Survey Research Methods (2018)
Vol. 12, No. 3, pp. 247-257
doi:10.18148/srm/2018.v12i3.7283

ISSN 1864-3361
http://www.surveymethods.org

Randomization in Small Sample Surveys with the Halton Sequence

Kien T. Le
Social and Economic Survey Research Institute
Qatar University
Doha, Qatar

Martha McRoy
Pew Research Center
Washington D.C., USA

Abdoulaye Diop
Social and Economic Survey Research Institute
Qatar University
Doha, Qatar

Randomization has been widely used in surveys for various purposes such as within household respondent selection, rotation of questions and answer choices, and split sample (or ballot) technique for survey experiments. The randomization is usually based on a random number generating process whereby the computer generates random numbers which are then used to classify respondents in different groups. In this study, we use an alternative randomization based on the Halton sequence. The method is used in a survey with a political experiment which requires randomization of the political candidate's characteristics. Our survey results demonstrate that the Halton sequence can be quite effective in randomly assigning respondents into groups, especially in surveys with small sample sizes.

Keywords: randomization, survey, experiment, halton sequence, political

1 Introduction

One important benefit of using computers in survey research is the integration of randomization in survey designs. In the split sample (or split ballot) technique commonly used to test survey designs or questions, the sample is randomly split into multiple groups and each group receives different survey designs or questions (Clarke, Kornberg, Mcintyre, & Bauer-Kaase, 1999; Dalal & Hakel, 2016; Fowler, 2014; Rosenthal & Hubble, 1993; Tourangeau & Smith, 1996; Trussell & Elinson, 1959). For within-household respondent selection, from the oldest Kish method to the recently developed minimally intrusive method, randomization is critical to ensure representative selection of a household member for the interview (Gaziano, 2005; Kish, 1949; Le, Brick, Diop, & Alemadi, 2013; Rizzo, Brick, & Park, 2004). In some surveys, randomization is used for controlling order effects through rotation of questions and answer choices (Diop, Guterbock, Kermer, & Le, 2008; Walle & Van Ryzin, 2011) and rationing for reducing survey length by asking different sections of the questionnaire to different subsamples of respondents (Kessler et al., 2004).

Contact information: Kien T. Le, Social and Economic Survey Research Institute, Qatar University, P.O. Box 2713, Doha, Qatar, Email: KienLe@qu.edu.qa

Randomization is usually based on a random number generating process whereby the computer generates (pseudo) random numbers which then determine the flow of the interviews. For example, in the split sample technique, the computer generates random numbers (uniformly distributed) in the unit interval (between 0 and 1), and then each respondent is assigned to one of the two groups based on their random numbers being less than or greater than 0.5. This randomization allows researchers to compare respondents' answers between two groups.

However, in some survey randomizations, depending on the realization of the random numbers, the classification may result in significantly uneven groups (i.e., groups with significantly different number of respondents). Fowler (2014) reported several experiments (using split sample technique) to compare different ways of wording for survey questions. In most of his experiments, the sample was split evenly across groups. However, in one experiment about health care appointments, the sample was not evenly split into two groups, 261 respondents in one group (46.6%) and 299 respondents in another group (53.4%). Diop et al. (2008) also conducted an experiment in a survey about public opinion towards immigrants. They wanted to evaluate if respondents' answers to a particular question would differ significantly with the use of either "residents" or "citizens". In their experiment, the sample (790 respondents) was split into two relatively uneven groups: 46.1% with "residents" and 53.9% with "citizens". Dalal and Hakel (2016) compared different methods for reducing distortion in self-report measures of sensitive questions. In one of their experiments about counterproductive workplace behavior, the sample was split into six groups with a relatively large difference between the smallest group (83 respondents or 14.1%) and the biggest group (128 respondents or 21.7%).

It should be noted that, statistical analysis comparing respondents' answers between groups is not compromised by the uneven groups. The comparison is still sound and valid, but the power to detect any statistically significant difference between groups could be reduced. One way to address the uneven groups issue is to impose some group quota system during the fieldwork. This system imposes a limit on the number of respondents in each group. Whenever the quota in a group is reached, the remaining respondents will not be assigned to this group, but to other groups. The quota system can control the size of groups, but it may slightly disrupt the randomization process if the early respondents differ from the late respondents. Another method, not used much in surveys but popular in clinical trial studies, is block randomization. This method randomizes respondents within blocks such that the same number are assigned to each group. In a study with only two groups, a block size of four is usually chosen and there would be six possible ways to assign respondents into a block.

In this study, we propose the use of the Halton sequence as another way of randomization. The idea behind this sequence is quite simple. Instead of covering the unit interval randomly, the numbers in the sequence follow a deterministic pattern to fill the interval more evenly. The sequence has been proven by mathematicians to be superior to the random number method in Monte Carlo simulations to calculate multi-dimension integration Bratley, Fox, and Niederreiter (1992), Kocis and Whiten (1997), Morokoff and Caflisch (1995). Applications of the Halton sequence have been found in traffic engineering to study travel modes (Bhat, 2001; Hensher & Greene, 2003; Milton, Shankar, & Mannering, 2008), in economics to study consumer choices (Jia, 2008; Train, 2003; Train & Winston, 2007), in public health to create a health index comparable across countries (Meijer, Kapteyn, & Andreyeva, 2008, 2011). These applied studies found that the Halton sequence achieves higher level of precision than the random number method, especially when a small number of simulations are used.

