Research Fundamentals: Study Design, Population, and Sample Size

Umair Majid, MSc [1] [2] [3]*
[1] Editorial and Advisory Board Member, URNCST Journal, Toronto, Ontario, Canada
[2] Curriculum Designer, Program Developer and Instructor, McMaster University, Hamilton, Ontario, Canada
[3] Director, The Methodologist (TMT), Mississauga, Ontario, Canada

*Corresponding Author: majidua@mcmaster.ca

Abstract
This is the second article of a three-part series that continues the discussion on the fundamentals of writing research protocols for quantitative, clinical research studies. In this editorial, the author discusses some considerations for including information in a research protocol on the study design and approach of a research study. This series provides a guide for undergraduate researchers interested in publishing their protocol in the Undergraduate Research in Natural and Clinical Sciences and Technology (URNCST) Journal.

Keywords: protocol; proposal; study design; population; study setting; sampling; sample size; undergraduate research; clinical research

Introduction
This article continues the discussion on writing a research protocol for a quantitative, clinical research study. In the second part of a three-part series, the author examines the components of a study’s design and approach to research inquiry. These components are the type of design, population of interest, study setting, recruitment, and sampling.

Study Design
The study design is the use of evidence-based procedures, protocols, and guidelines that provide the tools and framework for conducting a research study. The choice of the study design is a methodological decision made by the investigators before submitting the study for ethics review and starting data collection.

The study design is related to the philosophical orientation of the study and researcher because philosophical “assumptions drive methodological decisions” [1]. The study design is also a consequence of the research question, research objectives, phenomena of interest, population, and sampling strategies [2]. These components are integrated in such a way that their communion often suggests the nature of the study to be conducted. The nature of how these components align stems from the coherent narrative of the topic being studied, starting from pre-existing literature, to the rationale for the study, study approaches, the proposed study findings and the implications of those findings on principles and praxis. In clinical and epidemiological research, this narrative is constrained within a handful of study designs and approaches that can be meaningfully employed. These are classified into two categories: observational and experimental study designs.

Observational designs do not involve the overt manipulation or management of variables. Examples of these designs include cohort, case-control, and cross-sectional [3]. In observational studies, the investigators observe the context, environment, and behaviours in the real-world without participation or manipulation [4]. The Canadian Longitudinal Study on Aging (CLSA) is an excellent example of a prospective cohort study [5]. This research study is following approximately 50,000 men and women between the ages of 45 and 85 for at least 20 years to gather data on medical, psychological, and social factors relevant to how aging, disability, and disease affect Canadians. Figure 1 helps to visualize the differences between simple observational designs on a timeline.

Experimental designs, on the other hand, involve the manipulation and management of variables. Randomized controlled trials (RCTs) are an example of an experimental research study design [6]. In RCTs, investigators modify the levels or exposures of certain variables and observe their effects on clinical outcomes. For example, an RCT can modify the magnetic force exerted by repetitive transcranial magnetic stimulation to determine its effect on pain control in patients with chronic pain [7]. The primary advantage of using an experimental design over observational is that it provides stronger evidence of an association, and potential causality, between outcome and predictor variables through randomization and blinding. Table 1 provides simple
descriptions of different types of observational and experimental study designs.

**Randomization** refers to the process of assigning research participants randomly to either the treatment or control groups to equally distribute the demographic and clinical variables in the study sample [6]. These variables are known as **confounding factors**, and an equal distribution of these variables through randomization would remove their risk of influencing the study [8].

**Blinding**, on the other hand, is a methodological step that prevents research participants and the research team from having prior knowledge about the assignment sequence of research participants [6]. Such knowledge may unduly influence the study results, for example, some research participants who know that they are receiving a placebo treatment may experience worse clinical outcomes. This observation is also referred to as the **placebo effect** [9]. Moreover, some studies have observed a **trial effect** where research participants behave differently due to their involvement in a clinical trial [10].

Table 1: A Description of Basic Types of Observational and Experimental Study Designs

<table>
<thead>
<tr>
<th>Name of Design</th>
<th>Type of Design</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prospective Cohort</td>
<td>Observational</td>
<td>Follow a sample of research participants <em>from today to a later date.</em></td>
<td>[5]</td>
</tr>
<tr>
<td>Retrospective Cohort</td>
<td>Observational</td>
<td>Follow a sample of research participants <em>from an earlier date to today.</em></td>
<td>[11]</td>
</tr>
<tr>
<td>Case-Control</td>
<td>Observational</td>
<td>Study the history of research participants <em>from today to an earlier date.</em></td>
<td>[12]</td>
</tr>
<tr>
<td>Cross-Sectional</td>
<td>Observational</td>
<td>Study the characteristics of research participants <em>today.</em></td>
<td>[13]</td>
</tr>
<tr>
<td>Randomized Trials</td>
<td>Experimental</td>
<td><em>Blinding</em> and randomizing the distribution of variables in a participant sample.</td>
<td>[7]</td>
</tr>
</tbody>
</table>

Figure 1: Depicting the Timelines of Basic Observational Study Designs [14]
Population of Interest

The population of interest is the study’s target population that it intends to study or treat. In clinical research studies, it is often not appropriate or feasible to recruit the entire population of interest. Instead, investigators will recruit a sample from the population of interest to include in their study. In such cases, the objective of the research study is to generalize the study findings from the sample to the population of interest [15].

