How do leading methods mislead?

Maleki, Ammar

Publication date:
2021

Document Version
Other version

How do leading methods mislead?
Measuring public opinions in authoritarian contexts

Ammar Maleki
Assistant Professor of Comparative Politics, Tilburg University

Abstract

Telephone and face-to-face interviews are an established method of public opinion surveying in free societies. The high demand for understanding public opinions in influential authoritarian countries stimulates institutions and researchers to utilize such established sampling methods without considering their context-dependent limitations and misleading results. Given pervasive fears of state retribution, the challenge in societies under authoritarian regimes is measuring people's genuine opinions about sensitive topics. The accelerated rates of Internet penetration and social-media popularity, even in authoritarian contexts, create a window of opportunity for innovating the measuring of public opinions in authoritarian countries. This study presents an innovative usage of online surveys to measure Iranians' unmeasurable attitudes. After examining why conventional surveying methods mislead in authoritarian contexts, this article presents the results of an anonymous online survey conducted in 2021. A comparison of the online survey with the 2020 face-to-face survey conducted by World Values Survey (WVS) confirms the hypothesis that there is a match between answers to non-sensitive questions, and thus that the online survey's results are no less representative. A large discrepancy was found, however, between the answers to sensitive questions. This suggests that leading scientific methods may result in invalid findings about authoritarian states while anonymous online surveys can reveal a more realistic picture of what people think and demand.

Introduction

In April 2013, the World Values Survey (WVS) asked Egyptians whether they had, might, or would never, attend peaceful demonstrations. An overwhelming majority of 89% declared not to do so under any circumstances. Two months later, on June 30th, millions of Egyptians, estimated at least 14 million (to 33 million) (Alexander, 2013), took to the streets. That is, about 22% of the adult Egyptian population attended the demonstrations. How is it possible, then, that the WVS massively underestimated the Egyptian people's willingness to protest?

Public opinion research on political questions in authoritarian regimes is not impossible, but it is not a matter of following the survey routines in free and democratic societies. The main challenge is not to use rigorous sampling methods, as the WVS does, but to use an approach that will extract genuine answers from fearful, self-censoring respondents.

The established and leading methods of public opinion research emphasize the importance of probability sampling in which the randomization criterion is satisfied. In practice, probability sampling entails conducting random telephone surveys or face-to-face interviews at the respondents' addresses. The crucial challenge of these methods in societies under authoritarian regimes is that the respondents are not anonymous but (feel they) are recognizable when they are approached using their phone numbers or addresses. Thus, the problem of self-censorship or "preference falsification" (Kuran, 1997) arises particularly in answering the sensitive questions. Although giving ingenuine answers to sensitive questions is a global problem in survey research (Tourangeau and Yan, 2007), in authoritarian contexts many questions considered to be routine in democratic countries are perceived as sensitive and possibly dangerous to answer.

This article presents how the leading methods in survey research mislead researchers and policymakers regarding the real public attitudes of people living under authoritarian regimes. The aim is to introduce a paradigm shift of survey research in authoritarian contexts by arguing that "anonymity sampling" is more decisive than "probability sampling" alone in producing valid results. To do so, I show that well-designed online non-probability surveys can extract a more realistic picture of public opinion on sensitive questions in such societies than utilizing traditional probability surveys that provide answers with a high reliability but low validity.

The ever-increasing internet penetration and the popularity of social media in the world, including countries under authoritarian regimes, creates a unique opportunity for accessing the majority of a population using online platforms. The exponential growth in internet penetration is so fast that former criticisms regarding the coverage bias of online surveys are now less relevant.

Our case study shall be the Islamic Republic of Iran, where the current internet penetration rate is over 90% (Financial Tribune, 2020), with 77 million mobile internet subscribers and roughly 74% of Iranians over 18 using at least one social media platform (ISPA, 2021). That is, the vast majority of the literate adult population can be reached via social media to conduct anonymous surveys.

A recent study on the case of China using the results of an online survey in 2015 suggested the potentials and necessity of conducting and comparing online and offline surveys in countries under authoritarianism (Huang and Wang, 2021). To the best of my knowledge, currently there are no published experimental studies that directly compare the results of a conventional (telephone or face-to-face) probability survey with an online non-probability survey conducted at the national level in a society under an authoritarian regime. This article fills that gap.

Below, I begin with an overview of previous research on the challenges and potentials of non-probability online samples and the general problem of asking sensitive questions, particularly in authoritarian regimes. Next is a discussion on the importance of what I call "anonymity sampling" and its advantages over probability sampling in authoritarian contexts. Then, the hypothesis that there should be a match between non-sensitive questions – and thus similar reliability or representativeness – but a discrepancy in the results on sensitive questions, is corroborated by a comparison between the results produced by a 2021 online survey and a 2020 on-site survey conducted by the World Value Survey (WVS). In conclusion, I discuss remaining challenges and argue for the potentials that new online methods offer.

Online Surveys and Non-Probability Sampling

Online surveys are becoming the dominant mode of surveying in different fields of research. For instance, in policy research, it is estimated that the percentage of journal publications using

online surveys grew to six times that of the telephone-based survey publications in the past decade (Lehdonvirta et al., 2021).

There are also critical studies, which argue that online non-probability surveys cannot produce a similar data quality and thus cannot substitute traditional probability surveys (Bruggen et al., 2016; Callegaro et al., 2014; Langer, 2018; Yeager et al., 2011). However, counter studies show that online surveys with proper sampling and adjustments can generate results that are not too different (Roshwalb et al., 2016; Walter et al. 2019) or even as accurate as those from a probability survey (Ansolabehere and Schaffner, 2011), conducted as early as 2005 when the internet penetration rate was much lower than today (Duffy et al., 2005).

Online surveys are criticized as they are considered to be non-probability surveys. A probability survey denotes the sampling mechanism in which all units in the target population "have known and positive probabilities of inclusion" (Vehovar et al., 2016, p. 327). The randomized selection of respondents is the main feature of probability sampling. The notion of non-probability sampling is used when

1- there is coverage bias, meaning that some significant segments of the target population, like non-internet users, cannot be reached by the method;
2- data gathering is affected by non-random selection or self-selection bias, meaning that the respondents are those who share a special interest in the research topic. This bias can also exist with probability sampling when there is a low response rate, *de facto* blurring the lines between the two idealized conceptions of sampling (Vehovar et al., 2016; Lehdonvirta et al., 2021).

In many countries, the non-coverage bias for online surveys has become less relevant over time, as the internet penetration rate and the number of internet users have drastically increased since 2000 (Figure 1). On the other hand, there is a significant decline in the response rate to probability telephone surveys, from 36% in 1997 to 9% in 2016, according to the Pew Research Center (Keeter, 2018). Today, online surveys' topical self-selection bias, or opt-in bias, is comparable with the impact of the non-response bias, or opt-out bias, of telephone sampling. Accordingly, scholars argue that the distinction between probability and non-probability sampling no longer holds (Gotway Crawford, 2013; Rivers, 2013; Steinmetz et al. 2014) which suggests that "there is little practical difference between opting-out of a probability sample versus opting-in to a non-probability sample" (Riley et al., 2014: 1).

