



Measurement in a Post-NTO world



New Tariff Order

On February 1st, 2019, the Telecom Regulatory Authority of India (TRAI) implemented the New Tariff Order (NTO) with a requirement that migration needed to be completed by March 31st, 2019. The NTO has resulted in a situation where households may customize the channels or bouquets of channels they receive. Families also can select the 75 non-DD free to air (FTA) channels they receive as part of the payment of the Network Capacity Fee (NCF). This change resulted in a new environment where the specific channels received could dramatically differ from household to household.

New Tariff Order, those three words which pretty much changed the world for most of us: be it broadcasters, distributors, television viewers, marketers - almost everybody in our ecosystem. While NTO continues to be a very high decibel and high impact event in the history of television consumption, there have been other events (e.g., 'total digitization', DAS) in the past that shaped up television viewing as we know it.

Questions that are pertinent to ask:

- **During ground level changes like NTO, total digitization, etc. how does the sample remain representative?**
- **How have ground-level changes impacted BARC's measurement?**

Reception of a channel by a household is a necessary condition for the home to view the channel. As such, it is essential that the BARC India Television Measurement Panel correctly captures the right mix of households and their channel reception choices. It is, therefore, crucial to understand the principles of random sampling, which is the underlying principle driving BARC India's sample design, to assess whether BARC India continues to capture 'What India Watches' post-NTO.

This paper will explore some important concepts of random sampling as well as how random sampling performs against highly heterogeneous populations and, therefore, concludes that BARC India's measurement panel remains precise and robust.

Changing Television Distribution Landscape

TRAI's NTO provided consumers an opportunity to customize the actual channels they receive through their television service provider. Consumers would now pay for only those channels which they wanted to receive. The expected result was that many consumers would thereby scale back the number of channels they receive to be more fiscally prudent with their television expenditures.

This phenomenon of scaling back channels has been witnessed on the ground – as evident through the number of channels watched. The average number of channels viewed per TV Household has reduced post-NTO (Figure 1). At an All India level, pre-NTO¹ 35% of households watched 31 or more channels. This percentage decreased to 24% post-NTO². The proportion of households watching 1 to 15 channels increased from 21% to 30% over the same period.

