Respondent selection within the household - A modification of the Kish grid

Renata Nemeth

Hungary

Abstract: The problem of drawing a person from a household often occurs at the final stage of a survey design, e.g. in telephone surveys, after we contacted the households using random digit dialing. The Kish grid gives an algorithm for this random selection. We found that, contrary to the widely held opinion, the grid is not capable of providing representativeness by gender and age. This misconception stems from the fact that when the Kish grid was developed in the 1950's, both randomness and representativeness could be achieved using the method, due to the household structure of the USA. We show that this does not hold for today's Hungary. Finally, we suggest a modification of the Kish grid that is more appropriate for selecting a representative sample.

1 Introduction

Information on characteristics of populations is constantly needed nowadays. For reasons relating to time-limit and costs, this information is often obtained by use of sample surveys. A sample survey may be defined as a study involving a subset (*sample*) selected from a larger population. Characteristics of interest are measured on each of the sampled individuals. ¿From this information, extrapolations can be made concerning the entire population. The validity and reliability of these extrapolations depend on how the sample was chosen.

2 Reasons for sampling households

It is often best to draw the sample in two stages. These are designs in which primary sampling units are selected at the first stage, and secondary sampling units are selected at the second stage within each unit selected previously. The sampling designs considered in this paper are those in which households are selected at first, and then one adult member of each selected household is chosen.

When does the need for two-stage sampling arise, rather than selecting the respondents directly from the population? Lists of adults, from which the sample can be taken, are often not available. For example, the electoral register is usually a good quality database of addresses, but a poor quality database of individual adults. The register has many errors because of non-registration and population mobility. In practice, the register is used to construct a sample of flats or households, and the sample of adults is obtained at a second stage in some other way.

Another method involving respondent selection within household is called area sampling. It is used when the target population is located in a geographical region, such as a city. A frame for studying a population of a city may in the first stage consist of a list of districts, followed by a list of streets, followed by a list of blocks, then a list of households. And again, at the final stage, a sample of respondents is obtained from the sample of households.

The problem of translating a sample of the households into a sample of adult persons often arises in telephone surveys as well. The households are usually contacted by random digit dialing.

There is no need of selection if the respondent is uniquely defined e.g. as the head of the household. Suppose the household contains more than one member of the desired population. One may decide to include in the sample every member within the household. This may be a statistically inefficient procedure, unless one of these two conditions hold:

- There is seldom more than one member of the population in the household.
- If intra-class correlation within the household of the variables measured is of negligible size. Otherwise, the distribution is characterized by some homogeneity. Usually, the homogeneity of households is greater than in the case when individuals were assigned to them at random. Since homogeneity within sample clusters increases the variance of estimations, the sampler wants to reduce it in this case by selecting only one member per household (see Kish, 1965).

These conditions generally do not hold in surveys. Hence, there is a need for a procedure of selection that will translate a sample of households into a sample of the adult population. It is desired to make not more than one interview in every household. On the other hand, an interview in every sample household is desired in order to avoid futile calls on households without interviews. Finally, the procedure should be applied and checked without great difficulty. The simplest procedure we could apply is the uncontrolled selection in which the interview is conducted with whomever opens the door or answers the phone. A serious problem comes up in this case. The resulting sample will be made up of those persons more likely to be available at the time interviewers call or who are most willing to be interviewed. Experiences seen to date show that they are made up of women and older adults.

3 The Kish grid

The Kish grid gives a procedure of selection. The expression ”Kish grid” comes from the name of Leslie Kish, the Hungarian born American statistician. Kish was one of the world’s leading experts on survey sampling.

When creating the grid Kish intended to select persons within the household with equal probability. On the other hand the use of the grid can be checked easily contrary to e.g. a decision depending on the toss of a coin.

When applying the Kish grid, the interviewer at the first step uses a simple procedure for ordering the members of the household. A cover sheet is assigned to each sample household. It contains a form for listing the adult occupants (see Table 1), and a table of selection (see Table 2).

Table 1: Form for listing the adult occupants (see Kish, 1965)

Relationship	Sex	Age	Adult No.	Selection
Head	M		2	
Wife	F	40	5	
Head's father	M		1	
Son	M		3	
Daughter	F		6	
Wife's aunt	F	44	4	✓

The interviewer lists each adult on one of the lines of the form. Each are identified in the first column by his relationship to the head of the household. In the next two columns, the interviewer records the sex and, if needed, the age of each adult. Then the interviewer assigns a serial number to each adult. First, the males are numbered in order of decreasing age, followed by the females in the same order. Then the interviewer consults the selection table. this table tells him the number of the adult to be interviewed. In the example, there are six adults in the household, and selection table D tells to select adult number 4 (see Table 2).

