



**Scientific Institute of Public
Health (IPH)**

Health Interview Survey 2001

Protocol for the Sampling Design

Version July 20, 2001



Wetenschappelijk Instituut voor
Volksgezondheid
Dienst Epidemiologie

Demarest Stefaan
Tafforeau Jean
Van Oyen Herman



Limburgs Universitair Centrum
Center for Statistics

Bruckers Liesbeth
Molenberghs Geert
Tibaldi Fabián
Van Steen Kristel

1.	INTRODUCING THE HEALTH INTERVIEW SURVEY	4
1.1	Objectives of the survey	4
1.2	Target population, study population and the sampling frame	5
2.	SAMPLING DESIGN	7
2.1	Introduction	7
2.2	Sample Size	7
2.3	Stratified Clustered Multi-stage design	8
2.3.1	Motivation	8
2.3.2	Overview of the design	9
3.	NON RESPONSE	13
3.1	Non-response at household level	13
3.2	Non-response at Individual Level	15
4.	PROCEDURE IMPLEMENTATION	16
4.1	Stratification	16
4.2	Multi stage sampling	17
4.3	Systematic sampling	18
4.4	Clustering: Selection of a sets of households within each municipality	22
4.5	Advantages of the design	26
4.6	Activation of the households	26
4.6.1	Replacement of the households	28
4.6.2	Quarters 2-4	30
4.7	Selection of the household members	31
5.	CONTACTING THE RESPONDENTS	32
5.1	The announcement of the survey	33
5.2	Procedure for contacting a household	33
5.3	Selection of the respondents	36
5.3.1	Note on non-coverage	38
6.	INTERVIEWERS AND FIELDWORK	40
6.1	Training	40
6.2	Guidelines	41
6.2.1	Use of proxies	41
6.2.2	Institutionalized Elderly People	42
6.2.3	Interpreter	43
7.	QUALITY CONTROL BOARD	44
7.1	Indicators for the fieldwork	45

7.1.1	Management of the interviewers	46
7.1.2	Performance of the interviewers	46
7.1.3	Quality of the fieldwork	48
7.1.4	Progress of the fieldwork	48
8.	APPENDICES	51
8.1	Selected Sample	51
8.1.1	Introduction	51
8.1.2	The distribution over the regions	52
8.1.3	The distribution over the provinces	53
8.1.4	The systematic sampling of the municipalities	53
8.1.5	The selection	57
8.2	Graphs corresponding to the selection procedure	63
8.3	Indicators for the Fieldwork	69
9.	LIST OF ABBREVIATIONS	71
10.	REFERENCES	72

1. Introducing the Health Interview Survey

1.1 Objectives of the survey

The main objective of the Health Interview Survey is to give a description of the health status of the population in Belgium in general and of the three regional subpopulations (Flemish, Walloon and Brussels region) in particular. The idea is to obtain information on how people experience their health, to what extent they make use of health care facilities, and how they look after their health by adopting a certain life style or relying on preventive and other health services.

More specifically, the goals of the survey can be summarized as follows:

- Identification of health problems
- Description of the health status and health needs of the population
- Estimation of prevalence and distribution of health indicators
- Analysis of social (in)equality in health and access to the health services
- Study of health consumption and its determinants
- Study of possible trends in the health status of the population

A health interview survey provides one possible channel through which such information can be obtained. On the basis of this survey, assessing a large variety of personal, social and material characteristics, life habits and conditions, determinants for public health can be traced and identified. The ultimate goal of the health interview survey is to be an integrated instrument in decision making, while mapping out an adequate health policy. The aim is stating priorities in policy development and to monitor the progress of populations' health.

1.2 Target population, study population and the sampling frame

The objective of the Health Interview survey – i.e. to give a description of the health status of the population in Belgium-, leads to the broad definition of the target population as all people residing in Belgium. Due to (1) the selection of a sample frame and (2) practical considerations and decisions, not all persons belonging to this target population will or can be considered for the survey. This means that the study population – the population that can be defined accurately and reached in the study – does not cover the target population completely:

- (1) As a consequence of using the National Register as the sample frame, people not listed in the Register (homeless) are excluded from the survey but also new created households are sometimes not registered.
- (2) Not individuals, but households will be invited to participate in the survey. A household is defined as all people living at the address of the reference person. The selection of households is consequently a selection of reference persons. This implies that people not living with a reference person listed in the National Register are excluded from the survey. Furthermore, it was decided not to include ‘specific’ households, that is, households with a reference person living in
 - an institution, with the exception of institutions for elderly
 - a religious community or cloister with more than 8 persons
 - a prison
 - a psychiatric institution
 - a health institution

Excluding these households from the survey generates specific problems for the fieldwork, since only the ‘specificity’ of the household can be known. The National Register does not contain information whether the address is an institution, prison,... . These problems will be discussed in depth in Section 5.2.

Although the most actual version of the National Register is used, the situation of households can be different than the one used in the Register; one of the most important advantages is that the NIS has a direct access of the NR and this is a kind of constantly updated information. Even though when a change occurs in within the household it may take up to 1 month before the information in the Register is updated (these changes have to be reported to the local authorities and

transmitted to the Register). Given the use of a progressively deteriorating National Register copy, a verification process is necessary. This verification process will identify households that are no longer eligible or households, which need an update of the reference person or the address. This is done as late as possible, approximately one month before each quarter. Two criteria are checked (1) the vital status of the reference person and (2) the current address of the main residence of the reference person. This control is conducted in an automated fashion. Only in the case the reference person died in a household with two or more members and there is no partner (less than 1 percent of the selected households), a manual on-line search is necessary. The time needed for the control process is approximately 1 week.

Even when using the most actual version of the National Register, the real composition of a household can be different compared to the administrative composition. Next to the fact that it takes some time before the Register is adapted, remains the fact that households neglect, find it unnecessary or do not want to report changes. General guideline for the interviewers is that the real situation ALWAYS overrules the administrative situation. This means e.g. that

- when an 'administrative' member left the household for a period of at least one-year, he/she will be considered as not being a member of the household. This rule is not applicable for elderly people, who have their official address in the household while remaining in an institution for the elderly.
- compared to the administrative data, new members can have joined the household. As far as these are not considered as guests, they should be treated as being a member of the households.

2. Sampling Design

2.1 Introduction

Sample surveys can be distinguished from other statistical collections by their particular approach to two questions. The first concerns the units from which the population data are to be collected (the sample selection). The second relates to how to infer relevant conclusions, including estimates, concerning the population surveyed is from the data collected. Sampling theory is concerned with the answers to these two questions (Ref 10). In this chapter we will introduce and motivate the procedure used to select units from the target population which has been delineated in Section 4.

The results of sample surveys are always subject to some uncertainty because only part of the population has been included and because of errors of measurement. Simply increasing the sample size costs both in terms of time and money. Hence, the specification of the degree of precision wanted in the results is an important consideration.

2.2 Sample Size

The total number of successful interviews for the sample of 2001 is set to 10,000. This sample size is based on sample size calculations performed during pre-analyses for the health interview survey 1997 (Ref 16), taking into account specific budget constraints and the available logistic means.

On the basis of the preliminary reports and the analysis of the HIS 1997 (Ref 16), the efficiency obtained in estimation at the national and regional level appeared to be sufficient. It was however too small for estimation purposes at the provincial level. In order to answer specific requirements of the provinces in the HIS2001, provinces were encouraged to make extra funds available, enabling a province-specific analysis. To keep the fieldwork within limits, it was decided not to exceed 13,000 as a total number of interviews. Four provinces agreed to pay for the oversampling and to increase the number of interviews within their province. For the province of Antwerp the number to be oversampled was fixed to 350, for the province of

Limburg to 200, for the province of Luxemburg to 1000 and for the province of Hainaut to 500. As a result of this oversampling the final sample size, including the base sample of 10,000 interviews and the oversampling, is **12,050**.

2.3 Stratified Clustered Multi-stage design

2.3.1 Motivation

In the design of the sampling scheme both the coverage of the people in Belgium and the logistic feasibility of the fieldwork are important concerns. Even when a relatively exhaustive list is available (such as the National Register), a direct selection from this list would be too expensive, because the spread would be too wide. Cost savings may allow the investigators to use a larger sample size than they could use for a simple random sample of the same cost. Therefore, a **multiple stage design** is a preferable solution. In this design municipalities serve as primary selection units, while households within the municipalities and individuals within households are respectively second and third-stage units.

Municipalities are established administrative units, they are stable (in general those units do not change during the time the survey is conducted), and they are easy to use in comparison with other specialized sources of data related to the survey. Municipalities are preferred to regions or provinces, because the latter are too large and too few. The great variation in the size of the municipalities is controlled for by systematically sampling within a province (Section 2) with a selection chance proportional to their size.

Choosing a **stratified** sample instead of a simple random sample can be motivated as follows (Scheaffer Ref. **12**). Sample surveys displaying small variability among the measurements will produce small bounds on the errors of estimation. In other words, stratification may produce a smaller bound on the error of estimation than would be produced by a simple random sample of the same size.

This result is enforced if strata are largely homogeneous. In the HIS2001 there are two stratification levels (at the regional and the provincial level). Within a region, a proportional representation per province in the base sample of 10,000 is sought. A simple random sample of municipalities within a region would ascertain this condition from the sampling framework point of view. Resulting differences are regarded as purely random. However, stratifying proportionally over provinces controlled this random variation further.

Within each municipality, a sample of households is drawn so that groups of 50 individuals can be interviewed in total. Clustering also takes place at the household level since members of the same household are more alike than persons not belonging to the same household.

Whereas the stratification effects and the systematic sampling according to municipalities (Section 4.3) have the effect of increasing the precision, the **clustering effect** (selecting groups of 50) might slightly reduce precision, since units will resemble each other more than in a simple random sample. However, since stratification is based on unequal probabilities (to guarantee meaningful sample size per stratum) a slight decrease in overall efficiency is to be expected. The effects due to clustering and stratification observed in the HIS1997 are very mild and do not outweigh the advantages. This design choice enables indeed persons to be sampled from abbreviated listings and, hence, reduces the survey field worker's travel distance significantly. The effects observed in HIS97 are very mild. Nevertheless, this design choice enables persons to be sampled from abbreviated listings and hence survey field worker's travel is correspondingly reduced (Ref. 11).

In summary, in the light of the previous remarks multistage sampling is the appropriated way to get access to individuals. An overview of the steps in the selection procedures is given in the next section.

2.3.2 Overview of the design

The final sampling scheme, i.e. the mechanism to get a random sample of households and respondents is a combination of several sampling techniques: stratification, multistage sampling and clustering. The selection process consists of the following steps:

1. *Regional stratification.* Belgium is divided into 3 regions, the Flemish Region, the Walloon Region and the Brussels Region, for which the number of interviews has been predetermined. The reason for this stratification is to ensure that for each region inference is possible with nearly the same precision. The number of interviews to be

interviewed is fixed to 3500 for the Flemish and Walloon region and 3000 for the Brussels region. These figures do not include the oversampling.

2. *Stratification* at the level of *the provinces*. This second level of stratification is done to improve the quality of the sample over a simple random sample. In particular a balanced geographical spread is achieved. For the base sample, the sample size within the provincial stratification is proportional to the population size of the province. For the provinces agreeing with an oversampling, the number of interviews obtained via the proportional stratification is increased with the number of interviews the province is paying for. Furthermore, there is the special case of the province of Liege as the sample size of the German Community (which is geographically located within the province of Liège) has been predetermined. The province of Liège has been split into two strata: the German community and the rest of the province.
3. Then, within the strata, units are accessed in two (for the households (HH)) or three (for the individuals) stages:
 1. Municipalities are selected with a selection chance proportional to their size, within each stratum. These municipalities are called the *Primary Sampling Units* (PSU). To keep the fieldwork feasible, each time a PSU is selected a group of 50 individuals have to be interviewed successfully during the year 2001.
 2. Within each municipality an equi-probable sample of households, the *Secondary Sampling Units* (SSU), is drawn such that 50 individuals per group can be interviewed in total.
 3. Finally, within each household, at most four individuals, the *Tertiary Sampling Units* (TSU), are chosen. Only questioning the reference person within a household would not enable us to give a good picture of a household's health status. For household with four members or less, all members are selected. For households with at least five members, the reference person and his/her partner (if any) are selected. Among the remaining household members a selection is made according to the birthday rule (Section 4.7), so as to yield four household members selected. Interviewing more than four persons within a household is inefficient because of the familial correlation and the burden on the household would be too large.
4. To assure representativity over time, interviews are spread over the year so that each quarter is comparable in terms of number of successful interviews. The quarters are defined as follows: Q1: January-March; Q2: April-June; Q3: July-September and Q4: October-December.

All of these aspects and the practical implications to comply with this procedure will be discussed in the next Chapters.

To finish this chapter we present the results of the sampling procedure in Table 1. For each province the following items are displayed:

- the population size,
- the fraction relative to the total number of inhabitants of the region,
- the theoretical number of individuals to be interviewed according to the base sample,
- the corresponding number of groups of 50,
- the number of individuals to be sampled additionally (oversampling),
- the effective number of individuals to be interviewed (a multiple of 50),
- the corresponding number of groups of 50 individuals to be interviewed,
- and the probability of being selected (the sampling rate/1000).

From this table it can be deducted derived that the chance of being selected differs appreciably from region to region. In total 12,050 interviews will be done. Hence, as the total population of Belgium is about ten millions, the overall chance of being selected is almost 1.2/1000. In the Flemish region the relative chance is about a half (0.58/1000), in the Walloon region it is about 1.3/1000 (1.50/1000 versus 1.18/1000). In the Brussels region the chance that an individual is selected is approximately 3.1 out of 1000. Within each region the selection probabilities by province differ too, mainly because of the oversampling in some provinces. The chance of being selected is highest in Eupen (4.3) and Luxemburg (5.1) because of the (predetermined) oversampling.

It should be stressed that this difference in selection probability does not affect the representativity of the sample. By taking in each region "nearly" the same number of respondents, the precision and hence the quality of resulting inferences at regional level is made about equal. For the combination of the results to the national level some precision is lost but a valid estimate and inference can be obtained by weighting each region/province with weights inversely proportional to the selection probability.