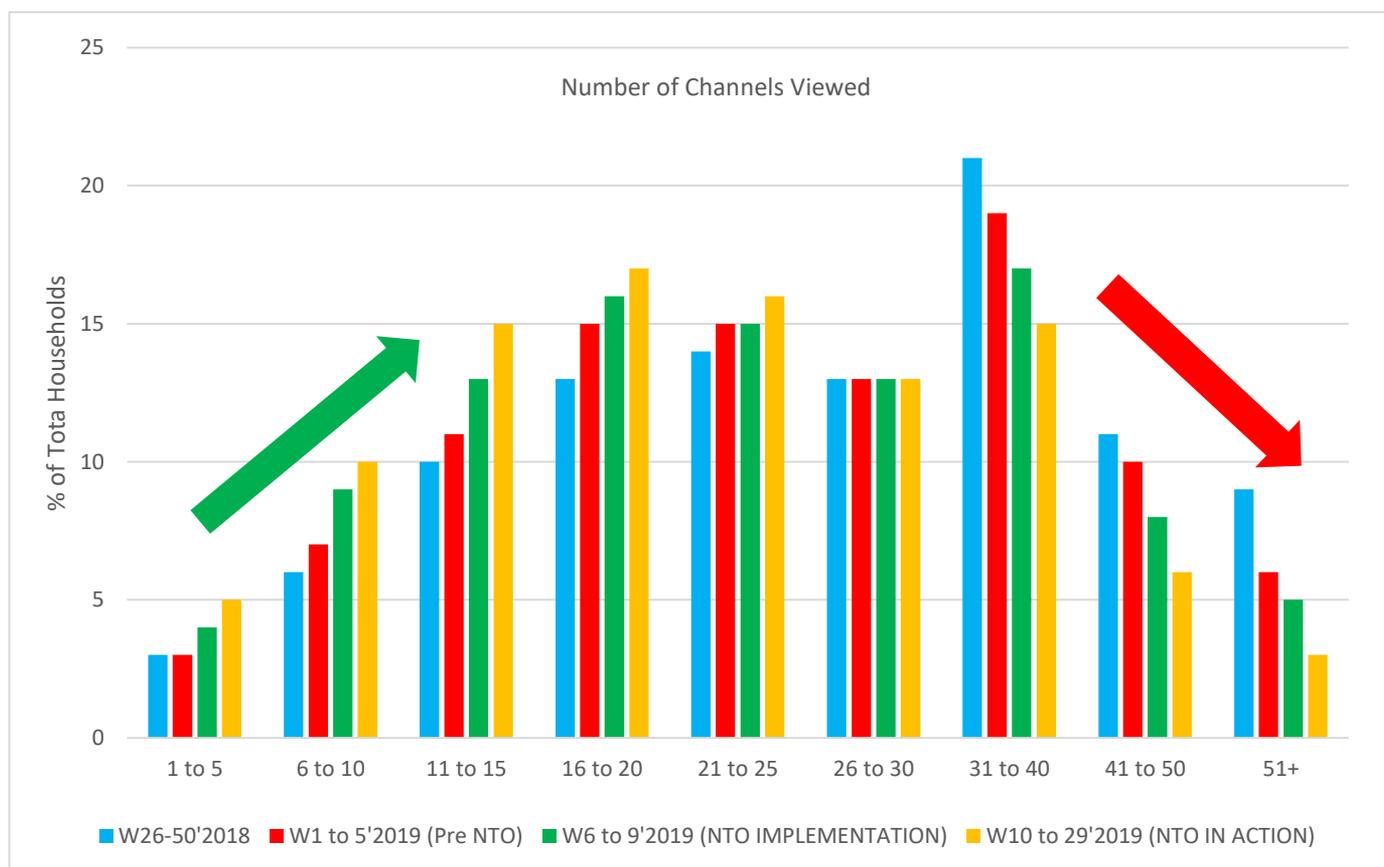


FIGURE 1

¹Weeks 1 to 5 2019

²Weeks 10 to 29 2019

As the average number of channels received by a household decreases, the variability between households in the channels received necessarily increases. Since channel reception is a necessary condition for channel viewership, this ultimately can result in increased heterogeneity in viewership within the country.

A census-based study would have given us an accurate view of ground reality. However, complete enumeration through census-based surveys is not only impractical but also imposes enormous costs that are both unsustainable and unnecessary if the nature and methods of statistical sampling are appropriately considered.

While channels watched is only a proxy for available channels, it is interesting to note that the changes in the BARC India television viewing panel mirror the expected behaviour on the ground. This phenomenon provides a degree of confidence that the panel is reflective of the ground reality.

BARC India has always emphasized on data being robust, and this brings us to the importance of accuracy and precision in our sample design - which results in this robust data.



Accuracy and Precision

Accuracy and precision can be how the quality of a survey is measured. These are sometimes also understood, or referred to as, validity and reliability. Both constructs refer to types of errors associated with the estimate of interest - in the case of BARC India, television viewing.

Accuracy focuses on systematic errors in measurement - or biases. These could be biases due to incomplete sample frames (e.g., the former service excluded households in rural India), biases due to technological limitations (e.g., an audio stream is required to capture an audio watermark), or processing errors.

Precision focuses on the error from only observing a portion (i.e., sample) of the population - often referred to as sampling error - where the sample does not correctly represent the population. In some instances, precision can be measured through the standard error. Estimates with smaller standard errors are more precise than those with more substantial standard errors.

These can be easily understood using the analogy of a dartboard (Figure 2).

Accuracy refers to how close the darts fall to the bullseye (i.e., the target), whereas precision refers to how consistently close the darts fall to one another.

A dart player can be either accurate or precise, both accurate and precise, or neither accurate nor precise.

Outside of technology and data production issues, accuracy is typically controlled through a robust sample design and sampling plan. Precision, on the other hand, is generally managed through sample size where larger sample sizes, all other things being equal, tend to produce more precise estimates than smaller samples.



