J. Statist. Res. Iran **16** (2019): 491–506
DOI: 10.52547/jsri.16.2.491

Split Questionnaire Design: An Application in Household Expenditure and Income Survey of Iran

Lida Kalhori[†,*], Asiyeh Abbasi[‡], Roshanak Aliakbari Saba[†] and Nasrin Ebrahimi[‡]

[†] Statistical Center of Iran
[‡] Statistical Research and Training Center

Received: 2021/22/04 Approved: 2021/13/09

Abstract. One of the challenges at National Statistical Organizations is conducting a survey with a long questionnaire and a large amount of items. Filling out a lengthly questionnaire is time consuming and leads to problems such as increased unresponsiveness rate. Especially during critical situations such as pandemic Covid_19, this would be more challenging. A solution to encounter this difficulty is using a split questionnaire design that is more flexible and taking less time for collecting data. In this paper, we propose a split questionnaire design for Household Expenditure and Income Survey(HEIS) which is one of the most important surveys conducted annually by Statistical Center of Iran and its questionnaire is really long and time consuming to complete.

Keywords. Split questionnaire design; regression model; HEIS.

MSC 2010: 62P99, 62J05.

[*] Corresponding author

1 Introduction

Conducting surveys with lengthy questionnaires is one of the most important challenges in the National Statistical Offices. Filling out lengthy questionnaires is not only time consuming and expensive, but also can have serious effects on data quality due to the increasing respondent burden. There are at least four steps involved in answering questionnaires, which make cognitive demands on respondents: comprehension of the question, recall of requested information from memory, evaluation of the link between the retrieved information and the question, and communication of the response (Tourangeau, 1984). Accordingly, massive questionnaires result in more measurement error and nonresponse (Adigüzel and Wedel , 2008). A quick solution may be reduction of the survey instrument, which may prevent achieving the expected descriptive and analytical results. So, to satisfy a survey's key objectives, reduction of the survey content may be highly undesirable. Thus, there is the need to balance the respondent burden and the survey's goals via the survey tool. Split Questionnaire Design (SQD) is a relatively new technique to reduce the mentioned barriers. Split a long questionnaire into different sets or blocks, and give each subset of these questions to different sub sample units in the survey could be a solution. SQD has not been widely used in practice by national statistical offices due to its complexity in the application. Navarro and Griffin (1993) studied the effect of shorter multiple questionnaires on improving coverage for the 2000 census, due to a very important goal of the 2000 census which was enhancing coverage and reduce the differential undercount. They propose five sampling matrix, but ultimately, matrices sampling was not used in the census.

Thomas et al. (2006) implemented matrix sampling in the National Survey of Medical Decisions. All survey respondents received screening questions, then ten more detailed questions were allocated to them based on their responses. To control the extent of burden, at most two of such modules were received by each respondent.

Household Expenditure and Income Survey (HEIS) is one of the most important surveys conducted annually in urban and rural areas by the Statistical Center of Iran. The HEIS main goal is to estimate the average income and expenditure for urban and rural households at provincial and national levels. The survey questionnaire is comprehensive and consists of four main parts that are: 1) household demographic information, 2) dwelling

characteristics and appliances, 3) expenditure on food and tobacco, and 4) expenditure on non-food and income. In practice completing such a detailed questionnaire is complicated both for the respondents and interviewer. To reduce the response burden, time, and cost of the survey we can gather less information from each sample unit through the implementation of the SQD and then apply the available information to provide desired estimates.
On the other hand, during a pandemic, NSOs are challenging for implementing household survey due to travel restriction and also fear of face-to-face interviewing and then some areas remained completely inaccessible due to insecurity, even most accessible areas held potential risks high non response rate. All of these situations leads us to think about a statistical approach to minimize the time for collecting data. In this paper, we suggest "SQD" as a tool that takes less time, more flexible to use CAPI (Computer Assisted Personal Interviewing) instead of face-to-face way for collecting data. In Section 2 a brief review of the previous studies in Split Questionnaire Design will present. An SQD for HEIS of Iran will study and present in Section 3, the paper will end by discussion and results.

