NATIONAL DRUG SURVEY 2014 - 2016

Chapter 6

Design

DESIGN

Design of a sample Survey is crucial for the successful conduct of Survey. There are many instances where sampling has to be done in the absence of complete information on the sampling frame, that is, when all the population units are either unknown or the effort needed to list all units is practically impossible or cost prohibitive. The Drugs Survey is one such instance where the sampling frame cannot be known at the time of planning.

Proper identification of the population and providing a pragmatic and meaningful statistical design for estimating the parameters of interest is important. This is the Second National Level Drugs Survey. The earlier one was conducted in 2009. The 2009 Survey experience has led to significant improvements in the design of the present Survey (1). In the present Survey, the main parameters of interest are the proportion of Drugs that are Not of Standard Quality (NSQ) and the proportion of Spurious Drugs.

SOURCE

It is important to assess possible difficulties/hurdles that may be faced while implementing any action plan. For this reason, a pilot study was conducted to understand the efforts needed in implementing the Survey and identify modifications required in the planned Drug Survey. This Survey is aimed at estimating parameters at three different Sources. The word "Source" is used to denote:

- Retail Outlet
- Government distribution points (State government Medical Store Depots, Civil/District Hospital Stores, CGHS dispensaries and ESI dispensaries)
- Ports (Air/Sea)

Of these, the first two are similar in their nature and Drugs were in their finished goods form whereas in case of imported Drugs, the majority of Drugs were in the form of Active Pharmaceutical Ingredients (APIs) and small quantity of finished formulations. Therefore, a separate sampling procedure was considered necessary for imported Drugs.

POPULATION

In the present Survey, as mentioned above the two sources - Retail Outlets and Government distribution points - were similar in their nature. There were about 7,00,000 retail outlets and about 10000 Government distribution points. As far as the imports stream was concerned, samples were to be drawn from import consignments from 8 airports/sea ports.

Molecules

The scope of the Survey is limited to either single ingredient molecules or fixed dose combinations falling under 15 therapeutic categories. In all, 224 molecules were identified for this Survey. For technical presentation, these molecules were coded as M1 to M224. Some of the molecules and their IDs are presented in Exhibit 6.1. For notational convenience, we shall use Mj for jth molecule, j = 1, 2,……, 224.

Exhibit 6.1

Drug Molecule	Drug Molecule ID
Paracetamol	M178
Amikacin	M013
Olanzapine	M172
Dextromethorphan	M069
Oxytocin	M176
Cephalexin	M041

Formulations

Under each molecule, there are a number of generic as well as branded Drugs available in different strengths and different dosage forms such as tablets, syrups, capsules, injections etc. These Drugs were referred to as formulations under the respective molecule. Based on the survey objectives, it was discussed and decided that two formulations were considered distinct if any of the following conditions hold:

- Brand names of the two formulations are different
- Strength/concentration of the two formulations are different
- Batch numbers of the two formulations are different

- Manufacturers of the two formulations are different
- The two formulations are obtained from two different Sources

All distinct formulations available (in adequate quantity for performing all tests) at all Sources at the time of drawing samples are considered as the population units. Clearly, it is not possible to know beforehand (at the time of planning) which formulations will be available at a Source at the time of drawing samples. In other words, population is not known at the time of selection of samples.

Quality of Drugs

All distinct formulations available (in adequate quantity for performing all tests) at all Sources at the time of drawing samples are considered as the population units. As per the Survey procedure, formulations were to be drawn from Sources according to a statistical design and the sampled formulations were to be subjected to physical inspection and laboratory tests. Based on physical inspection and subsequent verification, each sampled formulation was to be classified as Spurious or not. Further, based on laboratory tests formulations were to be classified as Spurious/NSQ or not.