Precision

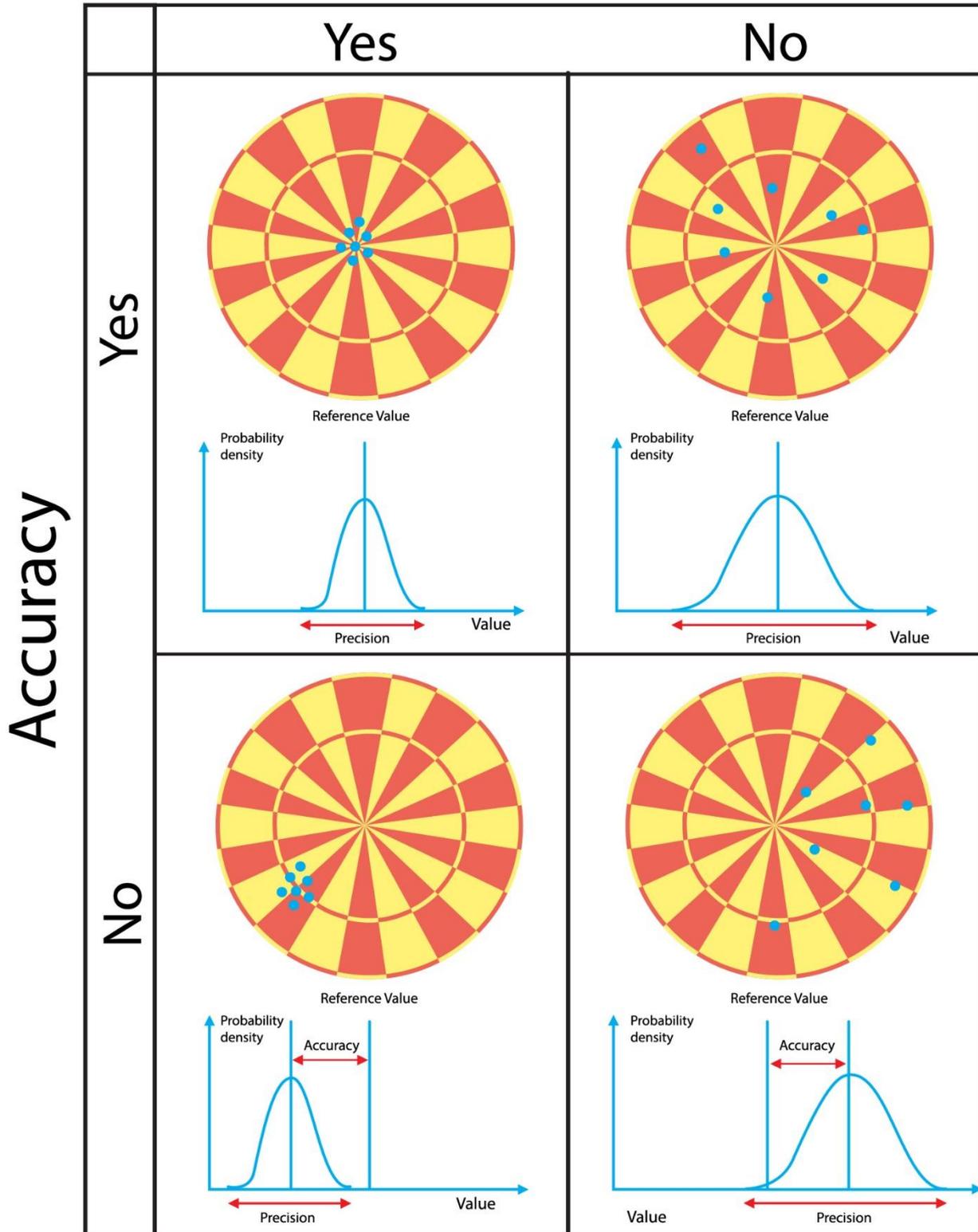


FIGURE 2: ACCURACY VS. PRECISION. BY ARBECK [CC BY 4.0 [HTTPS://CREATIVECOMMONS.ORG/LICENSES/BY/4.0](https://creativecommons.org/licenses/by/4.0)], FROM WIKIMEDIA COMMONS.

Random Probability Sampling

BARC India employs random probability sampling for its selection of panel households. A probability sample is one in which:

- a. Every sampling unit in the sampling frame has a known probability of selection; and
- b. The probability of selection for every sampling unit is greater than zero³.

The above is to say that every listed address in India has a known - and nonzero - chance of being selected for recruitment to the BARC India television panel.

A probability sample is very different from a non-probability, or convenience, sample where only certain sections of the population are included. In these cases, it is often difficult to understand which segments of the population might be missing and therefore entirely possible that changes on the ground may not be reflected in the population. An example would be opt-in samples where individuals join a panel through unprompted choice. It is often impossible to know what are the latent variables surrounding the choice of joining the panel, and therefore, cannot be determined how that sample might change reflective to the ground. In this example, opt-in could be through downloading a particular application on a smartphone. If the appeal of the

downloaded application is tied to a systematic bias, the sample may not behave in the same way as the general Indian population.

