Abstract
Once the research question and the research design have been finalised, it is important to select the appropriate sample for the study. The method by which the researcher selects the sample is the ‘Sampling Method’. There are essentially two types of sampling methods: 1) probability sampling – based on chance events (such as random numbers, flipping a coin etc.); and 2) non-probability sampling – based on researcher’s choice, population that accessible & available. Some of the non-probability sampling methods are: purposive sampling, convenience sampling, or quota sampling. Random sampling method (such as simple random sample or stratified random sample) is a form of probability sampling. It is important to understand the different sampling methods used in clinical studies and mention this method clearly in the manuscript. The researcher should not misrepresent the sampling method in the manuscript (such as using the term ‘random sample’ when the researcher has used convenience sample). The sampling method will depend on the research question. For instance, the researcher may want to understand an issue in greater detail for one particular population rather than worry about the ‘generalizability’ of these results. In such a scenario, the researcher may want to use ‘purposive sampling’ for the study.

Key Words: Non-probability sampling, sampling strategies, probability sampling

Introduction
The purpose of this section is to discuss various sampling methods used in research. After finalizing the research question and the research design, it is important to select the appropriate sample for the study. The method by which the researcher selects the sample is the “Sampling Method” [Figure 1].

Why do we need to sample?
Let us answer this research question: What is the prevalence of HIV in the adult Indian population?
The best response to this question will be obtained when we test every adult Indian for HIV. However, this is logistically difficult, time consuming, expensive, and difficult for a single researcher – do not forget about ethics of conducting such a study. The government usually conducts an exercise regularly to measure certain outcomes in the whole population – “the census.” However, as researchers, we often have limited time and resources. Hence, we will have to select a few adult Indians who will consent to be a part of the study. We will test them for HIV and present out results (as our estimates of HIV prevalence). These selected individuals are called our “sample.” We hope that we have selected the appropriate sample that is required to answer our research question.
The researcher should clearly and explicitly mention the sampling method in the manuscript. The description of these helps the reviewers and readers assess the validity and generalizability of the results. Furthermore, the authors should also acknowledge the limitations of their sampling method and its effects on estimated obtained in the study.

Types of Methods
We will try to understand some of these sampling methods that are commonly used in clinical research. There are essentially two types of sampling methods: (1) Probability sampling – based on chance events (such as random numbers, flipping a coin, etc.) and
(2) nonprobability sampling – based on researcher’s choice, population that accessible and available.

**What is a “convenience sample?”**

Research question: How many patients with psoriasis also have high cholesterol levels (according to our definition)?

We plan to conduct the study in the outpatient department of our hospital.

This is a common scenario for clinical studies. The researcher recruits the participants who are easily accessible in a clinical setting – this type of sample is called a “convenience sample.” Furthermore, in such a clinic-based setting, the researcher will approach all the psoriasis patients that he/she comes across. They are informed about the study, and all those who consent to be the study are evaluated for eligibility. If they meet the inclusion criteria (and need not be excluded as per the criteria), they are recruited for the study. Thus, this will be “consecutive consenting sample.”

This method is relatively easy and is one of the common types of sampling methods used (particularly in postgraduate dissertations).

Since this is clinic-based sample, the estimates from such a study may not necessarily be generalizable to the larger population. To begin with, the patients who access healthcare potentially have a different “health-seeking behavior” compared with those who do not access health in these settings. Furthermore, many of the clinical cases in tertiary care centers may be severe, complicated, or recalcitrant. Thus, the estimates of biological parameters or outcomes may be different in these compared with the general population. The researcher should clearly discuss in the manuscript/report as to how the convenience sample may have biased the estimates (for example: Overestimated or underestimated the outcome in the population studied).

**What is a “random sample?”**

A “random sample” is a probability sample where every individual has an equal and independent probability of being selected in the sample.

Please note that “random sample” does not mean arbitrary sample. For example, if the researcher selects 10–12 individuals from the waiting area (without any structure), it is not a random sample. Randomization is a specific process, and only samples that are recruited using this process is a “random sample.”

**What is a “simple random sample?”**

Let us recruit a “simple random sample” in the above example. The center only allows a fixed number of patients every day. All the patients have to confirm the appointment a day in advance and should present in the clinic between 9 and 9:30 a.m. for the appointment. Thus, by 9:30 a.m., you will all have all the individuals who will be examined day.

We wish to select 50% of these patients for posttreatment survey.

Steps:
1. Make a list of all the patients present at 9:30 a.m.
2. Give a number to each individual
3. Use a “randomization method” to select five of these numbers. Although “random tables” have been used as a method of randomization, currently, many researchers use “computer-generated lists for random selection” of participants. Most of the statistical packages have programs for random selection of population. Please state the method that you have used for random selection in the manuscript
4. Recruit the individuals whose numbers have been selected by the randomization method.