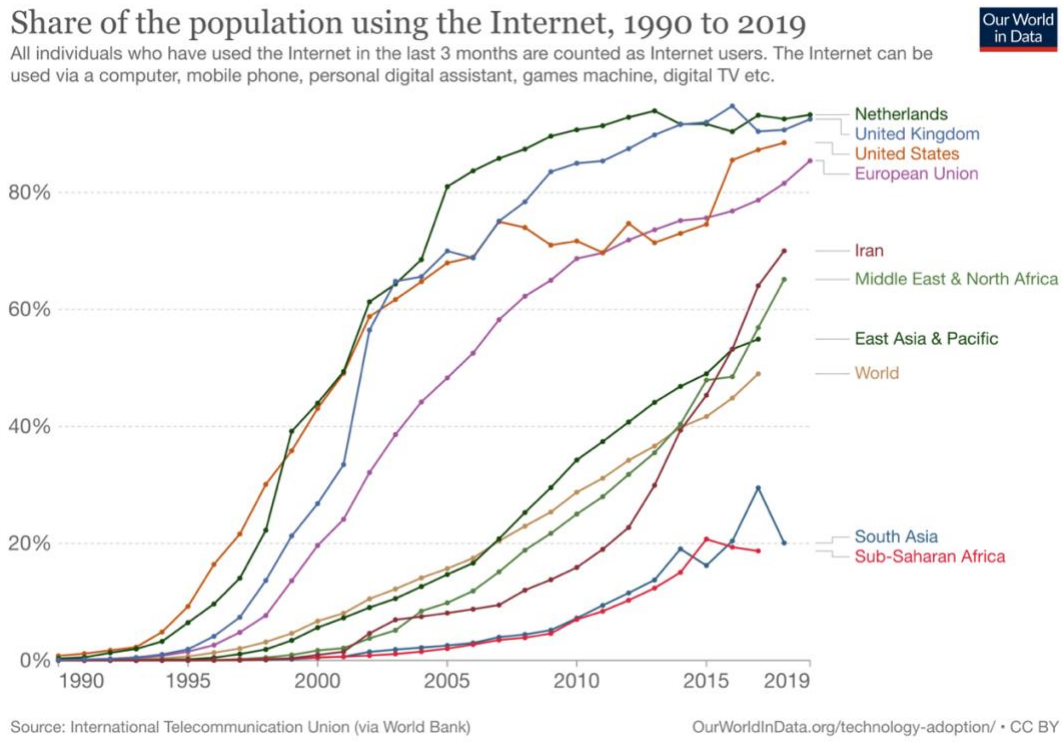

Figure 1: Internet penetration over past three decades (source: Our World in Data)

Non-Probability Surveys and Representativeness

There are several types of non-probability sampling. Those that apply also to online surveys are described as *convenience sampling*, *volunteer or self-selection sampling*, *river sampling*, *network sampling (or snowball sampling)*, *quota sampling* (or *sample matching*), and *panel sampling* (Baker et al., 2013; Elliott and Valliant, 2017; Lehdonvirta et al., 2021; Vehovar et al., 2016).

Using online panels is more common than other methods for national population sampling. However, online panels of well-known survey platforms do not necessarily outperform non-probability large-scale samples derived from river sampling. For example, Heen et al. (2014) showed an average discrepancy of up to 10% between official demographic data and the demographic characteristics of online panels of well-known platforms such as Survey Monkey, Qualtrics, and MTurk.

Of the abovementioned methods, some show greater potential to satisfy the randomization criterion for producing a representative sample from the target population. Ansolabehere and

Schaffner (2011) argue that an online opt-in survey, if carefully designed and conducted, can produce results that are very close to probability sampling. As Kish stressed decades ago,

"Great advances of the most successful sciences - astronomy, physics, chemistry - were and are, achieved without probability sampling. Statistical inference in these researches is based on subjective judgment about the presence of adequate, automatic, and natural randomization in the population … Probability sampling for randomization is not a dogma, but a strategy, especially for large numbers" (1965, 28-29, cited in Baker et al., 2013).

A crucial issue to consider, is that the data quality and results derived from non-probability surveys hinge on sampling design, effective spreading of the survey among the targeted population, and the weighting approach. When targeting a national population, randomization is the essential feature of probability sampling that can be achieved by spreading the non-probability sample as broadly as possible using a multitude of recruitment channels. This approach can assure, in some contexts and with the proper means, the survey to spread to the extent that it approximates an acceptable level of randomization. Using quota sampling based on the target population's socio-demographic features, at the recruitment and screening phase of a survey, we can profile non-probability samples to reflect the target population structure as much as possible. The sampling can be improved by involving attitudinal and behavioral along with socio-demographic variables when targeting relevant recruitment channels.

On the other hand, post-survey adjustments like weighting can adequately correct biases (Vehovar et al., 2016). Bruggen et al. (2016) compared 18 online non-probability panel surveys with probability surveys and concluded that weighting could not improve the estimates. Using 1550 samples and weighting only with three demographic variables (gender, age, region), Bethlehem (2015) came to the same conclusion about the inappropriateness of online self-selection surveys. He left a window of opportunity open, though, by adding that "weighting is only effective if proper weighting variables are used" (Bethlehem 2015, 230). Several studies, researching many surveys, show that using socio-demographic variables alone is insufficient for improving the accuracy of the weighted estimates. Involving attitudinal and political variables is repeatedly shown to be essential to decreasing bias (Baker et al., 2013; Börsch-Supan et al., 2004; Duffy et al., 2005; Lehdonvirta et al., 2021; Roshwalb et al., 2016). Thus, PEW has recently demonstrated that using non-demographic variables like political orientation can significantly improve estimates (Mercer et al., 2018).

Moreover, comparative studies show that complicated weighting methods are not more effective than the most common weighting approaches, like raking, when using the same adjustment variables (Mercer et al., 2018; Vehovar et al., 2016). Altogether, we can conclude that well-designed, well-spread, and well-weighted non-probability samples can in theory approximate representative samples.

Aside from challenges with the representativeness of online non-probability surveys, online surveys have advantages over traditional probability sampling. These advantages are not limited to cost-efficiency and convenience for researchers. Different studies show that online surveys are less susceptible to social desirability bias due to the absence of an interviewer, and accordingly, respondents are more sincere in their answers. Online surveys also fit better into people's busy lives in the present as they can fill the survey at their convenience, and give more thoughtful answers than when they answer in a hurry, sometimes giving random answers, to telephone interviewers (Duffy et al., 2014; Hanna et al., 2005).

Asking Sensitive Questions in Authoritarian Contexts

Sensitive questions are hard to survey in all societies. What is considered sensitive, however, varies from country to country. In free and democratic societies, questions about political orientations and preferences are not considered very sensitive, while similar questions can be perceived as very sensitive in societies under authoritarian regimes (Tannenberg, 2017; Truex and Tavana, 2019). The problem with sensitive questions is that respondents do not disclose their honest attitudes, resulting in "measurement error".

There are different underlying reasons for non-disclosure or self-censorship in answering sensitive questions. While the factor of "fear" is greater in authoritarian states, the challenges caused by "shy" respondents and "non-respondents" are more common in free societies. The latter explains why electoral surveys showed remarkable errors in predicting Trump's votes in the 2016 and 2020 presidential elections in the United States (Cohn, 2020; Keeter et al., 2020). Tannenberg argues that the self-censorship mechanism in authoritarian systems does not work by respondents refusing to answer sensitive questions or stating "Do not know". Instead, fearful respondents prefer to resort to the safest possible answer as "saying that you do not know or refuse to answer whether you trust the president or ruling party can likely raise suspicion"

(Tannenberg, 2017, p.17). In line with this observation, Benstead argues that "in an analysis of missing data, the Arab world does not stand out dramatically from other regions" (2018, p. 538). However, because Arab countries are dominated by authoritarian regimes, "citizens may feel compelled to participate in surveys and conceal their opinions for fear of retaliation or a loss of benefits if they do not support the regime" (2018, p. 537). The challenge of non-real answers is more pressing, in short, than counting missing data as an indicator of self-censorship in authoritarian regimes.