Table 1 : The distribution of the sample size by provinces.

Province	Total Population (1/1/2000)	%Pop within region	Theoretical Number of interviews	Base Sample	Over-sampling	Effective number of interviews	Number of Groups	Number of Municipalities	Selection rate per 1000
Antwerpen	1640966	27.7	969	950	350	1300	26	19	0.79
Flemish Brabant	1011588	17.1	598	600	0	600	12	12	0.59
Limburg	787491	13.3	465	450	200	650	13	13	0.83
East Flanders	1359702	22.9	803	850	0	850	17	17	0.63
West Flanders	1127091	19.0	665	650	0	650	13	13	0.58
Flemish Region	5926838	100	3500	3500	550	4050	81	74	0.68
Hainaut	1280427	39.3	1256	1250	500	1750	35	27	1.37
Liège (without the German community)	947787	29.0	929	950	0	950	19	19	1.00
Luxemburg	245140	7.5	241	250	1000	1250	25	21	5.10
Namur	441205	13.5	433	450	0	450	9	7	1.02
Walloon Brabant	347423	10.7	341	300	0	300	6	6	0.86
Walloon Region (without the German community)	3261982	100	3200	3200	1500	4700	94	80	1.44
Eupen	70472	1.1	300	300	0	300	6	6	4.26
Walloon Region	3332454	100	3500	3500	1500	5000	100	86	1.50
Brussels	954460	100	3000	3000	0	3000	60	18	3.14
BELGIUM	10213752	100	10000	10000	2050	12050	241	178	1.18

3. Non response

It is possible that an entire household, or a subset of it, will not be able or will not want to collaborate for various reasons. It is important to point out here that the participation in the survey is not compulsory. In this chapter we will define different household categories. In terms of the result of the contacting process we distinguish contactable and non-contactable households. From the response point of view we distinguish between participating and non-participating households.

3.1 Non-response at household level

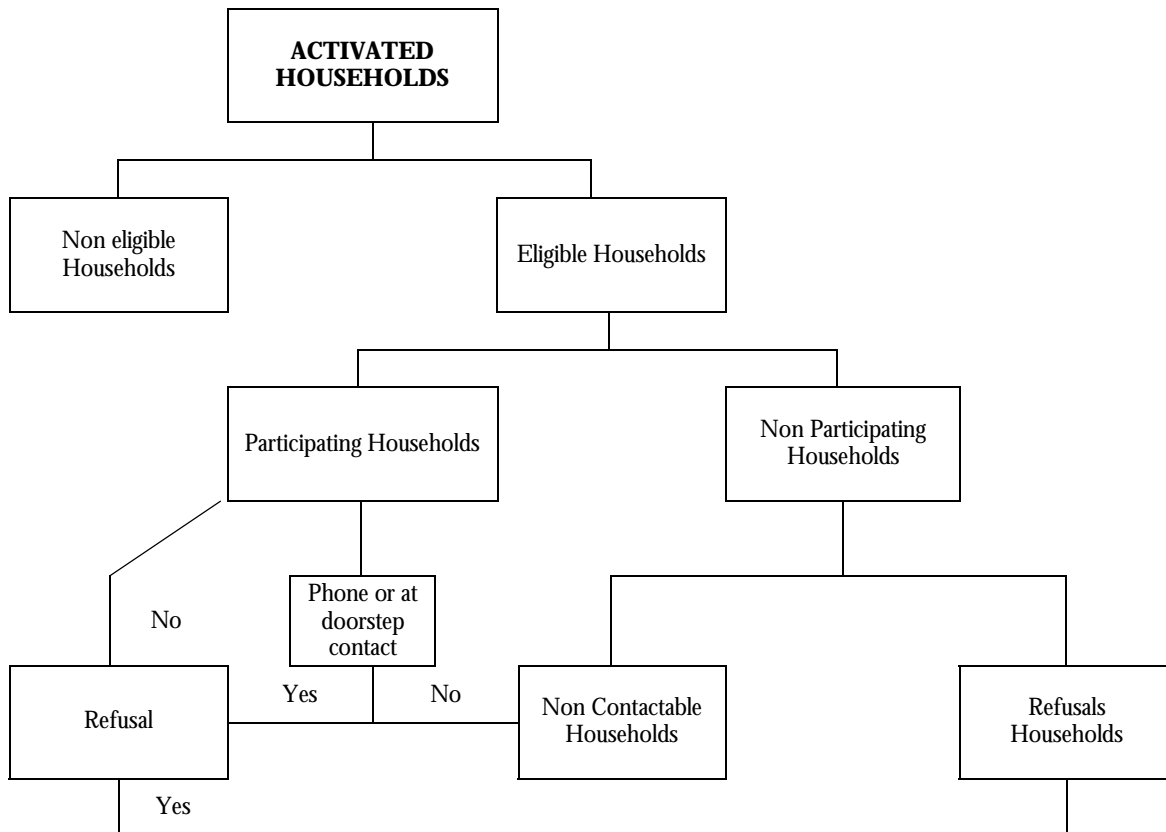
Once the household is selected, an invitation letter is sent and this household becomes an *activated* household. The interviewer has to try to establish a contact with these households. When no contact can be established, the household will be put in the category of non-contactable households. When it turns out that the household moved outside the PSU, it is classified as 'non-eligible'. In case a contact can be established, the interviewer will explain, the main ideas and objectives of the HIS and will seek collaboration, i.e. by making an appointment for an interview the interviewer will make an appointment in order to explain the main ideas and objectives of the HIS and he/she will seek collaboration. If the household effectively participates in the survey it will be called a *participating* household. In case the household does not accept it will be considered a *refusal* household.

We can also present the different categories from the response point of view:

1. **Participating household (PH):** An interview for at least one member of the household was obtained.
2. **Non-participating household (NPH):** This category includes the following two subcategories:

- **Non-contactable household (NCH):** The interviewer could not contact any of the household members. The address does not exist. Or it exists, but neither a telephone contact nor a doorstep contact could be established.
 - **Refusal household (RH):** In this case a contact took place, but the household refused to collaborate.
3. **Non-eligible household (NEH):** The household that is selected in the sample although it does not belong to the sampling frame. In general this type of household should be identified before the invitation letter is send out. However in some cases the letter can be sent out due to the lack of timelines in the update of the National Register. The following households will be classified as non-eligible:
- A one-person household where the reference person has died.
 - Households that moved out of the municipality.
 - Households that are re-selected in the samples for the 2nd, 3rd and 4th quarter and that were already selected for before.

The following diagram summarizes the defined categories for the households



3.2 Non-response at Individual Level

It is possible that within a participating household one or more members refuse to participate in the survey, in that case we will talk about non-response or refusal at an individual level. Concerning this type of non-response the rule is that within-household replacements are NEVER allowed because bias is very likely here (e.g., household members having less time would be more reluctant to participate and will be replaced by members having more time). In case the refuter agrees that a proxy answers in his place, the procedure must be followed as explained in Section 6.2.1.

4. Procedure Implementation

As outlined in Chapter 2 the final sampling of the households and respondents is a combination of many sampling techniques: stratification, multistage sampling and clustering. In this chapter we discuss more in detail the sampling techniques and we explain how the procedures were implemented.

The sampling of respondents takes place in the following steps:

- (1) Stratification by region and province,
- (2) Selection of the municipalities in each stratum,
- (3) Selection of a cluster of households within each municipality,
- (4) Selection of respondents within a household.

4.1 Stratification

In a simple random sample without replacement, each unit is selected at most once with equal probability. Although very simple and straightforward in theory, simple random sampling is seldom used in practice since:

- the sampling units are often unknown in advance;
- it can be very inefficient in statistical terms. Precision can be gained by subdividing a heterogeneous population into more homogeneous blocks and by taking a separate sample from each of them (**stratification**);
- it does not guarantee that some subgroups (e.g. the German speaking community) are present in a sufficiently large number.

For the Health Interview Survey, a base sample size was pre-specified for the three regions of Belgium (Flemish region, Walloon region and Brussels region): 3500, 3500, and 3000 respectively. It is decided to fix the number of interviews by province proportional to its population size. The German Community, however, will be oversampled. The stratification at the provincial level will prevent that some provinces are by chance over- or

underrepresented. This may be the case in a random sample for smaller provinces, as the probability with which this may occur is a function of the population size of the province.

The number of interviews to be realized as specified for the base sample is augmented with the requested number of oversampled interviews for four provinces (Antwerpen 350, Hainaut 500, Limburg 200, Luxembourg 1000), resulting in 12,050 successful interviews.

It is important to realize that, even though we distinguish for the sake of completeness the base sample from the additional sample resulting from the requested oversampling by four provinces, for all other purposes the sample will be considered as a whole.

To comply with this requirement, Belgium was subdivided into 12 strata:

- 5 provinces in the Flemish region (5)
- the Brussels region (1)
- 5 provinces in Walloon region (the province of Liège without the German Community in the district Eupen-Malmédy) and the German Community (5+1).

In order to keep the fieldwork manageable the number of interviews to be done in each municipality should be at least 50. Hence it was decided to work with multiples of 50 within each stratum. To comply also with this requirement some rounding was necessary and as a consequence the selection probabilities with respect to the base sample within each region are not perfectly equal. The results are presented in Table 1. For each stratum the following items are specified: the population, the fraction relative to the total number of inhabitants of the region, the theoretical number of individuals to be interviewed, the effective number of individuals to be interviewed (a multiple of 50), the corresponding number of groups of 50 individuals to be interviewed and the probability of being selected (the sampling rate/1000).

4.2 Multi stage sampling

Multistage sampling is the best way to get access to households/individuals. A direct selection of households/respondents from a list would be too expensive, because the spread and therefore the interviewers traveling costs would be too high.

Within a municipality, households are selected in a systematic way from a list ordered by statistical sector, size of household and age of reference person (see Section 4.2). Households are selected in a second stage, and are therefore called secondary sampling units (SSU).

The household itself can be considered as a cluster for the individual respondents. If the size of the family is smaller or equal than 4, the total cluster is taken, if not, subsampling is necessary (see Section 4.4). This third stage selection results in the respondents for the survey, the tertiary sampling units (TSU).

4.3 Systematic sampling

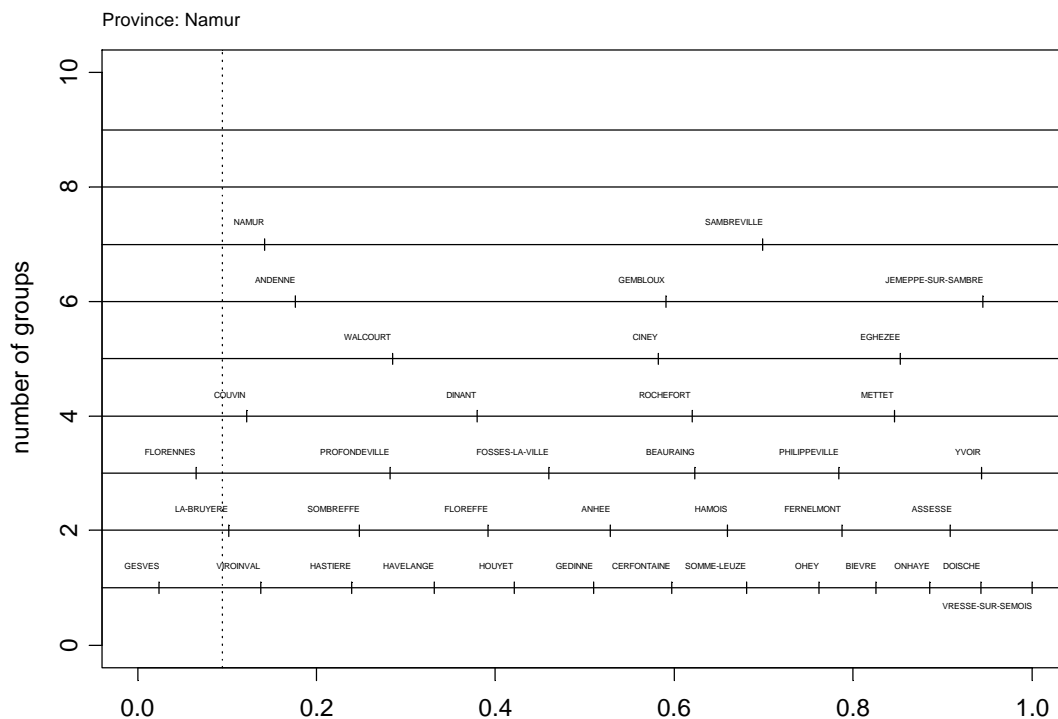
So in the first step municipalities (PSU) are selected within each stratum. To warrant the selection of important municipalities (metropolises and larger cities) but also of some smaller municipalities some additional rules should be built into the random selection:

- large cities and metropolises are included at least once with certainty in the sample. The number of times a city or a metropolis has to be included is determined by its actual population size.
- a similar remark holds for the smaller towns and villages. Also from this group municipalities should be included. By grouping smaller municipalities by ordering the whole set of municipalities communities according their population the representation of smaller communities out of the pool of smaller municipalities is ensured.

The above requirements are achieved by a weighted systematic sampling where the municipalities are ordered (from large to small) and expanded proportional to their size (so-called *area probability sampling*). As a consequence, the chance for a municipality to be selected is proportional to its number of inhabitants. The sample will contain municipalities ranging from large to small. This selection of municipalities is made for the whole year at once. In this way time is less confounded with place and it also facilitates the planning of the fieldwork.

Once the quota by province is defined (see Table 1), the sampling is straightforward. Within each province (region) the municipalities are ordered by size from large to small. To give the cities a weight proportional to their size, they are repeated in the list proportional to their population size. Hence the probability of selection is proportional to the size.

To make this procedure more transparent a graphical representation was developed. A segment proportional to its size represents each city. The sums of the segments represent the total population of the stratum. Then this line is cut into a number of pieces equal to the number of groups of 50 individuals one wants to draw from. These segments are put on top of each other. Then taking a systematic sample comes down to drawing an arbitrary vertical line. The process is illustrated for the province of Namur.