Survey Objectives

The primary objective of the Survey is to ascertain the extent of NSQ Drugs prevailing in the country. The word "extent" refers to quantity or percentage. However, it is not feasible to assess the quantity of NSQ Drugs prevailing in the country because, much of the finished products are in the inventories of dealers and clearing and forwarding agents, and not with the Sources. Sources can replenish their stock at short notices by just making a phone call. As a result of this, Sources maintain limited stocks depending upon the demand pattern. It is also possible that part of a formulation available at a Source is NSQ and the rest of it is of standard quality. Since the sample drawn will be only a part of the stock available at the Source, it is not possible to assess from the sampling at Source, the quantity of the formulation that is NSQ. Since, consumers get their medicines from Retail Outlets and Government distribution points, it makes sense to assess the quality and quantity of medicines available in the Sources. Hence, this survey is designed to estimate the extent of Drugs declared Spurious/NSQ in the sampled Drugs available at the Sources at the time of sampling and the scope is limited to the Drugs under the list of selected 224 single ingredient molecules or their fixed dose combinations.

Coverage Issues and Decisions

The availability of Drugs under a molecule depends on the disease burden in a particular area and cost of the Drugs under that molecule. For example, Drugs under paracetamol will be available in large quantities at most of the Sources, whereas Drugs under molecules used for cancer will be available in low quantities. Due to lack of prior information and the knowledge of the availability of formulations under different molecules and the complications involved in prioritizing the molecules, it was decided that equal importance would be given to all 224 molecules.

The primary objective was to assess the quality of Drugs at the national level. In India, there are 676 districts falling under 29 States and 7 Union Territories. It is known that Sources vary in their sizes. Here "size" refers to the variety and quantities of stocks available at a Source. No prior information was available about the size of Sources. In fact, getting the list of registered Outlets in the country itself turned out to be a gigantic task. It was also desired that in the total samples drawn, there should be at least one Source for each of the two streams (Retail Outlets and Govt. distribution points) from each district.

Based on practical considerations, it was decided that once a Source was visited for sample drawing, six formulations should be drawn from that Source. To maximize the coverage of molecules, it was further decided that the six formulations should be drawn from six different molecules.

STATISTICAL DESIGN

In many Surveys, the complete sampling frame is not known at the time of planning the Survey design. For example, in a socio-economic Survey of a region to estimate the average monthly income of households, the list of households is generally not available (2). In cases where the complete sample frame is not available, a common approach adopted is the multi-stage sampling. Multi-stage sampling is very useful and has been used in a number of Surveys. Lahiri [1954] discussed the use of multi-stage sampling in the Indian National Sample Survey (3); Roy and Singh used multi-stage sampling in estimation of variance components (4-6) to estimate the cost of road repairs on highways.

The Drug Survey is a typical problem for the application of multi-stage sampling. Unlike some of the Surveys cited above, the population in the Drug Survey is a dynamic one. That is, the availability of formulations at the Sources depends on the time at which the Source is visited for collection of samples.

Population Units

In order to determine whether a formulation is NSQ or not, certain minimum quantity is required. For instance, if the formulation is in tablets form, then 70 tablets are needed to perform all the tests required to determine the Spurious and NSQ characteristics of the Drug. For this reason, even when a formulation is available at a Source, it will not be considered as a population unit if the Source does not have adequate quantity for performing all the tests. Therefore, the population units for the Survey are taken as only those formulations under the 224 selected molecules available in adequate quantity at the Sources at the time of visiting Sources for sample collection.

Sampling Scheme

Given the constraints mentioned above, it was not possible to list the population units and draw a sample. The only way was to draw a random sample of Sources from the list of Sources and then from each selected Source, draw a random sample of formulations. Since it was required to draw samples from the molecules giving uniform priority to all 224 molecules, we had to select molecules first and then from the selected molecules select the formulations randomly. Thus, the sampling scheme involved a 3-stage sampling:

- **First Stage:** Sources were selected,
- **Second Stage:** molecules were selected from the selected Sources
- **Third Stage:** formulations were selected from the selected molecules

An important contribution of this work is in designing the sampling scheme so that all molecules are treated uniformly. One way of doing this is to list all molecules with at least one formulation with adequate quantity available at the Source selected, and then select six molecules from among them (if the available number is less than six, then select all the available ones). But this would be a laborious and time consuming activity. To avoid this problem and make the sampling procedure practically effective a new scheme was devised and is described in the next paragraph. The survey was conducted separately for each stream using this scheme.