In a research protocol of a clinical research study, it is important to describe the demographic characteristics of the population of interest including their age, ethnicity, socioeconomic status, education level, marital status, and work status. Reflecting on the characteristics of the “ideal” research participant is an important way to conceptualize the population of interest, eligibility criteria, study setting, and the sampling strategies that will optimize recruitment and retention.

The eligibility criteria determine whether or not an individual is qualified to be a participant in a research study. These criteria are determined a priori to the submission of an ethics application and start of data collection [16]. Eligibility criteria consist of inclusion criteria, which are the main characteristics of the population of interest. A potential research participant has to fulfill all criteria in order to participate in the study. Exclusion criteria, on the other hand, are characteristics that may interfere with data collection, follow-up, and safety of research participants [16]. If a potential participant fulfils any one of the exclusion criteria, then they are excluded from participation. Designing exclusion criteria require investigators to examine the literature on the topic and discern important variables and confounding factors that have shown to interfere with the study plan. Another way to develop exclusion criteria is to use the PICO(TS) components of the research study [2]. For example, in a research study looking at the effect of repetitive transcranial magnetic stimulation on patients with chronic pain, an exclusion criterion may be to exclude individuals who are older than 65 or younger than 20 because they may tolerate pain differently compared to the population of patients between ages 40 and 65. Eligibility criteria are usually formatted in a two-column table with inclusion criteria on the left side and exclusion criteria on the right. This is usually accompanied by a rationale for choosing the inclusion and exclusion criteria, and with the appropriate citing of previous research studies that have utilized similar criteria to guide their study.

Study Setting

The study setting is an important component of a research study. The nature, context, environment, and logistics of the study setting may influence how the research study is carried out. Investigators should record the characteristics, events, gatherings, and other features of a study setting before submitting their study for ethics review and beginning data collection. Observing a study setting before the start of data collection allows investigators to premeditate any practical challenges inherent in the organization, structure or layout of the study setting. In turn, this allows investigators to circumvent these challenges with appropriate strategies that can be included in the ethics applications, funding applications, and research protocols. Showing ethics officers and sponsors that the investigators have taken careful consideration of possible problems and challenges in the study setting or design may increase the likelihood of passing an ethics review and obtaining funding for a research study. Some examples of study locations for clinical research studies are inpatient bedrooms, hospital wards, operating rooms, and rehabilitation clinics.

The characteristics of the study setting deserve a separate section in a research protocol. Information that is pertinent to include in the research protocol about the study setting are the structure, layout, and organization of the setting, rationale for choosing this setting over others, external or online links that describe the setting if available, and any data from the literature on the setting. Keep in mind that a protocol’s discussion on the study setting has to be coherent with other parts of the research protocol. A protocol that appears incoherent is not considered good research practice, and in turn, may become an obstacle to obtaining ethics review and funding.

Sampling

Sampling is the process of selecting a statistically representative sample of individuals from the population of interest [16]. Sampling is an important tool for research studies because the population of interest usually consists of too many individuals for any research project to include as participants. A good sample is a statistical representation of the population of interest and is large enough to answer the research question [17].

In clinical research, there are different strategies that investigators can use to obtain a representative sample from the population of interest [16]. These strategies are referred to as sampling strategies, and the strategy employed in a research study depends on the characteristics of the population of interest, the desired power and significance level (discussed in the next section), and the research question. Table 2 describes some of the most commonly used sampling strategies in clinical research. The benefits and drawbacks of each sampling strategy are beyond the scope of this paper but can be found in other documents and articles published online.
There is a significance, there is a type II error. Type I error is also a binary number without a compound number within. In some cases, the investigators accept the H0. This conclusion does not mean that they are willing to tolerate type I and II errors. In other words, they must establish the thresholds for significance and power for their research study. A Primer to Statistics in Epidemiology

An organized research study contains a good research question and hypothesis. A hypothesis can be simple, comprising of one predictor and one outcome variable, or complex with multiple predictor variables [18]. In the real world, a hypothesis can be true or false, which is determined by the statistical significance of results. When considering the significance, there is a null hypothesis (H0), which assumes that there is no association between the predictor and outcome variables, and the alternative hypothesis (H1), which assumes that there is an association between the predictor and outcome variables. The statistical objective of a research study would be to reject the H0 in favour of the H1.