1. In the province of Namur nine groups of 50 individuals should be drawn, which is translated into 9 horizontal lines. On this horizontal lines the municipalities are represented by segments that can extend over many horizontal lines; e.g., Namur is so large in comparison with the other cities that it is distributed over more than two horizontal lines. The implication is that Namur as a PSU will be selected at least twice and there is also a chance it is selected 3 times. For the whole sample, the table below specifies the municipalities that are selected with certainty. First, the minimum number of groups of 50 that have to be included is given. The actual sample will then consist,

either of this number of groups, or of one more group. Whether one or the other number is chosen is probabilistic in nature and is the result of the actual sample selection. For a municipality to be selected with certainty, the population of the municipality needs to be at least equal to the population of the province divided by the number of groups of 50 to be interviewed in that province.

Table 2 : Municipalities that are selected at least once.

Municipality	Minimum number of times selected	Nr. of times Selected
Antwerpen	7	8
Mechelen	1	1
Leuven	1	1
Gent	2	2
Hasselt	1	1
Brugge	1	1
Liege	3	4
Seraing	1	1
Verviers	1	1
Charleroi	5	5
Mons	2	3
La Louviere	2	2
Tournai	1	2
Mouscron	1	1
Arlon	2	3
Marche en Femme	1	2
Aubange	1	1
Bastogne	1	1
Namur	2	3
Eupen	1	2
Brussels	All municipalities are selected at least once	

2. On the other hand one notes that all small villages of the province of Namur are located on the bottom line. Hence the procedure assures that one of these villages will be selected as they form one block. A possible issue with the smaller villages can be that there are not enough inhabitants. However, for all of them the number of inhabitants is larger than 50. The table below gives the five smallest villages in the Flemish region and the Walloon region, the only village with could give problems is Herstappe in the province of Limburg. However the selection probability is very small.

Code	Name	Pop.
73028	Herstappe	85
33016	Mesen	977
34043	Spiere-helkijn	1854
23009	Bever	1906
45062	Horebeke	1960
84029	Herbeumont	1436
84016	Daverdisse	1360
81013	Martelange	1453
82009	Fauvillers	1825
85047	Rouvroy	1896

3. Another point, which is illustrated, is that the systematic sampling cannot result in any arbitrary combination of the municipalities, a well-known property of systematic sampling. In fact, by fixing the order and working in a systematic way one restricts the

number of possibilities and the cities are grouped in loose “strata” from which just one will be selected. E.g., either Namur is selected a third time or Sambreville is selected. And if Namur is selected a third time, Andenne is also selected. Using a systematic sampling procedure, the universe of possible samples is reduced and the more extreme samples are ruled out.

4. Finally the vertical line on the plot is drawn at random by generating a random variable from a uniform distribution. The municipalities whose segment is crossed are selected. For the current line, the selected municipalities are listed in the following table which also indicates the number of times they were selected:

Nr	Code	Name
1	92003	Andenne
3	92094	Namur
1	92101	Profondeville
1	92141	La-bruyere
1	93014	Couvin
1	93088	Walcourt
1	93090	Viroinval

This systematic sampling procedure is repeated for each province. The corresponding lists of selected municipalities can be found in Appendix. Of the 589 municipalities, 178 are selected. The number of selected municipalities (PSU) is smaller than the number of groups (241) of 50 individuals because several municipalities are selected more than once.

4.4 Clustering: Selection of a sets of households within each municipality

At this stage, the municipalities in which the households (SSU) have to be sampled are known and the corresponding number of groups (or the number of respondents to be interviewed (a multiple of 50)) can be calculated because for each municipality the number of inhabitants is known, as well as the number of households. The expected number of households needed per group equals 50 divided by the mean household size. For example,

if the mean household size is 2.5 then the number of households needed is $50/2.5$, i.e., 20 households.

To plan the sampling of the households two further points should be taken into account:

1. Not all sampled households will result in an interview (about 50 %; cf. results of the HIS97). The reasons are variable not eligible (e.g. moved out the PSU), impossible to locate, refusal, or an underestimation of the household-size by the National Registry. To compensate for this, it is important to select more households than are actually needed to achieve the required number of successful interviews. In other words, *reserves* are needed.
2. The units of the sampling frame (the National Register) are households with a *variable number of members*. This is an issue because it is not the number of households but the number of individuals to be interviewed that is fixed. Hence, the number of households needed should be estimated. Additionally, one should take into account that a (small) fraction of the household members will refuse.

Let us discuss the solutions that will be implemented.

1. To tackle, at least partially, systematic trends in drop-out, it was decided not to replace the households in a simple random fashion, but to seek for a-priori matches based on:
 - a) the statistical sector within the municipality (and hence also the municipality);
 - b) the size of the household;
 - c) the age of the reference person.

To facilitate the organization of the fieldwork three matches per household (i.e. “reserves”) will be generated immediately. The so-resulting group of matched quadruples is called in this document as a ***cluster of households***.

2. To cope with the variable household size and to account for possible high rates of non-response, twice the expected number of clusters of households (i.e., a group of 4 “similar” households) will be sampled.

Hence the total sample size will be *8 times* as large.

To sample the households, their corresponding reference persons in the National Register are ordered hierarchically by:

- statistical sector,
- the size of the household,
- the age of the reference person.

In this way, households comparable to each other according to the first variable will be located close to each other on the list. To achieve the same desirable feature for the other variables, the order of the variable $j + 1$ is alternated when the level of the variable is changed from the current to the next level ($j = 1,2$). For example, in statistical sector 1, the household sizes are given in an increasing order, while they decrease in statistical sector 2. For household size 1 in statistical sector 1, the reference persons are listed according to increasing age, while the ages decrease for the households of size 2 in statistical sector 1. A schematic presentation is given in the following table

Table 3: The ordered sampling frame for the selection of the households

PSU	Statistical sector	Household size	Age reference person
PSU : X	1	1	youngest ... oldest
		2	oldest ... youngest
		3	youngest ... oldest
		4	...
		4+	
		4+	
	2	4	
		3	
		2	
		1	
	3	1	
		...	

From this list a clustered systematic sample is taken. We will explain here the procedure for the first quarter and point out the differences for the other quarters.

For each municipality, the **step size** for the systematic sampling is given by:

$$y = \frac{N - 3}{n}$$

with N the number of households within the PSU

n the number of clusters (quadruples) of households to be sampled:

$$n = m \cdot \frac{12.5}{\text{mean household size}} \cdot 2$$

m is the number of groups of 50 (succesfull) interviews that need be obtained for the PSU. The factor two in this expression accounts for the fact that the number of personal interviews can be estimated from the number of households, but not determined with certainty (variable household membership, within household refusal). Therefore a larger number of matched quadruples is selected to enable the inclusion of additional households.

Furthermore, instead of taking one household in each step of the sampling, the selected household and the three consecutive households are taken. In this way one obtains the four required matched households.

In practice the algorithm generates a random number x between 1 and y . The first cluster constitutes of the households x , $x+1$, $x+2$ and $x+3$, the second cluster contains the households $x+y$, $x+y+1$, $x+y+2$ and $x+y+3$, and more general, the n -th cluster contains the households $x + (n-1)y$, $x + (n-1)y + 1$, $x + (n-1)y + 2$ and $x + (n-1)y + 3$.

For most municipalities the step size will not be an integer. Therefore the first household of the n -th cluster is, is household number **ROUND**($x + (n-1)y$) on the list.

Suppose e.g. that $y=4.3$ and $x=2$. Then the clusters are composed as follows:

Cluster 1 : households 2, 3, 4, 5

Cluster 2 : households $6=\text{round}(2+4.3)$, $6+1$ $6+2$, $6+3$

Cluster 3 : households $11=\text{round}(2+2*4.3)$, $11+1$, $11+2$, $11+3$

Cluster 4 : households $15=\text{round}(2+3*4.3)$, $15+1$, $15+2$, $15+3$

Cluster 5 : households $19=\text{round}(2+4*4.3)$, $19+1$, $19+2$, $19+3$

Cluster 6 : households $24=\text{round}(2+5*4.3)$, $24+1$, $24+2$, $24+3$

... ..

The obtained sample of n households will be divided in m sub-samples (the number of groups of 50 interviews) such that the households of a group are located in the same statistical sector; this to limit the travel distances of the interviewers. Proximity of statistical sectors is not reflected in the alphabetical order of the municipalities. Therefore, for the municipalities where more than one group has to be interviewed the statistical sectors were manually ordered, such that sectors close to each other on the list are also geographically close. This procedure was only applied for the trimesters 2, 3 and 4.

After this sorting, the first n/m households of the sample are assigned to group 1, households $n/m+1$ to $2*n/m$ to group 2, etc.

For each group, a four-column table is formed where each row represents a cluster of the (four) households. There are as many rows as there are clusters selected (n/m). To prevent any order effect the households within each row are randomized. Also, the rows themselves are randomized. Then, there are no row-effects and it is possible to work from top to bottom until a sufficient number of interviews are realized

4.5 Advantages of the design

By taking a systematic sample from an ordered list, it is ensured that the characteristics of the sample will be close to that of the municipality with respect to the variables: statistical sectors, household size and age of the reference person. By making a list in advance, the organization of the fieldwork is facilitated because no algorithm is necessary to decide about the next replacement and all information about contacting is present.

4.6 Activation of the households

From this four-column table, households will be activated, this means that they are invited to participate in the survey by an invitation letter and that the contacting procedure can be started (see Section 4.6).

At the beginning of the quarter, households from the first column of the table will be activated. The number of households to be activated is based on a **cut-off point of 14 persons**. This means that the sum of the individuals in the households to be interviewed is closer to the cut-off point, with possible range 12 to 16. In case of equal distance to 14, there is a random choice necessary to either remain below 14 or to exceed it.

The different possibilities are displayed in Table 4. If the group of sampled households (in theory) results in 11 interviews, an additional cluster will always be selected, regardless of the size of the households in this new cluster. If in total already 13 individuals are sampled to be interviewed, the next cluster selected will be

- added to the sample if the household size is 1
- added to the sample with probability 0.5 if the household is of size 2
- not added to the sample if the household is of size ≥ 3 .

Table 4: Activation of the households

	Size of the household of the next cluster			
	1	2	3	4 ¹
Total (theoretical) Sample size ²				
11	add new cluster	add new cluster	add new cluster	Choice: 11 or 15 add new cluster
12	add new cluster	add new cluster	choice: 12 or 15 add new cluster	Choice: 12 or 16 random choice
13	add new cluster	Choice: 13 or 15 random choice	choice 13 or 16 do not add cluster	Choice 13 or 17 do not add cluster

If the activated household from the first column is not eligible or if it does not result in an interview, the next household (the second column) in the cluster is activated, and so on. When all the households of the cluster are used and further replacement is necessary (see Section 4.6.1), the first eligible household of the next available row is selected. Note that the newly activated and exhausted cluster are not matched with respect to the statistical sector, the household size and the age of the reference person.

¹ At maximum 4 persons of the same household will be interviewed.

² Of the already selected households.

4.6.1 Replacement of the households

In case the activated household is not eligible or does not result in an interview (household non-response), the replacement household is the next eligible household within the cluster (the second column in case the household of the first column does not result in an interview). Once a cluster is initiated, the algorithm continues to select the next eligible household within the cluster, *independently* of the number of successful interviews attained. So, once a cluster is started, all efforts are taken to have a successful interview for a household within the cluster. However, in the following two situations this rule will not be applied:

- when at the end of the calendar year the quota at the provincial and regional level is reached, an activated household that does not result in a successful interview does not need to be replaced with the next household of the cluster.
- when an interviewer drops out of the survey and he/she cannot be replaced within reasonable time limits, the activated households of the group will be deactivated. When a new interviewer takes over the group, new clusters of households are then activated.

When all the households of the cluster are used and further replacement is necessary, the first eligible household of the next available row is selected. The decision to activate this new cluster of households takes into account the number of interviews actually realized at that time. For this, a **cut-off point of 12.5** will be used. The different possibilities are displayed in Table 5

Table 5 : Replacements of the households

	Size of the household of the next cluster			
	1	2	3	4 ³
TOTAL NUMBER OF INTERVIEWS				
9	add new cluster	add new cluster	add new cluster	choice: 9 or 13 add new cluster (only if it is 13)
10	add new cluster	add new cluster	choice: 10 or 13 add new cluster	choice: 10 or 14 add new cluster
11	add new cluster	choice: 11 or 13 add cluster	choice: 11 or 14 Random choice	choice: 11 or 15 do not add cluster
12	choice: 12 or 13 random choice	do not add cluster	do not add cluster	do not add cluster

The first column in this table 'Total number of interviews' is the sum of two pieces: for the participating households: the number of realized interviews; for the households for which the final status is not yet known: the expected number of interviews.

Suppose e.g. the following situation. A marked cell means that the household is activated. At the beginning of the quarter four households are activated, resulting in an expected number of 13 interviews. If none of the households of the second cluster result in an interview, one has to decide if a new eligible cluster (n° 5) needs to be started. Therefore the number of expected interviews needs to be updated. For the first cluster the number of actual realized interviews will be taken into account instead of the expected number of interviews. So at this point in the survey the number of expected interviews is 9. Adding the expected number of interviews in the next eligible cluster brings this number to 12. Since 12 is closer to 12.5 than 9, this new cluster will be activated.

³ At maximum 4 persons of the same household will be interviewed.

Cluster n°	Cluster size	HH n°1	HH n°2	HH n°3	HH n°4	Number of realized interviews
1	4	Participating				3
2	3	Refusal	Refusal	Refusal	Refusal	0
3	2	Refusal	Refusal	Negotiation		
4	4	Refusal	Negotiation			
5	3					

4.6.2 Quarters 2-4

This procedure is repeated for the selection of households to be interviewed in the 2nd, 3rd and 4th quarter. The sample frame consists each time an “adjusted” copy of the National Registry of 1 January 2001. The NIS receives regularly an overview of changes in the Registry, such as changes in the address, moving houses, death of some one in a household, etc.. The selection of the sample will be based on the most recent update, by the NIS adjusted copy of the registry.