Consider the list of registered Outlets in the country, say, N in number. We used index i for the ith Outlet, i = 1, 2,..., N. Let mi be the number of molecules (from the list of 224 molecules) with at least one formulation (with adequate quantity for performing all tests) available in the Outlet i at the time of sample drawing. We used the index j for denoting the molecule ID. Let m_{ij} be the

number of formulations with adequate quantity under jth molecule of ith Outlet. We used the index k for the formulation number.

For ease of presentation, we shall introduce the following notation. Let Γ denote the set of all population units. A population unit $l \in \Gamma$ can be identified with a 3-dimensional integer vector (i, j, k) which stands for k^{th} formulation under j^{th} molecule available in the Outlet *i* at the time of Survey, k = 1, 2,. . . ., m_{ij}, j= 1, 2, 3 …..m_i , i = 1, 2, 3. . ., N. Note that the size of the population is $M = \sum_{i=1}^{N} \sum_{j=1}^{m_i} m_{ij}$ and we may write $\Gamma = \{1, 2, \ldots, M\}$. As our interest was the proportion of NSQ, for each $l \in \Gamma$ we associate a number Y_i which is 1 if formulation l is NSQ and 0 otherwise. Let $Y = \sum_{l \in \Gamma} Y_i$. Then $p = \frac{Y}{M}$ is the proportion of NSQ.

First stage of sampling

For the present sampling scheme, select n_O Outlets from N outlets using simple random sampling without replacement.

Second stage of sampling

Once outlets are selected, select molecules from each of the selected outlets. Suppose Outlet i is selected. According to our assumption, there are mi molecules at Outlet *i*. Let t_i = min (6, m_i). The decision is to draw six formulations from six different molecules. If t_i is less than 6, then only t_i molecules can be selected. Select t_i molecules randomly from m_i molecules giving equal opportunity to each of the m_i molecules.

Third stage of sampling

Suppose molecule j is selected from i^{th} outlet. There are m_{ij} formulations under this molecule. Select a formulation from the m_{ij} formulations giving equal opportunity to each of the m_{ij} formulations.

We denoted the set of population units selected in the sample by S and its size by n. Thus, S corresponds to n formulations and is a subset of Γ.

Note that m_i will be known when the Outlet i is visited, and m_{ij} will be known when the molecule j is selected from the i^{th} Outlet.

Implementing the Sampling Scheme

Implementing sampling scheme required identifying all the molecules at each sampled Outlet before selecting the molecules. This was a very cumbersome exercise and practically very difficult to implement. Actually, a pilot study was designed and carried out to assess the possible difficulties

and the extent of effort required in drawing the samples. We devised a mechanism to overcome the problem of listing all available molecules at a given Source. The procedure for this was as follows:

Let $\pi = (\pi 1, \pi 2, \pi 3, ..., \pi 224)$ be a random permutation of (1, 2, . . ., 224). For each Outlet, one such permutation is randomly and independently selected from the set of all possible permutations with uniform probability $\left(\frac{1}{224!}\right)$. Consider a sampled outlet whose permutation is $\pi = (\pi 1, \pi 2, \pi 3, ..., \pi 224)$. The procedure given below was adopted to draw samples from this outlet. Unless stated otherwise, by a "Formulation" we mean a formulation with adequate quantity for performing all tests.