In its purest form, sampling can be administered through a process known as *Simple Random Sampling (SRS)* where every sampling unit, or in the case of BARC, an address, has an equal probability of selection. That is to say, of the approximately 197 million television households in India, every household would have a probability of being selected for recruitment equal to roughly 1/197 million - or 0.00000005%.

The goal of sampling is to select a sample that is representative of the population. BARC, therefore, aims to have a panel which is a microcosm of India. Unfortunately, random samples can lead to errors in which the sample selected does not align with the population. This deviation is what is known as sampling error. This phenomenon can be illustrated

³Goodman, R., & Kish, L. (1950). Controlled selection – A technique in probability sampling. *Journal of the American Statistical Association*, 45(251), 350-372

through an example where a sample of four cards is randomly drawn from a deck of cards to estimate the percentage of Clubs within the deck. In this example, an ideal

sample would have precisely one Club - leading to an estimate of 25%, or 13 of the 52 cards. However, this will only happen for 43.9% of the times (Table 1).

Table 1
Probability of Drawing Clubs in a Four Card Hand

Number of Clubs	Probability
0	30.4%
1	43.9%
2	21.3%
3	4.1%
4	0.3%
Total	100.0%

A probability sample like this brings two significant advantages:

- a. The most probable outcome is the correct outcome, a hand with a single Club - occurring 43.9% of the time; and
- b. Due to the known probabilities, we can mathematically calculate a confidence interval around any of the possible estimates - allowing us some insight into the precision of our estimate.

In the above example, our expected value - or most likely outcome - is a hand with a single club. In this case, our estimate matches perfectly with the population - 1/4 of the deck being Clubs. While this perfect case is only expected to happen 43.9% of the times, we see that in 30.4% + 43.9%

+ 21.3% = 95.6% of the times, the resulting four card hand either perfectly matches the population (i.e., one Club), or only over- or under-states by a single Club. Deviances greater than one card (i.e., 3 or 4 Clubs in a hand) occur less than 1 out of 20 times.

Stratified Random Probability Sampling

Sampling accuracy can be improved (i.e., reduce sampling error) by employing sophisticated sampling processes and techniques such as stratification. In the case of stratification, the population is split into non-overlapping segments (i.e., strata) before sampling. These segments should have some degree of homogeneity within while having some degree of heterogeneity between segments. A random sample is then chosen from each segment.

To illustrate this, we can use the following example. Table 2 shows a scenario where we would like to sample from three classrooms in a school to estimate the proportion of Males within the school. Despite the school being 50% Male and 50% Female, the Male/Female ratio varies dramatically between classroom. Two approaches could be taken: (a) the school could be sampled as a single unit; or (b) the school could be stratified by classroom with a sample being taken from each class.

Table 2
School Population by Class and Sex

Count (Column%)	Classroom A	Classroom B	Classroom C	Total
Males	5 (25.0)	10 (50.0)	15 (75.0)	30 (50.0)
Females	15 (75.0)	10 (50.0)	5 (25.0)	30 (50.0)
Total	20 (100.0)	20 (100.0)	20 (100.0)	60 (100.0)

A sample of six is drawn. In the first approach, all six are sampled from the overall pool of sixty students - in other words, a Simple Random Sample is drawn. In the second approach, two are sampled from each class of 20. This latter procedure is known as Stratified Random Sampling. By comparing the results from the two methods, it is seen that the likelihood of obtaining an entirely representative sample (i.e., 3 out of the 6 sampled being

male) is higher in the Stratified Random Sample (Table 3). The likelihood of more extreme samples (i.e., the number of males being 0, 1, 5, or 6) also decreases by 5.0 percentage points. The result is a reduction in the variability of the samples, with the variance in possible outcomes reducing by 16.7% (i.e., decreases from a variance of 1.5 to 1.25). This phenomenon is well visualized by comparing the two probability distributions. (Figure 3)

Table 3
Probability of various outcomes for the number of males sampled in a sample of six

Number of Males in Sample	Approach 1: Simple Random Sampling	Approach 2: Stratified Random Sampling	Difference in Probability (pp)
0	1.6%	0.9%	-0.7
1	9.4%	7.6%	-1.8
2	23.4%	24.1%	+0.7
3	31.3%	34.8%	+3.5
4	23.4%	24.1%	+0.7
5	9.4%	7.6%	-1.8
6	1.6%	0.9%	-0.7
Total	100.0%	100.0%	0.0%