As four samples are taken, there is a chance that the same households are selected several times. To control for this, each sample is checked against the samples of the previous quarters and duplicates are removed, independently of the activation status of the household in the previous samples. A duplicate household receives the status of non-eligible. This procedure has the advantage that one does not need to remove households from the whole 2001 National Registry database.

This sampling procedure comes down to an elaborate version of sampling without replacement.

The number of households to be selected for e.g. the second quarter sample should take into account the risk that one has duplicates with the first quarter sample. This adjustment-factor will be PSU specific. Imagine a municipality of 1000 households. For each quarter 100 households are necessary to obtain the required number of successful interviews. The chance for a household to be selected in the first sample (1st quarter) is of

course 10%; and therefore the chance that it is not selected equals 90%. To adjust for the fact that the duplicates need to be removed from the second sample (2nd quarter), it is necessary to sample more than 100 households. In fact, the adjustment is the number of households to be selected divided by the probability that a household is not in the first sample; i.e. $100/90\%=111$.

So, 111 households are sampled where after the 11 households (10%), already in the first sample, are removed. The duplicates that are removed from the second sample are no longer eligible.

As a consequence of this procedure, the chance for a household to be selected for the 2nd quarter is slightly higher than the chance to be selected for the first quarter. In the above example we have that $P_1=10\%$ and $P_2= 111/(1000)=11.1\%$.

Activation and replacement of the households is done in the same way as for the first quarter.

4.7 Selection of the household members

At most four members of the activated household will be interviewed. Interviewing more persons is inefficient because of the familial correlation: members of the same family tend to resemble each other more closely than members from different households. By increasing augmenting the number of interviews, hardly any new information is obtained for the global sample. Furthermore, the burden on the household would be too large.

Hence, if a household consists of more than four members a selection rule is necessary. This selection should in principle be at random. Always selecting the reference person of the household may lead to bias since the reference person is not a random member of the household. He or she may have special characteristics. In the literature this person is sometimes denoted as the gatekeeper. Even including the partner may not totally compensate for this. The resulting bias can be removed by appropriate weights. The latter solution is followed here. The reasons why the reference person (and his/her partner, if any) is chosen, are as follows:

- It may be difficult to explain that the reference person will not be interviewed, while other members are interviewed.
- There is a general household questionnaire. This information on the HH-level

Therefore, the following selection rules are used within a household to select the individuals (TSU) to be interviewed.

1. In a household of no more than 4 members all individuals are interviewed.
2. In a household with 5 or more members only 4 members will be interviewed :
 - In a household with a reference person and a partner both the reference person and the partner are interviewed and only two additional individuals are selected using **the birthday rule**: the two or three persons having their birthday first, starting from the date of the interview session onwards, are included in the sample
 - In a household with a reference person without a partner, the reference person and 3 additional members, selected using the birthday rule, are interviewed.
 - If a selection is necessary, the reference person and his/her partner are selected automatically. A randomization will be done for the remaining persons only. The selection itself is based on the birthday rule.

5. Contacting the Respondents

Once a household is selected, it should be contacted to solicit its participation in the survey.

Three stages can be distinguished:

1. the announcement of the survey.
2. the procedure to contact a household.
3. the selection of respondents within a household.

These stages will be described in the following paragraphs.

5.1 The announcement of the survey

An invitation letter and an information leaflet is sent in advance to the household, selected by the National Institute of Statistics (NIS). This letter is addressed to the reference person. It explains the objectives of the survey and informs the reference person that an interviewer will contact the household within two weeks. In this letter a telephone number is where further information can be requested if necessary. Such a measure can increase the confidence in the soundness of the interview.

Once the invitation letter is sent, the household is **activated**. . At the same time the list of addresses of reference persons is sent to the interviewer (together with name, address and the number of members in the HH), and the process of contacting starts.

At the first contact (see Section 5.2) the interviewer must control whether the address corresponds to the indicated household. In case the household moved or an incorrect address was obtained, the NIS will try to trace the households' new address. If it is within the same municipality, the household will be re-contacted by a new invitation letter. If the household moved outside the municipality it is no longer eligible.

5.2 Procedure for contacting a household

When a household is activated the interviewer will seek contact with the household to explain the objectives of the Health Interview Survey and to get consent for cooperation.

This section outlines the procedure that the interviewer should follow in contacting a household.

At maximum *6 weeks* after activating the household, a final status should be obtained. This period of 6 weeks is split into *3 two-week intervals*. Every contact attempt in such an interval is registered using the communication form. This form has to be returned to the NIS secretariat every fortnight(2-week interval). It allows speeding up and improving the quality of the fieldwork. The rules that need to be followed while contacting a household are the same for every two-week interval. These guidelines are discussed in detail below.

One needs to try to contact the household at least 5 times within each two-week interval. These attempts are preferably done using different modes (face-to-face or at doorstep), at different times (morning, afternoon, evening) and at different days (week versus weekend).

While contacting the households the following rules need to be respected:

1. The first contact attempt needs to take place shortly after receiving the addresses and surely within the first time-interval. The mode of contact (phone or doorstep) is up to the interviewer. But preferably the first contact (attempt) is a doorstep contact. While there is some chance that this might disfavor those without a phone, it is known that, especially in urban areas, it is difficult to establish contact immediately at the doorstep, in particular during evening hours.
2. In case this first attempt does not result into an effective contact, the interviewer needs to re-try to contact the household at least additionally 4 times within the first time interval.
3. After the first interval it is up to the interviewer to decide about the actual status of the household.

- Participating**
- For all attempts that result in an actual contact it holds that
 - when the household agrees to participate, the interviewer makes an appointment for the actual interview; (PARTICIPATION STATUS).
 - When the household explicitly refuses to participate (REFUSAL STATUS), the interviewer needs to fill out two questions with respect to the housing (LO.01 and LO.02). In case of a face-to-face contact this can be done at once. When the contact was a phone contact, the interviewer needs to go physically to the house to fill out these questions.
- Refusal**
- Not contactable**
- In case none of the attempts result in a contact, the interviewer can classify the household as NOT CONTACTABLE only on the condition that he made at least 5 *attempts* to contact the household of which
 - ❖ at least one was a face-to-face (doorstep) contact attempt,
 - ❖ at least one was during the weekend,
 - ❖ at least two were outside working-hours during the week.

These conditions will be checked by the NIS secretariat (using the contact form). If they are not fulfilled the NIS can reject the decision of the interviewer and ask the interviewer to contact the household again at least 5 times in the next interval period. For this interval the same procedure and criteria apply as in the first interval.

In case no contact can be made, but the household could be located the housing questions (LO.01 and LO.02) have to be recorded by the interviewer.
- Stand by**
- When after five attempts the interviewer does not yet wish to classify the household as not contactable or when the household is on vacation for some time the interviewer can attempt to contact the household within the next two weeks. In such a case, the household is given a STAND BY STATUS.
4. For the *stand by* households of the first interval (the first 14 days), the interviewer gets the opportunity to undertake at least 5 new attempts within a second interval. The procedure for contacting the household and the rules for classifying the household are the same as in the first interval. .

5. If after these 28 days the final status of the household is not yet confirmed, the interviewer will be informed by the NIS that he gets a final interval to try another 5 contact to obtain the final status. If even these attempts do not result in a final status, thus after a total period of 6 weeks following the activation, the household will be de-activated by NIS, , and considered as not contactable.

There are however a few situations that we would like to highlight.

Once the interviewer has received the address, he tries to get in contact with the reference person, by telephone or by visiting the (official) address. If it becomes clear that this is the address of an institution (*with the exception of institutions for elderly*), the reference person/household is considered as non-eligible. No further contact with the reference person should be attempted.

If the interviewer finds out that the reference person has his/her official address in an institution for elderly people, the procedure described in Section 6.2.2 is applicable. The reference person is treated like a one-person-household. In special situations, e.g. service flats, residencies for elderly couples,... not only the reference person partner but also his/her partner spouse have to be interviewed.

In case one or more members of the household have left the household on a quasi-permanent basis, one must ask the question whether these members are institutionalized or not.

If they are not institutionalized (e.g. in case of a divorce, children who have left the household on a permanent basis,...), these – former – members are not considered as being members of the household. This implies that they are not taken into account when applying the birthday rule to select the household members for the interview (Section 4.7).

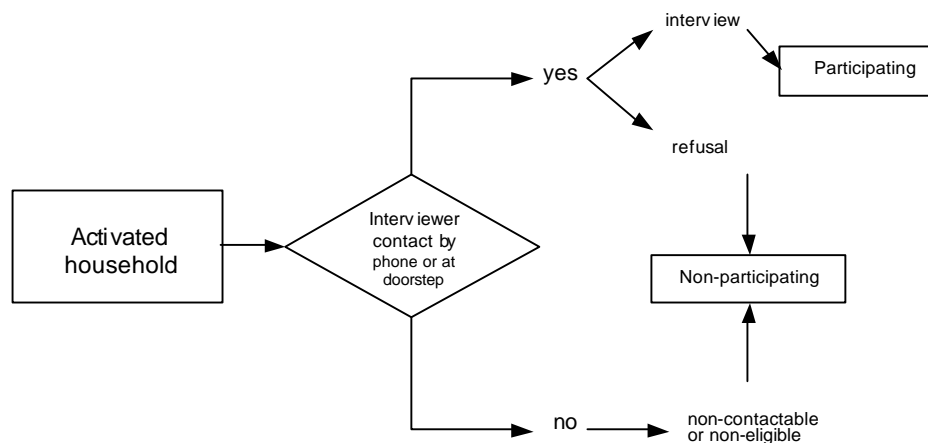
5.3 Selection of the respondents

In case of consent for cooperation, the interviewer has to record the number of household members during the first contact. If more than 4 persons belong to the household, a selection is made by applying the birthday rule, as explained in Section 4.7. It is important to point out that the reference person and his/her partner must always be interviewed. If necessary, one or more appointments should be made so that all the selected household members are interviewed. The number of interviews obtained has to be mentioned on the

communication form. The number of selected HH members that refuse to participate are mentioned on a summary form.

It is possible that an entire household, or a subset of it is not able or does not want to collaborate for various reasons. This has already been explained in previous Chapters.

The following flow chart summarizes the whole process of contacting the responders. In Chapter 3 we presented a chart of the full procedure.



When a household is non-participating, a replacement strategy is considered. To this end, each household is selected within a cluster of 4 households. This list, based on the National Register is ordered. When the first household is non-participating, it is replaced with the second one, and so on.. For each address, the interviewer has to give information on all steps undertaken in the process of contacting the households and the final outcome. An algorithm is developed to link the number of households to be contacted with the household size as known from the National Register and in a second phase with the real household size. This is necessary because on average 12.5 individuals have to be interviewed per quarter.. During the first quarter of 2001 the number of households that is activated corresponds to a number individuals that can be interviewed. This number is set to be 14. This figure results from the evaluation report of Van Steen *et al.* (2000)(Ref. 16) and will be updated for the following quarters, with the possibility of a differential cutoff value per region.

During the process of contacting the respondents several possibilities can occur before obtaining a successful interview. Once the household is activated, the interviewer has to establish the first contact according to the procedures already discussed.

An activated household is called a *stand-by* household up to the moment that the final situation is defined, i.e., up to the moment it becomes *participating*, *non-participating* or *non-contactable eligible*.

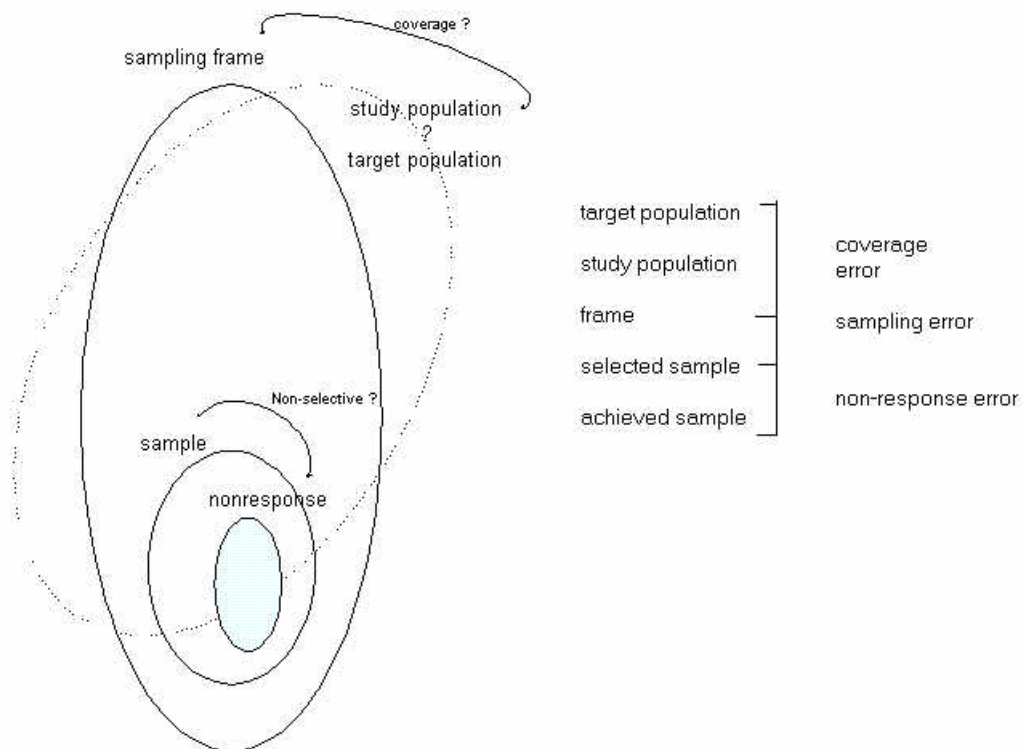
The period in which the potential respondent is in this *stand by* category cannot be more than 6 weeks. The information obtained during this whole period has to be recorded in the *communication form* in order to have a clear picture and to be able classify the household in the right category. In any case the final status will be set to be defined by 31/12/2001.