2 Split Questionnaire Design

The first attempt to split the questionnaire refers to Herzog and Bachman (1981) who divided the questionnaire into two parts and each part was devoted completely randomly for the respondent. Raghunathan and Grizzle (1995) introduced a developed method for designing a split questionnaire, where the questionnaire is split into components and individuals. An important issue in using this method was how the questionnaires should be designed and how the obtained data would be analyzed. Imputation of the missing parts was done by using Gibbs sampling.

Thomas et al. (2006) emphasize that the split parts should be created in such a way that each questionnaire includes items that are predictive of the excluded items, based on this a set of strongly correlated items are considered and divided into different sub-questionnaires. In result, the required information about the excluded items can obtain by multiple imputation. Thomas et al. (2006) also defined core items that should be included in all sub-questionnaires. Adigüzel and Wedel (2008) proposed the method of spliting the questionnaire using the Kolbeck-Liber distance as a criterion that achieves the optimal design to minimize information loss compared to the original questionnaire. The previous studies mostly concentrate on the

advantages of SQD in the reduction of the respondent burden, but Chipperfield and Steel (2009) and Chipperfield and Steel (2011) focus on the SQD from another point of view. The efficiency of a design can be measured by the cost required to meet constraints on the accuracy of estimates. For matrix survey sampling with overlapping subsets of questions, Merkouris (2015) proposed an efficient estimation method based on the principle of best linear unbiased estimation, generates composite optimal regression estimators of population totals. Kamgar and Navvabpour (2017) proposed a strategy to split long questionnaires and estimate the population parameters based on data collected from subquestionnaires. They examined the four methods of small-area estimation, Horvitz-Thompson estimation, regression method, and multiple imputation, and concluded that of these four methods if the appropriate auxiliary variables are used from the registers or previous census, the small-area estimation method gives us a more accurate estimate.
The housing and population census is conducted every then or five years, in results the available auxiliary variables from the census are not up to date. In this paper, we recommend a scenario for SQD in which the auxiliary information is provided from the respondent sample survey units. This is especially important in critical situations where conditions change rapidly and without a predetermined pattern. In addition to changes in lifestyles and household consumption patterns, the Covid 19 epidemic also brought about serious changes in the way surveys were conducted to create social distance and maintain health.

3 Design Split Questionnaire for HEIS of Iran

As mentioned before, methodological research demonstrates that data quality and response rate are related to the respondent burden. Certainly, a lengthy survey questionnaire tends to increase response burden. In recent years the HEIS faced several challenges for both interviewers and respondents. Providing detailed information about annual household income, purchased materials, and the amount of purchases during the last month is complicated. The respondent is expected to announce information for about 1200 items. In this regard, filling out a 68 pages questionnaire is time consuming, depending on the consumption pattern of the household it takes about 90 minutes on average to complete.

A straightforward solution for this concern may be to administer a shorter questionnaire. However, a new challenge emerges immediately about which

questions should be kept in the survey. Due to the large number of data users who regularly request more detailed information about a certain aspect, decision making is tough. Alternatively, creating a shorter questionnaire that still collects the necessary information from at least some of the sample members is helpful. These designs are often called split questionnaire design (Raghunathan and Grizzle , 1995) or multiple matrix sampling design (Shoemaker , 1973). To split the questionnaire, four issues should be considered: determining the core of the questionnaire, subquestionnaires, the sample units to complete the subquestionnaires, and estimation of population parameters.

The core of the questionnaire consists of items that are asked from all sample units. This information can be used as an auxiliary for imputation, or control and validation of other parts of the questionnaire. Such as asking about the type of housing unit occupation (ownership, rental, other) and amount of mortgage, or having a mobile or landline device. This information included in the core of the questionnaire makes it necessary to answer some sections or vise versa. For instance, when a household declares access to tap water or a telephone or internet connection, there should normally be a related charge. Or, as another example, characteristics of household members must be fully specified in order to gather proper information about expenses and income.