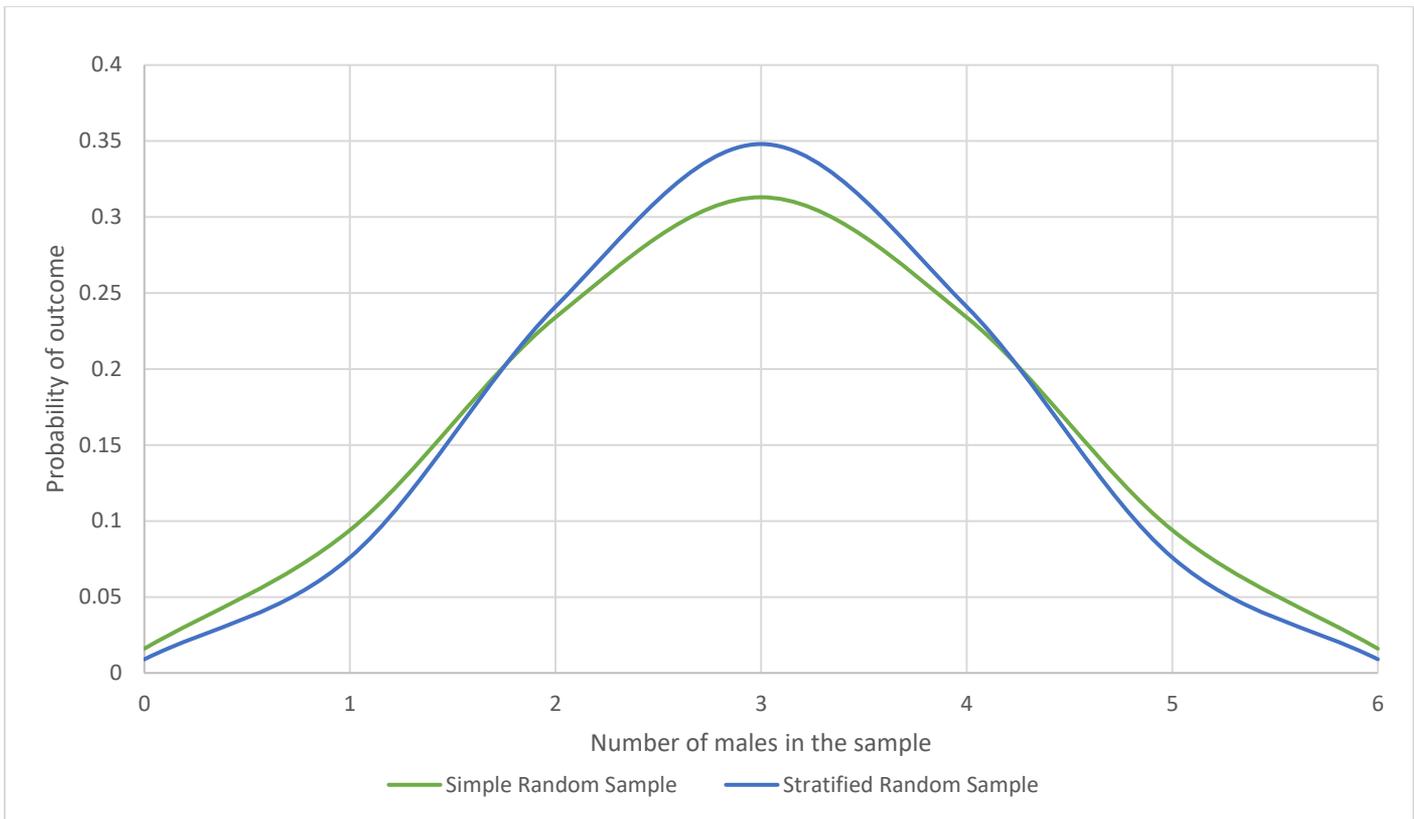


FIGURE 3: PROBABILITY DISTRIBUTIONS OF SAMPLE OUTCOMES BY SAMPLING APPROACH



As is demonstrated in the example above, Stratified Random Sampling scores over Simple Random Sampling. Therefore, by stratifying the sample, one can better control the possible sample outcomes, thereby ensuring a higher likelihood of a more representative sample and lower relative errors associated with the audience estimates. To capitalize on this phenomenon,

BARC India utilizes a sophisticated sample design and sampling procedure for the management of their television viewing panel. BARC India stratifies the panel against three primary control variables and four secondary control variables (Table 4). These seven variables have been identified as having the highest impact on television viewing behavior.