Kuran calls the main challenge of surveying public opinions in authoritarian regimes "preference falsification", which is "the act of deliberately misrepresenting one's genuine wants under perceived social pressure" (1997, p. 3). Recent studies in authoritarian contexts reveal how the preference falsification mechanism conceals the genuine attitudes of citizens in these societies and provides an opportunity for autocrats' propaganda based on inflated percentages. For example, according to Kalinin, based on an analysis of the 2012 Russian presidential elections, "the autocrat is strategically interested in boosting preference falsification and organizing the proportionate amount of election fraud in a given country so as to provide himself with the most favorable and, importantly, credible electoral outcome" (2018, p. 95).

In another study by Kalinin (2016), he estimated that the magnitude of the preference falsification (or social desirability bias) reaches 15% in Putin's electoral support and 20% for voter turnout. He argues that these inflated ratings serve as a valuable source for the authoritarian regime's propaganda as they permit election administrators to announce the desired final outcome, matching the results predicted by pre-election polls. Another study, conducted in early 2015 and that used a pair of list experiments and confirmed Putin's high popularity rate, suggested that up to 9% of Russian respondents hide their opposition to Putin when asked directly (Frye et al., 2017).

In another study, Jiang and Yang (2016) found evidence of systemic preference falsification as their analysis revealed a noticeable divergence between expressed and actual regime support among Chinese respondents. Interestingly, they found that "contrary to the conventional view that citizens who are wealthier, better educated, and more informed are more likely to voice out criticisms about the authority, the analyses show that they are in fact among the groups that tried to hide their true beliefs to the greatest extent when the potential cost of expressing dissent

increases" (Jiang and Yang 2016, p. 626). Comparing direct questioning and indirect list experiments, Robinson and Tannenberg (2019) also found systemic preference falsification for regime support among Chinese people, with the level of self-censorship being about 25%. In congruence with Jiang and Yang (2016), Robinson and Tannenberg also found similar results concerning the subgroups who self-censor themselves the most.

In a multilevel analysis of 36 African countries, Tannenberg recognized that "fear of the government induces a substantial and significant bias on questions regarding the citizen-state relationship in more autocratic countries, but not in more democratic countries" (2017, p. 2). The evidence for a mechanism of preference falsification has also been reported for respondents in Zimbabwe under Mugabe (Garcia-Ponce and Pasquale, 2015). A study in Egypt also shows that "implicit attitudes" are far different from "explicit attitudes" about the popularity of an authoritarian leader. While in implicit questions the positive attitudes toward Egyptian autocrat, El-Sisi, are about 58%, the surveys asking explicit questions reported the popular approval of more than 79% (Truex and Tavana, 2019).

Furthermore, another study in the UK by Dayan et al. (2007) shows a strong correlation between the sensitivity of questions and the disclosure level in traditional, telephone and face-to-face, surveys. We can conclude, therefore, that the established survey methodologies, using probability sampling via telephone or face-to-face interviews, give the wrong estimates about the real public attitudes on sensitive socio-political questions in authoritarian states. Dayan et al.'s study additionally reveals that self-disclosure in online surveys is higher than offline modes regardless of the sensitivity level. Other studies also show that online surveys decrease self-censorship and social-desirability bias because of the advantage of self-administration (Tourangeau and Yan, 2007), and with the promise of anonymity, increase self-disclosure (Abrajano and Alvarez, 2019; Davidowitz, 2017; Hanna et al., 2005; Heerwegh, 2009; Holbrook and Krosnick, 2010).

Anonymity Sampling vis-à-vis Probability Sampling

Most of, if not all, the studies that emphasize the superiority of probability surveys to non-probability surveys have analyzed and compared surveys that were conducted in free and democratic societies: Bethlehem (2015) and Bruggen et al. (2016) examined surveys in the Netherlands, Yeager et al. (2011) and Dutwin and Buskirk (2017) in the USA, Lehdonvirta et

al. (2021) in Finland, and Callegaro et al. (2014) and Langer (2018) compared surveys in several democratic countries. The first and most crucial assumption in these studies is that people express their honest attitudes in traditional probability surveys. Based on this assumption, the results of probability surveys are assumed as a benchmark, and the results of non-probability surveys are compared with the benchmark survey and evaluated accordingly.

However, to arrive at valid results, people living in authoritarian regimes should be considered "hard-to-survey" from the outset. Tourangeau (2014) categorizes hard-to-survey populations as those who are "hard-to-sample", "hard-to-identify", "hard-to-reach", "hard-to-persuade", and "hard-to-interview". We may add another category of populations who can be sampled, identified, reached, persuaded, and interviewed: those who are "hard-to-reveal" their true opinions due to the fear of state repression. If people in authoritarian regimes fall under this sub-category of "hard-to-survey" populations, we should design an appropriate survey by recruiting methods used for sampling that secure the respondents' anonymity. Among these methods are snowball sampling and online chain referral sampling (Khoury, 2020).

I call the proposed method "anonymity sampling", that is, to employ anonymous non-probability sampling and to generate sufficient chains to reach the national population. My proposed sampling method matters, given that many studies use and cite the survey results of prestigious institutes like the WVS or GALLUP, which conducted traditional probability surveys in authoritarian states. Relying on the results of these surveys conducted with rigorous "probability sampling" could be more problematic and misleading than relying on the results of non-probability surveys with "anonymity sampling" in which respondents would answer honestly to the questions.

The distinction between probability sampling versus anonymity sampling in societies governed by authoritarian states is tied, respectively, to the concepts of reliability versus validity of survey results. Probability sampling has a rigorous methodological ground that can fulfil the reliability of results, which due to self-censorship leads to measurement error in authoritarian contexts. Reliable results are therefore not by definition also valid. On the other hand, with "anonymity sampling", the validity of results can be better fulfilled while the reliability of the results could be a challenge. The latter can be treated and fulfilled if an anonymity sampling would be well-designed, conducted, and adjusted.

AAPOR's report on non-probability sampling indicates that “fit for purpose” is an important concept for judging the data quality of surveys (Baker et al., 2013). This concept may justify the use of online non-probability sampling in authoritarian regimes more than in free societies. A survey conducted by probability sampling (on sensitive questions) cannot be considered a valid benchmark in societies under authoritarian regimes, given its significant measurement error. On the other hand, online non-probability surveys will have random sampling errors. In such contexts, we should examine which of the two errors is more significant and harder to remove – the point being that the sampling bias of non-probability surveys can be treated and reduced with the strategies and adjustments mentioned earlier, while the measurement error caused by preference falsification in non-anonymous probability samples could hardly be removed. Therefore, anonymous online surveys could be the most practical tools or “fit for purpose” in authoritarian contexts in which other survey methods can hardly overcome the preference falsification problem.

