



+31 153010030



www.sample.solutions



info@samplesolutions.eu



@sample_solutions

ISSUED BY
**Sample
Solutions**
DATE UPDATED
30/03/2018

Sampling Methodology for low incidence audience

Methodology for targeting hard to reach
audience with a migration background

INTRODUCTION TO THE METHODOLOGY



Often when telephonic consumer fieldwork is conducted in Europe, companies can select between targeted B2C sample and national representative RDD sample. For targeting a specific part of the population, it is more efficient to make use of lifestyle sample which included listed telephone numbers which contain aspects like age, gender and other lifestyle traits. Nevertheless, this has as an advantage that a bias is introduced due to the fact that the listed lifestyle sample is built up. Another approach would be RDD sample which however would yield in lower incidence rates and higher cost during fieldwork.



HIGH COVERAGE

Method can be used for more than 60+ countries worldwide.



INOVATIVE

A breakthrough in our in-house R&D



REPRESENTATIVE

Through combination of Big Data & Probability sampling, the desired population can be

FLAGGING USING ONOMASTIC PROCESSING

Sample Solutions makes use of the onomastic approach which takes into account the first name and surname of different ethnic groups. By using the most frequent first names and surnames of a certain group it is possible to narrow down the target population from the general country population. An additional step can include an analysis of combined names and removal of similar names (surnames that appear in the local population but also the targeted ethnic group). This method is then applied to residential listings which contain all listed mobile and landline phone number for a specific country. A subset of the residential telephone sampling frame is thus selected and used as sampling frame.

For example, each name combination uses a probability indicator of being muslim such that a threshold can be set. Eg. a value of 0.7 would give people with an estimated 70% probability of being muslim or higher. Lower values are ignored.

Full name	First Name Score	Last Name Score	Combined Score
FirstName LastName	$0 < 1$	$0 < 1$	$0 < 1$

A machine learning algorithm then compares existing data sets to provide the combined score.

LIST ASSISTED **APPROACH**

The list assisted approach makes use of white pages data which exists for most of Western European countries with moderate coverage of households.



For the case of Germany, the database consists of 20.9 million records. Germany having an average household size of 2.1 it can be seen that more than 50% of the population is covered within this frame (considering the population of 81.4 million).

The first names and surnames are then processed via the the onomastic filter to account likelihood for being a immigrant household.

BIGDATA LINKAGE



Many current discussion boards, social media platforms and listing pages make use not only of the email address but also the mobile phone numbers, especially with regard to TFA. Therefore it is possible to connect information from public information to phone numbers which can help in profiling a cell phone rdd sample. Additionally in many cases a full name is provided thus that the onomastic filter can be applied



ADVANTAGES AND DISADVANTAGES

In this section the various advantages and disadvantages of the two approaches are outlined.

Advantages

Looking at the listed sample from white pages, it will be biased towards mid-age ranges (30+) as the younger population will be underrepresented in the phone book. Furthermore most of the numbers will be landline numbers which will have a higher proportion of female respondents.

Disadvantages

Regarding the BigData Linkage, all sample will be cell phone based and also will have a slightly higher proportion of males while having a younger target group. Furthermore only people with internet activity will be in the sampling frame.



+31 153010030



www.sample.solutions



info@samplesolutions.eu



@sample_solutions



COVERAGE

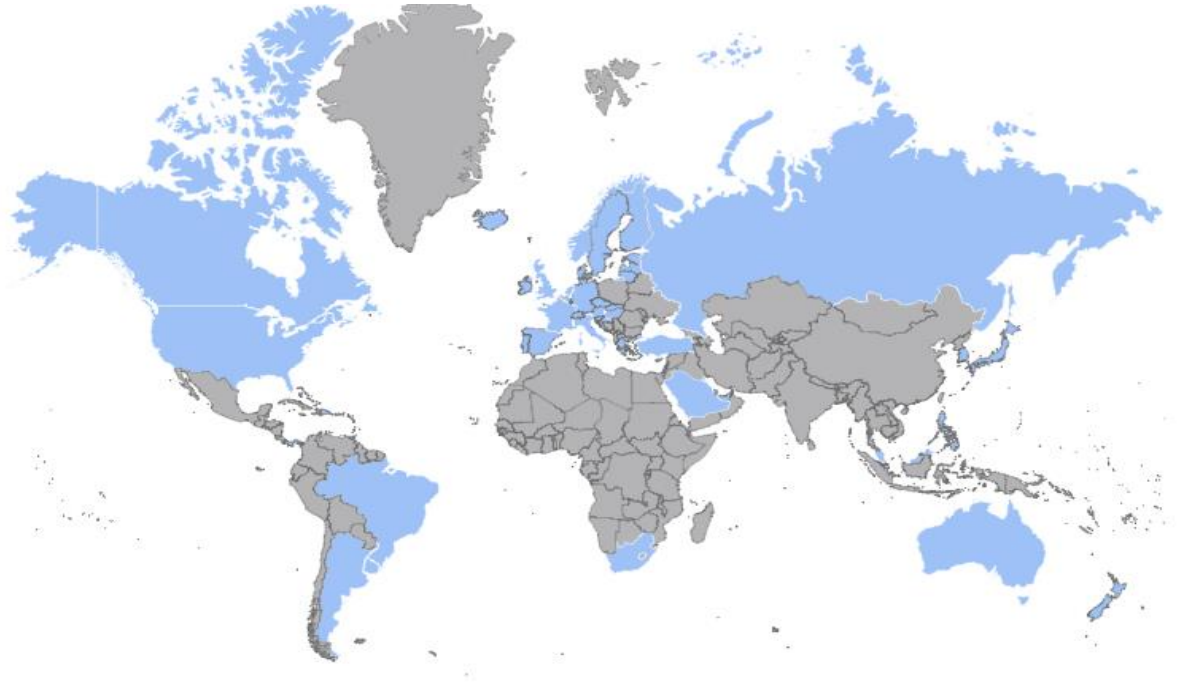
For Sampling immigrant population in different countries, Sample Solutions uses two source to provide largest possible coverage.

The first source is based on listed directories where available (white pages data). The second approach uses a proprietary technology in which RDD cell phone number are matched with BigData from various online sources such as discussion board, listing pages and social media profiles.

In order to sample immigrant audiences in Europe it is essential to propose a method to define the target population since a concise sampling frame does not exist. Sample Solutions makes use of the onomastic approach which takes into account the first name and surname of different ethnic groups. By using the most frequent first names and surnames of a certain group it is possible to narrow down the target population from the general country population. Sample Solutions Europe uses the 1000 most frequent surnames and first names. In some cases however the data is limited to 100 most frequent surnames and first names. An additional step can include an analysis of combined names and removal of similar names (surnames that appear in the local population but also the targeted ethnic group). This method is then applied to residential listings which contain all listed mobile and landline phone number for a specific country. A subset of the residential telephone sampling frame is thus selected and used as sampling frame.

In the previous step the creation of a sampling frame is elaborated on. This sampling frame is a subset of the general population but with characteristics of specific surnames and first names and should be representative in terms of geography, age and gender when compared to the general telephone directory. It needs to be noted that a bias for the telephone directory nevertheless exists (gender, age or education).

Each record in the target sampling frame then receives a randomized number. When drawing a sample from this target population – a random selection is made thus that in terms of eg. Geography, the created sample should be representative of the target audience.



Countries that can be covered by this methodology