Table 4
BARC India stratification variables

Primary Control Variables	Secondary Control Variables
<ul style="list-style-type: none"> • State Group • Town Class • NCCS 	<ul style="list-style-type: none"> • Household size • Languages spoken at home + Language most often spoken at home • Education of the highest educated individual in the households • Mode of signal reception (MOSR)

By controlling the sampling processes in such a way, BARC India can increase the likelihood that the panel remains representative of the Indian TV owing population - thereby minimizing relative errors and improving the precision of television audience estimates. The panel sampling procedures and panel representativeness has been audited by CESP, a global audit company specializing in audience measurement and research audits, and is found to be at least on par, if not exceeding, with global standards.

Do Samples Capture on the Ground Behavior?

Due to their dynamic nature, real-time ground level changes like NTO cannot be factored in during the sampling process. However, the multiple control variables do ensure that the sample remains largely representative. To illustrate this, let's add another variable: *subjects chosen by students*, to the school example stated earlier. The number of students choosing Science, Math, and History varies significantly between classrooms and gender.

Table 5 shows a scenario where we would like to sample from three classes in a school to estimate the proportion of students choosing Math within the school.

A sample of twelve is drawn. In the first approach, all twelve are sampled from the overall pool of sixty students - in other words, a Simple Random Sample is drawn.

In the second approach, four are sampled from each class of 20. As we have seen in the earlier illustration, this procedure is known as Stratified Random Sampling.

In the third approach, two boys and two girls are sampled for each class of 20. This procedure is Stratified Random Sampling with two variables. By comparing the results from the three methods, it is seen that the likelihood of obtaining an entirely representative sample is higher in the Stratified Random Sample with two variables (Table 6). This phenomenon is well visualized by comparing the three probability distributions (Figure 4).



Table 5
School Population by Class and Sex and Subjects Chosen

Count		Subjects Chosen							
		<u>S</u> <u>M</u> <u>H</u>	<u>S</u> <u>M</u>	<u>S</u> <u>H</u>	<u>M</u> <u>H</u>	<u>S</u>	<u>M</u>	<u>H</u>	<u>Total</u>
Classroom A	Male	1	1	1	0	1	1	0	5
	Female	1	1	3	3	1	1	5	15
	<i>Total- Classroom A</i>	2	2	4	3	2	2	5	20
Classroom B	Male	2	2	1	0	2	1	2	10
	Female	1	2	1	3	2	1	0	10
	<i>Total- Classroom B</i>	3	4	2	3	4	2	2	20
Classroom C	Male	3	3	0	4	2	3	0	15
	Female	1	0	2	0	1	0	1	5
	<i>Total- Classroom C</i>	4	3	2	4	3	3	1	20
Total Classrooms	Male	6	6	2	4	5	5	2	30
	Female	3	3	6	6	4	2	6	30
	<i>Total</i>	9	9	8	10	9	7	8	60

S = Science; M = Math; H = History



Table 6

Probability of various outcomes for the number of students choosing the subject Math in a sample of twelve

<u>Took Math</u>			
Number of Students in Sample	Approach 1: Simple Random Sampling	Approach 2: Stratified Random Sampling	Approach 3: Stratified Random Sampling (2 variables)
0	0.0%	0.0%	0.0%
1	0.1%	0.0%	0.0%
2	0.5%	0.4%	0.2%
3	2.2%	1.9%	1.3%
4	6.3%	6.0%	5.1%
5	13.3%	13.2%	13.1%
6	20.3%	20.8%	22.2%
7	22.7%	23.5%	25.4%
8	18.6%	18.9%	19.4%
9	10.8%	10.5%	9.7%
10	4.2%	3.8%	3.0%
11	1.0%	0.8%	0.5%
12	0.1%	0.1%	0.0%
<i>Total</i>	<i>100.0%</i>	<i>100.0%</i>	<i>100.0%</i>