Measuring Attitudes in an Authoritarian Regime: The Case of Iran

The Islamic Republic of Iran is known as an authoritarian regime, being ranked 152 among 167 countries in The Economist’s Democracy Index (Democracy Index, 2020). The country’s press freedom is ranked 174 among 180 countries by Reporters Without Borders (2021). In such an authoritarian theocratic state, people must be cautious about what to say since expressing critical comments or disclosing unacceptable (lack of) beliefs can result in harsh punishments and can even have fatal consequences. There are hundreds of prisoners of conscious in the country, including civil activists, political dissidents, those who changed their religious beliefs, and citizens who merely voiced criticisms on social media. The country also has the worst record in the number of executions in the world, being second only to China – a far larger country – in the total number. Recent executions included people who participated in the nationwide protests in 2019, an abduction abroad and subsequent hanging of a well-known journalist, and even an execution for the consumption of alcoholic beverages (IHR, 2021).

Given the context, we can imagine how measuring Iranians’ genuine opinions about sensitive social and political questions, including those on religion, would be difficult. In this context, the respondents often only dare to express their honest attitudes on sensitive questions if they can do so anonymously. Thus, we can hypothesize that conventional probability surveys would

suffer from preference falsification in the same way that studies conducted in other authoritarian regimes demonstrated. Considering the drastic growth of internet penetration in Iran, today over 90%, anonymous online sampling can be used to meet the surveying challenge. According to the state-backed Iranian Students Polling Agency (ISPA), in 2021 about 74% of Iranians over 18 use at least one social media platform (ISPA, 2021). On the other hand, the 2016 census showed that about 15% of the adult population is illiterate (Statistical Center of Iran, 2017). Assuming that the illiterate population cannot use social media, 87% of the Iranian adult literate population use social media. Therefore, the coverage bias for online surveys should be negligible for the literate population in Iran.

Below I present the methodology and results of a national online survey I directed, conducted by [removed for the sake of anonymity], and compare its findings with the results of a national, face-to-face probability survey conducted by the World Values Survey in 2020. The aim is to evaluate our measurement method and sampling quality by comparing the results from both probability and non-probability, on-site and online, surveys. We asked both "non-sensitive" and "sensitive" questions to test the hypothesis that there should be a match between answers to non-sensitive questions but a discrepancy between the answers to sensitive questions.

Data gathering and sample refining

[removed for the sake of anonymity] conducted the national online survey I discuss in this article from May 27 to June 3, 2021. The multiple chain referral sampling method was used to gather data from different demographics and socio-political groups. This method can be considered a combination of river sampling (online opt-in samples gathered through social media), network sampling (starting "seeds" to distribute the survey in specific networks), and quota sampling (real-time monitoring of samples' socio-demographic distribution, starting new seeds for balancing the sample). The survey was spread through individuals and social-media groups, channels, and pages representing radically diverse social layers of social and political perspectives. The survey was shared with and by online pages and channels belonging to specific groups such as minorities' and pro-regime groups' networks, as well as a mass audience consuming social, political, and entertainment contents. The survey link was spread in a targeted way by more than 40 popular Instagram pages and ten public Telegram channels ranging between 10 and 100 thousand followers. The pages and channels that shared the survey

and belonged to a general audience had 1 to several million followers. These measures increased the sample size and were successful in pulling samples from all different targeted socio-political and demographic strata.

For monitoring, matching, and weighting samples, we used five demographic variables, namely sex, age group, level of education, province, and urban/rural areas. We also used three political variables: respondents' voting behavior in the 2017 presidential election in Iran, and, based on the WVS, also the importance of politics in life and the Principlists' social base (those who strongly support the regime).

As a benchmark survey, we have used the seventh wave survey conducted by the World Values Survey in March-April 2020 in Iran. WVS's probability survey was conducted on-site through face-to-face interviews at respondent's home or place of residence (WVS, 2021).

We added to our online survey several sensitive and non-sensitive questions, copied verbatim from the WVS survey (which was conducted in Persian). These attitudinal questions were added next to questions covering demographic and socio-economic information as well as electoral behavior. For the external validity check, the survey asked questions about the employment status, the household income level, and respondents' home language to compare the results with the benchmarks of official data, the WVS results, and Ethnologue's research on spoken languages in Iran. One of the online survey's questions was designed to detect random responses and bot submissions. The samples with a wrong response to this question and samples with contradictory answers were excluded from the final sample (for example, those who declared that they had not reached voting age in the 2017 presidential election but also chose their age as over 30, or those who declared that they live in Iran in one question but selected outside Iran in another question). Multiple verification methods were used to ensure that the refined sample only includes the respondents who live in Iran. The verification tests showed that around 1%, either intentionally or unintentionally, reported false information about being inside or outside Iran. The refined sample size for respondents inside Iran was 68,271 respondents. All results in this article are extracted from this refined sample.

Given the fact that only literate individuals can answer online surveys, we define our target population as literate adults above 19 who reside in Iran. We extracted the demographic data

of the target population from the most recent national census that was held in Iran in 2016 (Statistical Center of Iran, 2017).

Weighting and matching

As expected, our online samples are biased toward men, those with higher education, middle-age groups, and urban and capital residents. However, the large sample size gathered from the multiple chain referral sampling ensured a diversity of respondents from a cross-section of regions, age groups, educational levels, and political orientations. To balance and adjust the samples, we employed two different methods, "raking weighting" and "sample matching," for generating a representative sample from the refined sample. Using iterative proportional fitting, commonly known as raking weighting, "a researcher chooses a set of variables where the population distribution is known, and the procedure iteratively adjusts the weight for each case until the sample distribution aligns with the population for those variables" (Mercer et al. 2018, p. 11). A study conducted by the PEW Research Center shows (Mercer et al., 2018) that the raking weighting method is among the most effective and reliable for samples derived from online surveys. As Mercel et al. (2018) suggest, a variable reflecting respondents' political behavior, was introduced in our survey and used for weighting. This variable was the respondents' electoral behavior in 2017 Iran's presidential election whose results were not disputed so that its formal outcomes can be used as the benchmark data. In addition, we weighted for respondents' reported importance of politics, as well as the Principlist social base, using WVS's results. Because they represent the ruling group, we conjecture that Principlists are more comfortable giving their true opinion in on-site surveys.

Sample matching is a methodology "for selection of representative samples from nonrandomly selected pools of respondents" (Rivers, 2006, p. 5). In the proximity matching method, each respondent is given a distance (or proximity) score based on the measured variables in the target population. For each variable used for matching, a distance function, usually just the absolute value of the difference, is defined for each particular attribute. The overall distance between a member of the target sample and a respondent from the opt-in sample is the sum of the individual distance functions on each attribute. An iterative process of selecting/dropping respondents from the sample based on the distance scores repeats/iterates until the sample matches the target population (Rivers, 2006; YouGov, 2018).

The refined sample drawn from respondents living inside Iran was weighted based on sex, age group, education level, province, urban/rural areas, respondents' voting behavior in the 2017 presidential election and "importance of politics" in life. The 2016 National Population and Housing Census (Statistical Center of Iran, 2017) and the official Labor Force Survey Results (Statistical Center of Iran, 2021) were used to extract the target population characteristics.

Regarding the "design effect", which is defined as the extent to which the sampling error in a survey departs from the sampling error that can be expected under simple random sampling, we calculated the "effective sample size" as suggested by Kish (1960). Kish's effective sample size is calculated based on the aggregation of the weight of each sample and estimates the sample size required to achieve the same level of accuracy if that sample was a simple random sample.

After weighting, the effective sample size of our survey became 6,038 which is very high and reliable. Tables 1 to 4 show the distribution of demographic and political variables in our online and WVS's on-site samples.

