>SARA</b>			<b>upcoming</b>				
25 names - 27,243 children			9 names - 19,229 children				
20% male - 80% female			92% male - 8% female				
0.6% 1987 < 0.6% 2001			0.3% 1987 < 0.8% 2001				
In this group, the cluster SARA is special as it includes names from both Western and Arabic origin, which originates in the fact that <i>Sara</i> and <i>Leila</i> belong to the name inventory of both cultures. <i>Heidi</i> is in the cluster because of a strong attraction to <i>Sara</i> .							
	L	R(%)	N		L	R(%)	N
SARA	0.4	30	15,410	MOHAMMED	0.8	4	14,362
ANTONY	0.6	36	7044	MUHAMMAD	0.8	2	2503
BLAKE	0.6	25	2456	KIRAN		1	1469
SASHA	0.7	36	2333	FAISAL		2	895
<b>Sara</b>		f	4687	<b>Mohammed</b>		m	11,116
Yasmin		f	2840	Mohammad		m	3246
				<b>Muhammad</b>		m	1783

TABLE 5 (Continued)

Muhammed	m	720	<b>ARABIC NAMES 3</b>				
<b>Kiran</b>	f	737	22 names - 12,364 children				
<b>Faisal</b>	m	532	71% male - 29% female				
<b>ARABIC NAMES 2</b>			0.3% 1987 > 0.3% 2001				
21 names - 13,246 children							
56% male - 44% female							
0.2% 1987 < 0.5% 2001							
	L	R(%)	N				
MOHAMED	0.9	2	4144	ALI	0.5	5	2864
MARIAM	0.7	2	3785	AMIR	0.5	2	2845
HAMZA	0.8	2	2004	IMRAN	0.9	2	2761
JAMAL	0.7	2	1276	SALMA	0.9	2	2318
AMINA	0.6	1	1185	PRIYA	0.3	1	1576
KHALID	0.7	2	852	<b>Ali</b>		m	1375
<b>Mohamed</b>		m	1383	Omar		m	1102
Ahmed		m	1031	<b>Amir</b>		m	576
Fatima		f	867	Adil		m	528
Maryam		f	863	<b>Imran</b>		m	944
<b>Mariam</b>		f	751	Usman		m	707
Zainab		f	705	Umar		m	691
Ibrahim		m	646	<b>Salma</b>		f	649
Zahra		f	633	Sadia		f	567
Hasan		m	536	<b>Priya</b>		f	717
Zain		m	514	<b>ARABIC NAMES 4</b>			
<b>Hamza</b>		m	819	<b>unclassified</b>			
Bilal		m	638	6 names - 1,881 children			
Anisa		f	547	26% male - 74% female			
<b>Amina</b>		f	760	0.1% 1987 > 0.0% 2001			
					R(%)	N	
				NAZIA	1	765	
				FARZANA	1	622	
				NADEEM	1	4940	

## Discussion and further research

Given names of children can reveal cultural, ethnic, linguistic, and socioeconomic backgrounds of parents, but the relationships are usually complex, hidden, and noisy. And although some relationships are much stronger than others, in all cases extreme care should be taken in their interpretation. We had the advantage that we could base our analysis on the names of children born in the same family, rather than a mere



list of names of children. This enabled us to reveal structures in naming that are otherwise very difficult to unravel. Whereas ethnic and linguistic backgrounds of exogenic names can be inferred from popularity in cultures, countries, or linguistic communities from which they originate or where they currently have a high frequency, this is much less obvious for cultural and socioeconomic backgrounds in endogenic names. The self-organizing methodology we propose has the advantage that it does not require the assumption of explicit underlying factors beforehand. The interpretation of the results comes afterwards, and could be based on correlations with factors like income, educational level of the parents, or geodemographic spread. The advantage is that these correlations need not to be based on individual names, but on the aggregated level of name groups, which statistically makes a much stronger case.

The interpretation of name groups in the present study is still impressionistic. Its validity resides in the plausibility of the results, the reason why we included the full lists in Tables 4 and 5. Further research is under way that will link socioeconomic information and names of children, both available at the family level. The power of using name groups in relation to geodemographic spread has been demonstrated already by Bloothoof et al. (2004). In contrast to individual names, name groups have a sufficient frequency at the level of postal code area (with a total of 3961 areas in the Netherlands) to define a reliable profile of their presence. Using factor analysis on these profiles, a limited number of most characteristic name group profiles can be distinguished. By attaching the best fitting of these profiles to each postal code area, a map of the Netherlands can be drawn that highlights naming preferences. The same can be done for the UK.

There is a correspondence between the analysis of the linked information of given name and surname in the same person and CEL factors as performed by Tucker (2003), and the linked information of given names of children in the same family in our study. In Tucker's case, the knowledge of onomastic experts on the origin (or usage) of first names could be used to obtain probabilities on the otherwise unknown ethnic and linguistic background of surnames. However, the combination given name and surname can be seen as a pair of linked information, just like the pair(s) of given names in the same family. With minor adaptations, the self-organizing clustering we applied is methodologically applicable to any combination of types of names, including the combination of given name and surname. As in our case, only when it comes to the interpretation of resulting clusters *ex hoc* expert knowledge will be indispensable.