In some cases, random variations in the sample may yield results that appear statistically significant but do not reflect real associations in the population. When the study findings reflect random variations, then a statistical error has occurred. There are two types of statistical errors that can occur in a research study, which are considered probabilities of making an incorrect conclusion. A type I error occurs when the investigators reject the H0 when it is true in the population of interest. Type I error is also referred to as the level of statistical significance (α). On the other hand, a type II error (β) occurs when the investigator does not reject the null hypothesis when it is untrue in the population of interest. The compound (1 - β) of the type II error is referred to as power, which is the probability of rejecting the null hypothesis given that it is untrue in the population of interest [17].

Example 1: Errors, Significance and Power

Type I Error (α): The probability that the null hypothesis is true, but the investigator incorrectly rejects it

Type II Error (β): The probability that the null hypothesis is untrue but the investigator incorrectly accepts it

Power (1 - β): The probability that the null hypothesis is untrue and the investigator correctly rejects it

The p-value is the probability of obtaining the study results because of random variations in the population of interest. If this probability is small and less than the predetermined significance level (p < α), then the H0 can be rejected in favour of the H1. This conclusion assumes that there is an association that truly reflects the population of interest. On the other hand, if the p-value is higher than the predetermined significance level (p > α), the investigators cannot reject the H0. This conclusion does not mean that the investigators accept the H0 or reject the H1. Instead, it means that the study findings are more likely due to random variations and therefore, may not truly reflect real associations in the population of interest.

### Table 2: Descriptions and Examples of Different Sampling Strategies

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convenience</td>
<td>The most accessible or available participants.</td>
<td>Recruit research participants from the first 40 patients who enter the emergency department at a hospital.</td>
</tr>
<tr>
<td>Simple Random</td>
<td>Everyone in the population of interest has an equal chance to being selected as a research participant.</td>
<td>In a study of older individuals with cardiac concerns in metropolitan Toronto, simple random sampling may occur by contacting all general practitioners in the area, and then selecting participants using a randomization algorithm.</td>
</tr>
<tr>
<td>Stratified Random</td>
<td>Simple random sampling but within predefined subgroups (e.g., nationality).</td>
<td>In a study of the newcomer experiences in housing programs, stratified sampling divides the population of newcomers based on nationality, and perform simple random sampling within each nationality.</td>
</tr>
<tr>
<td>Cluster</td>
<td>Simple random sampling but within naturally occurring subgroups (e.g., patients admitted to different departments of one hospital).</td>
<td>In a study of patient satisfaction in one hospital, cluster sample may include the administration of a survey in different departments of the hospital and comparing the differences between them.</td>
</tr>
</tbody>
</table>
Example 2: P-value
Research Project 1: $\alpha = 0.05$
- $H_0$: No association
- $H_A$: Association
- $p = 0.04$
- $p < \alpha$; reject the $H_0$

The probability of getting the results due to random variation is 4%, which is lower than the predetermined significance level ($\alpha = 0.05$). The results from the sample are unlikely due to random variations in the population of interest. Therefore, reject the $H_0$ in favour of $H_A$.

Research Project 2: $\alpha = 0.05$
- $H_0$: No association
- $H_A$: Association
- $p = 0.10$
- $p > \alpha$; do not reject the $H_0$

The probability of getting the results due to random variation is 10%, which is higher than the predetermined significance level ($\alpha = 0.05$). The results from the sample may be due to random variations in the population of interest. Therefore, do not reject the $H_0$.

Sample Size
One of the objectives of sampling in epidemiological studies is to obtain a statistically representative sample from the population of interest such that the inferences and study findings from the sample represent real associations in the population of interest. The sample size of a research study should have adequate power and significance [22], allowing the investigators to be confident that the study findings cannot be attributed to random variations in the population of interest. In this way, computing the sample size becomes an important step in clinical, quantitative studies.

When computing the sample size of a research study, the first step is to consult a statistician to ensure that the computations use appropriate statistical methodologies. The next step is to set the significance and power levels depending on the characteristics of the research study.

These are usually set to 0.05 and 0.80, respectively [18], however it may differ depending on the discipline, methodology, number of research participants and the research question. The next step is to determine whether the research study needs a one- or two-sided statistical test. Generally, two-sided tests are usually employed because of a statistical uncertainty that the results can go either in the positive or negative direction. For example, after diagnosis of a chronic medical condition, patients may experience an increase or decrease in psychological well-being [23]. However, in studies where there is a logical rationale for the study results to deviate in one or the other direction, then a one-sided test should be used [18]. For example, in a study of the deleterious effects of carbon monoxide exposure on the heart function of infants, the results will be in the negative direction because investigators can assume that no research participant will benefit from carbon monoxide exposure.