5.3.1 Note on non-coverage

The target population (i.e., the population determined by the general objectives of the study) may not be the same as the study population (i.e., the population that can be defined accurately and reached in study). This creates a discrepancy from the very start. Defining the population and delineating the population are two separate issues.

Furthermore, budget and time constraints enforce a tradeoff between:

1. The original objectives of the study; especially with respect to the population subgroups or the domains of public health for which results need to be extrapolated from the study;
2. The feasibility (a function of resources, such as cost and available methodology). It may be clear that when some groups are excluded from the study it is not allowed to extrapolate analysis results for those specific groups.



A further feature of surveys is that it is generally not possible to take directly a random sample from the target population. A sampling frame needs to be constructed. The simplest implementation is a list that enumerates the entities of the target population as completely as possible. In our case such a list is not present or inaccessible and the population should be approached in a more sophisticated fashion. As a consequence, the sampling frame seldom coincides with the study population. Well-defined inclusion/exclusion criteria give an idea of under-coverage (individuals not reached by the survey) and over-coverage (individuals that have probability greater than zero to be selected but that do not belong to the target population). Compared with under-coverage, the problem of over-coverage is the least problematic: one can always exclude a subset of units from the analysis.

Units are selected from the sampling frame according to an elaborate sampling mechanism. This is necessary, but not sufficient, to guarantee that the sample selected is representative/random, and is not guided by the (implicit) preferences of the researcher.

6. Interviewers and Fieldwork

The training of the interviewers is a very important issue. As fieldworkers, they have to retrieve the data from the households.

6.1 Training

Basically 250 experienced and well-trained interviewers are needed in order to collect the data. Normally these interviewers are engaged to be active during the whole year of the fieldwork. As it is anticipated that some interviewers will drop out during the course of the year, replacements must be available. The aim is to construct a pool of 400 interviewers who will be activated in the beginning or during the year of the fieldwork.

Preferably By preference, these interviewers already have experience in large-scale surveys. If not, a minimal educational level (at least higher secondary education) is required. All interviewers have to follow an in depth training of at least 2 hours. During the training sessions, the aim of the survey, the rules for contacting the households, the content of the questionnaires,... are explained. Special attention is paid to the communication between the secretariat of the survey and the interviewers. In order to speed up the proceedings, regular (weekly) contacts with the interviewers are crucial. As the replacement of a refusing household can only be based on information provided by the interviewers, a fast and clear communication is essential.

If an interviewer stops for any reason all his assigned households have to be reassigned to another interviewer. It can happen that the new interviewer cannot finish all the households in the required period of 3 to 4 weeks, and in that case the remaining HHs are deactivated. The situation is similar at the end of the Survey when the activated households that are not interviewed are automatically deactivated. More details can be found in the Interviewer's Guide (See Website for details <http://www.iph.fgov.be>).

6.2 Guidelines

If the interviewer cannot establish a contact, either because:

- the address is wrong (other name on door bell or mailbox);
- the house is abandoned;
- nobody is at home.

he/she should try to get the correct address (neighbors, ...).

- (i) Verify with neighbors or municipality services.
- (ii) Try to collect as much information as possible:
 - a. Description of the house (villa, flat, ...) or premises: LO.01, LO.02.
 - b. Time at which someone of the household may be contactable.
 - c. Location of job.
 - d. How many persons live in the household.
 - e. Whether the reference person is an elderly (institutionalized or not – has its consequences).
 - f. Check the telephone directory book.

The household should be re-contacted by a letter at the new address if it is in the same PSU. If the new address is outside the PSU, the household is no longer eligible.

6.2.1 Use of proxies

Proxies are allowed in the following situations:

1. By default for
 - (a) a person younger than 15 years.
 - (b) an individual who is too sick or is mentally disabled.
 - (c) an elderly person, living in an institution, with his/her official address within a non-institutionalized household.
2. When the target person cannot be reached for an extended period (at least more than 1 month). This situation may occur if the target person is away on a long journey or is hospitalized.
3. When the target person refused to participate but agrees on having a proxy answering for him/her.

6.2.2 Institutionalized Elderly People

The interviewers need to be instructed motivated very carefully as to the importance of interviewing the institutionalized people.

Regarding this point two scenario's can be distinguished:

1. the institutionalized people are officially registered in an institution (their official address is the address of the institution)
2. the institutionalized people are not officially registered in an institution (their official address is an address of a household)

Scenario 1.

Basically, the institutionalized people are considered as 'one-person-households' and are eligible to be interviewed. Other residents of the institution are not eligible for an interview. In special cases, e.g. when a eligible resident lives together with a spouse within the institution, it is to be considered as a 'two-persons-household'. Using a nurse or another caretaker as a proxy is allowed.

Scenario 2.

When an elderly person has his/her official address within a household, but is in fact institutionalized, the normal rules of selection household members for an interview is applied. If the elderly person is selected for an interview, a complete proxy interview is obliged. Next to the 'normal' questioning, supplementary information must be gathered on the name and address of the institution. If, due to circumstances, a proxy interview is not possible, at least the name and address of the institution must be gathered. In a next phase, the NIS will send an introduction letter to the institutionalized person, to invite him/her for an interview within the institution. Using a nurse or another caretaker belonging to the institution as a proxy is allowed. If all goes well, two sources of information will be available: information based on the interview with a proxy (within the household) and information based on the interview with the elderly person (or proxy within the institution). Both pieces of information are kept in separate files as two records. In the basic analysis, the information gathered in the direct interview of the elderly overrules the information gathered in the interview with the proxy within the household.

6.2.3 Interpreter

For people living in Belgium who do not speak one of the three official national languages (Dutch, French or German), no special measures are taken and no special questionnaires are planned (e.g. immigrants). These cases will be regarded as *non-contactable*, due to language problems.

However, when one of children or another household member does speak one of the interview languages, he/she can act as an interpreter for the selected person and an interview should be obtained. The interpreter only translates the questions to the selected person and his/her answers and does NOT serve as a proxy. . The interpreter is not a proxy.

Note that in Belgium on average 10% of the population are immigrants, with important differences between their regions of residency (Flemish region: 4.5%, Walloon region: 11.4%, Brussels region: 28.5%).

7. Quality Control Board

(Comité d'Accompagnement / Begeleidingscommissie)

The CA/BC has an advisory authority with respect to the Director of the Survey. In general, the CA/BC is qualified to guide and to critically evaluate all phases of the fieldwork. This encompasses (a) preparation of the fieldwork, (b) the collection of the data and finally (c) the analyses of the data. The agenda of the CA/BC is linked with the subsequent logical steps undertaken in the HIS2001.

The CA/BC is seen as a critical reflection chamber and gives advice to the director of the survey. This advice is given on the basis of standardized and periodic information passed on by the director and coordinator of the survey.

The commission meets every month, starting from 6 months before the start September until 3 months after termination of the fieldwork. In emergency situations, the frequency of the meetings may be increased. Its tasks are

- defining the evaluation instruments / indicators of the fieldwork
- evaluating the feasibility and the practical implications of the implementation of the instruments
- reviewing whether the instruments are correctly used
- advising on the results produced by the instruments

The stages are:

a) Preparation of the fieldwork:

- defining “non-participating”, “non-responding”, “non-available”, “non-contactable”, “(de-)activated”, ... households;
- developing scenario that may occur in practice and, investigating whether all phases of the (pre-)fieldwork are well-documented;
- determining the actions to be undertaken when preset deadlines are not met and /or when procedures outlined in the protocol are violated

b) Collection of the data / interviewers:

- reviewing the progress of data entry, of data base setup and data consistency checks;
- reviewing the interviewers performance by means of communication forms and progress report forms.

c) Analyses:

- checking whether the analysis plan is feasible (in terms of time and objectives of the survey, in terms of methodology and available data).

Several sources of information are required for the CA/BC:

- progress report (bi-weekly): reporting on the number of realized interviews, the number of refusing households and the number of households with a standby status;
- overview of non-response, non-participation, refusal, non-contactable and (de-) activated households (at the national level, at the regional level, at the provincial level, at the level of municipalities);
- overview of accrual rates;
- balance sheet of realized interviews and number of interviewers still needed to attain the preset quota for the quarter;
- general informative reports from the director of the survey, including (methodological) questions that were raised during the practical implementations.
- Depending on the progress of the survey, the CA/BC is entitled to review its own job description. This may imply extension, reduction, and amendment of its tasks.

7.1 Indicators for the fieldwork

On the agenda of the past CA/BC meetings possible indicators for evaluating the quality and progress of the fieldwork and interviewers were discussed. A summary of these discussions is presented here.

7.1.1 Management of the interviewers

Basically 250 experienced and well-trained interviewers will perform the interviews and collect the data from the households. These interviewers are engaged to be active during the whole year. It can be foreseen that some interviewers will drop out during the year. Therefore, a pool of about 400 interviewers is composed, from which (replacement) interviewers can be activated during the year of the fieldwork.

The following indicators are used to manage the interviewers:

- turn-over: this is the ratio of the number of interviewers that stopped during the fieldwork and the number of interviewers ever activated.
- the number of groups for which the interviewer needs to be replaced.
- the number of groups for which the interviewer is on holidays.
- the number of groups for which the communication form is missing.

Note that the number of interviewers that need to be replaced and the number of groups that need a replacement interviewer are not necessarily the same. It depends on the capacity of the (replacement) interviewer.

- When an interviewer drops out, the replacement interviewer will finish the started group. However, when it is impossible to find a new interviewer within 3 weeks, the households that were not interviewed are then classified as 'de-activated households'; and a new set of households will be activated once an interviewer is found. The de-activated households will be informed that no interviewer will pass by via a letter from the central secretariat.

The number of households to be activated for this new group takes into account the number of interviews the first interviewer has already undertaken;

7.1.2 Performance of the interviewers

In evaluating the performance of the interviewer two aspects can be distinguished: (1) are the households contacted according to the guidelines, (2) is the interviewer able to convince the households to participate in the survey. The following indicators – that can

be constructed from the communication form - will be used to evaluate the performance of an interviewer.

Completed households

- I_1 : the proportion of households interviewed

Not participating households

- I_2 : the proportion of refusals
- I_3 : the proportion of non-contactable households
- I_4 : the proportion of de-activated households

Stand by households

- I_5 : the proportion of stand-by households

For all these calculations the number of activated households, that should be contact by the interviewer, will be used as the denominator.

For each interviewer values for these indicators can be calculated. For I_2 , I_3 and I_5 extremes in each region (or province) can be detected by using tools like deciles, box plots, etc.

In addition the following indicators will be calculated to allow us to evaluate the performance of each interviewer separately.

- I_6 : number of households classified as non-contactable, while the procedure was not properly applied.
- I_7 : number of households that are not contacted at the end of the first 14-day interval. This information will also be presented as a proportion: the proportion of the activated households for a specific period, which were not contact within 14 days.
- I_8 : (Binary) Indicator whether or not one third of the quota is reached half way the trimester. So, for each group, 4 interviews need to be done at week 6 (after 3 communication forms).

The first time that the guidelines are violated, so when $I_6 = 1$ or $I_7 = 1$, a letter will be send to the interviewer to inform him about the violation and to ask him to correct the mistake. In

case both indicators I_6 and I_7 equal one, or when one of the indicators is larger than 1, the central secretariat will contact the interviewer by phone to find out what is going on. Half way the trimester, all interviewers with $I_8=0$ will be contacted by phone to motivate them to speed up the fieldwork so that at the end of the trimester the quota can be reached.

7.1.3 Quality of the fieldwork

To evaluate the quality of the fieldwork we will look at transgressions regarding the contact procedures.

The number of households that are not yet contacted at the end of the first 14-day interval, and the proportion with respect to the number of activated households.

The number of households classified as non-contactable, but for which the procedure of contacting was not correctly applied, and the proportion with respect to the total number of non-contactable households.

7.1.4 Progress of the fieldwork

The performance of the interviewers is reflected in the progress of the fieldwork. Therefore indicators, similar to the ones used to evaluate the performance of the interviewers, are used to evaluate the progress of the fieldwork.

Every 14 days the following indicators will be presented, based on all data collected since the start of the survey.

Completed households

I_1 : the proportion of households interviewed

Not participating households

I_2 : the proportion of refusals

I_3 : the proportion of non-contactable households

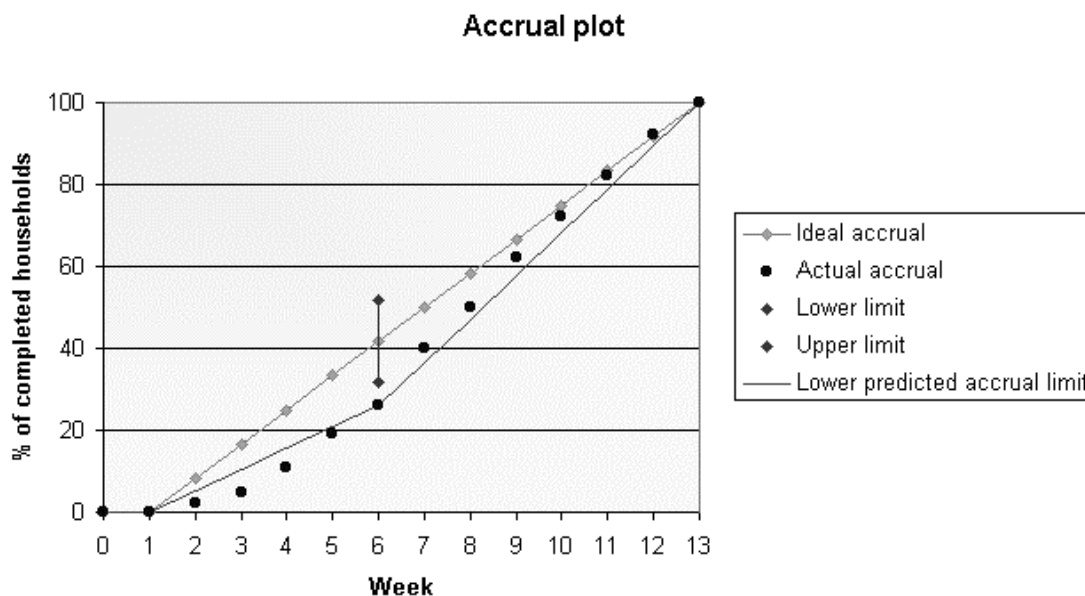
I_4 : the proportion of de-activated households

Stand by households

I_5 : the proportion of stand-by households

The results should be presented bi-weekly in table and graph format. In the table the ratio 'refusal/ (refusals + responders)' and 'completed/target' will also be displayed. An example of a table and a figure are given in Appendix 8.1.5.