Our approach is strongly motivated by the opportunity to let the data speak for themselves. The current procedures are mathematically well defined — much more solid than in the original publications of Bloothoof (2001, 2002). But, to be successful, we learned that the analysis procedure should be very carefully tuned. A few notes:

- (1) The analysis should be limited to names with relatively high frequency, and for each name there should be enough names of brothers and sisters to identify statistical relationships reliably. Once name groups have been identified on such a solid basis, names with lower frequencies could be associated to

these, although uncertainties and misallocations grow rapidly with decreasing numbers.

- (2) Because of the severe conditions on name frequencies in our analysis, many CEL types that are identified for the UK by Mateos et al. (2007) do not show up in our analysis. A huge majority (136 out of 186) of their CEL types have a frequency of less than 10,000 people. It is likely that names in these CEL types cannot be distinguished by a statistical analysis of systematic co-occurrences, but only under the availability of expert knowledge and additional information. On the other hand, the largest CEL types found by Mateos were England (31 million), Scotland (4.7 million), Ireland (3.2 million) and Wales (3.1 million), on a total sample of 46.4 million people. It is within these major CEL types that our analysis provides details related to subcultures.
- (3) The name groups have a considerable overlap, with the exception of the culturally very distinct groupings of Western and Arabic and Turkish names. Besides this major division, there are almost no names that are only found in combination with names from their own group. The most complicating are the popular names, which are found in combination with virtually all other names (within the same culture), and obscure otherwise clear relationships. Our approach concentrated on local relations (the nearest neighbours only), utilizing the relatively few prototypical parents that fully follow a subculture naming pattern for their children.
- (4) The way the relation between names is defined, is central to any analysis method. The ideal relation measure is symmetric between names and insensitive to the popularity of the names. We believe that our attraction measure fulfils these requirements quite well, although there remains a difficulty in the estimation of the relevant part of the population that would consider particular names for their children.
- (5) The name space consists of clusters of names of very different sizes, from strongly related name pairs to much bigger groups of names. It is important not to lose sight of the smaller clusters that often demonstrate interesting details. A hierarchical structure might best represent the supposedly layered structure of the name space. The standard hierarchical cluster techniques fail, however, to produce meaningful results.
- (6) The large variation in the size of name clusters does not make it attractive to use a k-nearest neighbours approach since a distance measure built on a fixed number of neighbours endangers the identification of the smaller clusters. A carefully tuned exponential Minkowski weighting of the attraction between a name and a name cluster solves the issue in an elegant way.

Whereas this study has a methodological focus, the application on a large population dataset from both the Netherlands and the UK also allows for some contrasting observations. A very important observation is that the method seems to work equally well for the Netherlands and the UK. Of course, the detailed results are quite different, but the idea that we may learn from parental preferences holds for both countries, and this suggests a general applicability.

As expected, in both countries, clusters emerge of names belonging to specific CEL groups, both at the cluster level and the grand cluster level. In addition, clusters emerge with defining characteristics that are not so obvious from standard classifications or social stratifications of society into groups. For instance, for the UK, beside clusters of typical Scottish, Welsh, Irish, Italian, French, Hebrew, and Arabic names, we found clusters of gems name for girls (*Jade, Amber, Crystal, Coral*) and a group of names ending in -y for boys (*Danny, Ricky, Tommy*); for the Netherlands, beside clusters of Frisian, Nordic, French, Italian, English, Hebrew, Turkish, and Arabic names, a cluster for flower names for girls (*Iris, Fleur, Roos*) and classical Dutch names (*Floris, Laurens, Maurits*) emerged.

In the UK the group DANIEL, containing 146 traditional names, covered in the 1980s still half of the total sample, but this dramatically reduced to 13% within one generation. A comparable decline of traditional names is observed in the Netherlands, where it had an earlier start. The clusters JOHANNES and JAN, with 197 traditional names, show a decline from 12.3 to 5.7% between 1985 and 2004, but these traditional names had a coverage of over 80% in the middle of the twentieth century.

The decline of traditional names has opened the way for a rich and much more varied naming pattern in both countries, which very clearly indicates different motives and backgrounds of parents. Common tendencies in the Netherlands and the UK are the emergence of clusters with trendy short or abbreviated names, names of gems and from nature for girls, alongside names connected with regional identity and language (Frisian in the Netherlands, Celtic/Gaelic in the UK), foreign names (although they form a substantial larger part in the Netherlands than in the UK), names from the Koran or Bible, and Hebrew names. An interesting feature in both countries is that several clusters have a dominating gender. That implies that parents have a strong gender-related preference. When they name their daughters after flowers, this preference is much stronger than a likely much more varied (and therefore individually less probable) choice for names for their sons.