In our study, the method is applied to a Computer Assisted Telephone Interviewing (CATI) survey in Qatar which requires randomization for a political experiment. In the experiment, respondents were randomly assigned into various groups with different candidate characteristics. Researchers then use respondents' answers to understand how people choose among various candidate characteristics. We used the Halton sequence for the randomization, but also created random numbers for comparative purposes. The objective of the study is to evaluate the performance of the Halton sequence relative to that of the random number method using two criteria. The first one is the number of respondents assigned to various groups. Our survey results indicate that groups are much more evenly distributed with the Halton sequence than the random number method, resulting in more precise estimates for the Halton sequence. The second criterion is the randomness in assigning respondents into groups. This is the key requirement for any survey experiments. Since the Halton sequence is a deterministic process, there may be a concern that respondents are not assigned randomly into groups. However, in our survey, the Halton sequence performs as well as the random number method on this criterion.

The rest of this study is organized as follows. First, we will discuss the survey design, providing detailed information about the political experiment and the randomization used in the survey. Second, further explanation of the Halton sequence will be presented together with its relative advantage compared to the random number method. Third, the performance of the two methods in our survey experiment are compared. Particularly, we will look at how evenly and how randomly respondents are assigned into groups. Finally, the study ends with the conclusion section which summarizes the results and discusses the limitations of the study.

2 Survey Design

The Qatar Quarterly Survey (QQS) is an ongoing research project whose goal is to provide unbiased data on questions regarding the social and economic situation in Qatar. It consists of a 15 to 20-minute telephone survey. The target population for the survey consists of adults (18 or older) who currently live in Qatar. To reach this population, the Social and Economic Survey Research Institute (SESRI) at Qatar University works with local cell phone providers to develop a cell phone sample using the listed dialing technique. As the proportion of adult Qataris with a cellular phone is about 98%, the survey sample is expected to have excellent coverage and representation of the target population.[1] All the phone numbers in the sample are randomly ordered when loaded into the BLAISE system, a computer-assisted interviewing program, for dialing.[2]

For every phone number in the sample, there are seven attempts to complete the interview. The phone calls are made over different times during the day and different days of the week to maximize the chances of making contact with respondents. For phone numbers with break-off and soft re-

[1] This number is based on the face to face survey conducted by SESRI in 2015.

[2] BLAISE is developed by Statistics Netherlands and is designed for use in official statistics. Many survey organizations around the world (e.g. University of Michigan Survey Center) are using this system in their surveys.

Table 1
Calling dispositions

Disposition	Frequency
Completed	889
Not completed	3700
Eligible	521
Ineligible	2460
Unknown eligibility	719
Raw response rate (%) (RR1)	42
Adjusted response rate (%) (RR2)	53

fusal, dedicated interviewers would try to contact and convert them to completed interviews. The following table shows the disposition of all dialed phone numbers during this survey.

On the basis of Table 1, response rates are calculated. We report two response rates in the last two rows of the Table. First, the raw response rate is the ratio between the number of completes or partials and total sample sizes after excluding ineligibles: RR1 = $\frac{C}{C+E+\mathrm{UE}}$ where C is the number of completes or partials, E is the number of eligible responses, and UE is the number of unknown eligibility. Second, the adjusted response rate is RR2 = $\frac{C}{C+E+e\cdot\mathrm{UE}}$, where e is the estimated proportion of eligibilities which is given by this expression $e = \frac{C+E}{C+E+\mathrm{IE}}$, where IE is the number of ineligibles.

The questionnaire for the survey consists of a demographic section, three core sections, and one rotating section in the survey. The questions in the core sections are used to evaluate the perceptions of respondents on various issues including the quality of life, consumer confidence, and issues related to migrant workers. The core sections are repeated every quarter and are used to monitor trends over time. Meanwhile, the rotating section is used to ask questions related to current situations/hot topics in the country. In the past QQS surveys, the rotating section covers one of the following topics in each survey: traffic accidents and traffic law, summer vacation plans, and health insurance. The questions in the rotating section are usually not repeated.