An accrual plot, showing the 'ideal accrual' and the actual accrual is used to decide whether or not the interviewers need be motivated/contacted to speed up the fieldwork and to make sure that at the end of the trimester the quota is reached.



The above picture is an example for the first trimester. The survey started in fact in the second week of January. Ideally the accrual is equally spread in time. In practice however we expect the accrual to be a bit slower at the start. At week 6, the ideal and actual accrual is compared to see if action needs to be taken. In the ideal situation in fact 41.7% of the households should have been interviewed at this point in time. We do however allow a deficit of 10% before taking action. If less than 31.7% of the households are interviewed at week 6, the central secretariat contacts the interviewers. For every activated household, the interviewers should try to obtain the final status as soon as possible; and for the households that are willing to participate the actual interviewing should also be completed as soon as possible. A straight line from the actual accrual at week 6 and the quota to be reached at week 13 can be drawn. To 'ensure' that the quota is reached, the accrual rate from week 6 on should be at least according to the purple line. In practice, however we expect/hope that the observed accrual rate is in the triangle formed by the two lines in the accrual plot.

Not all 'types' of households are as willing to participate, or as easy to contact. E.g. people in rural areas are maybe easier to convince to fill out the questionnaire, elderly people living alone in a big city are probably more difficult to convince, etc. However, our sample should be representative for the Belgian population and therefore all efforts should be done to contact/interview also these households. For this reason it is interesting to know which types of households need special attention. Therefore the proportion of household non response (in terms of refusals and not-contactable) will be presented by e.g. size of the household, age of the reference person, province, moment of contact and mode of contact.

8. Appendices

8.1 Selected Sample

8.1.1 Introduction

As discussed before, the selection of the respondents is a threestep process:

1. The municipalities are selected (PSU).
2. Then 50 households are selected in each chosen municipality (SSU).
3. Finally, within a selected household, at most 4 persons are selected for the interview (TSU).

In this chapter, the selection of the municipalities is described.

An important decision in this respect was to take the municipalities (589 in total, with an average of 17221 inhabitants) and not the old municipalities, which are more homogeneous (about 2500 in total with on average 4000 inhabitants). Although this last option seems attractive, it was rejected, as this information is not (directly) available in the sampling frame.

The basic idea was to subdivide Belgium into strata and to take a systematic sample of the municipalities within each stratum. For the three different regions the number of interviews was: 4050 (3500+550) for the Flemish region, 4700 (3200+1500) for the Walloon region, 300 for the German Community and 3000 for the Brussels region.

However, there are other additional requirements:

- The German community should be oversampled (300 interviews). To guarantee this, the German community is considered to be a specific stratum within the Walloon Region.
- For the provinces, oversampling is allowed. Four provinces subscribed to this (Antwerpen, Hainaut, Limburg, Luxembourg). Their oversampled interviews are added to the base sample, which is determined proportional to their size.

These two points lead to make a subdivision of 12 strata: the 10 provinces, Brussels and the German Community.

8.1.2 The distribution over the regions

To make fieldwork feasible it is chosen to have in each primary sampling unit at least 50 interviews (a group). Hence in total 241 groups should be selected in the first step. As some municipalities might be selected more than once the number of groups may be greater than the number of PSUs (municipalities) selected. However the sampling probabilities are not equal. In the first place the number of interviews in each region is fixed: 3500+550 for the Flemish region, 3500+1500 for the Walloon region and 3000 for Brussels region. Also it is specified that an oversampling should be taken of 300 in the German Community.

To achieve these boundary conditions, Belgium is subdivided into 4 parts (within brackets the number of groups to be selected is given):

- Flemish region: 4050 (81)
- Walloon region : 5000 (94 + 6 (Eupen))
- Brussels region : 3000 (60)
- Total: 12050 (241)

If we relate these figures to the population size of each region (see table below, figures of INS, January 1st 2000) it is seen that the probability of being selected differs considerably from one geographical area to the next (e.g., oversampled versus non-oversampled provinces). In other words, a differential selection rate for each of the three regions is seen. Currently, the Belgian population size is about 10 million people. Hence on average the chance of being selected is 1.2/1000. However in the Flemish region as a whole this is about 0.68/1000, for the Walloon provinces (minus Eupen), it is 1.44/1000, for Brussels this is about 3 times the global chance and for the German Community it is about 4 times as large (see Table 1). Note that the rates are considerably different for the oversampled provinces (Antwerpen, Hainaut, Limburg, Luxembourg).

These differential selection probabilities do not affect the overall validity of statistical inference based on the HIS, provided a correct system of weights is implemented.

8.1.3 The distribution over the provinces

In principle, the subdivision proposed above is sufficient to select the municipalities within each region.

However

- to guarantee that the number of interviews realized in each province is (nearly) proportional to the population and
- to allow that in some provinces an oversampling is possible,

it was decided also to fix also in advance the number of PSU to be selected in each province.

This comes down to making 12 strata: the 10 provinces, Brussels and Eupen (the German Community of Belgium).

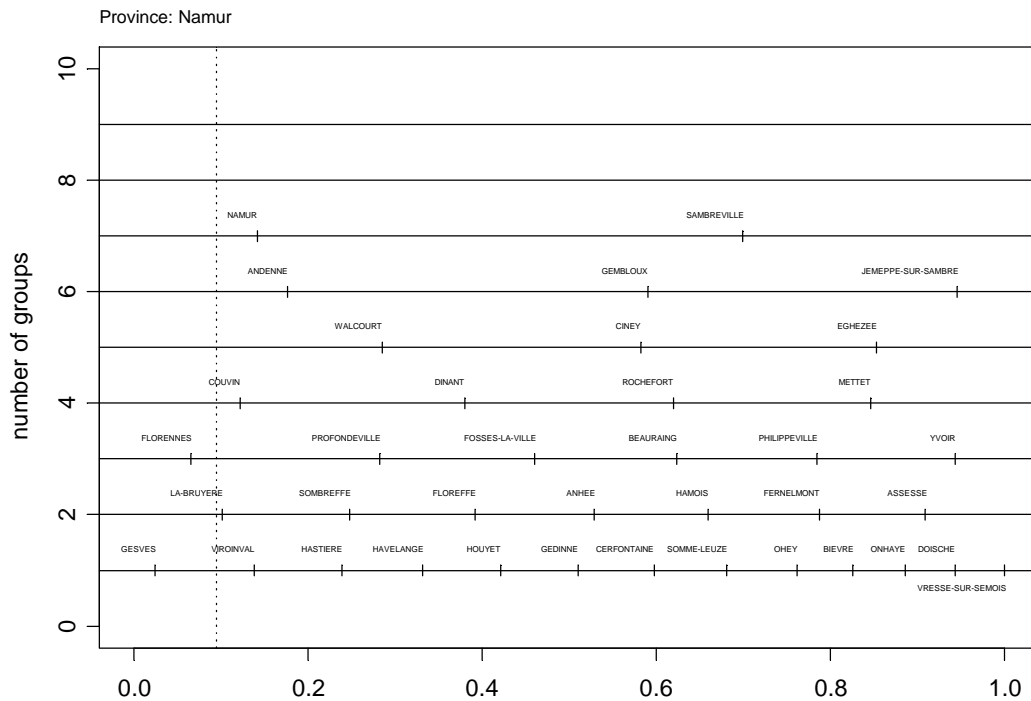
It is impossible to assign the number of PSU such that the probability of selection is totally equal. In Flemish region the probability varies between 0.58/1000 (West-Flanders) and 0.83/1000 (Limburg) where an oversampling of 200 was applied. This is the optimal solution and in fact all are close to the overall average (0.68). The implication is that a reweighing will be necessary to reconstruct the total for the region. A similar remark holds for the Walloon provinces (without Eupen). Here, the figures vary between 0.86/1000 (Walloon Brabant) and 5.10/1000 (Luxembourg) where an oversampling of 1000 interviews was considered. This last situation is not a problem and it will be taken into account by the weighting schemes developed for the estimation procedure. For reference, see Table 1.

8.1.4 The systematic sampling of the municipalities

Once the quota by stratum (by province) is defined, the sampling is straightforward. Within each province (region) the cities are ordered by size from large to small. To give the cities a weight proportional to their size, they are repeated in the list proportional to their population size. Hence the probability of selection is proportional to the size.

To make this procedure more transparent a graphical representation was developed. A segment proportional to its size represents each city. The sums of the segments represent the total population of the stratum. Then this line is cut into a number of pieces equal to the number of groups of 50 individuals one wants to draw from. These segments are put on top of each other. Then taking a systematic sample comes down to drawing an arbitrary vertical line.

The process is illustrated for the province of Namur in the following graph.



1. In the province of Namur nine groups of 50 individuals should be drawn, which is translated into 9 horizontal lines. On this horizontal lines the municipalities are represented by segments that can extend over many horizontal lines; e.g., Namur is so large in comparison with the other cities that it is distributed over more than two horizontal lines. The implication is that Namur as a PSU will be selected at least twice and there is also a chance it is selected 3 times. The same happens in other provinces

and below a summary is given of the cities, which are selected at least once (see table in previous section).

2. On the other hand one notes that all small villages of the province of Namur are located on the bottom line. Hence the procedure assures that one of these villages will be selected as they form one block or “stratum”. A possible issue with the smaller villages however can be that there are not enough inhabitants. However, for all of them the number of inhabitants is larger than 50. The table below gives the five smallest villages in the Flemish region and the Walloon region, the only village with could give problems is Herstappe in the province of Limburg. However the selection probability is very small.

Code	Name	Pop.
73028	Herstappe	85
33016	Mesen	977
34043	Spiere-Helkijn	1854
23009	Bever	1906
45062	Horebeke	1960
84029	Herbeumont	1436
84016	Daverdisse	1360
81013	Martelange	1453
82009	Fauvillers	1825
85047	Rouvroy	1896

3. Another point, which is illustrated, is that the systematic sampling cannot result in any arbitrary combination of the municipalities, a well-known property of systematic sampling. In fact, by fixing the order and working in a systematic way one restricts the number of possibilities and the cities are grouped in loose “strata” from which just one will be selected. E.g., either Namur is selected a third time or Sambreville is selected. And if Namur is selected a third time also Andenne is selected. Using a systematic sampling procedure, the universe of possible samples is reduced and then the more extreme samples are ruled out.

4. Finally the vertical line on the plot is drawn at random by generating a random variable from a uniform distribution. These municipalities whose segment is crossed are selected. For the current line, the selected municipalities are listed in the following table where is also indicated the number of time they were selected:

Nr	Code	Name
1	92003	Andenne
3	92094	Namur
1	92101	Profondeville
1	92141	La-Bruyere
1	93014	Couvin
1	93088	Walcourt
1	93090	Viroinval

Above procedure can be repeated for each province. The corresponding lists can be found in the next section.

8.1.5 The selection

By the sampling scheme described above a random sample was selected. The result is given below. If a municipality (PSU) is selected more than once it is indicated in the “Nr.” column. After this list, the corresponding graphical representations are included.

ANTWERPEN

Nr	Code	Name
8	11002	Antwerpen
1	11013	Edegem
1	11016	Essen
1	11029	Mortsel
1	11030	Niel
1	11037	Rumst
1	11039	Schilde
1	11040	Schoten
1	11052	Wommelgem
1	11053	Wuustwezel
1	11056	Zwijndrecht
1	12014	Heist-Op-Den-Berg
1	12025	Mechelen
1	12026	Nijlen
1	12029	Putte
1	13008	Geel
1	13025	Mol
1	13035	Ravels
1	13046	Vosselaar

VLAAMS BRABANT

Nr	Code	Name
1	23002	Asse
1	23016	Dilbeek
1	23025	Grimbergen
1	23039	Kapelle-Op-Den-Bos
1	23062	Overijse
1	23086	Ternat
1	23101	Sint-Genesius-Rode
1	24014	Boortmeerbeek
1	24062	Leuven
1	24109	Tremelo
1	24133	Linter
1	24134	Scherpenheuvel-Zichem

LIMBURG

Nr	Code	Name
1	71004	Beringen
1	71016	Genk
1	71017	Gingelom
1	71022	Hasselt
1	71034	Leopoldsburg
1	71053	Sint-Truiden
1	71057	Tessenderlo
1	71066	Zonhoven
1	72020	Lommel
1	72039	Houthalen-Helchteren
1	72040	Meeuwen-Gruitrode
1	73001	Alken
1	73083	Tongeren

OOST VLAANDEREN

Nr	Code	Name
1	41002	Aalst
1	41027	Herzele
1	43002	Assenede
1	44011	Deinze
1	44019	Evergem
2	44021	Gent
1	44034	Lochristi
1	44043	Merelbeke
1	44045	Moerbeke
1	44048	Nazareth
1	44072	Waarschoot
1	45041	Ronse
1	46003	Beveren
1	46014	Lokeren
1	46020	Sint-Gillis-Waas
1	46021	Sint-Niklaas

WEST VLAANDEREN

Nr	Code	Name
1	31005	Brugge
1	31022	Oostkamp
1	31033	Torhout
1	31043	Knokke-Heist
1	32006	Houthulst
1	32011	Kortemark
1	34022	Kortrijk
1	34040	Waregem
1	35002	Bredene
1	35005	Gistel
1	35013	Oostende
1	36008	Izegem
1	37012	Ruiselede