Step 1. Ask if the outlet has any formulation(s) under $M_{\pi 1}$. If the answer is yes, then select one formulation from $M_{\pi 1}$ randomly giving equal probability to every formulation under $M_{\pi 1}$ as described earlier. If the outlet does not have formulations under $M_{\pi 1}$ to sample, explore the availability of $M_{\pi 2}$.

Step 2. After exploring $M_{\pi 1}, M_{\pi 2}, \ldots, M_{\pi i}$, $1 \leq i < 224$ explore $M_{\pi(i+1)}$. If the outlet has formulation(s) under $M_{\pi(i+1)}$, then select one formulation from $M_{\pi(i+1)}$ randomly giving equal probability to every formulation under $M_{\pi(i+1)}$. If the Outlet does not have formulations under $M_{\pi(i+1)}$, then go to the next formulation in the sequence.

Step 3. Repeat Step 2 until six formulations are sampled or all molecules are exhausted. That is, explore molecules as per the sequence π and stop as soon as six formulations are sampled.

This procedure avoided listing all molecules which would have been a cumbersome exercise. However, this procedure was equivalent to listing all molecules in the Outlet and choosing $t_i = \min(6, m_i)$ from them giving equal probability to all the m_i molecules. Clearly, this procedure improved the efficiency of sampling as there was a better scope for verifying availability of formulations under the molecules as the availability was verified molecule by molecule one at a time.

ESTIMATION OF PARAMETERS

The main parameters of interest in this Survey are the proportions of Spurious and NSQ formulations. Since we adopted the three stage sampling for selection of samples, it is necessary to carefully look into the estimation of parameters, namely the proportions of Spurious and NSQ formulations. The common sense estimator uses the simple sample averages as estimators. That is, if the total number of formulations drawn is n, and the number of NSQ formulations found

in them is s, then one would tend to use $100\ s/n$ as the estimator for proportion NSQ. In general, this turns out to be a biased estimator. However, when the sample sizes are large, which is the case in this drugs survey, the bias tends to be negligible. This aspect was studied at length and verified that the biases of the common sense estimator are negligible.

The survey design captured two parameters in the second and third stage of the design. In the second stage, the serial number v is captured at each source (outlets / govt. distribution points). This number v is the serial number in the list of 224 molecules generated for the source (using Part-C of data form) at which the sample for the sixth molecule was obtained. If v is equal to 224, then one would know exactly how many molecules from the list of 224 molecules were there at the source. As the sampling procedure was developed to avoid exploring all the molecules at the sources, one would not know how many molecules were there in the sources. This parameter is estimated from the serial number v. It may be noted that v is a number that lies between 6 and 224 for any source. If v is large, it means that one has to explore large number of molecules in the list provided for the respective source (Part-C of the data form) which in turn would indicate that the source had small number of molecules from the list to sample. Thus, larger the v, smaller is the number of molecules in the source. When v is less than 224, then the number of molecules is estimated as 1344/v.

Since the common sense estimator is biased we proposed an alternative estimator. This estimator for proportion NSQ uses the number of molecules in the sources. Consequently, we considered three estimators for estimating the proportion of NSQ/spurious drugs. These are $\hat{q}$, the common sense estimator, $\hat{p}_2$, our proposed estimator when the exact number of molecules at each sampled source is determined, and $\hat{p}$, our proposed estimator when the number of molecules is estimated using v mentioned in the previous paragraph. The formulae for these three estimators are given below:

$$\hat{q} = \frac{\text{Number of NSQ formulations in the sample}}{\text{Number of formulations in the sample}}$$

$$\hat{p}_2 = \frac{n_o \sum_{ijk \in S} m_i m_{ij} Y_{ijk}}{(\sum_{i:ijk \in S} m_i)(\sum_{ij:ijk \in S} m_{ij})}$$

$$\hat{p} = \frac{n_o \sum_{ijk \in S} \hat{m}_i m_{ij} Y_{ijk}}{(\sum_{i:ijk \in S} \hat{m}_i)(\sum_{ij:ijk \in S} m_{ij})}$$

where n_o is the number of sources in the samples, m_i is the number of molecules in the j^{th} source, $\hat{m}_i$ is the estimate of m_i estimated using v mentioned above, m_{ij} is the number of formulations under the j^{th} molecule of i^{th} outlet of the

sample, and y_{ijk} is 1 if sampled formulation ijk is NSQ, 0 otherwise, and S stands for the collection of samples ijk.