The next steps include discerning the types, nature, and quantity of clinical outcomes to be included in the statistical computation of sample size. The investigators need to determine whether or not each clinical outcome follow a normal distribution and if they are binary or continuous. This information is often obtained from previous studies in the same or similar populations of interests or pilot studies on the research question of interest. After making this decision, the investigators determine the size of the difference they hope to detect from their research study by answering:

1) How large of a difference would impact patients’ lives and/or clinical practice?
2) How large of a difference are we expecting from this research study?

Once these considerations are made, the investigators are ready to compute their sample size calculation. Depending on the answers to the questions above, the formula for the sample size will be different [17]. Considering the factors that affect sample size while consulting a statistician is an important step for sample size determination. Some factors that may influence the sample size of a research study are shown in Table 3.

![Table 3: Factors that Affect the Sample Size of a Research Study](image)

**Table 3: Factors that Affect the Sample Size of a Research Study**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Effect on Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decreasing the significance level</td>
<td>↑</td>
</tr>
<tr>
<td>Increasing the power</td>
<td>↑</td>
</tr>
<tr>
<td>Decrease the size of difference to detect</td>
<td>↑</td>
</tr>
<tr>
<td>Higher variability in outcome(s)</td>
<td>↑</td>
</tr>
<tr>
<td>Higher expected loss-to-follow-up</td>
<td>↑</td>
</tr>
<tr>
<td>More than one primary hypothesis</td>
<td>↑</td>
</tr>
</tbody>
</table>

Conclusion
This article continued the discussion on the components of a clinical research protocol. In the second article of a three-part series, the author discussed the characteristics of the study design and approach including the population of interest, study setting, sampling strategies,
and computing the sample size. In the next article, the author will provide some considerations for data extraction, data management and undergoing an ethics review.

Conflicts of Interest
The author(s) declare that they have no conflicts of interest.

Ethics Approval and/or Participant Consent
This study was approved by the University of Somewhere research ethics board. All study participants gave written informed consent to take part in this study.

Authors’ Contributions
UM wrote the entire publication based on his experiences and knowledge in addition to the methodological literature on writing effective manuscripts, proposals and protocols.

Acknowledgements
None.

Funding
This study was not funded.

References


Article Information
Managing Editor: Desmond She
Article Dates: Received Dec 03 17; Published Jan 10 18

Citation
Please cite this article as follows:
DOI Link: https://doi.org/10.26685/urncst.16

Copyright
© Umair Majid. (2018). Published first in the Undergraduate Research in Natural and Clinical Science and Technology (URNCST) Journal. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Undergraduate Research in Natural and Clinical Science and Technology (URNCST) Journal, is properly cited. The complete bibliographic information, a link to the original publication on http://www.urncst.com, as well as this copyright and license must be included.

Do you research in earnest? Submit your next undergraduate research article to the URNCST Journal!
| Open Access | Peer-Reviewed | Rapid Turnaround Time | International |
| Broad and Multidisciplinary | Indexed | Innovative | Social Media Promoted |
Pre-submission inquiries? Send us an email at info@urncst.com | Facebook, Twitter and LinkedIn: @URNCST
Submit YOUR manuscript today at https://www.urncst.com!
INTRODUCING KNOWLEDGE & DISCOVERY

The Official Publication of the URNCST Journal

Want to Learn How to Research in Earnest?
With all the necessary strategies, integrated with relevant and relatable personal experiences, this strategy manual is the essential guide to gaining research experience if you are planning for, or pursuing, your undergraduate education in the sciences.

Beyond Undergraduate
An ideal teaching resource for faculty members and graduate students who teach research methods and mentor undergraduates.

Knowledge and Discovery is comprehensively written to provide you with everything you need to:

✓ Understand the undergraduate research opportunities available to you
✓ Find the ideal supervisor and prepare your best research application (with resume, cover letter, and other written template examples)
✓ Master the interview process (with a list of commonly asked interview questions, and additional preparatory material)
✓ Optimize your research experience, and acquire funding you can use to pay off your undergraduate tuition
✓ Present and publish your research and seek opportunities for advancement
✓ Build your academic network and create professional relationships

Plus… so much more!

Everything you need to know about undergraduate research in 12 comprehensive chapters.

ISBN: 978-1775149101 | Copyright © 2017 by the Undergraduate Research in Natural and Clinical Science Journal