Table 1: Distribution of demographic variables in different samples and benchmarks

		Online				WVS	
	Demographic variables	**Unweighted**	**Weighted**	**Matching**	**2016 Census***	**Literate**	**All**
		%	%	%	%	%	5
Sex	Female	22.0	47.0	46.9	47.0	48.7	48.9
	Male	78.0	53.0	53.1	53.0	51.3	51.1
Age	Between 20 and 29 years old	17.4	30.1	30.1	30.1	29.3	28.0
	Between 30 and 49 years old	66.4	51.1	51.1	51.1	50.5	49.2
	At least 50 years old	16.2	18.8	18.8	18.8	20.2	22.7
Education	High school diploma and lower	15.4	72.3	72.2	72.3	60.7	62.4
	University education degree	84.6	27.7	27.8	27.7	39.3	37.5
Region	Rural Areas	4.1	21.2	21.2	21.2	25.4	26.0
	Urban Areas	95.9	78.8	78.8	78.8	74.6	74.0
Province	East Azerbaijan	3.0	4.8	4.8	4.8	4.6	4.7
	West Azerbaijan	1.6	3.6	3.6	3.6	2.7	2.7
	Ardabil	0.8	1.5	1.5	1.5	2.7	2.7
	Isfahan	6.8	6.9	6.9	6.9	5.7	5.9
	Alborz	4.7	3.8	3.8	3.8	3.4	3.3
	Ilam	0.8	0.7	0.7	0.7	1.3	1.3
	Bushehr	1.0	1.5	1.5	1.5	2.7	2.7
	Tehran	38.9	19.1	19.1	19.1	15.9	15.9
	Chaharmahal and Bakhtiari	0.7	1.1	1.1	1.1	0.7	0.7
	South Khorasan	0.7	0.9	0.9	0.9	2.1	2.1
	Razavi Khorasan	7.1	7.8	7.8	7.8	7.4	7.3
	North Khorasan	0.7	0.9	0.9	0.9	0.6	0.7
	Khuzestan	4.0	5.4	5.4	5.4	6.9	6.7
	Zanjan	0.9	1.3	1.3	1.3	1.8	2.0
	Semnan	0.8	1.0	1.0	1.0	0.6	0.7
	Sistan and Baluchistan	0.8	2.1	2.1	2.1	2.7	2.5
	Fars	5.5	6.3	6.3	6.3	7.6	7.3
	Ghazvin	1.2	1.6	1.6	1.6	2.0	2.0
	Qom	1.8	1.6	1.6	1.6	1.5	1.5
	Kurdistan	1.2	1.8	1.8	1.8	2.0	2.0
	Kerman	1.5	3.6	3.6	3.6	4.3	4.7
	Kermanshah	1.6	2.4	2.4	2.4	2.0	2.0
	Kohgiluyeh and Boyer-Ahmad	0.8	0.8	0.8	0.8	0.0	0.0
	Golestan	1.0	2.2	2.2	2.2	3.4	3.3
	Gilan	2.8	3.5	3.5	3.5	3.3	3.3
	Lorestan	1.4	2.0	2.0	2.0	1.3	1.3
	Mazandaran	3.1	4.5	4.5	4.5	2.8	2.7
	Markazi	1.3	1.8	1.8	1.8	2.7	2.7
	Hormozgan	1.0	2.0	2.0	2.0	0.7	0.7
	Hamadan	1.3	2.1	2.1	2.1	2.0	2.0
	Yazd	1.3	1.4	1.4	1.4	2.7	2.7
	Effective sample size	68271	6038	2000	-	1426	1499

* Population of literate individuals above 19 years old (from the 2016 Iran census)

Table 2: Distribution of political variables in different samples and benchmarks

		Online			
	Electoral behaviour (2017 elections)	**Unweighted**	**Weighted**	**Matching**	**Official Results ***
		%	%	%	%
Those who voted for	Hassan Rouhani (+ Hashemitaba)	61.2	42.2	42.3	42.2
	Ebrahim Raisi (+ Mir-Salim)	12.9	28.8	28.7	28.8
	I did not vote (+ I cast a blank vote)	25.9	29.0	29.0	29.0
	Effective sample size	68271	6038	2000	-

* reference: Ministry of Interior (2017)

Table 3:Distribution of important of politics in life.

	Online			WVS	
For each of the following aspects, indicate how important it is in your life.	**Unweighted**	**Weighted**	**Matching**	**Literate**	**All**
Politics	%	%	%	%	%
Very important	52.8	22.4	22.4	22.4	22.6
Rather important	35.9	36.2	36.2	36.2	35.6
Not very important	7.4	19.6	19.6	19.6	19.3
Not at all important	3.3	21.6	21.6	21.6	22.2
Don't know	0.5	0.2	0.2	0.2	0.3
Effective sample size	68271	6038	2000	1426	1499

Table 4: Distribution of Principlists' social base

	Online			WVS	
	Unweighted	**Weighted**	**Matching**	**Literate**	**All**
Political orientation	%	%	%	%	%
Principlists	11.0	15.0	15.0	16.1	15.6
Effective sample size	68271	6038	2000	1426	1499

Main findings

comparing socio- political variables

To test the accuracy of our online survey and the adjusted samples, we have first compared the online survey results with some benchmark data as external validity checks. We use Iran's official statistics, Ethnologue, and WVS figures as external evidence.

Table 5 compares the employment-to-population ratio in the online/weighted/matched samples with that of the target population based on the official workforce statistics on both urban and rural levels. While the online sample includes only literate individuals above 19 years old, the official statistic of those formally employed reflects both literate and illiterate populations of the labor force who are above 15. As reported in the Labor Force Survey Results (Statistical Center of Iran, 2021), in urban areas, the employment rate of literate people is higher than that of the illiterate and barely literate population. As can be seen, the weighting and matching adjustments could remove the bias to a great extent and approximate the employment ratio to that of the target population.

Table 5: Employment Rate

	Online			**Benchmark**	**WVS**	
	Unweighted	**Weighted**	**Matching**	**Workforce statistics***	**Literate**	**All**
Employed	%	%	%	%	%	%
Whole country	65.9	41.8	40.8	36.9	39.6	38.5
Urban areas	66.7	42.0	43.1	36.0	39.7	38.8
Rural areas	48.4	41.2	31.9	40.1	39.2	37.6
Effective sample size	68271	6038	2000	-	1426	1499

* Workforce statistics for individuals above 15 years old in Winter 2020 (Statistical Center of Iran (2021).

As another external validity check, we compare the household income distribution, based on income deciles, of the online survey with the adjusted samples (that were not weighted for income). Based on the definition, an income decile is each of ten equal groups into which a population can be divided according to income distribution. Thus, each income decile should include 10% of the population.

Table 6: Household income distribution (based on the official exchange rate of the time)

	Online				WVS	
	Unweighted	Weighted	Matching	Benchmark	Literate	All
Income level of different deciles *	%	%	%	%	%	%
1st decile (household monthly income below 380 $)	4.9	14.2	16.1	10	21.7	22.8
2nd decile (household monthly income between 380 $ and 499 $)	4.0	9.3	9.1	10	6.7	6.7
3rd decile (household monthly income between 499 $ and 618 $)	4.1	7.6	7.8	10	12.2	11.9
4th decile (household monthly income between 618 $ and 736 $)	5.8	10.4	9.6	10	14.7	14.1
5th decile (household monthly income between 736 $ and 879 $)	6.4	8.3	7.5	10	24.5	24.2
6th decile (household monthly income between 879 $ and 1022 $)	8.4	9.7	9.3	10	6.1	5.9
7th decile (household monthly income between 1022 $ and 1211 $)	14.2	13.6	13.7	10	6.9	6.7
8th, 9th, or 10th decile (household monthly income above 1211 $)	52.1	26.9	26.9	30	6.2	6.3
Effective sample size	68271	6038	2000		1426	1499

* The distribution of wealth in the society for each decile should be 10%. Reference for household monthly income: Eghtesad News (2020).