It may be noted that the estimator $\hat{p}_2$, cannot be used for the results of this survey because the m_i s are not determined at the sources. However, we used this estimator in the simulation exercises (see the next paragraph) we conducted to evaluate the other two estimators $\hat{q}$ and $\hat{p}$.

Using simulation exercises, we simulated a number of populations with specified NSQ proportions ranging from 3% to 16%, with varying molecule and formulation numbers. For each of these populations, we applied the 3-stage sampling design with varying sample sizes ranging from 1000 to 6000 (number of sources) and estimated the NSQ proportions using the three estimators $\hat{q}$, $\hat{p}_2$ and $\hat{p}$, and compared them with the actual (specified) NSQ proportions to assess the biases. The results of the simulation exercises are summarized in Exhibit 6.2. From the results of the simulation experiments, we find that, though the common sense estimator is biased (statistically significant), the bias is practically negligible. At a sample size of 1000 sources, the two estimators $\hat{q}$ and $\hat{p}$ are both equal to 0.03 and the actual population proportion is also 0.03. Therefore, the performance of common sense estimator is quite satisfactory under the proposed design.

While the proportions are estimated separately for each stream (retail outlets, government sources and ports), one would like to see the overall picture. That is, one wants to know what is the NSQ/spurious proportions in the three streams put together. Since the population sizes are different, then one should use the weighted means for estimating the overall proportions. This is illustrated as follows. Let the number of formulations in the populations of three streams be m_R, m_G and m_P for retail outlets, government sources and ports respectively. If $\hat{p}_R$, $\hat{p}_G$ and $\hat{p}_P$ are the estimates of proportion for retail outlets, government sources and ports respectively, then an estimate of overall proportion is given by ($m_R \hat{p}_R + m_G \hat{p}_G + m_P \hat{p}_P$ are the estimates of proportion for retail outlets, government sources and ports resp)/($M_R + M_G + M_P$). In this survey, we estimated M_R, M_G and M_P using the serial number v mentioned earlier.

Sample Size

Before studying the properties of the estimators proposed above, the sample sizes considered for the Survey were needed to be figured out. In determining the sample sizes, two aspects were considered - coverage of the population and the standard error of the estimator. The primary interest was in estimation of percentage of NSQ Drugs and percentage of Spurious Drugs. Estimator

was to be obtained for each Source - the Retail Outlets and the Government distribution points separately. Though estimates were required at the national level, it was desired that at least one representation should be there from each district.

One broad guideline for determining the sample size was to examine the standard error using simple random sampling approach. Suppose p is the proportion of NSQ Drugs in the population and a sample of size n was taken using simple random sampling (with or without replacement does not affect estimation error when the sample size is negligible compared to population size (10) and estimate p using the simple average $\sum_{(ijk)\in s} Y_{ijk}/n$, where S is the set of sampled formulations and n is the number of formulations in S. Then the standard error of this estimator is equal to $\sqrt{\frac{p(1-p)}{n}}$. Error of estimation can be taken as plus/minus two times the standard error. Since p is unknown, we can assess the error of estimation by varying p in its possible range. The possible range for p was taken to be 0.0001 to 0.15 and the error of estimation was computed. Exhibit 6.3 presents the error in terms of proportions for n = 1000, 2000, 4000, 6000 and p in the range of 0.0001 to 0.15. Based on these observations and cost and effort considerations, it was decided that 6000 retail outlets and 1500 Government distribution points should be surveyed. Since the plan was to collect six formulations from each sampled outlet, the targeted number of formulations for retail outlets was 36000. Though the targeted size was 36000 samples, the actual figure was likely to be smaller as some of the outlets may not be accessible (either outlets do not exist or closed at the time of Survey or the number of molecules available at the Outlets is less than six). This aspect was also taken into account while deciding on the sample size.