Political experiment. In the June 2015 QQS survey,[3] the rotation section was used to conduct a political experiment about the municipal elections in Qatar. There were 512 respondents (a response rate of 56%) who participated in the survey but only 490 of them reached the experiment section (which is in the middle of the interview). The aim of the political experiment is to analyze respondents' perceptions of potential political leaders regarding five characteristics: gender, potential tribe affiliation (based on the candidate's family name), level of education, amount of work experience, and the candidate's religious advocacy. Respondents were read a brief statement about a fictional candidate with randomly assigned characteristics and then were asked various questions regarding the competency of the candidate on various political topics, such as the candidate's ability to improve education and promote economic development. The five characteristics were governed by five random variables with values ranging from 0 to 1. The five variables are as follows:

- Candidate's gender: There are two possibilities for gender, male or female. If the random variable governing candidate gender is less than 1/2 then the fictional candidate presented to the respondent is a male. Otherwise the candidate is a female.
- Candidate's name: There are six possibilities for name, from name#1 to name#6.[4] If the random variable governing candidate name is less than 1/6, then the name of the fictional candidate is name#1. If the random variable is more than or equal 1/6 but less than 2/6 then the candidate name is name#2. In a similar way, the candidate name can take on name #3 to #6 depending on the value of the random variable.
- Candidate's education: There are three possible levels, Bachelor's degree, Master's degree, or education level was not mentioned. Both the Bachelor's and Master's degrees were stated to be in engineering from Qatar University. As before, the education of the candidate depends on the value of the random variable. It is Bachelor degree, Master degree, or not mentioned if the random number is less than 1/3, from 1/3 to less than 2/3, or from 2/3 to 1, respectively.
- Candidate's work experience: There are two possibilities. The candidate was either stated to be currently working as a planning coordinator for the Ministry of Municipal and Urban Planning or the work experience was not included in the statement. The work experience of the candidate is one of these two possibilities depending on the random variable (less than $\frac{1}{2}$ or not).
- Candidate's religious advocacy: There are two possible levels. The candidate was either stated to "promote Qatari religious and cultural values" or the religious advocacy sentence was not included in the statement. Depending on the random variable (less than $\frac{1}{2}$ or not), one of the two levels is selected for the candidate.

Each respondent was presented with one candidate with a particular characteristic (e.g., a female candidate with name#2 "Noor Al-Ghanim", having a bachelor degree, working as a planning coordinator, and promoting religious and cultural value). By comparing respondents' answers between different candidate characteristics, researchers are able to understand respondents' preferences. For example, researchers

[3]Qatari nationals, 18 years of age or older, are able to register to vote and to participate in the municipal elections, which are held every 4 years.

[4]The 6 names for male candidate are Khalid Abdulla, Khalid Al-Ghanim, Khalid Al-Majed, Hussein Abdulla, Hussein Al-Ghanim, Hussein Al-Majed. The 6 names for the female candidate are Noor Abdulla, Noor Al-Ghanim, Noor Al-Majed, Zaneb Abdulla, Zaneb Al-Ghanim, Zaneb Al-Majed.

are sometimes interested in how people in an Islamic country in general, and Qatar in particular, prefer a male candidate over a female candidate. This can be done by comparing respondents' answers between the group of respondents who receive a male candidate and the group of respondents who receive a female candidate. For this purpose, a proper randomization of candidate characteristics is critical to ensure the validity of this comparison.

Randomization. Randomization involves assigning respondents to various groups by chance. This ensures that each respondent has an equal chance of being assigned to any group in a manner that does not depend on any experimental outcome. The purpose of the randomization is to create multiple groups that include respondents with similar characteristics (e.g., similar education and age) so that the groups are equivalent.[5] Therefore, any differences in respondents' answers between groups can be confidently attributed to the experimental procedures.

In surveys, the randomization is usually based on a random number generating process whereby the computer generates (pseudo) random numbers (uniformly distributed) between 0 and 1, and then the sample is split into various groups based on these random numbers. In the case of our political experiment for example, based on the random numbers for gender and names, the sample can be split into 12 groups (2 for gender and 6 for name). According to the Law of Large Numbers, the sample would be evenly split across these groups if the sample size approaches infinity. However, there are finite sample sizes for all surveys. In some surveys, the sample size may not be large enough for the law to apply. In these cases, depending on the realization of the random numbers, researchers may end up with significantly uneven groups. When comparing mean difference between groups, the variance of the difference gets bigger as the group sizes become more uneven.

In this study, the Halton sequence is used in place of the random numbers. Since the sequence has been shown to be superior to the random numbers in Monte Carlo simulations (Bratley et al., 1992; Kocis & Whiten, 1997; Morokoff & Caflisch, 1995), we believe that it can also perform better in survey randomization. Particularly, we think the sequence is better than the random number method in evenly splitting the sample, while ensuring the randomness in the assignment of respondents.

3 Halton Sequence

The sequence was first proposed by Halton (1964). It has been used extensively in the Monte Carlo simulation to calculate multi-dimensional integration. Theoretical studies in mathematics have demonstrated faster convergence rate and superior accuracy of the Halton sequence compared to the random numbers in the simulation (Bratley et al., 1992; Kocis & Whiten, 1997; Morokoff & Caflisch, 1995). At

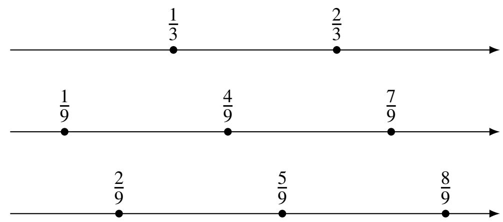

Figure 1. Example of the Halton sequence with prime number 3

the same time, empirical studies have shown that for a relatively small number of simulations, the estimates based on the Halton sequence achieve a much higher level of precision than those based on the random numbers (Bhat, 2001; Train, 2003).