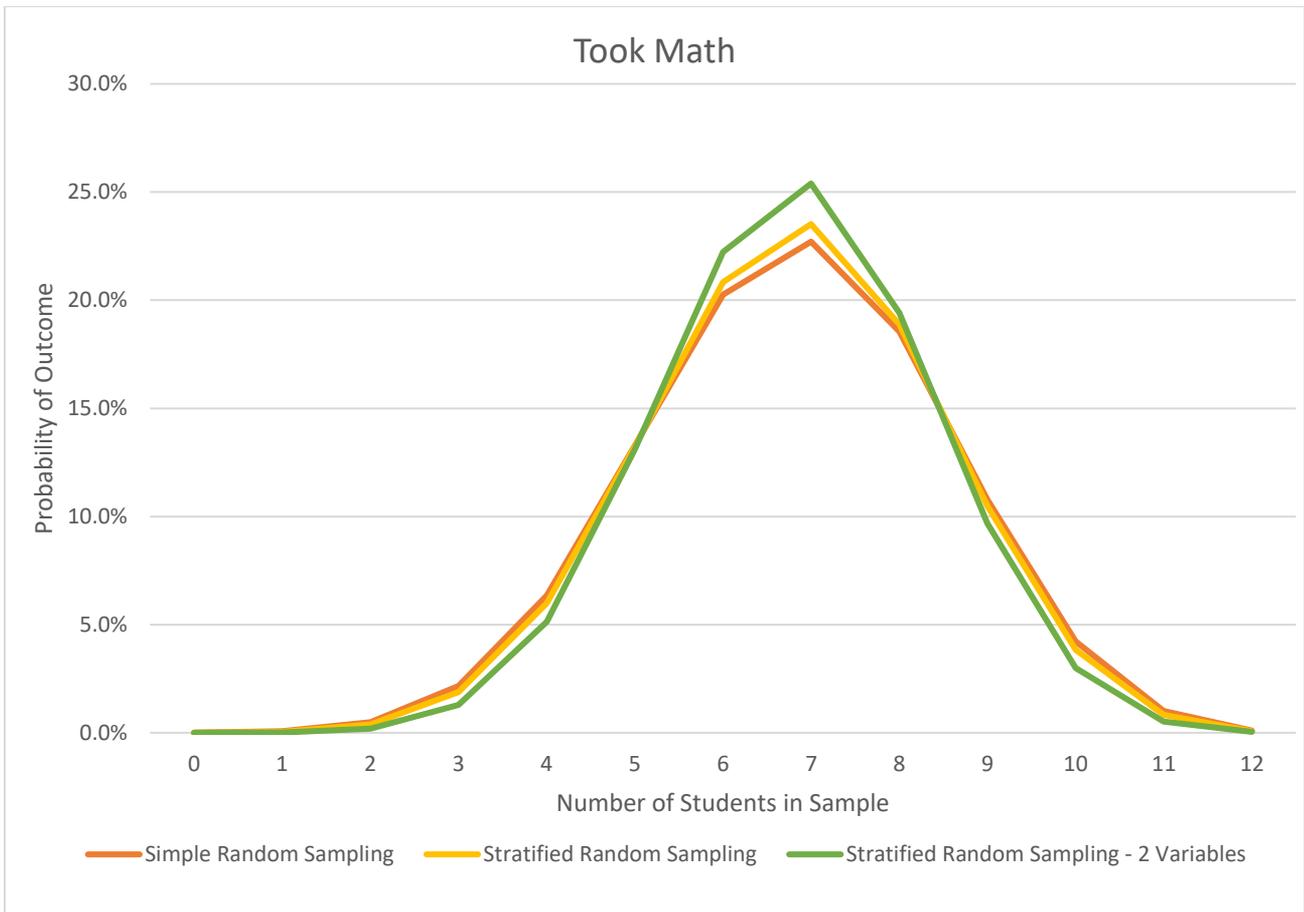


FIGURE 4: PROBABILITY DISTRIBUTIONS OF SAMPLE OUTCOMES BY SAMPLING APPROACH

We have thus seen how with the right sampling process, the sample continues to remain representative despite changes in the universe.



How Does a Random Sample Perform Under Pressure of Increased Heterogeneity?

To further better understand how random samples can capture the on-ground behavior, we can extend the above examples to sampling households with subscriptions to particular channels. Assume there is a channel which 5% of the households have chosen to subscribe. We can then create a data set of 197 million households where exactly 5% have subscribed to the channel - this data set forms our population. A random sample of 50,000 households can be drawn from the population, and the percentage of households in the sample subscribing to the channel can be analyzed. By repeating this sampling process multiple times – e.g. 1,000 times – we can view the behavior of sampling under this scenario. This approach is known as Statistical Simulations.

