

How to Align Specific Designs of Quantitative Research Studies with Relevant Research Questions and Appropriate Statistics?

Muhammad Saaq

Plastic Surgeon, NIRM Islamabad

ABSTRACT

The Problem and Gap: Design and execution of quantitative research requires aligning specific design elements with robust research question(s) and appropriate statistics. The educational researchers often face lack of guidance in this regard.

The Hook: Adequate know-how of the essential design elements and their appropriate alignment with research question and relevant statistics shall help the researchers to plan and pursue quantitative research more methodically.

Methodology: The existing relevant published literature was searched systematically. The databases and search engines of the PubMed and Google Scholar were searched. The finally included publications were employed to answer the research question posed by the current review.

Results: The most commonly reported quantitative research designs included descriptive surveys, descriptive analytical studies, correlational studies, experimental studies, case-control studies and cohort studies. There was also a plethora of literature on questionnaire development and its validation.

Conclusion: The current review is expected to serve as a comprehensive guide for future researchers in planning and executing their quantitative research more thoroughly and robustly.

Keywords: Quantitative research; Research question; Descriptive statistics; Inferential statistics.

Correspondence:

Dr Muhammad Saaq
Plastic Surgeon, NIRM Islamabad
muhammadsaaq5@gmail.com

Article info:

Received: April 16, 2024
Accepted: August 11, 2024

Funding Source: Nil

Conflict of interest: Nil

Cite this article. Saaq M. How to Align Specific Designs of Quantitative Research Studies with Relevant Research Questions and Appropriate Statistics? J Fed Med Coll 2024. 2024; 1 (1): 28-37.

Introduction

Quantitative research studies represent the most common form of research pursuits which the clinicians as well as basic sciences faculty commonly undertake. The primary focus of these studies is to quantify the collection, analysis and interpretation of scientific data which the doctors routinely come across in their professional practice. These studies typically involve subjects or patients and are based on deductive reasoning with the aim to test a hypothesis. In quantitative studies, data are

collected through measuring things or certain clinical conditions. These measurements are done through the use of questionnaires, proformas or certain machine-based tools. The collected data are subsequently analysed through numerical comparisons and statistical inferences. Reporting of the data is thus performed through statistical means. The researchers ensure quality of rigor and trustworthiness of these studies through the use of randomization, reliability and validity.¹⁻³

There are certain key statistical concepts that underpin the quantitative research. Normal



distribution, and alpha and beta errors are central in this regard. Also, it is crucial to understand which tests are commonly employed for determining the p-values in cases of parametric and non-parametric data.

A normal distribution refers to any observation that will tend to cluster around the mean. It is a symmetrical, bell-shaped probability distribution with mean (it is symbolized by the Greek letter mu " μ ") and standard deviation (it is symbolized by the Greek letter sigma " σ "). If observations follow a normal distribution, the interval ($\mu \pm 2\sigma$) contains 95% of the observations. The major importance of the normal distribution is its role it plays in statistical inference. It forms the basis for such inferences even when the population is not normally distributed. The statistical inference generally involves mean values of a population, not values related to individuals.⁴

Understanding the alpha and beta errors is crucial as these not only impact the hypothesis rejection or acceptance but also the calculation of sample size for a given research study. The α -error is also called type-I error. In a comparative trial, the investigator typically tries to reject the null hypothesis and accept the alternate hypothesis. The probability of rejecting the null hypothesis when it is true is called type 1 error or α -error. So, by definition when there is no difference between the two groups, the probability of incorrectly concluding that there is a difference is called the α -error. The α -error is essentially the p -value and denotes the probability of incorrectly finding a statistically significant effect when the scenario is actually just the opposite. The α -error or p -value serves as the comparison point for statistical tests when determining the significance of the relevant findings (i.e., whether accept or reject the null hypothesis (H_0)). The p -value is often set at 5% ($\alpha = 0.05$). As this value decreases, the sample size needed to detect a significant effect increases.^{4,5}

The β -error is also called type-II error. When there is a difference between two groups, the

probability of incorrectly concluding that there is no difference is called the β error. The β -error is reflective of the statistical power of the study. It is related to the sample size of the study. Therefore, the statistical power is $= 1 - \beta$. Given this fact, studies with small sample size have less power than large studies. So by taking large sample size or combining many small studies, one can increase the overall statistical power of analysis. Ideally, both α and β should be set at zero to eliminate the possibility of false positive and false negative results, however it is practically not possible and so they are made as small as possible.⁴⁻⁷

Determination of p-values is an important aspect of the quantitative research studies. Different tests are employed for determining the p-values for parametric and non-parametric data. The commonly used parametric tests include one way analysis of variance (ANOVA), t-tests, paired t-test, unpaired t-test and Pearson correlation test. The Non-parametric tests include Chi squared test, Mann Whitney U test, Kruskal Wallis test, Wilcoxon Rank Sum test and Spearman correlation test. The parametric tests rely on the statistical distribution in data whereas the non-parametric tests don't rely on any such distribution. The non-parametric tests don't make any assumptions and are primarily aimed at the measuring the central tendency with the median values.^{7,8}

The researchers strive hard to generate results which are authentic, valid and reliable. These features ensure generalizability and general applicability of the of study results. To achieve these goals, the researcher employs sound methodology as well as correct estimation of sample size. There are different authentic formulae and even liberally online calculators which guide this process of sample size calculation. Such calculations take into account some important parameters such as the significance level, effect size, power of the study, and margin of error. A plethora of such formulae

is available to guide the sample size calculation for different types of quantitative study designs.^{1,9}

The current study was undertaken to find the various quantitative research designs and their alignment with appropriate research questions and relevant statistics.