The formula to generate the sequence can be found in any of these studies, but the sequence is best explained by the following example. We start with a prime number – a natural number with no divisors other than 1 and itself –, say 3. The Halton sequence for this prime number is created in the following way. First, the unit interval from 0 to 1 is divided into 3 parts and the dividing points 1/3 and 2/3 become the first two elements in the sequence as shown in the first line in Figure 1.

Next, we take each of the three parts and divide them into three parts. The dividing points become the next elements in the sequence: 1/9, 4/9, 7/9 (the second line in Figure 1), and then 2/9, 5/9, and 8/9 (the third line in Figure 1). Now, there are nine parts in the unit interval, we then divide each of nine parts into three parts as before and the dividing points become the next elements of the sequence. This process is repeated until a sufficient number of elements for the sequence is reached. The resulting sequence based on prime number 3 is (1/3, 2/3, 1/9, 4/9, 7/9, 2/9, 5/9, 8/9, ...). Similarly, Halton sequences can be created for other prime numbers. In our political experiment, we created five Halton sequences to assign the five candidate characteristics based on these prime numbers 2, 3, 5, 7, and 11. Many statistical software packages such as Stata, SAS, and Matlab include built-in routines to generate the Halton sequence.[6] The assignment of candidate characteristics using these Halton sequences can be done in the same way as for random numbers (discussed above in the Political Experiment section).

Figure 2 shows a scatter plot of the two Halton sequences generated with prime numbers 2 and 3 (on the left) and the

[5]In the randomization, groups are supposed to be similar not only in respondent characteristics but also in all other aspects. However, only information about respondent characteristics are available and can be used for comparison.

[6]We use Stata built-in function HALTON to generate the Halton sequence.

two random numbers on the right (there are 1000 points in each plot). As illustrated in this Figure, the Halton sequences cover the unit interval space much more evenly than the random numbers and this explains why our survey sample can be more evenly split with the Halton sequence than with the random number method. We will formally show this in the survey result section.

It should be noted that the Halton sequence is deterministic and follows a pattern when filling the unit interval. To take advantage of this pattern, it is necessary to use consecutive elements in the sequence. In our survey, we created five Halton sequences based on the first five prime numbers (any prime numbers can be used here) and the first 490 elements in these sequences are used for the 490 respondents who reached the political experiment segment of the survey.

Studies using the Halton sequences caution about the potential correlation problem between sequences if a relatively large number of sequences (ten sequences or more) are required. One way to avoid this correlation, in this case, is to drop the first 10 or 20 elements in all sequences (Bhat, 2001; Train, 2003). Another more sophisticated method to deal with this problem is the scrambled Halton sequence (Braaten & Weller, 1979; Vandewoestyne & Cools, 2006). In our political experiment, the Halton sequences perform well since only five sequences are needed.

4 Survey Results

The Halton sequence is used in the experiment to split the sample into groups. At the same time, we used the random numbers generated from BLAISE to assess how the sample would have been split in the absence of the Halton sequence. We then compared the performances between the Halton sequence and the random number method using two criteria: (i) how evenly the sample was split across groups and (ii) how randomly respondents were assigned to groups. For the first criterion, we simply compare the number or proportions of respondents across groups. A good randomization should give similar proportions across groups. As for the second criterion, we check the respondent characteristics (e.g., gender and age) across groups. If the respondents are randomly assigned to groups, then there should not be any significant differences in respondent characteristics across groups.

4.1 How evenly is the sample split across groups?

Table 2 shows the proportions of respondents across groups for the five characteristics: candidate gender, name, education, experience, and religious advocacy. Columns two and three show the number of groups and the corresponding expected proportions of respondents in each group. For example, there are two groups (male and female) for "candidate gender" so the expected proportion is $\frac{1}{2}$ = 50% for each group. Similarly, there are three groups (Bachelor, Masters, education not mentioned) for "candidate education", so the expected proportion is $\frac{1}{3}$ = 33% for each group.

The last two columns are the proportions of respondents based on the Halton sequence and BLAISE random numbers. For ease of presentation, we only show the minimum proportion (the proportion with the lowest number of respondents), the maximum proportion (the proportion with the highest number of respondents), and the proportion range in parentheses (the difference between the minimum and the maximum proportions). For example, in the third row of Table 2 about "candidate education", the proportions of respondents in the three groups using the Halton sequence are 33%, 33.1%, and 33.9%, so the minimum and the maximum proportions are 33% and 33.9% and the proportion range is 0.9%. These are the numbers reported in the third row of Table 2 for the Halton sequence. We will focus on the proportion ranges since they can be used to show how the sample is split evenly into groups, the larger the range the more uneven the groups.[7]

According to Table 2, the proportion ranges from the Halton sequences are smaller than those from the random number method in most of the candidate characteristics. Especially with "candidate gender" and "religious advocacy", the proportion ranges are 0% and 1.6% with the Halton sequences and 8.2% and 7.8% with the random number method. Only with "candidate experience" the range is slightly bigger with the Halton sequence (0.8%) than that of the random number method (0.4%). On average (see the last row of the Table), the proportion range using the Halton sequence is 0.8% while this number using the random number method is 4.2%.