Statistical Simulations are a widely accepted means of assessing the performance of a method. They bring a particular advantage as they allow the statistician to control various inputs to understand how the method may react under different scenarios. In this case, the simulations provide an understanding of the sampling distributions under multiple scenarios.

Similar analyses using Statistical Simulations are found in many peer-reviewed academic journals.

To illustrate the impact of declining availability of a channel in the population and its effect on samples, a statistical simulation was conducted against a population with a channel availability in 5.00%, 2.50%, 1.00%, 0.50%, 0.25% and 0.01% of households.

Various measures of central tendency (i.e., mean, mode, median) were calculated across the 1,000 samples at each population availability level as well as the 10th and 90th percentiles.

For each of the simulations, all three measures of central tendency either perfectly match or are very close to the population proportion (Table 5). This result suggests that the samples on average are highly representative of the sample regardless of the availability of the channel since this phenomenon remains consistent for all simulated channel penetrations - from a high of 5% to a low of 0.01%.

When viewing the 10th and 90th percentiles of the 1,000 samples (i.e., the lowest and highest estimated proportions - occurring 20% of the time), the range decreases relative to the proportion. In the case of a population proportion of 5.00%, the mean of the samples was 5.00% with the 10th and 90th percentiles being 4.88% and

5.13% respectively. This finding means that 80% of the samples had a proportion of households receiving the channel within 13 basis points of the actual population value. As the population proportion decreases, that range reduces, reaching a range of 0.01 percentage points for a population proportion of 0.01%.

Table 7
Simulation results

Population Proportion	Mean of samples	Mode of samples	Median of samples	10 th percentile of samples	90 th percentile of samples
5.00%	5.00%	4.97%	5.00%	4.88%	5.13%
2.50%	2.50%	2.47%	2.50%	2.42%	2.59%
1.00%	1.00%	0.99%	1.00%	0.95%	1.05%
0.50%	0.50%	0.50%	0.50%	0.46%	0.54%
0.25%	0.25%	0.25%	0.25%	0.22%	0.28%
0.01%	0.01%	0.01%	0.01%	0.00%	0.02%

The above simulation demonstrates the effectiveness of random sampling, even in the case of niche or low availability channels. In each of the case, the sample effectively captures the necessary number of households with access to the channel. This phenomenon was viewed in over 6,000 independent samples. All simulated cases used Simple Random Sampling - the most basic form of sampling. Therefore, results for the BARC panel - which uses a far more sophisticated sampling procedure - can only be more precise.

Does BARC India's Panel Continue to be Robust Post the New Tariff Order?

Sample surveys are a widely used technique to understand the characteristics of a population adequately. Samples offer a cost-effective and operationally-effective means of capturing information such as television viewing. While bringing many advantages, a sample - and the information it provides - is only as good as the accuracy and precision with which it reflects the population. By using techniques such as probability sampling, the possible degree of error associated with a viewing audience can be quantified and thereby understood. Through this, it can be seen that estimates closer to reality are more likely, and extreme estimates - while possible - are far less likely. There are many sampling techniques - such as those employed by BARC India - that make precise viewing estimates much more likely.

The TRAI NTO has increased the fragmentation of television viewing. The availability, and thereby possible reach, of individual channels has thus been reduced. It is natural to question how such a shift in the ecosystem could impact sampling and thereby impact television estimates.

The above examples help understand that a correctly controlled sample can indeed mirror reality.

There are also other factors which can help ensure that BARC India's panel households reflect the reality of the ground such as panel rotation. Due to various panel rotation factors (i.e., panel churn, forced turnover), BARC India is continuously recruiting new households into the panel. Each of these new households are randomly sampled from the ground under new ground realities thereby allowing the panel to naturally evolve with the on-ground changes. Historically too, BARC India's panel has stood the test of ground level changes. Case in point being- the true reflection of Digitization implementation delays in DAS I, DAS II, DAS III, and DAS IV areas in the BARC Panel data.



This goes to support that despite changes in the distribution ecosystem, BARC India's television viewing estimates continue to be robust and precise, thanks to the robustness of the sampling methodology and process. The Indian television and advertising industries can remain equally confident in the quality of BARC data post-NTO as pre-NTO. BARC continues to deliver "What India Watches" effectively.

In its journey of continuous improvement, BARC India has commissioned the next Broadcast India Study which determines the number of television owning households in the country and captures any change in the factors determining television viewing. The panel will undergo a change basis the findings of the study.

We continue to report What India Watches.