Exhibit 6.2
Bias and Standard Errors of Estimators of p and q

Sample Size	NSQ Proportion								Chance of buying NSQ			
			$\hat{p}$			$\hat{p}_2$			$\hat{q}$			
	Expected	Actual	Average	Std. Error	Bias	Average	Std. Error	Bias	Actual	Average	Std. Error	Bias
1000	0.03	0.030	0.030	0.0038	0.000	0.029	.0037	-0.001	0.030	0.030	0.002	0.0000
	0.07	0.070	0.066	0.0061	-0.004	0.067	0.0057	-0.003	0.070	0.070	0.000	-0.0002
	0.10	0.100	0.096	0.0062	-0.004	0.096	0.0063	-0.004	0.100	0.100	0.004	0.0002
	0.13	0.130	0.125	0.0077	-0.005	0.124	0.0078	-0.006	0.130	0.129	0.005	-0.0008
	0.16	0.160	0.154	0.0085	-0.006	0.155	0.0074	-0.005	0.160	0.160	0.005	0.0001
2000	0.03	0.030	0.029	0.0023	-0.001	0.029	0.0025	-0.001	0.030	0.030	0.002	-0.0002
	0.07	0.070	0.068	0.0045	-0.003	0.067	0.0039	-0.003	0.070	0.070	0.003	-0.0001
	0.10	0.100	0.096	0.0051	-0.004	0.096	0.0045	-0.004	0.100	0.100	0.003	0.0000
	0.13	0.130	0.126	0.0057	-0.004	0.124	0.0046	-0.006	0.130	0.130	0.003	0.0002
	0.16	0.160	0.153	0.0058	-0.007	0.154	0.0061	-0.006	0.160	0.160	0.004	0.0001
4000	0.03	0.030	0.029	0.0020	-0.001	0.029	0.0017	-0.001	0.030	0.030	0.001	0.0001
	0.07	0.070	0.068	0.0032	-0.002	0.067	0.0028	-0.003	0.070	0.070	0.002	0.0001
	0.10	0.100	0.096	0.0030	-0.004	0.096	0.0032	-0.004	0.100	0.100	0.002	0.0001
	0.13	0.130	0.125	0.0034	-0.005	0.124	0.0039	-0.006	0.130	0.130	0.002	0.0001
	0.16	0.160	0.153	0.0044	-0.007	0.153	0.0039	-0.007	0.160	0.160	0.003	0.0000
6000	0.03	0.030	0.029	0.0016	-0.001	0.029	0.0017	-0.001	0.030	0.030	0.001	0.0000
	0.10	0.100	0.096	0.0032	-0.004	0.096	0.0026	-0.004	0.100	0.100	0.002	-0.0002
	0.13	0.130	0.125	0.0032	-0.005	0.124	0.0030	-0.006	0.130	0.130	0.002	-0.0002
	0.16	0.160	0.154	0.0034	-0.006	0.153	0.0029	-0.007	0.160	0.160	0.002	-0.0003
	0.07	0.070	0.067	0.0026	-0.003	0.067	0.0023	-0.003	0.070	0.070	0.002	-0.0001

Exhibit 6.3
Calculation of Standard Errors

n=1000		n=2000		n=4000		n=6000	
P	2 X SE	P	2 X SE	P	2 X SE	P	2 X SE
0.0001	0.001	0.000	0.000	0.0001	0.000	0.0001	0.000
0.0005	0.001	0.001	0.001	0.0005	0.001	0.0005	0.001
0.0010	0.002	0.001	0.001	0.0010	0.001	0.0010	0.001
0.0100	0.006	0.010	0.004	0.0100	0.003	0.0100	0.003
0.0500	0.014	0.050	0.010	0.0500	0.007	0.0500	0.006
0.1000	0.019	0.100	0.013	0.1000	0.009	0.1000	0.008
0.1500	0.023	0.150	0.016	0.1500	0.011	0.1500	0.009
	0.023		0.016		0.011		0.009

Note: n = sample size; p = proportion of NSQ formulations; SE = Standard Error.