Table 2: One of the eight selection tables (see Kish, 1965)

1.5cm

Selection table D	
If the number of adults in household is:	Select adult numbered:
1	1
2	2
3	2
4	3
5	4
6 or more	4

Selection table D is only one from the 8 types (see Table 3). One of the 8 tables (A to F) is printed on each cover sheet. The cover sheets are prepared to contain the 8 types of selection tables in the correct proportion, e.g. table A is assigned to one-sixth of the sample addresses. The aim is to reach equal selection probabilities within household without the necessity of printing many more forms. Table 4 shows the selection probabilities. It can be seen that the chances of selection are exact for all adults in households with 1, 2, 3, 4 and 6 adults. As numbers above six are disallowed, there are some adults who are not represented. Moreover, there is an overrepresentation of number five in the households with five adults.

Table 3: Summary of eight selection tables (see Kish, 1965)

Proportion of assigned tables	Table number	If the number of adults in household is: 1	2	3	4	5	6 or more
		Select adult numbered:					
1/6	A	1	1	1	1	1	1
1/12	B1	1	1	1	1	2	2
1/12	B2	1	1	1	2	2	2
1/6	C	1	1	2	2	3	3
1/6	D	1	2	2	3	4	4
1/12	E1	1	2	3	3	3	5
1/12	E2	1	2	3	4	5	5
1/6	F	1	2	3	4	5	6

Table 4: Summary of selection probabilities

Adult numbered	If the number of adults in household is: 1	2	3	4	5	6 or more
1	1	1/2	1/3	1/4	1/6	1/6
2		1/2	1/3	1/4	1/6	1/6
3			1/3	1/4	1/4	1/6
4				1/4	1/6	1/6
5					1/4	1/6
6						1/6
7 or more						0

It may be noted that the procedure was modified several times by many researchers. Kish himself suggests modifying the tables for special reasons. In paper-and-pencil interview the interviewer uses the grid as described above. In computer-assisted telephone or personal interviews, the tables are randomly assigned to the households by the computer in their prescribed proportions. The researchers stick to ordering the persons by sex and age, though they have the technical background for generating random number. By using random numbers, it would be possible to select a person from the set of the previously identified adults. Although nobody expresses it explicitly, they consider the sample to be *representative* by sex and age with the use of the original Kish grid. This representativity would much less be expected if the applied procedure was e.g. identifying the adults by first name, then selecting one of them by generating a random number.

4 ”Representativity”

We mentioned that representativity is a desirable character of a sample. It refers to the similarity between the sample and the population in some characteristics of interest. Why is it desirable to reproduce the distribution of certain population characteristics in the sample? Suppose there is a high positive correlation between the characteristic to be estimated and a different one. The more representative the sample is by the latter one, the more reliable the estimation of the former one will be. (The *reliability* of an estimator is evaluated on the basis of its variance.)

It is a standard practice to evaluate the sample by its representativity in order to support the validity of the extrapolations or estimations. We attempted to take into account the accessible literature about samples obtained by using the Kish grid. When evaluating the representativity of their samples, Hungarian researchers often refer to the undersampling of males and overrepresentation of elderly people (see ISSP Család II, 1994, Táblaképek az egészségrõl, 2000, Egészségi Állapotfelvétel, 1994).

According to the comments this deviation stems from problems that occur when putting the interview into practice, e.g. males are undersampled because they are more difficult to find at home, and are less willing to participate. Later we will give some theoretical evidence that explains the representation problems without considering these assumptions.

It is important to mention that according to Kish’s own words, he used the variables sex and age only for ordering the household members. He did not aim explicitly to reproduce the sex and age distributions. At the same time, however, he expected the sample to be representative. In the article published first about the grid, Kish checked the distribution of the respondents, and explained the males’ underrepresentation with practical problems mentioned above: they are more difficult to find at home etc. [Kish, 1949]. Although he emphasized the fact that the grid is for random selection within household, he was the first not to make a distinction between randomness and representativity.

4.1 The cause of non-representativity - assumption

When households are selected with equal probabilities, and the selection probabilities within household are equal, then the chance of selection of a single adult becomes inversely proportional to the number of adults in the household. Hence overall selection probabilities are not equal.