The sub-questionnaires can be identified in several ways, perhaps the most important step in splitting a questionnaire is specifying the subquestionnaires. In the case of HEIS some sections can be selected completely at random, or important items are included. For example, the required items for calculating the consumer price index (CPI) are preserved as the core of the questionnaire. One approach is to break the questionnaire into sub-sections that do not overlap and devote them to the independent sample units at random. This approach is useful for estimating the total mean, and nonresponse is considered as unit nonresponse. In result, there are not enough sample size for the sub questionnaires. Another way to split a questionnaire is to create sub-questionnaires that generally contain some identical sections. In this case, nonresponse is considered as item nonresponse and common methods can be used to improve the data quality.

As mentioned before, the questionnaire of HEIS consists of four main parts, demographic characteristics of household members, characteristics of housing unit and facilities, household expenses, and household income. A List of household members and characteristics such as gender, age, level of edu-

cation, and activity status are included in the demographic part. Characteristics of housing unit and facilities part consist of eight sub sections and contain information such as how to occupy a housing unit, area of housing unit, and available appliances and facilities. The household expenditure section consists of thirteen subsections and is the most important and longest part of the questionnaire. Its subsections include food, tobacco, clothing and footwear, housing, furniture and equipment, medical care and health services, transportation and communication, miscellaneous household goods, recreation and services, supply and sale of durable household goods, other costs, and transfers. Expenses related to the supply and sale of durable household goods, and transfers are asked about the last twelve months and other expenses are asked about the last month. The main and longest part of the questionnaire is the third part, in which household expenses and amount of consumption are asked separately according to the six digit COICOP (Classification Of the Individual Consumption Purpose) codes. There are more than 1000 six digit item codes are presented in the HEIS questionnaire. Obviously asking all of this detailed information, in addition to the high response burden and required time, can also lead to non-sampling errors due to memory recall error and fatigue. In addition, different importance of the product yielding to infrequent report of many of the expenditure items that HEIS collects information for in the Survey.

In this paper, we use the data of urban areas of HEIS in 2017 year, which includes 18701 sample units. We aim to implement the split questionnaire design on these data through the following scenario.

According to the results of HEIS during previous years, some frequently repeated items are reported by at least 90% of households. But in some sections, information for luxury goods is provided for a low percentage of households. Therefore, during the statistical period of one month about thirty percentage of information is available in average. In result, thirty percent of all sample units are selected randomly and their completed information will remain completely unchanged, to evaluate the results of splitting the questionnaire design and compare in to the current method. For the 70 percent of the remaining sample units, four main parts of the questionnaire are assumed as the core section and the whole information is kept in the study. Therefore, the core part is consists of the demographic characteristics of household members, characteristics of housing unit and facilities, household's expenses on food, and household income. The other parts of the HEIS questionnaire will consider for splitting. Each of the remaining parts is

Table 1. Percentage of completed sections of the split questionnaire of HEIS

	Demographic information	Housing	Food	Tobacco	Clothing and Footwear	Miscellaneous	$\cdots$	Transfer	Income
Household1	___	___	___	___	___	___	___	___	___
Household2	___	___	___			___	$\cdots$		___
Household3	___	___	___	___					___
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$		$\vdots$
Percentages of Sample units	100%	100%	100%	36%	36%	36%	$\cdots$	36%	100%

assumed as one of the split parts and six percent of households are selected at random to complete the sub questionnaire. It refers to 6% of the total population, but their selection is made from the remaining 70% of sample units. Thus, in split sections, there is a total of 36% of the information. 30% of which is related to the retained section and 6% of which is related to the section that is randomly selected to complete the split questionnaire. The sections of questionnaire related to the expenses on tobacco, clothing, and footwear, transfer and miscellaneous will call sub questionnaires in the following parts. Note that in this study we work on the data set obtained from the 2017 survey, so the information of six percent of households are preserved and the others will be cleared from the data set. The structure of data for all sample units is shown in Table 1. In other words, after vacating some parts of the questionnaire, the data set is made above:

Obviously, in the defined design for the split questionnaire, auxiliary information is available through the core of the questionnaire, which is the main source of survey data and is up to date in comparison to the previous census. Using the information obtained from thirty percent of the sample units whose complete information is available, we will fit a regression model to find out the relationship of auxiliary variables and household expenditure for each sub questionnaire. Assume that T_k is the total expenditure of a household for the kth subquestionnaire. Using data from households whose information was completed for the kth part, a regression model was fitted to evaluate the relationship between the desired expenditure and available covariates. The applied regressors included province, household size, number of literate, educated, and employed household members. Number of the household members having income, education and activity status of head of household, area of housing unit, household food expenditure, and household income.