SAMPLING FROM PORTS

With regard to the stream of imports, much of the Drugs are received in the form of APIs. To a small extent, Drugs are also received in the form of finished formulations. The Source for drawing samples was the import consignments. That is, samples were to be drawn as soon as the consignments were received. Preliminary analysis of the past data revealed the following observations:

1. By and large, the imported medicines (APIs) came in the powder form packed in bags of 25 Kgs or 50 Kgs.

2. It may not be possible to draw samples at ports due to vulnerability of Drugs getting contaminated. Therefore, samples had to be drawn at the location/destination of consignments.

3. A total of 887 consignments were received over a period of six months during September 2014 to March 2015 from the ports. Of these 511 were from airports and 376 were from sea ports. Exhibit 6.4 and 6.5 highlight the summary of past data on number of consignments.

4. The variation in the quantity of each consignment was high. The summary statistics for the quantity are presented in Exhibit 6.6 and 6.7. About 80% of the consignments had the quantity of less than 5,000 Kgs. The average quantity per month was 6,834 Kgs and the median was 610 Kgs.

Exhibit 6.4
Number of consignments from all ports

Sep	Oct	Nov	Dec	Jan	Feb	Mar
5	109	171	183	143	116	160

Exhibit 6.5
City-wise number of consignments

Mumbai	Chennai	New Delhi	Hyderabad	Ahmedabad	Kolkata
619	125	87	47	8	1

Exhibit 6.6
Descriptive Statistics: QUANTITY

Month	N	Mean	Q1	Median	Q3
Dec	182	4960	50	500	1741
Feb	116	2646	64	500	2000
Jan	143	5056	100	610	1600
Mar	160	3858	150	1000	3000
Nov	170	10931	300	1000	4250
Oct	109	14565	50	250	3000
Sep	5	10515	513	1000	25275

Exhibit 6.7
Descriptive Statistics (Over 6 Months): QUANTITY

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
QUANTITY	885	6834	25177	0	100	610	2500	326213

It was clear from the past data that the scope for drawing big number of samples did not exist. Assuming that there were about 100 consignments per month, it was expected to get about 200 samples over a period of two months. The standard error (possible error in the estimate), against different NSQ percentages is shown in Exhibit 6.8 for two sample sizes, namely, 200 and 300 consignments. The values in the table below can be interpreted as follows:

Suppose the true NSQ percentage was 9% and sample size was 200 consignments. Then it can be expected with a high probability that the difference between the estimated NSQ percentage and the true NSQ percentage (9%) will be at most 4. In other words, the estimated NSQ percentage could be expected to be within 5 to 14% when the true NSQ percentage was 9%. If sample size was 300 consignments, then this estimated percentage could be between 6 and 12%, a marginal reduction.

Exhibit 6.8
Estimation of Errors

For sample size 200				
Percentage NSQ	2%	4-6%	8-10%	12-16%
Error of estimation	2%	3%	4%	5%
For sample size 300				
Percentage NSQ	2-4%	6-10%	12-16%	
Error of estimation	2%	3%	4%	

In order to see the internal consistency, that is, the consistency of quality within a consignment, it was decided to take multiple samples from each consignment. Keeping practical constraints and the size variation of quantity of each consignment in view, it was considered to be appropriate to take three samples from each consignment. Since consignments come in packets of size 25 or 50 kgs (in majority of the cases), it was decided to take a sample of three packets (or bags) using simple random sampling from each consignment.