HAINAUT

Nr	Code	Name
1	51004	Ath
1	51012	Brugelle
1	51065	Frasnes-Lez-Anvaing
1	52010	Chapelle-Lez-Herlaimont
5	52011	Charleroi
1	52012	Chatelet
1	52015	Courcelles
1	52021	Fleurus
1	52022	Fontaine-L'evêque
1	52043	Manage
1	52048	Montigny-Le-Tilleul
1	53014	Boussu
1	53028	Frameries
1	53039	Hensies
3	53053	Mons
1	54007	Mouscron
1	54010	Comines-Warneton
1	55004	Braine-Le-Comte
2	55022	La-Louvière
1	55050	Ecaussinnes
1	56001	Anderlues
1	56011	Binche
1	56085	Estinnes
1	56086	Ham-Sur-Heure-Nalinnes
1	57018	Celles
1	57064	Peruwelz
2	57081	Tournai

LIEGE

Nr	Code	Name
1	61003	Amay
1	62009	Aywaille
1	62011	Bassenge
1	62015	Beyne-Heusay
1	62022	Chaufontaine
1	62027	Dalhem
1	62051	Herstal
4	62063	Liege
1	62093	Saint-Nicolas
1	62096	Seraing
1	62120	Flemalle
1	63020	Dison
1	63035	Herve
1	63079	Verviers
1	64025	Fexhe-Le-Haut-Clocher
1	64063	Remicourt

LUXEMBOURG

Nr	Code	Name
3	81001	Arlon
1	81004	Aubange
1	82003	Bastogne
1	82005	Bertogne
1	82032	Vielsalm
1	82037	Gouvy
1	82038	Sainte-Ode
1	83012	Durbuy
1	83028	Hotton
1	83031	La-Roche-En-Ardenne
2	83034	Marche-En-Famenne
1	83055	Manhay
1	84010	Bouillon
1	84033	Léglise
1	84035	Libin
1	84043	Neufchateau
1	84059	Saint-Hubert
1	84077	Libramont-Chevigny
1	85009	Etalle
2	85045	Virton
1	85046	Habay

NAMUR

Nr	Code	Name
1	92003	Andenne
3	92094	Namur
1	92101	Profondeville
1	92141	La-Bruyere
1	93014	Couvin
1	93088	Walcourt
1	93090	Viroinval

BRABANT WALLON

Nr	Code	Name
1	25014	Braine-L-Alleud
1	25048	Jodoigne
1	25072	Nivelles
1	25084	Perwez
1	25091	Rixensart
1	25112	Wavre

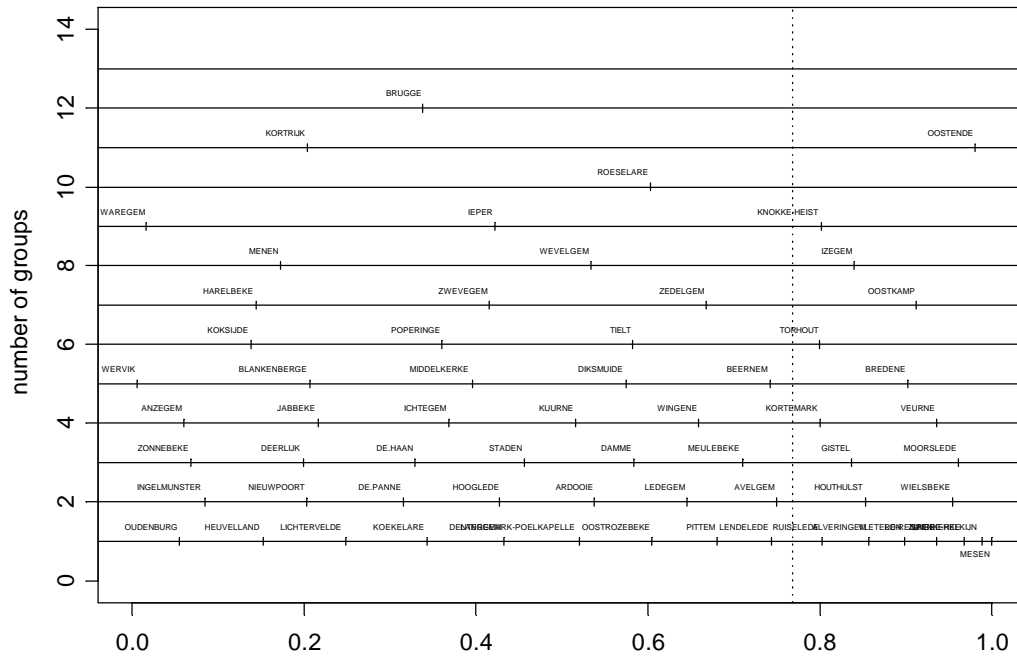
BRUSSELS

Nr	Code	Name
5	21001	Anderlecht
2	21002	Auderghem
2	21003	Berchem-Sai
8	21004	Bruxelles
2	21005	Etterbeek
1	21006	Evere
3	21007	Forest
1	21008	Ganshoren
4	21009	Ixelles
3	21010	Jette
1	21011	Koekelberg
5	21012	Molenbeek-S
2	21013	Saint-Gille
1	21014	Saint-Josse
7	21015	Schaerbeek
5	21016	Uccle
2	21017	Watermael-B
3	21018	Sint-Lambre
3	21019	Sint-Pieter

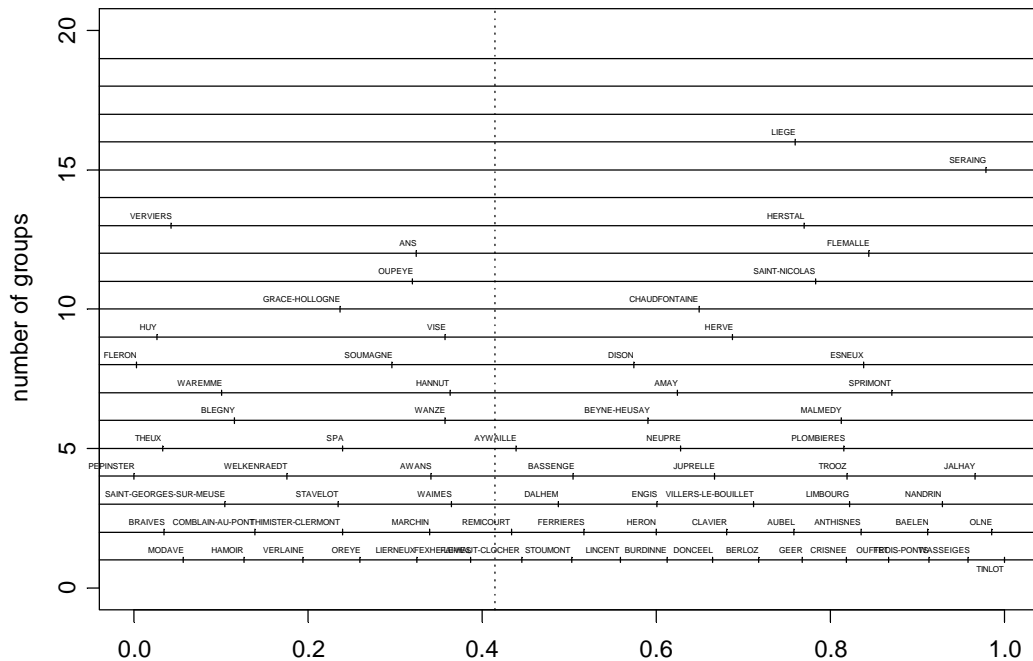
GERMAN COMMUNITY

Nr	Code	Name
1	63001	Ambleve
1	63013	Butgenbach
2	63023	Eupen
1	63040	La-Calamine
1	63067	Saint-Vith