Table 3 is similar to Table 2 except that the groups in Table 2 are created from one characteristic while the groups in Table 3 are created from the interaction of two characteristics (column 1 shows all possible interactions). For example, the first row shows the interaction of "candidate gender" with two groups and "candidate name" with six groups. The total number of groups for this interaction is $2 \cdot 6$, which is 12, and the corresponding expected proportion in each group is $\frac{1}{12}$ = 8.3%. The results in Table 3 support our finding from Table 2. In all interactions, the proportion ranges in the Halton sequence are smaller than those in the random number method. In some interactions such as "gender × work experience" and "gender × religious advocacy", the proportion ranges are very large with the random number method, 7% and 8%, respectively, while the corresponding numbers using the Halton sequence are 0.6% and 1%. On average (see the last row), the proportion range using the Halton sequence is 0.9%, while this number using the random number method

[7] In addition to the proportion ranges, one can also compare the actually proportions to the expected proportions. We can see that the proportions from the Halton sequence are closer to the expected proportions than the random number method.

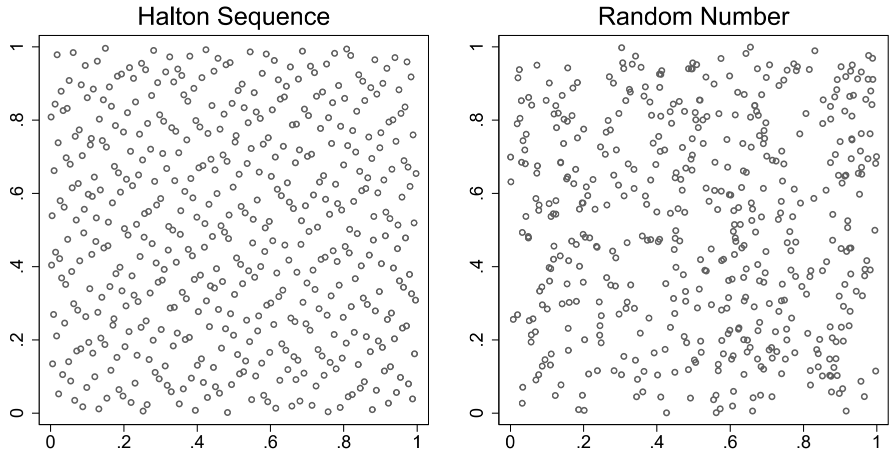

Figure 2. Plot of the Halton sequences and Random numbers

Table 2
Proportions of respondents across groups: One candidate characteristic

Candidate characteristics	Number of groups	Expected proportion (%)	Halton sequence min (%)	Halton sequence max (%)	Halton sequence range (%)	Random number min (%)	Random number max (%)	Random number range (%)
Gender	2	50.0	50.0	50.0	0.0	45.9	54.1	8.2
Name	6	16.7	16.3	17.1	0.8	15.1	18.2	3.1
Education	3	33.3	33.0	33.9	0.9	32.6	34.1	1.5
Experience	2	50.0	49.6	50.4	0.8	49.8	50.2	0.4
Religious Advocacy	2	50.0	49.2	50.8	1.6	46.1	53.9	7.8
Avg. proportion range					0.8			4.2

is 5.0%.

Both Table 2 and Table 3 demonstrate that the sample is split much more evenly with the Halton sequence than with the random number method. This result comes as expected. As shown in Figure 2, the Halton sequence is a deterministic process and covers the unit interval more evenly than the random number method and that explains why the Halton sequence performs better in splitting the sample.

4.2 How randomly are respondents assigned into groups?

The purpose of randomization is to create multiple groups that include respondents with similar characteristics so that the groups are equivalent and any differences in respondents' answers between groups can be confidently attributed to experimental procedures. Since the Halton sequence is a deterministic process, there may be a concern that the respondents are not randomly assigned into groups. In this section, we compare respondents' characteristics across different groups.

There are four respondent characteristics available in the data: being male, being young (less than 30 years old), achieving higher education (bachelor or above), and being currently married. Table 4 shows the descriptive statistics of these characteristics in the data.[8] About half (51.9%) of the respondent are male and 36.7% of them are less than 30 years of age. There are 29.5% of respondents with higher education and 58.1% are currently married.

In order to compare respondent characteristics across groups, we use the one-way ANOVA test with the null hypothesis being that there are no differences in respondent characteristics across groups.[9] Table 5 shows the results of the test. In this Table, the first column shows the candidate characteristic that is used to split the sample and the second column shows the respondent characteristic that is used to compare across groups. For example, in the first row, candidate gender is used to split the sample into two groups: male candidate and female candidate, and then we compare the proportions of young respondents between these two groups. The last two columns show the F-statistics (from one-way ANOVA) with the p-values in the parentheses for both methods. For example, in the first row, when comparing proportions of young respondents, the p-values are greater than ten

[8]Note that there is rounding in the reported numbers.