The process is described in Figure 2.

**What is a major issue with this recruitment process?**

As you may notice, “only males” have been recruited for the study. This scenario is possible in a simple random sample selection.

This is a limitation of this type of sampling method – population units which are smaller in number in the sampling frame may be underrepresented in this sample.

**What is “stratified sample?”**

In a stratified sample, the population is divided into two or more similar groups (based on demographic or clinical characteristics). The sample is recruited from each stratum. The researcher may use a simple random sample procedure within each stratum.

Let us address the limitation in the above example (selection of 50% of the participants for postprocedure survey).
Steps:
1. Make a list of all the patients present at 9:30 a.m.
2. Divide the list into two strata: Males and females
3. Use a “randomization method” to select three numbers among males and two numbers among females. As discussed earlier, the researcher may use random tables or computer generated random selection. Please state the method that you have used for random selection in the manuscript.
4. Recruit the individuals whose numbers have been selected by the randomization method.

The process is described in Figure 3.

Thus, with this sampling method, we ensure that people from both sexes are included in the sample. This type of sampling method is used for sampling when we want to ensure that minority populations (in number) are adequately represented in the sample.

Kindly note that in this example, we sampled 50% of the population in each stratum. However, the researcher may oversample in one particular stratum and under-sample in the other. For instance, in this example, we may have taken three females and three males (if want to ensure equal representation of both). All this should be discussed explicitly in methods.

What is a “systematic sample?”
Sometimes, the researcher may decide to include study participants using a fixed pattern. For example, the researcher may recruit every second patient, or every patient whose registration ends with an even number or those who are admitted in certain days of the week (Tuesday/Thursday/Saturday). This type of sample is generally easy to implement. However, a lot of the recruitments are based on the researcher and may lead to selection bias. Furthermore, patients who come to the...
hospital may differ on different days of the week. For example, a higher proportion of working individuals may access the hospital on Saturdays.

This is not a “random sample.” Please do not write that “we selected the participants using a random sample method” if you have selected the sample systematically.

Another type of sampling discussed by some authors is “systematic random sample.” The steps for this method are:
1. Make a list of all the potential recruits
2. Using a random method (described earlier) to select a starting point (example number 4)
3. Select this number and every fifth number from this starting point. Thus, the researcher will select number 9, 14, and so on.

Please note that the “skip” depends on the total number of potential participants and the total sample size. For instance, you have a total of fifty potential participants and you wish to recruit ten participants, do not skip to every 10th patient.

Aday (1996) states that the skip depends on the total number of participants and the total sample size. For instance, you have a total of fifty potential participants and you wish to recruit ten participants, do not skip to every 10th patient.

In the above example, it will be 50/10 = 5
Thus, using a random table or computer-generated random number selection, the researcher will select a random number from 1 to 5
The number selected in two
The researcher selects the second patient
The next patient will be the fifth patient after patient number two – patient number 7
The next patient will be patient number 12 and so on.

What is a “cluster sample?”

For some studies, the sample is selected from larger units or “clusters.” This type of method is generally used for “community-based studies.”

Research question: What is the prevalence of dermatological conditions in school children in city XXXXX?

In this study, we will select students from multiple schools. Thus, each school becomes one cluster. Each individual child in the school has much in common with other children in the same school compared with children from other schools (for example, they are more likely to have the same socioeconomic background). Thus, these children are recruited from the same cluster.

If the researcher uses “cluster sample,” he/she also performs “cluster analysis.” The statistical methods for these are different compared with nonclustered analysis (the methods we use commonly).

What is a “multistage sample?”

In many studies, we have to combine multiple methods for the appropriate and required sample.

Let us use a multistage sample to answer this research question.

Research question: What is the prevalence of dermatological conditions in school children in city XXXXX? (Assumption: The city is divided into four zones).

We have a list of all the schools in the city. How do we sample them?

Method 1: Select 10% of the schools using “simple random sample” method.

Question: What is the problem with this type of method?

Answer: As discussed earlier, it is possible that we may miss most of the schools from one particular zone.

However, we are interested to ensure that all zones are adequately represented in the sample.

Method 2:
1. Stage 1: List all the schools in all zones
2. Stage 2: Select 10% of schools from each zone using “random selection method” (first stratum)
3. Stage 3: List all the students in Grade VIII, IX, and X (population of interest) in each school (second stratum)
4. Stage 4: Create a separate list for males and females in each grade in each school (third stratum)
5. Stage 5: Select 10% of males and females in each grade in each school.

Please note that this is just an example. You may have to change the proportion selected from each stratum based on the sample size and the total number of individuals in each stratum.