If the selection probability is the function of the household size, and the household size is not independent from the members’ demographic characteristics, then the sampling design itself is the source of the representation problems. The sample would not be representative even if we could obtain perfectly random household sample and 100 percent response rate. Kish found that the samples obtained by using the grid show close agreement with the population data on important demographic characteristics. Kish developed the grid in the 1950’s USA. He emphasized the relatively low variance of the selection probabilities. That was because of the high concentration within a small range of household sizes: over 70 percent of households contained two adults (see Table 5).

Our results so far show that representativity is the function of the current household

Table 5: Household structure, 1957, USA (see Kish, 1965)

Number of adults in the household	1	2	3	4	5	6 or more
Proportion	14.6	73.0	9.0	2.8	0.4	0.2

structure, and the grid's performance depends on where and when it is used. It is worth making a comparison between the current Hungarian household structure and the one observed by Kish. In today's Hungary, 26 percent of the households are one-person-household, that is 2 times greater than the one examined by Kish. This difference for itself is so significant that the question arises whether to accept the grid without modification.

4.2 The cause of non-representativity - proof

To put assumptions into a concrete form, we determined the exact connection between the grid's performance and the population household structure. The required information on current Hungarian population is not available. That is the reason why we worked with a sample from a great national household survey[1]. The data contain information on each member of the sample households, so we could use it as a population for further sampling. In the following we will refer to it as "*pseudopopulation*". Table 6 shows age and sex distribution in the pseudopopulation.

Table 6: Pseudopopulation, age and sex distribution (n=4188)

	Sex		
Age	Male	Female	Total
18-39	19.79	20.51	40.31
40-59	14.66	17.60	32.26
60+	10.94	16.50	27.44
Total	45.39	54.61	100.00

We tested the use of the grid by this pseudopopulation, concerning the age and sex distribution in the samples. The *expected* sex and age proportions of the sample can be formulated as follows. Let p_{kl} denote the selection probability of the adult l living in a household of size k ($k = 1, \ldots, 6, l = 1, \ldots, k$), supposing the household is already selected. As we choose households with equal probabilities, the chance of choosing a household of size k equals to the proportion of these households. Let H_k denote this value. The expected sex and age-group joint distribution can be given by a 3×2 matrix, denoted by a. $a[11]$ is the proportion of young males, $a[21]$ is the proportion of middle-aged males etc., $a[32]$ is the proportion of elderly females.

[1]The TÁRKI Social Research Databank made the database of "Hungarian Household Panel IV"available for us.

We need the information on households' composition: after selecting the person number l within a household of size k, what is the probability of his or her being a male or female, of his or her being young or middle-aged or elderly. Let a_{kl} be 3×2 matrix ($k = 1..6$, $l = 1..k$) In the above way, $a_{kl}[11]$ denotes the proportion of young males among the persons numbered l living in a household of size k, $a_{kl}[21]$ is the proportion of middle-aged males etc.

The expected age and sex joint distribution is the function of the other parameters, see Equation 1. H_k, a, and a_{kl} are known input parameters. They come from the information about pseudopopulation.

$$a[ij] = \sum_{k=1..6} H_k \left(\sum_{l=1..k} p_{kl} a_{kl}[ij] \right), i = 1, 2, 3, j = 1, 2 \tag{1}$$

Substituting the known parameters, we obtain the expected distribution shown in Table 7.

Table 7: Expected sample, age and sex distribution

	Sex		
Age	Male	Female	Total
18-39	17.27	19.47	36.74
40-59	12.92	17.04	29.96
60+	11.87	21.43	33.30
Total	42.06	57.94	100.00

It is seen that the expected sample differs from the population in sex and age distributions. Firstly, the elderly people are oversampled, especially the women. (It is worth mentioning, that in the current population of Hungary, great proportion of the one-person-households consists of an older female occupant, and the quarter of all households is one-person-household, so it can be concluded that it is more likely to select an elderly female in this way than by simple random sampling.) Secondly, males appeared to be underrepresentated. Our experiences are similar to those obtained from real surveys.

5 Modification of the Kish grid

In this section we present a modification of the Kish grid. We intended to receive a representative or at least more representative expected sample. We modified the grid by modifying the selection tables. This modification method is not unprecedented in the literature: Kish himself already suggested modifying the tables when needed.

All the sampling features are fixed, i.e. the following conditions hold:

- each household has the same chance of selection

- one and only one interview per household is made
- the selection tables are based on a list of the household members
- this ordering is made by sex and age
- the population to be surveyed is the previously mentioned pseudopopulation
- 12 selection tables are used (Obviously, the more tables are used, the finer probabilities can be achieved, that is the closer agreement between the sample and the population can be obtained. This is why we limited the number of the tables.)
- the same rules are applied to households with 6 ore more members.