In the next step, the obtained model was implemented to estimate the

total expenditure of the kth subquestionnaire for sample units whose data were not gathered due to splitting the questionnaire, using the available auxiliary variables in the core section. It is worth to note that the missing cases are created to simulate the real situation when using the Split Questionnaire Design. All households (not only 30%) were considered to estimate the parameters, the weighted average of expenditures in each part was assumed as the population mean.
For the sake of brevity results of the regression models are not presented, but evaluation of the fitted model for each sub questionnaire will present in the following part.

3.1 Evaluation of the Design Used to Split the HEIS Questionnaire

To measure the accuracy and precision of the expenditure estimate of the different sections of the questionnaire, we repeat 1000 times, the sample unit selection steps as shown in Table 1. That is, thirty percent of sample units out of 18701 sample households are randomly selected 1000 times and the whole information is preserved for them. Among the seventy percent of the remaining households, information of the core part is kept for all of them and six percent of available information from sub questioners are maintained at random. The estimated expenses of each subquestionnaire from 1000 replications have been evaluated with two indicators of percentage of Coefficient of Variation (CV) and Relative Bias. It is worth to note that, actually each of the 1000 sample is a random sub-sample of the HEIS framework. Therefore, the weights of the HEIS is usable to generalize the result from sub-sample to the population. The coefficient of variation is calculated as follow:

$$\bar{\bar{\theta}}_j = \frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i x_{ij} \qquad j = 1, \ldots, 1000 \tag{1}$$

$$\hat{\theta} = \frac{1}{1000} \sum_{j=1}^{1000} \bar{\bar{\theta}}_j \qquad \hat{\mathrm{MSE}}\left(\hat{\theta}\right) = \frac{1}{1000} \sum_{j=1}^{1000} \left(\bar{\bar{\theta}}_j - \hat{\theta}\right)^2 \tag{2}$$

$$\hat{\mathrm{CV}}\left(\hat{\theta}\right) = \frac{\sqrt{\hat{\mathrm{MSE}}\left(\hat{\theta}\right)}}{\hat{\theta}} \times 100 \tag{3}$$

where for each sub questionnaire x_{ij} is the total expenditure of the ith household in the jth replication, n is the number of samples in the jth replication, w_i is the ith household weight in HEIS, $\bar{\bar{\theta}}_j$ mean of expenditure of all households in the jth replication of each sub questionnaire, $\hat{\theta}$ the total average of expenditure for each sub questionnaire in all 1000 replications. Percentage of Absolute Relative bias (RelBias) is calculated as follow:

$$\text{Absolute Relative Bias}(\theta) = \left|\frac{\theta - \hat{\theta}}{\theta}\right| \times 100 \tag{4}$$

where, for each sub questionnaire, $\hat{\theta}$ is the estimate of expenditure from split design and θ is the estimate from whole data of Household Expenditure and Income survey in urban areas in 2017. The coefficient of variation in 1000 replications of the simulation study for non food expenditures as sub questionnaires are shown in Table 2.

Table 2 represents the CV and absolute value of relative bias for the estimation of non food expenses of urban areas at the national level. Results of Table 2 show that estimating the average of expenditures of different sections of the questionnaire using a split questionnaire design is accurate but bias is relatively high, especially for the housing, transportation, medical care and health services, and miscellaneous goods and services sections. Figure 1 shows the estimates of parameters obtained from the complete and split questionnaire design.

The outcome demonstrates that the applied auxiliary variables in the regression model does not work properly to estimate the average of expenditures for all sub questionnaires. In results, we suggest applying distinct appropriate model for each sub questionnaire based on its exclusive characteristics.
Due to the importance of medical care and health services expenditures, we will focus on identifying covariates affecting the household medical and health care expenditures.