We have studied the existence of name groups within a limited period of about twenty years. Even in such a relatively short period the increase and decline of groups could be observed. This indicates that, just like names themselves, also name groups have a life cycle. To better understand the origins of name groups, knowledge of their dynamics is necessary: when and where did a name group arise, when was the period of maximum popularity, did a name group emerge from a disappearing group, and so on. We recently acquired a corpus of the given names of the full population of the Netherlands (16 million) from the civil registration, covering dates of births throughout the twentieth century, and including family relations and places of birth. With this rich source at hand we hope to unravel the dynamic mechanisms of name groups in more detail.

## Acknowledgement

We wish to express our gratitude to the Social Security Bank in the Netherlands and the HM Revenue and Customs in the United Kingdom for making available the

invaluable corpora of first names. The opinions expressed in this article are those of the authors and cannot be held to represent the views of the data providers.

## Notes

- <sup>1</sup> Lloyd et al. (2004) use the 1881 Census of Population survey and the 1998 Electoral Register in Britain to find out the geographical origin of surnames and the geographical movement of names in time. Dividing Britain in postcode areas, they calculate indices of surnames to map the geographical distribution of (groups) of surnames over the country and show that particular (indigenous) surnames are highly concentrated in some parts of the country. Combining these findings with the information contained in name datasets of the US, Canada, Australia, and New Zealand makes it possible to track migration flows of the past.
- <sup>2</sup> Interestingly, the same type of methodological difficulties (among which decisions with respect to accuracy, coverage, normalization of scores, and setting thresholds for including names) encountered in the studies mentioned above appear in our research.
- <sup>3</sup> Each of the 1409 names can make a name pair with any of the 1408 names, so the maximum number of name pairs is  $1409 \times 1408$ , amounting to 1.96 million name pairs.

## Bibliography

- Bloothoof, Gerrit, 2001. 'Naming in Dutch Families Between 1983 and 1999,' *Naamkunde*, 33: 1–47 (in Dutch).
- Bloothoof, Gerrit, 2002. 'Naming and Subcultures in The Netherlands,' *Proceedings of the International Conference of Onomastic Sciences*, ed. E. Brylla and M. Wahlberg, Uppsala, Part 2, 53–62 (published 2006).
- Bloothoof, Gerrit, van Nifterick, Emma, and Gerritzen, Doreen, 2004. *On First Names — How The Netherlands Gets its First Names*, Utrecht: Het Spectrum (in Dutch).
- Everitt, Brian, Landau, Sabine, and Leese, Morven, 2001. *Cluster Analysis*, Oxford: Oxford University Press.
- Galbi, Douglas A., 2002. 'Long-Term Trends in Personal Given Name Frequencies in England and Wales,' Working paper version 1.1, Federal Communications Commission, Washington DC, USA.
- Fryer, Roland G., and Levitt, Steven D., 2004. 'The Causes and Consequences of Distinctively Black Names,' *The Quarterly Journal of Economics*, 119(3): 767–805.
- Harding, Seeromanie, Dews, Howard, and Ludi Simpson, Stephen, 1999. 'The Potential to Identify South Asians Using a Computerised Algorithm to Classify Names,' *Population Trends*, 97: 46–49.
- Gorsuch, Richard L., 1983. *Factor Analysis*, New Jersey: Erlbaum.
- Lauerdale, Diane S., and Kerstenbaum, Bert, 2000. 'Asian American Ethnic Identification by Surname,' *Population Research and Policy Review*, 19: 283–300.
- Lloyd, Daryl A., Webber, Richard, and Longley, Paul A., 2004. 'Surnames as Quantitative Evidence Resource for the Social Sciences,' <http://www.casa.ucl.ac.uk/surnames/papers.htm>
- Mateos, Pablo, Webber, Richard, and Longley, Paul, 2007. 'The Cultural, Ethnic, and Linguistic Classification of Populations and Neighbourhoods Using Personal Names,' CASA working paper 116, [http://www.casa.ucl.ac.uk/working\\_papers/paper116.pdf](http://www.casa.ucl.ac.uk/working_papers/paper116.pdf)
- Mateos, Pablo, 2003. 'A Review of Name-based Ethnicity Classification Methods and Their Potential in Population Studies,' *Population, Space and Place*, 13: 243–263.
- Tucker, D. Kenneth, 2003. 'Surnames, Forenames, and Correlations,' *Dictionary of American Family Names*, ed. P. Hanks, New York: Oxford University Press, xxiii–xxvii.

## Notes on Contributors

Loek Groot is associate professor in economics of the public sector. His research interests include multidisciplinary economics, sport economics, social justice, and the

interaction between social security and the labor market. Currently he is working on a project on the transnational distribution of public goods.

Gerrit Bloothoof is researcher at the Utrecht Institute of Linguistics. His interests are in speech technology, phonetics, and onomastics. He has published on nominal record linkage, corpus-based methods for name standardization, naming after relatives, and was co-author of a landmark book on first names in the Netherlands. His current focus is on the methodology of studies on full population data from the civil administration.

Correspondence to: Gerrit Bloothoof, Utrecht Institute of Linguistics-OST, Utrecht University Trans 10, 3512 JK Utrecht, The Netherlands.  
Email: [Gerrit.Bloothoof@let.uk.nl](mailto:Gerrit.Bloothoof@let.uk.nl)