The problem is to make selection tables those yield a sample giving close agreement with the pseudopopulation data. The modification can be simplified: instead of determination of the tables, it is enough to determine the selection probabilities.

Our aim is to obtain a representative expected sample, which is as close to the distribution given by table 6 as possible. Let A denote the 3×2 matrix describing the sex and age joint distribution in the pseudopopulation, $A[11]$ equals to the young males proportion etc. Using the notation of Equation 1 the problem is as follows. H_k's and a_{ij}'s are given parameters, and a is to be determined as the functions of p_{kl}'s, so as to reproduce A. Equation 2 is to be solved.

$$\sum_{i=1,2,3,j=1,2} |a[ij] - A[ij]| = 0, \tag{2}$$

with constraints:

$$\begin{aligned} &\textstyle\sum_{j=1..i} p_{ij} = 1, \forall i \\ &p_{ij} > 0, \forall i, j \\ &p_{ij} = \tfrac{k_{ij}}{12}, \forall i, j, \; where \; k_{ij} \in \mathbb{N} \end{aligned} \tag{3}$$

The constraints make the solution to meet the conditions: one and only one person per households is needed, and 12 tables are used that means probabilities are given in $\frac{1}{12}$. The model is a nonlinear equation, with inequality and integer constraints. We applied the Solver software of Microsoft Excel to solve the equation. The problem has no solution.

This raises the question whether there would be a solution if the limitation of the number of the tables did not hold. Using more tables has a practical disadvantage: it implies increasing costs. Moreover, the sample size itself is an upper limit of the tables. Apart from this, the theoretical problem is worth considering. In this case, the integer constraint is to be omitted from (3). The problem does not have a solution in this way either.

It is impossible to obtain a perfectly representative sample. Let us modify the problem instead: which selection table yields a sample that is the closest to the pseudopopulation. A distance function is needed to find the closest solution that minimizes the distance function. We used two functions according to two different approaches. The first one is similar to the Pearson-chi-sqare, Equation 4 shows function f to be minimized.

$$f(a) := \sum_{i=1,2,3,j=1,2} \frac{(a[ij] - A[ij])^2}{A[ij]} \tag{4}$$

The idea of using the other distance function comes from weighting which is widely used in survey statistics. Weights are generally used to improve the precision of the estimate. Poststratification is a weighting method that produces a sample in which each stratum is represented in its proper proportion. In our case, strata are defined as the 6 cells of the sex and age group crosstable. Postratification weight for a given person in a given stratum is defined as the proportion of the population stratum divided by the proportion of the sample stratum. The disadvantage of using the poststratification is that in some cases it increases the variance of the estimation. Increase in variance is a monotonic function of the sum of the weights squared. This implies the following approach: to find the selection table that yields a sample with the minimal sum of postratification weights squared. Equation 5 shows function g to be minimized.

$$g(a) := \left(\frac{1}{n}\right) \sum_{k=1..n} w_k^2 = \sum_{i=1,2,3,j=1,2} \frac{A[ij]^2}{a[ij]}, \tag{5}$$

where n is the sample size.

As we mentioned when finding the solution of the equation with absolute values, the constraints can be determined in two different ways. If they include the integer constraint, then the use of 12 tables are assumed. Otherwise, the number of the tables is not limited, therefore selection probabilities can be any real numbers between zero and one. Combining the two dimensions four problems appear: let us find the minimum value of function f or g, with or without the integer constraint.

A model, in which the objective function or any of the constraints is not a linear function of the variables, is called a nonlinear programming (NLP) problem. In our case, inequality and integer constraints are added to the model. The Weierstrass's theorem states that a real valued continuous function on a closed bounded set assumes a maximum and a minimum value. In our case the conditions of Weierstrass's theorem meets, but the determination of the minimum is not a simple mathematical problem. Apart from special cases the nonlinear optimization problems have numerical solutions. We applied the Solver software of Microsoft Excel to find the minimums.

Table 8-12 contain the results.

We can observe some expected trends in all the four cases. For example $p_{21} \approx 2/3$, that affects against the male underrepresentation that we found when using Kish grid (since p_{21} is the selection probability of the first adult in a two-persons-household, and the first one tends to be male because of the ordering procedure.).