Table 2. $\hat{CV}$ and RelBias in Estimation of Mean of Expenses of each Sub Questionnaire in 1000 Simulation.

Types of non_Food Expenditures Sub Questionnaire	$\hat{CV}$	AbsRelBias
Tobacco	0.035	80.301
Clothing and Footwear	0.020	45.974
Housing,Water,Waste water	0.010	5.818
Furniture, Furnishing, Equipment and Services	0.029	0.027
Medical Care and Health Services	0.018	24.697
Transportation	0.018	4.289
Communication	0.035	1.019
Entertainment and Cultural Services	0.037	24.240
Training and Education	0.027	56.595
Prepared Foods and Hotel	0.029	45.986
Miscellaneous Household good and services	0.015	1.920
Transfer	0.045	1.280

3.2 Factors Affecting the Household Medical and Health Care Expenditures

Based on the information obtained from the full survey of household income and expenditure in 2017, we know that 32 percent of households do not have reported medical or health care expenses. Therefore, the main point before estimating medical and health care expenditure is whether the household has any costs in this area or not? The answer to this question can be found in the model proposed by Grossman (1972). He identified education, age, and wage rates as factors influencing health expenditures. Parker and Wong (1997) have studied the estimation of household monetary health care expenditures, using the data available from the Mexican National Survey of Income and Expenditures of 1989. They applied economic and demographic characteristics of the household as explanatory variables and found that the household income level and insurance as the most effective variables on health care expenses.

In this study, having health expenditure is modeled through logistic regression model. Assume that Y_i is a binary response variable, where the values 0 and 1 demonstrate that the household have or do not have medical expences respectively. Suppose that the response variable for the ith sample

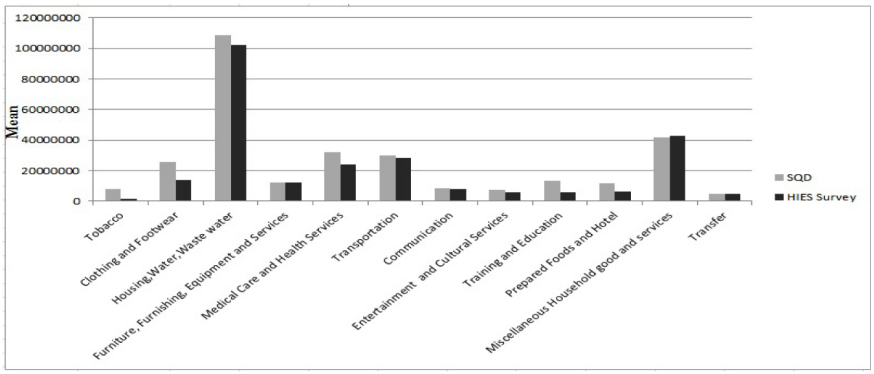

Figure 1. Estimation of the average of different sections of the complete and split questionnaire.

unit, $i = 1, \ldots, n$, follows the binomial distribution, that is $Y_i \sim Bin\,(p_i)$:

$$\text{logit}\,(p_i) = \text{Ln}\left(\frac{p_i}{1-p_i}\right) = \beta_{0i} + x_{1i}\beta_{1i} + \cdots + x_{ki}\beta_{ki} \qquad i = 1, \ldots, n. \quad (5)$$

where p_i is the probability that the ith sample household has medical or health care expenditure, $x = (x_{1i}, \ldots, x_{ki})$ and $\beta = (\beta_{0i}, \ldots, \beta_{ki})$ are the vector of covariates and regression parameters respectively. The model fitting is done using the whole data obtained from urban areas of HEIS 2017. Gender, age, activity status, level of education of head of household, household size, existence of elderly member(over 60 years old) or child(lower than 10 years old), quantile of household expenditure, and status of medical insurance are the covariates included in the model. The odds ratios of having medical or health care expenditures are presented in Table 3.