Table 6 shows how weighting could effectively approximate the online sample to the target values, even better than the matching method. Moreover, we can see that the weighted sample is much closer to the benchmark values (10%) than the WVS sample. The latter is highly biased toward the first and fifth deciles. *Prima facie*, the figure of 6% for the highest three income deciles, who make up 30% of the population, implies a highly biased sample. These biased results of the WVS sample could be a sign of measurement error due to the presence of an interviewer, which may make people hide their actual household income level.

We have used the distribution of respondents' self-reported home language as another external check. Table 7 presents the home language distribution in online samples and the Ethnologue database (Eberhard et al., 2021) and in the WVS sample. The weighted and matched samples are similar, and close to the benchmark figures of the Ethnologue statistics, with differences lower than 3%.

These results show that the external validity checks confirmed the effectiveness of the weighting and matching methods and approved the quality of the online sample.

Table 7: Home language distribution

	Online			Benchmark	WVS	
	Unweighted	Weighted	Matching	Ethnologue (2019)	Literate	All
Home language	%	%	%	%	%	%
Farsi	79.1	66.1	66.4	63.3	61.2	60.8
Azerbaijani / Turkic	7.9	12.6	2.5	13.6	16.8	17.2
Kurdish	3.6	5.3	5.9	5.8	5.5	5.8
Luri	3.9	5.0	5.2	3.6	3.9	4.0
Gilaki	1.0	1.9	1.8	3.0	6.4*	6.3*
Mazandarani	1.0	1.7	1.3	2.8	-*	-*
Balochi	0.4	1.4	1.3	1.5	-	-
Arabic	0.4	0.9	0.6	3.5	1.3	1.2
Laki	0.6	0.9	0.8	1.5	-	-
Turkmeni	0.2	0.6	0.7	0.9	-	-
Tati	0.2	0.3	0.3	0.1	-	-
Armeni	0.1	0.2	0.2	0.1	-	-
Other	1.7	3.0	3.0	0.3	4.8	47
Effective sample size	68271	6038	2000		1426	1499

* In the WVS sample, Gilaki and Mazandarani was asked as one group.

Comparing the results of non-sensitive and sensitive questions

The online survey asked a standard WVS question about how much people in the society trust each other. Table 8 shows that the results of the weighted and matching sample are close to the results of the WVS survey with about a 2.5% difference.

Table 8: Trust in people

	Online			WVS	
Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?	Unweighted	Weighted	Matching	Literate	All
	%	%	%	%	%
Most people can be trusted	20.6	11.6	11.6	14.1	15.0
Need to be very careful	79.4	88.4	88.4	85.9	85.0
	68271	6038	2000	1426	1499

Another way to validate the online survey's results is by comparing with the WVS's results for non-sensitive questions, which according to my hypothesis should overlap. On the other hand, we should expect to find diverging answers to sensitive questions. In the online survey used as

an example in this article, the first six questions on what is "important in life" from the 2020 WVS survey were asked using the same Persian wordings (WVS has published the original questionnaires in the surveyed languages). In tables 9 to 13, the answers on the importance of five different subjects are shown for both the online survey and the face-to-face WVS survey. The tables present the estimates of unweighted and weighted samples and the results of matching samples. Regarding the WVS, we present the results of both the whole sample and the sample of only literate respondents, as the latter is best comparable with our online survey's target population.

Table 9 shows the importance of family for Iranians in five different samples. As the variance in answers to this question is low, the weighted and matching figures are not significantly different from the unweighted results. All figures are close to the WVS figures.

Table 10 presents a better picture of the impact of weighting and matching adjustments regarding the importance of friends in life. The discrepancy between the unweighted online sample and the WVS sample on "net importance of friends in life" was about 14%. The weighted sample has lowered these differences to about 5%.

Table 9: Important in life: Family.

For each of the following aspects, indicate how important it is in your life.	Online			WVS	
	Unweighted	**Weighted**	**Matching**	**Literate**	**All**
Family	%	%	%	%	%
Very important	91.3	90.7	91.9	94.0	93.9
Rather important	7.3	6.3	5.1	5.1	5.2
Not very important	0.7	1.0	0.9	0.8	0.8
Not at all important	0.4	1.7	1.9	0.1	0.1
Don't know	0.3	0.4	0.4	-	-
Important	98.6	97.0	97.0	99.1	99.1
Not important	1.1	2.7	2.8	0.9	0.9
Net important	97.5	94.3	94.2	98.2	98.2
Effective sample size	68271	6038	2000	1426	1499

Table 10: Important in life: Friends.

For each of the following aspects, indicate how important it is in your life.	Online			WVS	
	Unweighted	Weighted	Matching	Literate	All
Friends	%	%	%	%	%
Very important	32.7	26.7	27.1	28.6	29.7
Rather important	57.1	58.4	58.6	54.4	53.1
Not very important	8.3	10.3	9.5	11.2	10.9
Not at all important	1.5	3.8	3.6	5.7	6.1
Don't know	0.5	0.9	1.2	0.1	0.2
Important	89.8	85.1	85.7	83.0	82.8
Not important	9.8	14.1	13.1	16.9	17.0
Net important	80.0	71.0	72.6	66.1	65.8
Effective sample size	68271	6038	2000	1426	1499

Regarding the importance of leisure time, as seen in Table 11, weighting and matching could not improve the estimates of unweighted figures that already closely resembled the WVS figures.

Table 11: Important in life: Leisure time.

For each of the following aspects, indicate how important it is in your life.	Online			WVS	
	Unweighted	Weighted	Matching	Literate	All
Leisure Time	%	%	%	%	%
Very important	47.4	46.4	46.3	40.2	40.4
Rather important	45.2	43.1	43.4	51.5	51.1
Not very important	5.9	6.9	6.9	5.7	5.7
Not at all important	0.8	2.3	2.0	2.6	2.6
Don't know	0.7	1.4	1.4	-	0.2
Important	92.6	89.5	89.7	91.7	91.5
Not important	6.7	9.2	8.9	8.3	8.3
Net important	85.9	80.3	80.8	83.4	83.2
Effective sample size	68271	6038	2000	1426	1499

According to Table 12, weighting and matching could improve the estimates of unweighted figures to approximate the WVS figures. While the discrepancy between the unweighted online sample and the WVS sample on "net importance of work in life" was about 3.5%, the weighted and matching samples have lowered these differences to about 0.5%.

Altogether, for non-sensitive items, we see from tables 9 to 12 that the discrepancies between the figures of adjusted online samples and the WVS sample are between 1 to 5 percent.

Table 12: Important in life: Work.

For each of the following aspects, indicate how important it is in your life.	Online			WVS	
	Unweighted	**Weighted**	**Matching**	**Literate**	**All**
Work	%	%	%	%	%
Very important	82.4	81.6	82.6	78.0	77.5
Rather important	15.8	14.3	13.5	18.6	18.6
Not very important	1.2	1.9	1.7	1.9	1.9
Not at all important	0.3	1.2	1.1	1.3	1.7
Don't know	0.4	0.9	1.2	0.2	0.3
Important	98.2	95.9	96.1	96.6	96.1
Not important	1.5	3.1	2.8	3.2	3.6
Net important	96.7	92.8	93.3	93.4	92.5
Effective sample size	68271	6038	2000	1426	1499

On the other hand, as shown in Table 13, there is a significant difference between the results of online and WVS surveys on the importance of religion. Religion is, based on contextual knowledge, a very sensitive topic in Iran, due to its political theocracy. The figures show that while 92% of the respondents in a face-to-face survey declared that religion is "very important" or "rather important" in their life, the results of the online survey showed that religion is important for only 53% to 56% of Iranians according to the weighted and matched figures.