The optimal sex ang age group distributions (matrix a) compared to the one belonging to the Kish grid shows that we managed to improve the young people and the females agreement with the population data, while other cells show some change for the worse.

The four solutions do not differ from each other, either regarding a or p_{ij}'s. This means it is not worth using more than 12 tables. Moreover, the return value of function g at the optimum place of function f is very close to the real optimum value of g - and vice

Table 8: Optimization - results

Original Kish grid									
	Function value when substituting the Kish grid	Matrix a (expected sex/age %)		p_{ij}'s					
	f : 0.021573301	17.27	19.47	p_{21}	1/2	p_{31}	1/3	p_{41}	1/4
	g : 1.018901199	12.92	17.04	p_{22}	1/2	p_{32}	1/3	p_{42}	1/4
		11.87	21.43			p_{33}	1/3	p_{43}	1/4
								p_{44}	1/4
				p_{51}	1/6	p_{61}	$1/6$		
				p_{52}	1/6	p_{62}	$1/6$		
				p_{53}	1/4	p_{63}	$1/6$		
				p_{54}	1/6	p_{64}	$1/6$		
				p_{55}	1/4	p_{65}	$1/6$		
						p_{66}	$1/6$		

versa, i.e. the optimal tables are close to each other whether measured by f or measured by g. We can say that the optimal methods perform well from both points of view.

Table 13 presents the modified selection table obtained by the function f with the integer constraint.

The main results of our work are as follows.

- The samples obtained in Hungary by using the Kish tables differ from the population in sex and age group distributions. We proved that this is caused by the sampling method and not by practical problems. The literature does not prove this connection, nor does it normally mention it.

- The grid is successfully modifiable if our aim is to adjust the sample to the population.

- The problem we treated is of international significance. The trends observed in Hungarian household structure are global trends. Size of households is currently decreasing, the proportion of the single persons is on the rise.

There are further problems to be considered. We limited our analysis to the representativeness by sex and age. It should be useful to take into account the distributions of other characteristics when using the grid. At the same time, the distributions of other characteristics need checking when using the modified tables. Obviously, improving the sex and age adjustment does not mean that the sample shows agreement to the population with respect to other variables.

Change in selection probabilities implied by the modification needs further consideration. The variability of the probabilities can result in an increase of the estimation variance.

Table 9:

Optimization of function f

Constraints	Optimal value	Matrix a (expected sex/age %)		p_{ij}'s					
$\sum_{j=1..i} p_{ij} = 1, \forall i$	0.013938589	18.56	19.94	p_{21}	2/3	p_{31}	2/12	p_{41}	1/12
$p_{ij} > 0, \forall i, j$		12.81	16.16	p_{22}	1/3	p_{32}	1/12	p_{42}	3/12
$p_{ij} = \frac{k_{ij}}{12}, \forall i, j,$		13.11	19.42			p_{33}	9/12	p_{43}	1/12
$where\ k_{ij} \in \mathbb{N}$								p_{44}	7/12
				p_{51}	1/12	p_{61}	1/12		
				p_{52}	8/12	p_{62}	7/12		
				p_{53}	1/12	p_{63}	1/12		
				p_{54}	1/12	p_{64}	1/12		
				p_{55}	1/12	p_{65}	1/12		
						p_{66}	1/12		

The optimal solution was derived from the pseudopopulation. It would be worth developing the study with the real population as a starting point in order to support the generalization of the results.

Table 10:

Optimization of function f									
$\sum_{j=1..i} p_{ij} = 1, \forall i$	0.013216638	18.75	20.15	p_{21}	0.6539	p_{31}	0.2351	p_{41}	0.0100
$p_{ij} > \frac{1}{100}, \forall i, j$		12.93	15.84	p_{22}	0.3461	p_{32}	0.0100	p_{42}	0.3728
		13.07	19.26			p_{33}	0.7549	p_{43}	0.0100
								p_{44}	0.6072
				p_{51}	0.0100	p_{61}	0.0100		
				p_{52}	0.8877	p_{62}	0.9500		
				p_{53}	0.0100	p_{63}	0.0100		
				p_{54}	0.0823	p_{64}	0.0100		
				p_{55}	0.0100	p_{65}	0.0100		
						p_{66}	0.0100		

Table 11:

Optimization of function g									
Constraints	Optimal value	Matrix a (expected sex/age %)		p_{ij}'s					
$\sum_{j=1..i} p_{ij} = 1, \forall i$	1.012869186	18.54	19.60	p_{21}	2/3	p_{31}	2/12	p_{41}	2/12
$p_{ij} > 0, \forall i, j$		13.14	16.07	p_{22}	1/3	p_{32}	1/12	p_{42}	3/12
$p_{ij} = \frac{k_{ij}}{12}, \forall i, j,$		13.23	19.42			p_{33}	9/12	p_{43}	1/12
$where\ k_{ij} \in \mathbb{N}$								p_{44}	6/12
				p_{51}	1/12	p_{61}	1/12		
				p_{52}	7/12	p_{62}	7/12		
				p_{53}	1/12	p_{63}	1/12		
				p_{54}	2/12	p_{64}	1/12		
				p_{55}	1/12	p_{65}	1/12		
						p_{66}	1/12		

Table 12:

Optimization of function g									
$\sum_{j=1..i} p_{ij} = 1, \forall i$	1.012269154	18.63	19.93	p_{21}	0.6574	p_{31}	0.2543	p_{41}	0.0270
$p_{ij} > \frac{1}{100}, \forall i, j$		13.11	15.91	p_{22}	0.3426	p_{32}	0.0100	p_{42}	0.3866
		13.23	19.19			p_{33}	0.7357	p_{43}	0.0100
								p_{44}	0.5744
				p_{51}	0.0100	p_{61}	0.0100		
				p_{52}	0.5594	p_{62}	0.7638		
				p_{53}	0.0100	p_{63}	0.1962		
				p_{54}	0.4106	p_{64}	0.0100		
				p_{55}	0.0100	p_{65}	0.0100		
						p_{66}	0.0100		

Table 13: The modified selection tables

		If the number of adults in household is:					
Proportion of	Table	1	2	3	4	5	6 or more
assigned tables	number	Select adult numbered:					
1/12	1	1	1	1	1	1	1
1/12	2	1	1	1	2	2	2
1/12	3	1	1	2	2	2	2
1/12	4	1	1	3	2	2	2
1/12	5	1	1	3	3	2	2
1/12	6	1	1	3	4	2	2
1/12	7	1	1	3	4	2	2
1/12	8	1	1	3	4	2	2
1/12	9	1	2	3	4	2	3
1/12	10	1	2	3	4	3	4
1/12	11	1	2	3	4	4	5
1/12	12	1	2	3	4	5	6

References

D. Binson, J.A. Canchola, and J.A. Catania. Random selection in a telephone survey: A comparison of the kish, next-birthday, and last-birthday methods. *Journal of Official Statistics*, 16:53–59, 2000.

R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls, and J. Waksberg, editors. *Telephone Survey Methodology*. John Wiley and Sons, Inc., New York, 1988.

ISSP. ISSP 1994, Család II. TÁRKI, 1994. `http://www.tarki.hu`.

J.M. Kennedy. A comparison of telephone survey respondent selection procedures, 1993. Available at `http://www.indiana.edu/csr/aapor93.html`.

L. Kish. A procedure for objective respondent selection within the household. *Journal of the American Statistical Association*, pages 380–387, 1949.

L. Kish. *Survey Sampling*. John Wiley and Sons, Inc., New York, 1965.

Egészségi Állapotfelvétel 1994 - Életmód, kockázati tényezõk. Központi Statisztikai Hivatal.

Magyar Statisztikai Évkönyv 1993. Központi Statisztikai Hivatal, Budapest, 1994.

Mikrocenzus, 1996 A népesség és a lakások jellemzõi. Központi Statisztikai Hivatal, Budapest, 1996a.

Mikrocenzus, 1996 Fõbb eredmények. Központi Statisztikai Hivatal, Budapest, 1996b.

Magyar Statisztikai Évkönyv 1999. Központi Statisztikai Hivatal, Budapest, 2000.

P.J. Lavrakas. Telephone survey methods: Sampling, selection and supervision. *Applied Social Research Methods Series*, 7, 1993.

P.S. Levy and S. Lemeshow. *Sampling of Populations*. John Wiley and Sons, Inc., New York, 1999.

Táblaképek az egészségrõl - A veresegyházi példa. MTA Szociológiai Kutatóintézet - Fekete Sas Kiadó, 2000.

R.W. Oldendick, G.G. Bishop, S.B. Sorenson, and A.J. Tuchfarber. A comparison of the kish and last birthday methods of respondent selection in telephone surveys. *Journal of Official Statistics*, 4:307–318, 1988.

T. Piazza. Respondent selection for CATI/CAPI (Equivalent to the Kish Method). Available at `http://srcweb.berkeley.edu:4229/res/rsel.html`.