The results show that the odds ratio of having medical expenses is not significant for the gender, age, and activity status of head of the household. But the odds ratio of having medical expenses is reduced when the head of household is more educated. They may be less likely to incur heavy medical costs due to awareness and preventative care. Parallel with an increase in household size and level of expenditure quantile, the chance of having medical of health care expenditure is raised up. In addition, the existence of elderly member and insurance factors cause a higher amount of household spending in the health sector. Therefore, it can be concluded that it is better to

Table 3. Estimates and 90% Confidence Interval for the Odds ratio of the fitted model.

		90% Confidence Interval		
	Odds ratio	Lower	Upper	P-Value
Intercept	1.65	1.31	2.08	0.00
Gender (Head of Household)				
Female	1.02	0.87	1.20	0.77
Age(Head of Household)				
Over 40 years	1.06	0.95	1.19	0.30
Activity Status of Head Of Household (Unemployed)				
Employed	0.91	0.81	1.03	0.14
Education Status of Head of Household (Illiterate)				
Without University Degree	0.82	0.71	0.94	0.01
University Degree	0.68	0.57	0.82	0.00
Household Size (1 or 2)				
3 or 4	1.18	1.01	1.38	0.03
More than 4	1.35	1.13	1.61	0.00
No Existence of elderly or child Household Member				
Elderly Member	1.36	1.19	1.55	0.00
Child	0.67	0.55	0.82	0.00
Quantile of Household Expenditure (First quantile)				
Second Quntile	1.57	1.39	1.79	0.00
Third Quantile	1.85	1.62	2.12	0.00
Forth Quantile	2.29	1.97	2.66	0.00
Fifth Quantile	3.12	2.61	3.73	0.00
Medical Insurance (Without Medical Insurance)				
Having Medical Insurance	1.41	1.24	1.60	0.00

determine whether the household having medical expenses before fitting the model to estimate the household medical and health care expenditures using the split questionnaire design. The same thing is trivial for the other sub questionnaires, and each of which needs its special model.

4 Discussion and Results

The Household Expenditure and Income Survey is conducted annually by the Statistical Center of Iran. The data are crucial to the computation of the Consumer Price Index (CPI) and in the study of National Accounts. Researchers also use HEIS data to study the consumption pattern of households. The data collection methods used in the HEIS program have not been fundamentally redesigned in several decades. However, the challenge of completing a long questionnaire has been increased in recent years. Especially, due to the completion of questionnaires through telephone interviews due to the outbreak of pandemic Covid 19 disease, the challenge of obtaining information from households are facing more trouble and it is necessary to find a solution to encounter the problem.

A split questionnaire design is proposed in this paper to reduce the response burden. In the proposed design in each round of the survey, the whole questionnaire is completed by thirty percent of sample units and for the seventy remaining percent, the questionnaire is split into core part and sub questionnaires. The core part consists of the demographic information of household members, characteristics of housing unit and facilities, household's expenses on food, and household income. Each of the remaining parts are assumed as one of the split parts and six percent of households are selected at random to complete the sub questionnaire. So, sections of the questionnaire related to the expenses on medical and health care, tobacco, clothing and footwear, transfer, and miscellaneous are assumed as sub questionnaires. Using the auxiliary information available through the core of the questionnaire, we estimate the mean of expenditure for each sub questionnaire at the national and province levels. Evaluation of the results demonstrate that estimation of the average expenditures of different sections of the questionnaire using a split questionnaire design is accurate but a little biased, especially for medical care and health services, transportation, and miscellaneous goods and services. This is due to the inconsistent pattern of household consumption. Medical and health care expenditures is studied specifically and the important covariates related to having medical expenses are identified. In order

to reduce bias in the proposed SPD, we recommend that according to the information available from the auxiliary variables in the core part for each sub-questionnaire, at the first step it should be predicted whether the household has expenses in that section or not.
The expenditure variable in different parts of the questionnaire can be zero, so for each household we must first know whether there was a cost in the desired part or not. Therefore, knowing that the household has expenditures in each split-questionnaire can be helpful in later stages. For imputation of such variables, this point should also be taken into account and the amount of this variable should be imputed in two stages. First having / not having expenses and second, impute the expenditure value to achieve a more desirable result. If the household does not have a cost in one section according to its own statement, instead of imputing the cost value for that household, it is better to considered zero value to reduce bias.
In literatures multiple imputations are used to estimate the population parameters in a SQD. But it is expected to reduce the accuracy of estimates due to the high volume of missing data caused by split questionnaire design. Due to the difference in household consumption pattern and the factors affecting it in each expenditure segment, discovering and using a suitable regression model for each split part and using the up-to-date auxiliary information available to each household from the core part of the questionnaire can improve the quality of results. Comparing and evaluating the efficiency of the method used in this paper and multiple imputations can be the subject of another study. Although splitting the questionnaire can reduce the response burden and cost of the survey, providing an operational solution requires extensive research in this area. Therefor, development of studies and drilling down into this field would be the subject of future research to accomplish a really practical split design.