[9]The ANOVA is usually used to compare mean values (or proportions) for more than 2 groups since the simpler t-test can be used for 2 group comparison.

Table 3
Proportions of respondents: Interaction of two candidate characteristics

Interaction of two Candidate characteristics	Number of groups	Expected proportion (%)	Halton sequence min (%)	Halton sequence max (%)	Halton sequence range (%)	Random number min (%)	Random number max (%)	Random number range (%)
Gender × Names	12	8.3	8.2	8.6	0.4	6.1	10.0	3.9
Gender × Education	6	16.7	16.5	16.9	0.4	13.5	19.8	6.3
Gender × Work experience	4	25.0	24.7	25.3	0.6	21.6	28.6	7.0
Gender × Religious advocacy	4	25.0	24.5	25.5	1.0	22.2	30.2	8.0
Names × Education	18	5.5	5.1	6.1	1.0	4.3	7.1	2.8
Names × Work experience	12	8.3	8.0	8.8	0.8	6.5	9.6	3.1
Names × Religious advocacy	12	8.3	7.6	8.8	1.2	5.7	9.8	4.1
Education × Work experience	6	16.7	15.9	17.1	1.2	14.9	19.2	4.3
Education × Religious advocacy	6	16.7	16.1	17.1	1.0	14.7	19.4	4.7
Work experience × Religious advocacy	4	25.0	24.1	25.5	1.4	22.2	28.0	5.8
Avg. proportion range					0.9			5.0

Table 4
Descriptive statistics of respondent characteristics

Respondent characteristics	Proportion (%)	SD (%)
Males	51.9	50.0
Young respondents	36.7	48.3
Higher education	29.5	45.6
Currently married	58.1	49.4

percent in both methods, suggesting no differences between groups. In other words, both the Halton sequence and the random number method have randomly assigned respondents into the two groups: male candidate and female candidate. Similarly, for the rest of the Table, almost all the p-values (except one) are greater than ten percent, indicating no significant differences in respondent characteristics across groups. The only exception is the p-value in row four (with the asterisk) when comparing proportions of currently married using the random number method. This p-value is 0.066 which is less than 10% but greater than 5% significance level.

Table 6 is similar to Table 5, except that we use interaction of two candidate characteristics to create groups (first column shows all possible interactions). To save space, we only present items with significant results in the ANOVA test (p-value less than 10%). For example, in the interaction of "gender × name" we do not find any significant differences in any of the respondent characteristics across groups, so nothing is presented for this interaction. However, for the interaction of "gender × education", we find significant differences in the proportions of male respondents between groups created from the random number method, so the test result is presented for this one. The two methods can be compared by evaluating the results contained in the last two columns in the Table.

In the Halton sequence column, there are two items with significant difference between groups created by the interaction of "Education × Experience" and interaction of "Education × Religiosity". Also in the random number method column, there are two items with significant difference between groups created by the interaction of "Gender × Education" and "Experience × Religiosity". Note that these items are statistically significant at the 10% but not 5% significance level. However, researchers should be cautious when comparing respondent's answers between groups with these significant items.

The results from Table 5 and 6 indicate that respondent characteristics are pretty similar across groups (except for some items) in both the Halton sequence and the random number method. Therefore, researchers can use the experiment to compare respondents' answers across group. As mentioned above, there might a concern about the randomness using the Halton sequence since it is a deterministic sequence. However, the results show that the Halton sequence is performing as well as the random number method with regards to randomly assigning respondents.

4.3 Simulation

Our comparison between the Halton sequence and random number method is based on our actual survey data using one realization of the random numbers. If the survey is repeated, there would be a different realization of the random numbers. Also, the assignment of the respondents into groups using the Halton sequence would be different since the phone numbers are randomly loaded into the system. Therefore, the comparison results between the two methods can be quite different. To address this issue, we conduct the following simulation

Table 5
One-way ANOVA: One candidate characteristic

Candidate characteristics	Respondent characteristics	Halton sequence F-statistic	Halton sequence p-value	Random number F-statistic	Random number p-value
Gender	Young	0.31	0.581	0.14	0.712
	Male	0.03	0.856	0.06	0.806
	Higher Education	1.95	0.163	0.21	0.647
	Currently Married	0.98	0.322	3.38	0.066*
Names	Young	1.12	0.349	0.14	0.984
	Male	0.22	0.952	1.85	0.102
	Higher Education	1.30	0.262	1.21	0.304
	Currently Married	0.38	0.865	1.37	0.234
Education	Young	0.88	0.416	1.29	0.277
	Male	1.45	0.236	0.37	0.688
	Higher Education	0.16	0.856	0.01	0.989
	Currently Married	0.87	0.420	0.26	0.769
Experience	Young	1.14	0.286	2.52	0.113
	Male	0.50	0.482	1.59	0.151
	Higher Education	0.37	0.546	0.52	0.471
	Currently Married	0.13	0.718	0.03	0.861
Religious advocacy	Young	0.09	0.768	0.66	0.417
	Male	1.51	0.220	1.74	0.187
	Higher Education	0.47	0.494	0.36	0.551
	Currently Married	2.13	0.145	0.21	0.651