What are other types of sampling methods?

Although these are the common types of sampling methods that we use in clinical studies, we have also listed some other sampling methods in Table 1.

Conclusion

• It is important to understand the different sampling methods used in clinical studies. As stated earlier, please mention this method clearly in the manuscript
• Do not misrepresent the sampling method. For example, if you have not used “random method” for selection, do not state it in the manuscript
• Sometimes, the researcher may want to understand an issue in greater detail for one particular population rather than worry about
the “generalizability” of these results. In such a scenario, the researcher may want to use ‘purposive sampling’.

Financial support and sponsorship
Nil.

Table 1: Some other types of sampling methods

| Type of Method                  | Description                                                                                                                                                                                                 |
|--------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Purposive sampling             | Often used in qualitative studies (such as those conducting Focus Group Discussions (FGD) and In-depth Interviews). The researcher selects the researcher subjects purposefully to answer the specific research questions. For example, FGD of Care-givers of HIV infected individuals In-depth interviews of Medical Officers of Civil Hospitals to understand the issues of drug shortage in these hospitals. |
| Voluntary Sampling             | This is also a non-probability form of sampling. Here the study is advertised and individuals who volunteer to participate in the study will be recruited. This type of sample is used for some socio-behavioural, psychological, and even pharmacological studies. It is important to remember that volunteers may not necessarily be representative of the population. For example, in opinion and behaviours surveys people who volunteer may have strong opinions. |
| Quota Sampling                 | This is a non-probability form of sampling. In this type of sample, we select designated number of individuals based on study requirements. For example, 10 women aged 20-25 in upper socio-economic status and 10 women aged 20-25 in middle SES, and 10 women aged 20-25 in lower SES. |
| Snowball sampling              | This is used to access hard to reach or marginalised populations. In this method, the researcher identifies on individual from this group. This individual then identifies some more individuals with similar characteristics and refers them to the researcher. For example, The researcher wants to study HIV prevalence in People Who Inject Drugs (PWID). The researcher does not know how to reach the population, since the population is mostly hidden. Thus, the researcher identifies one PWID. This PWID agrees to be a part of the study and refers two other PWIDs (since this individual knows other PWIDs) to the researcher. Only one PWID agrees to be a part of the study. This recruited PWID refers two more PWIDs for this study. |
| Respondent Driven Sampling (RDS) | This type of sampling was described by Douglas Heckathorn (1997). This is relatively new method of sampling. It combines ‘snowball sampling’ and mathematical modeling. It is very specific method and the researcher is expected to use the analyses methods for RDS [http://www.respondentdrivensampling.org/] We encourage the readers to familiarise themselves with this method by reading the details in the above mentioned website. |

Conflicts of interest
There are no conflicts of interest.

Bibliography

1. Aday LA. Designing & Conducting Health Surveys. 2nd ed. San Francisco, USA: Jossey-Bass; 1996.
2. Atkinson R, Flint J. Accessing Hidden and Hard-to-Reach Populations: Snowball Research Strategies; 2001. Available from: http://srhu.soc.surrey.ac.uk/SRU33.pdf. [Last accessed on 2016 Aug 07].
3. Heckathorn DD. Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. Social Problems 1997;44:174-99.
4. Heckathorn DD. Snowball Versus Respondent-Driven Sampling. Sociol Methodol 2011;41:355-66.
5. Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. Designing Clinical Research. 2nd ed. Philadelphia, USA: Lippincot Williams & Wilkins; 2001.
6. Miller RL, Brewer JD. The A-Z of Social Research a Dictionary of Key Social Science Research Concepts. London, Thousand Oaks, New Delhi: Sage; 2003.
7. Patton MQ. Qualitative Research and Evaluation Methods. London, Thousand Oaks, New Delhi: Sage; 2002.
8. Porta M. A Dictionary of Epidemiology. 6th ed. New York, USA: Oxford University Press: 2014.
9. Statistics Canada. Non-probability sampling; 2013. Available from: http://www.statcan.gc.ca/edu/power-pouvoir/ch13/nonprob/5214898-eng.htm. [Last accessed on 2016 Aug 14].
10. Statistics Canada. Probability Sampling; 2013. Available from: http://www.statcan.gc.ca/edu/power-pouvoir/ch13/prob/5214899-eng.htm. [Last accessed on 2016 Aug 14].
11. Shott S. Statistics for Health Professionals. Philadelphia, USA: W.B. Saunders Company; 1990.
12. Vogt WP. Dictionary of Statistics and Methodology: A Nontechnical Guide for the Social Sciences. London: Sage; 1999.
13. Respondent Driven Sampling, http://www.respondentdrivensampling.org/. [Last accessed on 2016 Aug 15].