Procedure for Drawing Samples from Imports

In view of the above analysis, it was proposed to take three representative samples from each consignment (of the listed 224 molecules) from all the ports for two or three months for the purpose of this Survey. Further, with regard to drawing of samples, it was noted that sampling may not be done in some cases due to the nature of the Drugs (e.g. sterile Drugs). These difficulties and accordingly the procedure for drawing samples from imports are summarized below:

1. To draw sample from every consignment imported through notified ports in respect of finished Drugs formulations and APIs pertaining to the list of 224 molecules identified for the Drugs Survey. This was to be done for a period of three months.

2. Most of the consignments received were in the form of APIs and the type of packaging is either plastic containers or polyethylene bags of 25 kg or 50 kg medicine in powder form. Occasionally the imports were in the form of finished formulation. Medicine in API form is of two types: sterile and non-sterile. For the non-sterile APIs, the samples could be drawn at the ports at the time of receiving the consignments. The sampling of sterile APIs cannot be done at the ports. The sampling had to be done by taking the containers to the manufacturer's site and thereafter draw the samples under sterile conditions.

3. If the consignment was that of finished goods, then a sample with adequate quantity or minimum quantity were drawn as in the case sampling from Retail Outlets /Government distribution points. But the difference here was that a sample was to be drawn from each formulation of each molecule provided that the molecule was in the list of the 224 molecules. If a formulation had different batch numbers, then one sample was drawn from each batch.

4. If the consignment was in the form of APIs, a sample was drawn from each API. Thus, for instance, if there were 10 different APIs from the list of 224 molecules in one consignment, then 10 samples of APIs were drawn. If there were different batches within an API, then one sample was drawn from each API batch. The following method was used for drawing a sample from an API:

Procedure for drawing a sample from an API: For the purpose of this sampling, it was assumed that the product (API) is packed in containers. If the API was in different forms or in different batches, then samples were drawm in a manner so that that there was one sample from each form and each batch. A container could be a plastic container or it could be a bag of 25 or 50 kgs. Once the containers of an API of a particular form and particular batch were identified, the containers were numbered as 1, 2,….. n, where n was the total number of containers. Three random numbers were drawn from 1 to n without replacement. A mobile app was provided for drawing these random numbers. If the random numbers were i, j and k, then adequate sample quantity was taken from each of the containers that were numbered as i, j and k respectively.

SUMMARY

In this chapter, the problem of designing the national Drug Survey to estimate the proportion of Drugs which are not of standard quality has been discussed. The same procedure was also to be applied for estimating the proportion of Spurious Drugs. The problem was challenging because the sampling frame was complex and could not be known completely. Since, we were interested

in estimating the proportion, we needed to estimate both population total and the population size. Adopting a 3-stage sampling procedure, we designed a special sampling scheme and proposed an estimator for estimating the population proportion. A significant contribution of this work was that our sampling scheme drastically reduced the sampling effort.

REFERENCES

(1) Murthy, G. S. R., and A. L. N. Murthy [2009] Determination of sample size and samples for estimation of counterfeit Drugs, Technical Report, Indian Statistical Institute, Hyderabad, 2009.

(2) Murthy, M. N. Sampling: theroy and methods, Statistical Publishing Society, Calcutta (1967) pp.318.

(3) Lahiri,D.B. Technical Paper on Some aspects of the development of sample design, Sankhya,14 (1954), pp.264-316.

(4) Roy, J. A note on estimation of variance components in multistage sampling with varying probabilities, Sankya,17 (1957) pp.367-372.

(5) Singh, D. Estimates of variance components in nite population, J. In. Soc. Agr. Stat., (1958) pp.10-15.

(6) Adams, T. M. Estimating cost per lane mile for routine highway operations and maintenance, Project report, University of Wisconsin, Madison, (2011).