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 6
One-way ANOVA: Interaction of two candidate characteristics

Candidate characteristics	Respondent characteristics	Halton F-statistic	Halton p-value	Blaise F-statistic	Blaise p-value
Gender × Education	Young	-	-	2.03	0.073*
Education × Experience	Currently Married	2.08	0.067*	-	-
Education × Religiosity	Male	2.12	0.062*	-	-
Experience × Religiosity	Male	-	-	2.52	0.057*

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 7
Simulation: Average proportion ranges

Gap	N simulations	Mean (%)	min (%)	max (%)
One Characteristic	10000	4.0	0.9	9.3
Two Characteristic	10000	4.4	1.9	8.2

study.[10]

1. Randomly order respondents in the sample.
2. Assign respondents to groups using the Halton sequences.
3. Generate random variables to assign respondents to groups using these random variables.
4. Recalculate average proportion ranges for the two methods (last row in Table 2 and Table 3) to assess how evenly the sample is split.
5. Recalculate the number of significant items for the two methods (items with asterisks in Table 5 and Table 6) to assess how randomly respondents are assigned to groups.

[10]We would like to thank the anonymous reviewer for pointing out this issue and his/her suggestion to use simulation to address this issue.

Table 8
Simulation: Number of significant items with one candidate characteristic

	Blaise		Halton	
N significant items	Frequency	Proportion (%)	Frequency	Proportion (%)
0	1225	12.2	1301	13.0
1	2710	27.1	2754	27.5
2	2749	28.5	2812	28.1
3	1868	18.7	1764	17.6
4	985	9.6	901	9.0
5+	463	4.6	468	4.7
Total	10000	100.0	10000	100.0

Table 9
Simulation: Number of significant items with two candidate characteristics

	Blaise		Halton	
N significant items	Frequency	Proportion (%)	Frequency	Proportion (%)
0	622	6.2	595	5.9
1	1283	12.8	1249	12.5
2	1502	15.0	1532	15.3
3	1452	14.5	1606	16.1
4	1370	13.7	1375	13.7
5	1145	11.5	1108	11.1
6	831	8.3	859	8.6
7	664	6.6	614	6.1
8	421	4.2	393	3.9
9	275	2.8	265	2.6
10 +	435	4.4	404	4.0
Total	10000	100.0	10000	100.0

6. Repeat the above steps 10000 times.

Table 7 shows the results in step 4 for 10000 simulations. Since the Halton sequences are fixed processes, the proportions of respondents in each groups are unchanged across simulations (0.8% for one candidate characteristic and 0.9% for two candidate characteristics). However, the results for the random number method vary significantly from simulation to simulation, with a minimum of 0.9% and 1.9% to a maximum of 9.3% and 8.2% for one and two candidate characteristics, respectively. On average, the proportion ranges using the random method are 4% and 4.4% for one and two candidate characteristics, respectively. These are significantly higher than those from the Halton sequence method (0.8% and 0.9%), reaffirming our previous result that the Halton sequence can split the sample more evenly than the random number method.

Table 8 and Table 9 show the number of significant items in step 5. We can see that there are not much difference between the two methods in terms of the Frequency (or the Percentage) for each and every Number of significant items. This simulation result reaffirms our previous conclusion that the Halton sequence, though deterministic, can assign respondents to groups as randomly as the random number method

5 Conclusion

Due to its simplicity, the random number method is usually chosen for randomization in surveys. This method can split the sample evenly in a large sample surveys according to the Law of Large Number. However, in surveys with small sample sizes, the method may not work as well since the realizations of the random numbers may not come as researchers want. In some cases, the resulting groups can be quite uneven. Although the statistical analysis comparing across these groups are still sound and valid, the precision of some estimates is lower with uneven groups. This study makes use of an alternative randomization based on the Halton sequence. The benefit of using the sequence is the theoretical support and empirical applications of the sequence in other fields such as economics and engineering.

The sequence is used in a survey with a political experiment which requires random assignment of respondents into various groups. In the experiment, five random variables are used to represent the five candidates' characteristics: candidate gender, name, education, experience, and religious advocacy. Respondents are randomly assigned to various groups based on these random variables. The performance of the Halton sequence is assessed by evaluating how the sample is split across groups and how the respondents are randomly assigned into groups. The survey results demonstrate the advantage of the Halton sequence in splitting the sample across groups. Regardless of how the groups are created, either from one candidate characteristic or interaction of two characteristics, the sample is more evenly split with the Halton sequence than with the random number method. In terms of randomness, the Halton sequence performs as well as the random number method even though it is a deterministic process.