Table 13: Important in life: Religion.

For each of the following aspects, indicate how important it is in your life.	Online			WVS	
	Unweighted	Weighted	Matching	Literate	All
Religion	%	%	%	%	%
Very important	23.0	33.0	35.8	69.4	70.5
Rather important	17.7	19.6	20.5	22.8	22.0
Not very important	18.1	14.7	13.6	3.5	3.3
Not at all important	39.2	30.3	27.6	4.2	4.1
Don't know	2.0	2.4	2.6	0.1	0.1
Important	40.7	52.6	56.3	92.2	92.5
Not important	57.3	45.0	41.2	7.7	7.4
Net important	-16.6	7.6	15.1	84.5	85.1
Effective sample size	68271	6038	2000	1426	1499

Given the observed similarities between the online survey results and the WVS survey, the huge discrepancy between the two surveys' figures on religion cannot be assigned to sampling bias, but reveals the preference falsification problem. The Iranian regime considers atheism, heresy, and conversion to constitute severe crimes for which a person can be executed. In such a context, it is not strange that people would answer falsely about the importance of religion in their life in a face-to-face interview (many of which were conducted with third persons present). The question to what extent the figures from the online survey are realistic and accurate should be examined by replicating and repeating a similar research. However, the presented results leave no doubt that the level of difference corroborates the severity of the problem of self-censorship and preference falsification.

This finding is crucial for survey research scholars who conduct surveys in societies under the yoke of authoritarian regimes. The figures reported by the reputable WVS are widely used by sociologists and political scholars across the world. Moreover, the well-known polling agency PEW has echoed Iran's census, claiming that 99.5% of the population identify as Muslim (Pew Research Center, 2013). Reproducing and using these numbers gives a wrong picture of a diverse society and results in false forecasts about the socio-political developments in that country. Such false figures mislead observers who are struck by "unanticipated revolutions",

as Kuran argued, that may occur "because preference falsification concealed the opposition developing under the surface" (1997: 20).

Conclusion

This study examined the application and accuracy of online non-probability sampling in a society under an authoritarian regime. Comparing the results of our online survey with official statistics and a benchmark face-to-face survey conducted by the reputable WVS, we found that weighted and matched samples could approximate the benchmark figures on demographic data and non-sensitive attitudinal questions. Our findings also corroborate the hypothesis that genuine answers to sensitive questions cannot be measured using the leading methods of probability surveys that work in free and democratic countries.

We observed a discrepancy of about 40% between on-site and on-line samples in estimating the importance of religion in a country governed by a theocratic regime. Other evidence and external validity checks have suggested that this discrepancy cannot be caused by a "sampling error" of a non-probability online survey but is the consequence of the "measurement error" of probability sampling using a rigorous and established method in the incompatible context of an unfree society. I conclude that in such societies, where people self-censor in answering sensitive questions, the measurement error of traditional probability sampling is more critical than the sampling error of non-probability online surveys.

That does not take away the fact that such online surveys require a unique team with contextual knowledge, access to (mass) media, and a high response rate of the population, in Iran caused by the democratic deficit, that is able to generate a high quality and large sample. When these requirements are met, anonymity sampling leads to more valid results than probability sampling concerning crucial social and political questions. Moreover, we know that due to the low response rate and high opting-out in phone surveys, probability sampling is nowadays by definition not feasible using conventional survey methods. In reality, we have different non-probability samples with different data qualities based on their approximation to a randomized sample. This article showed that the randomization of online opt-in surveys can be greatly improved, whereas the problem of preference falsification cannot be removed in authoritarian contexts using traditional survey methods.

Thanks to the large sample size and using non-demographic variables (political behavior, the importance of politics, and the Principlists' social base) besides demographic factors, we could approximate a representative sample by applying two adjustment approaches of weighting and matching. Also, our large sample size ensured that after applying heavy weighting, the resulting "effective sample size" was still large enough to stand up to significant testing (cf. Duffy et al. 2014).

The findings of this study should be replicated and examined with more surveys and more comparative tests. For now, this study is hopefully an eye-opener for scholars and policy-makers who use the data extracted from the traditional surveying methods in authoritarian contexts. Even the reputable WVS's results should be scrutinized and never taken at face value. No matter how rigorous the leading traditional surveying methods are, they do not yield valid results in frightening authoritarian contexts.

References

Abrajano, M., and Alvarez, R. M. (2019). Answering questions about race: how racial and ethnic identities influence survey response. *American Politics Research, 47*(2), 250-274.

Alexander, R. (2013, July 6). Counting crowds: Was Egypt's uprising the biggest ever?, BBC News. https://www.bbc.com/news/magazine-23312656

Ansolabehere, S., and Schaffner, B. F. (2011). Re-examining the validity of different survey modes for measuring public opinion in the US: Findings from a 2010 multi-mode comparison. In *66th AAPOR Annual Conference, Phoenix AZ*.

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., ... and Tourangeau, R. (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of survey statistics and methodology*, *1*(2), 90-143.

Benstead, L. J. (2018). Survey research in the Arab world: Challenges and opportunities. PS: Political Science and Politics, 51(3), 535-542.

Bethlehem, J. (2015). Sunday shopping: the case of three surveys. Survey Research Methods, 9(3), 221-230.

Börsch-Supan, A., Elsner, D., Faßbender, H., Kiefer, R., McFadden, D., and Winter, J. (2004). How to make internet surveys representative: A case study of a two-step weighting procedure. *MEA Discussion Papers*, 067-04.

Bruggen, E., van den Brakel, J. A., and Krosnick, J. (2016). *Establishing the accuracy of online panels for survey research* (report no. 4). The Netherlands: Statistics Netherlands.

Callegaro, M., Villar, A., Yeager, D., and Krosnick, J. A. (2014). A critical review of studies investigating the quality of data obtained with online panels based on probability and nonprobability samples. In *Online panel research: A data quality perspective* (pp. 23–53). John Wiley and Sons, UK.

Cohn N. (2020, November 10). What Went Wrong With Polling? Some Early Theories. The New York Times. https://www.nytimes.com/2020/11/10/upshot/polls-what-went-wrong.html

Dayan, Y., Paine, C. S., and Johnson, A. J. (2007). Responding to sensitive questions in surveys: A comparison of results from Online panels, face to face, and self-completion interviews. *World Association for Public Opinion Research, Berlin*.

Democracy Index. 2020. "Democracy Index 2020: In sickness and in health?" The Economist. https://www.eiu.com/n/campaigns/democracy-index-2020/.

Duffy, B., Smith, K., Terhanian, G., and Bremer, J. (2005). Comparing data from online and face-to-face surveys. *International Journal of Market Research*, *47*(6), 615-639.

Dutwin, D., and Buskirk, T. D. (2017). Apples to oranges or gala versus golden delicious? Comparing data quality of nonprobability internet samples to low response rate probability samples. *Public Opinion Quarterly*, *81*(S1), 213-239.