References

Adigüzel, F., and Wedel, M. (2008). Split Questionnaire Design for Massive Surveys, *Journal of Marketing Research*, **45**, 608-617.

Chipperfield, J.O., and Steel, D.G. (2009). Design and Estimation for Split Questionnaire Surveys, *Journal of Official Statistics*, **25**, 227-244.

Chipperfield, J.O., and Steel, D.G. (2011). Efficiency of Split Questionnaire Surveys, *Journal of Statistical Planning and Inference*, **141**, 1925-1932.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health, *Journal of Political Economy*, **80**, 223-255.

Herzog, A.R., and Bachman, J.G. (1981). Effects of Questionnaire Length on Response Quality, *Public Opinion Quarterly*, **45**, 549-559.

Ioannidis, E., Merkouris, T., Zhang, L.C., Karlberg, M., Petrakos, M., Reis, F., and Stavropoulos, P. (2016). On a Modular Approach to the Design of Integrated Social Surveys, *Journal of Official Statistics*, **32**, 259-286.

Kamgar, S., and Navvabpour, H. (2017). An Efficient Method for Estimating Population Parameters Using Split Questionnaire Design, *Journal of Statistical Research of Iran JSRI*, **14**, 77-99.

Merkouris, T. (2015). An Efficient Estimation Method for Matrix Survey Sampling, *Survey Methodology*, **41**, 237–262.

Navarro, A., and Griffin, R.A. (1993). Matrix Sampling Designs for the Year 2000 Census, *In American Statistical Association*, Proceedings of the Section on Survey Research Methods, 480-485.

Parker, S.W., and Wong, R. (1997). Household Income and Health Care Expenditures in Mexico, *Health Policy*, **40**, 237-255.

Peytchev, A., and Peytcheva, E. (2017). Reduction of Measurement Error due to Survey length: Evaluation of the Split Questionnaire Design Approach, *In Survey Research Methods*, **11**, 361-368.

Raghunathan, T.E., and Grizzle, J.E. (1995). A Split Questionnaire Survey Design, *Journal of the American Statistical Association*, **90** ,54-63.

Savojipour, S., Assari Arani, A., Agheli, L., and Hassanzadeh, A. (2018). The Determinants of Urban Families' Health Expenditure, *Journal of Econimic Poliy*, **10**, 25-52.

Shoemaker, D.M. (1973). A Note on Allocating Items to Subtests in Multiple Matrix Sampling and Approximating Standard Errors of Estimate with the Jackknife, *Journal of Educational Measurement*, **10**, 211-219.

Thomas, N., Raghunathan, T.E., Schenker, N., Katzoff, M.J., and Johnson, C.L. (2006). An Evaluation of Matrix Sampling Methods Using Data from the National Health and Nutrition Examination Survey, *Survey Methodology*, **32**, 217.

Tourangeau, R. (1984). Cognitive Sciences and Survey Methods, *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*, **15**, 73-100.

Lida Kalhori
Statistical Research and Training Center,
Tehran, Iran.
email: *kalhori@srtc.ac.ir,lidakalhori@yahoo.com*

Asiyeh Abbasi
Statistical Center of Iran,
Tehran, Iran.
email: *asieh_abasi@yahoo.com*

Roshanak Aliakbari Saba
Statistical Research and Training Center,
Tehran, Iran.
email: *roshanaksaba22@gmail.com*

Nasrin Ebrahimi
Statistical Center of Iran
Tehran, Iran.
email: *nsrebr@gmail.com*