In our survey with a small sample size, there is clear benefit of the Halton sequence over the random numbers. However, we expect this benefit to diminish as the sample size increases. More studies are needed to validate the benefit of the Halton sequence in small sample size surveys and also to identify at what sample size this benefit would disappear. One limitation of our study design is that we are unable to compare respondents' answers in the political experiment between the Halton sequence and the random number method. This comparison can be done by splitting the sample into two groups, one with the Halton sequence and the other with the random number method. With this design, the two methods can be compared not only in the respondent assignment but also in the respondents' answers.

Acknowledgement

This publication was made possible by Program grant # NPRP9-190-5-022 from the Qatar National Research Fund (a member of Qatar Foundation). The findings herein reflect the work, and are solely the responsibility, of the authors.

References

Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research*, *35B*, 677–695.

Braaten, E. & Weller, G. (1979). An improved low-discrepancy sequence for multidimensional quasi-Monte Carlo integration. *Journal of Computational Physics*, *33*, 249–258.

Bratley, P., Fox, B., & Niederreiter, H. (1992). Implementation and tests of low-discrepancy sequences. *ACM Transactions on Modeling and Computer Simulation*, *2*, 195–213.

Clarke, H., Kornberg, A., Mcintyre, C., & Bauer-Kaase, P. (1999). The effect of economic priorities on the measurement of value change: New experimental evidence. *American Political Science Review*, *93(3)*, 637–647.

Dalal, D. & Hakel, M. (2016). Experimental comparisons of methods for reducing deliberate distortions to self-report measures of sensitive constructs. *Organizational Research Methods*, *19(3)*, 475–505.

Diop, A., Guterbock, T., Kermer, D., & Le, K. (2008). Prince William County Citizen Satisfaction Survey: Report of results.

Fowler, F. (2014). The case for more split-sample experiments in developing survey instruments. In P. S., J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Method for testing and evaluating survey questionnaires* (pp. 173–188). New York: Wiley.

Gaziano, C. (2005). Comparative analysis of within-household respondent selection techniques. *Public Opinion Quarterly*, *69*, 124–157.

Halton, J. H. (1964). Algorithm 247: Radical-inverse quasi-random point sequence. *Communication of the ACM*, *7*, 701–702.

Hensher, D. & Greene, W. (2003). The mixed logit model: The state of practice. *Transportation*, *30(2)*, 133–176.

Jia, D. (2008). What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry author(s). *Econometrica*, *76(6)*, 1263–1316.

Kessler, R., Berglund, P., Chiu, W., Demler, O., Heeringa, S., Hiripi, E., … Zheng, H. (2004). The US National Comorbidity Survey Replication (NCS-R): Design and field procedures. *International Journal of Methods in Psychiatric Research*, *13(2)*, 69–92.

Kish, L. (1949). A procedure for objective respondent selection within the household. *Journal of the American Statistical Association*, *44*, 380–387.

Kocis, L. & Whiten, W. (1997). Computational investigation of low-discrepancy sequences. *ACM Transactions on Mathematical Software*, *23*, 266–294.

Le, K. T., Brick, J. M., Diop, A., & Alemadi, D. (2013). Within-household sampling conditioning on household size. *International Journal of Public Opinion Research*, *25*, 108–118.

Meijer, E., Kapteyn, A., & Andreyeva, T. (2008). Health indexes and retirement modeling in international comparisons.

Meijer, E., Kapteyn, A., & Andreyeva, T. (2011). Internationally comparable health indices. *Health Economics*, *20(5)*, 600–619.

Milton, J., Shankar, V., & Mannering, F. (2008). Highway accident severities and the mixed logit model: An exploratory empirical analysis. *Accident Analysis & Prevention*, *40(1)*, 260–266.

Morokoff, W. & Caflisch, R. (1995). Quasi-Monte Carlo integration. *Journal of Computational Physics*, *122*, 218–230.

Rizzo, L., Brick, J., & Park, I. (2004). A minimally intrusive method for sampling persons in random digit dial surveys. *Public Opinion Quarterly*, *68*, 267–274.

Rosenthal, M. & Hubble, D. (1993). Results from the National Crime Victimization Survey (NCVS) CATI experiment.

Tourangeau, R. & Smith, T. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, *60*, 275–304.

Train, K. E. (2003). *Discrete choice methods with simulation*. Cambridge: Cambridge University Press.

Train, K. E. & Winston, C. (2007). Vehicle choice behavior and the declining market share of U.S. automakers. *International Economic Review*, *48*, 1469–1496.

Trussell, R. & Elinson, J. (1959). *Chronic illness in a rural area*. Cambridge, MA: Harvard University Press.

Vandewoestyne, B. & Cools, R. (2006). Good permutations for deterministic scrambled Halton sequences in terms of L2-discrepancy. *Journal of Computational and Applied Mathematics*, *189(1–2)*, 341–361.

Walle, S. & Van Ryzin, G. (2011). The order of questions in a survey on citizen satisfaction with public services: Lessons from a split-ballot experiment. *Public Administration*, *89(4)*, 1436–1450.