Eberhard, D. M., Simons, G.F., and Fennig C. D. (2021). Ethnologue: Languages of the World 24th edn. Dallas: SIL International. Online version: http://www.ethnologue.com.

Eghtesad News (2020, December 5). Household deciles based on monthly income (in Persian). Eghtesad News. https://www.eghtesadnews.com/fa/tiny/news-377097

Elliott, M. R., and Valliant, R. (2017). Inference for nonprobability samples. *Statistical Science*, 249-264.

Financial Tribune (2020, November 29). Iran Internet Penetration Rate Reaches 94 Percent. Financial Tribune. https://financialtribune.com/node/106324.

Frye, T., Gehlbach, S., Marquardt, K. L., and Reuter, O. J. (2017). Is Putin's popularity real?. *Post-Soviet Affairs*, *33*(1), 1-15.

García-Ponce, O., and Pasquale, B. (2015). How political repression shapes attitudes toward the state: Evidence from Zimbabwe. Unpublished working paper.

Gotway Crawford, C. A. (2013). Comment. Journal of Survey Statistics and Methodology, 1, 118-124.

Hanna, R. C., Weinberg, B., Dant, R. P., and Berger, P. D. (2005). Do internet-based surveys increase personal self-disclosure?. *Journal of Database Marketing and Customer Strategy Management*, *12*(4), 342-356.

Heen, M. S., Lieberman, J. D., and Miethe, T. D. (2014). A comparison of different online sampling approaches for generating national samples. *Center for Crime and Justice Policy*, *1*(9), 1-8.

Heerwegh, D. (2009). Mode differences between face-to-face and web surveys: an experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research*, *21*(1), 111-121.

Holbrook, A. L., and Krosnick, J. A. (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly*, *74*(1), 37-67.

Huang, Y., and Wang, L. (2021). Political Values and Political Trust in the Digital Era: How Media Engagement Divides Chinese Netizens. International Journal of Sociology, 51(3) 197-217.

IHR (2021). Annual Report On The Death Penalty In Iran 2020. Iran Human Rights. https://www.iranhr.net/media/files/Rapport_iran_2021-gb-290321-BD.pdf

Jiang, J., and Yang, D. L. (2016). Lying or believing? Measuring preference falsification from a political purge in China. Comparative Political Studies, 49(5), 600-634.

Kalinin, K. (2016). The social desirability bias in autocrat's electoral ratings: evidence from the 2012 Russian presidential elections. *Journal of Elections, Public Opinion and Parties*, *26*(2), 191-211.

Kalinin, K. (2018). Linking Preference Falsification and Election Fraud in Electoral Autocracies: The Case of Russia. *Political Studies*, *66*(1), 81-99.

Keeter S., Kennedy C. and Deane C. (2020, November 13). Understanding How 2020 Election Polls Performed And What It Might Mean For Other Kinds Of Survey Work. Pew Research Center. https://www.pewresearch.org/fact-tank/2020/11/13/understanding-how-2020s-election-polls-performed-and-what-it-might-mean-for-other-kinds-of-survey-work/

Keeter, S. (2018). Evidence about the accuracy of surveys in the face of declining response rates. In *The palgrave handbook of survey research* (pp. 19-22). Palgrave Macmillan, Cham.

Khoury, R. B. (2020). Hard-to-Survey Populations and Respondent-Driven Sampling: Expanding the Political Science Toolbox. *Perspectives on Politics*, *18*(2), 509-526.

Kish, L. (1965). Survey Sampling. John Wiley, New York.

Kuran, T. (1997). Private truths, public lies: The social consequences of preference falsification. Cambridge, MA: Harvard University Press.

Langer, G. (2018). Probability versus non-probability methods. In *The Palgrave handbook of survey research* (pp. 351-362). Palgrave Macmillan, Cham.

Lehdonvirta, V., Oksanen, A., Räsänen, P., and Blank, G. (2021). Social media, web, and panel surveys: using non-probability samples in social and policy research. *Policy and Internet, 13(1),* 134-155.

Mercer, A., Lau, A., and Kennedy, C. (2018). For weighting online opt-in samples, what matters most. *Pew Research Center*.

Ministry of Interior (2017). Results of 2017 presidential elections in Iran. https://bit.ly/3h4JJrd

Pew Research Center (June 11, 2013). Iranians' Views Mixed on Political Role for Religious Figures. Pew Research Center.

Reporters Without Borders (2021). 2021 Worls Press Freedom Index. https://rsf.org/en/ranking

Riley, W., Hays, R. D., Kaplan, R. M., and Cella, D. (2014). Sources of comparability between probability sample estimates and nonprobability web sample estimates. In *Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference.*

Rivers, D. (2013). Comment. Journal of Survey Statistics and Methodology, 1, 111-117.

Rivers, D. (2006). Sample matching: Representative sampling from Internet panels. Polimetrix White Paper Series.

Robinson, D., and Tannenberg, M. (2019). Self-censorship of regime support in authoritarian states: Evidence from list experiments in China. Research and Politics, 6(3), 1-9.

Roshwalb, A., Lewis, Z., and Petrin, R. (2016). The Efficacy of Nonprobability Online Samples. *JSM Proceedings, Survey Research Methods Section*, 3657-3666.

Statistical Center of Iran (2017). Census 2016: Detailed Results. https://www.amar.org.ir/english/Population-and-Housing-Censuses/Census-2016-Detailed-Results

Statistical Center of Iran (2021). A Selection of Labor Force Survey Results: Winter 2020. https://www.amar.org.ir/Portals/1/releases/lfs/LFS-Winter-1399.pdf

Steinmetz, S., Bianchi, A., Tijdens, K., and Biffignandi, S. (2014). Improving web survey quality: Potentials and constraints of propensity score adjustments, in Online Panel Research: A Data Quality Perspective, 273-298. New York: John Wiley and Sons Ltd.

Stephens-Davidowitz, S. (2017). Everybody Lies: what the internet can tell us about who we really are. Bloomsbury Publishing, New York.

Tannenberg, M. (2017). The autocratic trust bias: Politically sensitive survey items and self-censorship. V-Dem Working Paper, 49.

Tourangeau, R. (2014). Defining Hard-to-Survey Populations. In *Hard-to-Survey Populations*, edited by Roger Tourangeau, Brad Edwards, Timothy P. Johnson, Kirk M. Wolter, and Nancy Bates, 3–20. Cambridge: Cambridge University Press.

Tourangeau, R., and Yan, T. (2007). Sensitive questions in surveys. *Psychological bulletin*, *133*(5), 859.

Truex, R., and Tavana, D. L. (2019). Implicit Attitudes toward an Authoritarian Regime. *The Journal of Politics*, *81*(3), 1014-1027.

Vehovar, V., Toepoel, V., and Steinmetz, S. (2016). Non-probability sampling. In C. Wolf, D. Joye, T. Smith, and Y. Fo (Ed.), The Sage handbook of survey methods (pp. 329-345). Thousand Oaks, CA: Sage.

Walter, S. L., Seibert, S. E., Goering, D., and O'Boyle, E. H. (2019). A tale of two sample sources: Do results from online panel data and conventional data converge?. *Journal of Business and Psychology*, *34*(4), 425-452.

WVS (2021), What We Do: Fieldwork and Sampling, https://www.worldvaluessurvey.org/WVSContents.jsp

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., and Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public opinion quarterly*, *75*(4), 709-747.

YouGov (2018). YouGov Sampling Methodology. www.colorado.edu/lab/aprl/sites/default/files/attached-files/yougov_sampling_2018_0.pdf