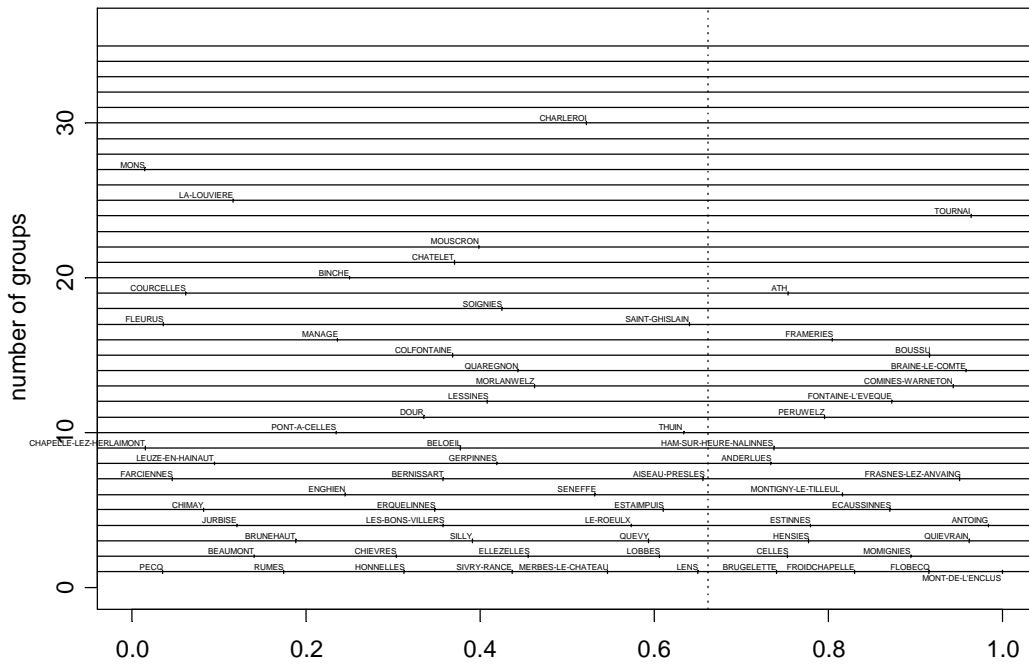
Province: West Flanders



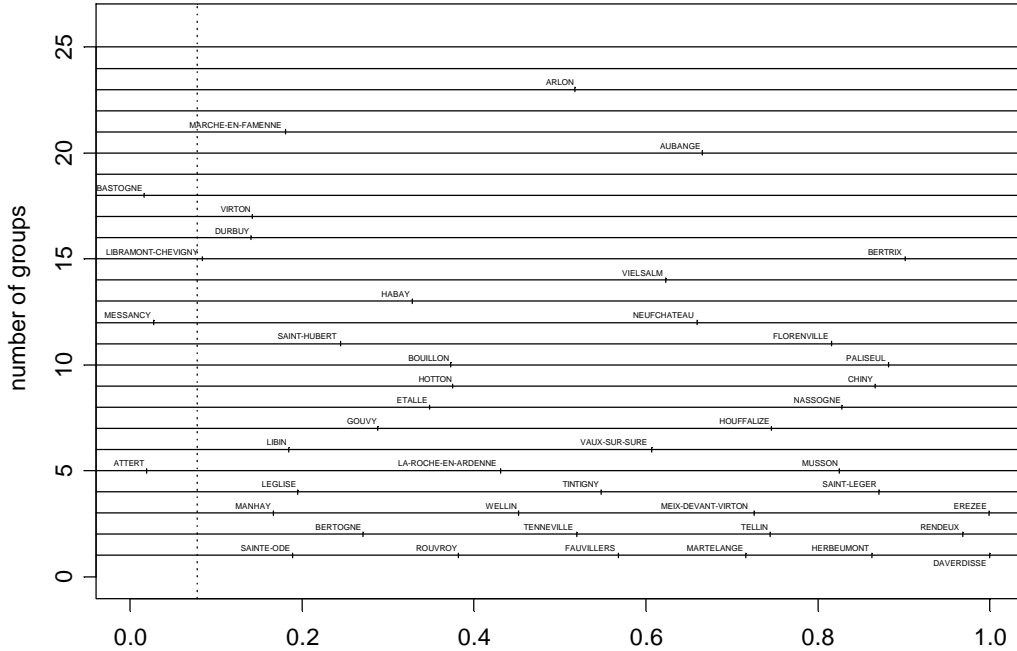
Province: Liege



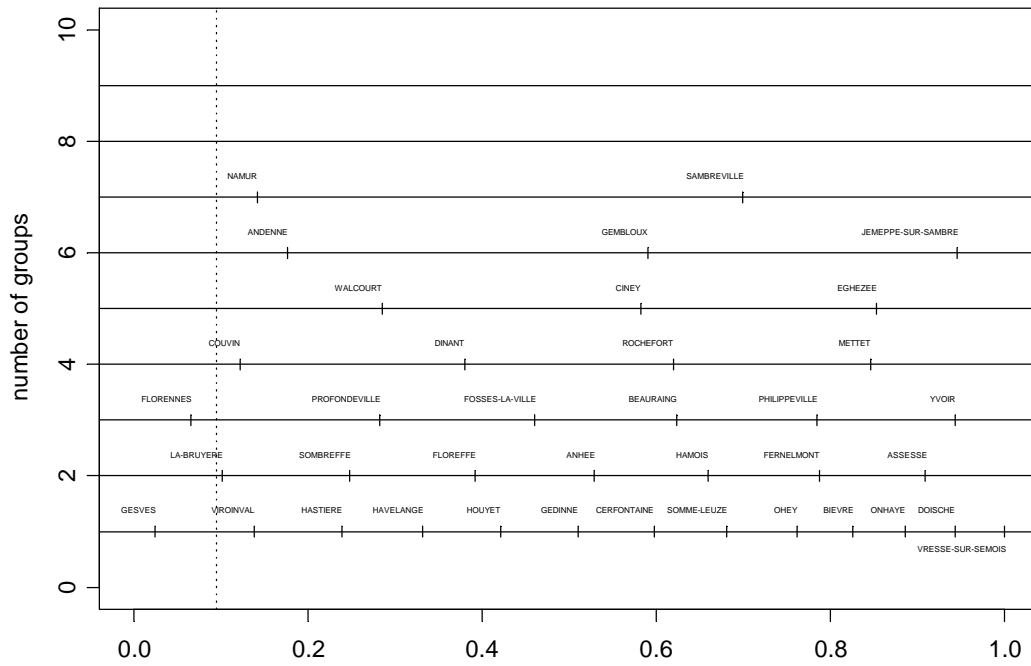
Province: Hainaut with 500 oversampling



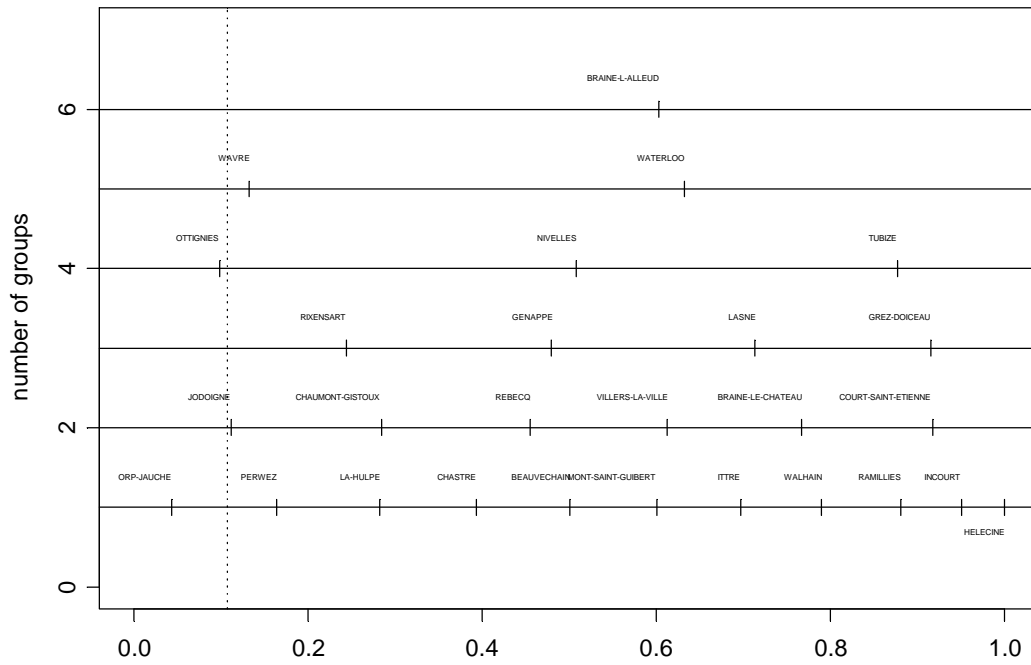
Province: Luxembourg with 1000 oversampling



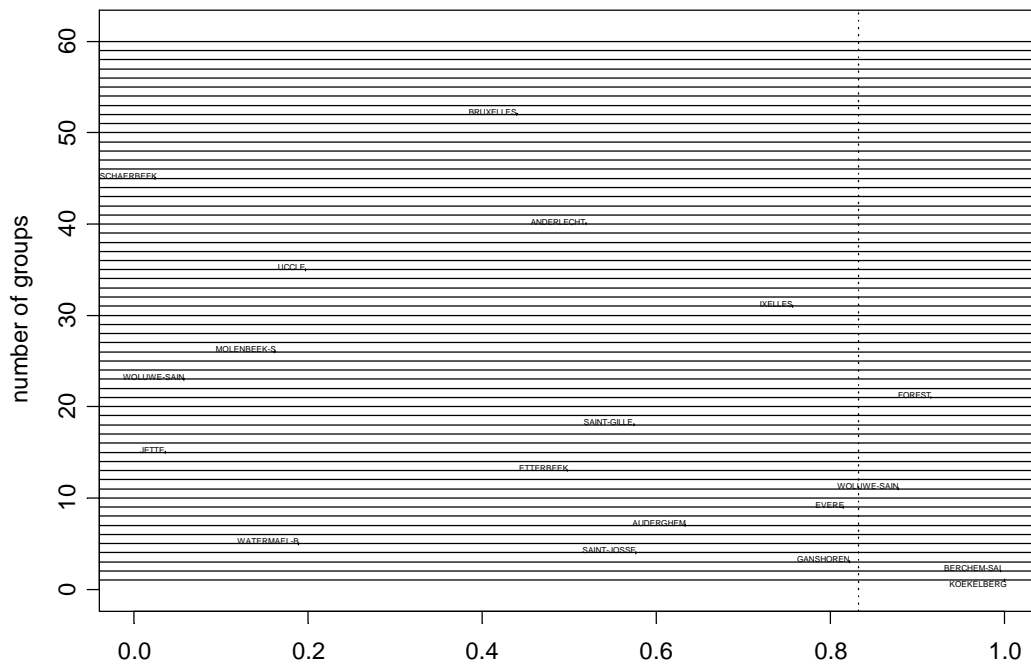
Province: Namur



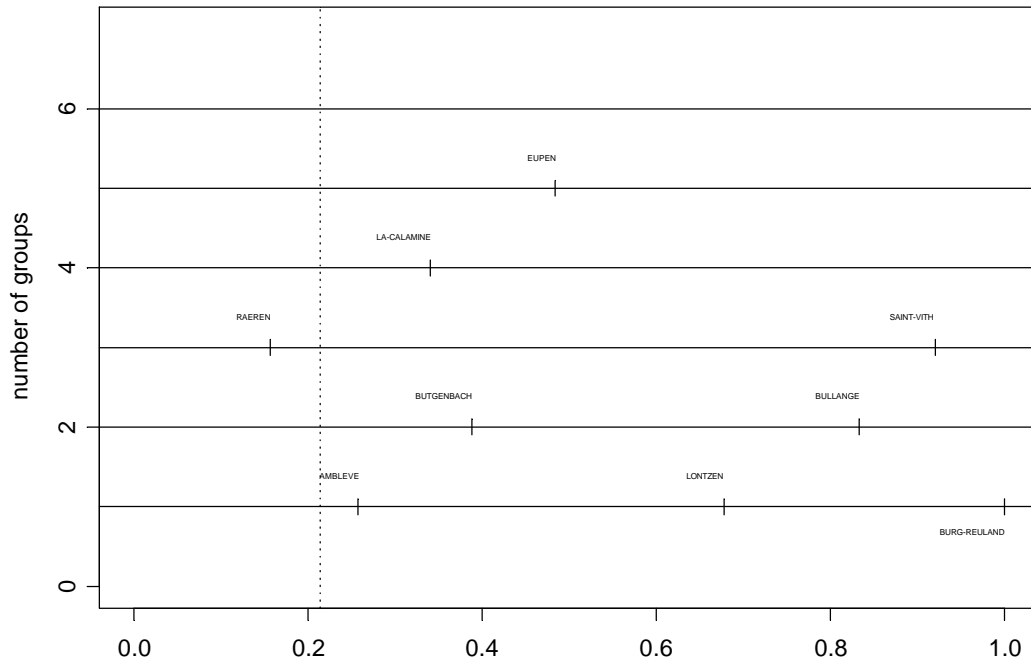
Province: Wallon Brabant



Province: Brussels



Province: Eupen



8.3 Indicators for the Fieldwork

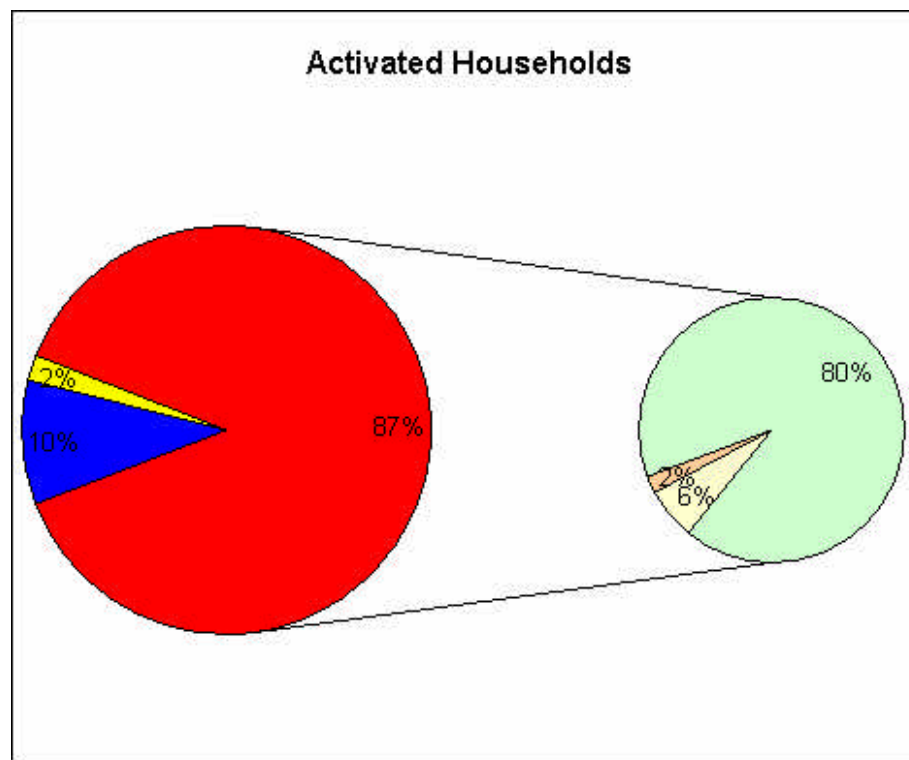
		Belgique	Bxl.	VI.	Wal.	Antw.	Brab.Fl.	Brab. Wal.	Fl. Occ.	Fl. Or.	Hainaut	Liege	Limb.	Lux.	Namur
Target	nb of HH														
Activated	nb of HH														
	% of target														
Completed	nb of HH														
	% of activated														
	% of target														
Not-participating	nb of HH														
	% of activated														
Refusals	nb of HH														
	% of activated														
	% of contacted HH ⁴														
Non-contactable	nb of HH														
	% of activated														
De-activated	nb of HH														
	% of activated														
Stand-by	nb of HH														
	% of activated														

⁴ Number of HH that refuse to participate divided by the number of households that could be contacted (completed and refused).

Belgique

End of Week 6

TARGET	nb of HH	3012,00
ACTIVATED	nb of HH	15000,00
COMPLETED	nb of HH	1500,00
	% of activated	10,00
	% of target	49,80
DE-ACTIVATED	nb of HH	300,00
	% of activated	2,00
STAND-BY	nb of HH	12100,00
	% of activated	80,67
REFUSAL	nb of HH	850,00
	% of activated	5,67
NOT-CONTACT.	nb of HH	250,00
	% of activated	1,67



9. List of Abbreviations

HH	: Household
PSU	: Primary sampling unit
SSU	: Secondary sampling unit
TSU	: Tertiary sampling unit
HIS	: Health Interview Survey
NR	: National Register
NIS	: National Institute of Statistics
PH	: Participating Household
NPH	: Non Participating Household
NCH	: Non contactable Household
RH	: Refusal Household
NEH	: Non eligible household
RP	: Reference person
HIS97	: Belgian Health Interview Survey 1997
HIS2001	: Belgian Health Interview Survey 2001

10. References

1. Burzykowski T., Molenberghs G., Tafforeau J., Van Oyen H., Demarest S. and Bellammamer L., Missing data in the Health Interview Survey in Belgium, *Arch. Public Health*, (1999), **55**: 1-13.
2. Cochran W. G., Sampling techniques - third edition, New York: Wiley, (1977).
3. Foreman E. K., Survey sampling principles, New York: Marcel Dekker, inc., (1991).
4. Graubard B. I. and Korn E. L., Modelling the sampling design in the analysis of health surveys, *Stat. Meth. Med. Res.*, (1996), 5: 263-281.
5. Hermans H., Lambert, M., Reginster, G., Tafforeau, J. and Van Oyen, H. Naar en gezondheidsenquête door middel van interview in België. Brussels. IHE, 1995;**1-193**.
6. Kish L., Survey Sampling, New York: Wiley, (1995).
7. Lemeshow, S., Levy, P. Sampling of Populations: Methods and Applications. 3rd Edition, Wiley 1999.
8. Molenberghs, G. (1994). Health Interview Survey: Statistische Beschouwingen, *Scientific Institute of Public Health, Technical Report*.
9. Quataert, P., Van Oyen, H., Gegevensinzameling i.v.m. middelengebruik d.m.v. CATI. IHE/Episerie N° 6. C.O.O.V., Instituut voor Hygiëne en Epidemiologie, 1995.
10. Quataert, P., Van Oyen, H., Tafforeau, J., Schietecatte, E., Lebrun, L., Bellamammer, L., Molenberghs, G. (1997). The Health Interview Survey 1997: Protocol for the Selection of the Households and the Respondents. *Scientific Institute of Public Health Episerie 12*.
11. Renard D, Molenberghs G, Van Oyen H and Tafforeau J (1998). *Investigation of the clustering effect in the Belgian Health Interview Survey 1997*, *Archive of Public Health*, **56**, 345-361.
12. Scheaffer, R.L., Mendenhall, W., Ott L. (1990). *Elementary Survey Sampling* Boston: PWS-Kent.
13. Tellier V., Demarest S., Leurquin P., Tafforeau J., Van der Heyden J., Van Oyen H., La santé de la population en Belgique et à Bruxelles. Enquête de Santé par Interview, Belgique, 1997, Brussels: Centre de Recherche Opérationnelle en Santé Publique, Institut Scientifique de la Santé Publique, (1998).

14. Van der Heyden J., Demarest S., Leurquin P., Tafforeau J., Tellier V., Van Oyen H., De gezondheid van de bevolking in het Brussels Gewest. Samenvatting. Gezondsheidenquête, België, 1997, Brussels: Centrum voor Operationeel Onderzoek in Volksgezondheid, Wetenschappelijk Instituut Volksgezondheid-Louis Pasteur, (1998).
15. Van Oyen H., Tafforeau J., Hermans H., Quataert P., Schiettecatte E., Lebrun., Bellamammer L., The Belgian Health Interview Survey, Arch. Public Health, (1997), 52: 79-82.
16. Van Steen, K., Burzykowski, T., Molenberghs, G. (1999) Evaluation of the design and conduct of the Belgian Health Interview Survey 1997 – Final Report.
17. Vehovar, V. (1999).Field substitution and unit nonresponse. *Journal of Official Statistics*, **15**, 335--350.