Methodology

- Search strategy:
A systematic search strategy was used to find answer(s) to the following research question: *How to align specific design of quantitative research study with relevant research questions and appropriate statistics?* The key terms were defined and relevant databases were searched to find out the published literature. The search engines included PubMed and Google Scholar. Manual search for relevant other articles was also undertaken.
- Key terms used:
Maximum possible key terms were used for the literature search. The search strings employed for the PubMed were as follows: (quantitative research study* OR quantitative research designs* OR quantitative research*) AND (research question*) AND (statistic*).
- Review period:
The literature published between Jan 01, 1974 to Feb 29, 2024 was reviewed.
- Inclusion criteria:
All publications relevant to the research question, published between 1974 and 2024 were included.
- Exclusion criteria:
The exclusion criteria were the various forms of grey literature and publications in non-English language.
- Studies selected:
Relevant articles were selected through the phases of identification, screening, eligibility determination and final inclusion in the literature synthesis. (Figure-1). The finally included articles were thoroughly reviewed to

find answer(s) to the research question of the current research.

Results & Discussion

In the following discussion, the most common types of quantitative research designs are comprehensively described with easy and practical examples in the form of phrasing an appropriate research question, calculation of the sample size and brief outline of the most relevant statistics necessary for the process of data analysis and drawing logical conclusions.

1. DESCRIPTIVE SURVEYS¹⁰⁻¹³:

i. Research question:

What are the attitudes and practices of the First, Second and Third year MBBS students regarding problem-based learning (PBL) at the Medical College?

ii. Sample size Calculation:

As in this survey we are surveying a finite population (i.e., students of the First, Second and Third year MBBS classes of the particular Medical College), the sample size shall be calculated using the following formula:

$$n = N \times \frac{Z^2 \cdot r(100-r)}{E^2 + Z^2 \cdot r(100-r)}$$

$$x = Z(c/100)2r(100-r)$$

$$E = \text{Sqrt} \left[\frac{(N-n) \cdot x}{n(N-1)} \right]$$

Where:

n is the sample size

N is the total population size

E is the margin of error

r is the fraction of responses that we are interested in (at least 50%)

Z(c/100) is the critical value for the confidence level c.

Taking the total population of the three classes as 300 students, and keeping 95% Confidence

interval (CI) with 5% margin of error, the sample size will be 169 respondents.

iii. Most relevant tests for undertaking statistical analysis:

Descriptive and inferential statistics shall be employed for analyzing the data collected through the survey questionnaires. The analyses shall thus include measuring the response rates, wave analysis for response bias and descriptive analyses of the included items. The research question shall be tested using inferential statistics.

The Descriptive statistics shall cover the following fundamental aspects in particular:

- Descriptive analyses of the data to determine general trends of the responses. The numerical data shall be expressed as mean \pm Standard Deviation (SD) whereas the categorical data shall be expressed as frequency and percentages. These analyses will cover the following aspects:
- Demographic profile of the participants.
- Descriptive statistics (mean, median, variance, and range) for each question.
- Analysis of answers (significant statements and themes) given by the respondents to the descriptive questions that were posed in the questionnaire.

The Advanced statistics/ tests will cover the following aspects:

- Scales shall be developed by combining questions. (i.e., correlate items using the statistical procedure of factor analysis).
- The reliability of the scores on the scales shall be ensured using a coefficient of internal consistency.
- The validity of the scores on scales shall be ensured using factor analysis.
- With inferential statistics, data shall be analyzed (i.e., for comparing subgroups and

relating variables) to address the research question.

2. DESCRIPTIVE ANALYTICAL STUDIES ^{9,10,14-16}.

i. Research question:

What is the diagnostic accuracy of elevated total leucocyte count (TLC) for predicting acute appendicitis among patients presenting with pain of the right iliac fossa?

ii. Sample size Calculation:

For determining sensitivity, the sample size will be calculated using the formula:

$$n S_n = T_+ + F_ / p$$

Where:

$$T_+ + F_ = (Z_{\alpha/2})^2 \times \{ S_n (1 - S_n) \} / E^2$$

T₊ represents true positive,

F₋ represents true negative,

S_n represents sensitivity

Z_{α/2} is the standard value for the corresponding level of confidence. At 95% Confidence Interval, it is 1.96

E is accuracy, usually taken as 0.05.

p is the prevalence of the disease in the test population.

Taking sensitivity as 92% and prevalence as 10%, the sample size calculations would be as follows:

$$T_+ + F_ = (1.96)^2 \times \{ 0.92 (1 - 0.92) \} / 0.05^2 \\ = 3.841 \times 0.0736 / 0.0025 = 113$$

So n S_n = 113/0.10 = 1130.79 or 1131 patients

For determining specificity, the sample size shall be calculated using the formula:

$$n S_p = T_+ + F_ / 1-p$$

Where:

$$T_+ + F_ = (Z_{\alpha/2})^2 \times \{ S_p (1 - S_p) \} / E^2$$

S_p represents specificity

Taking specificity as 43% and prevalence as 10%, the sample size calculations would be as follows:

$$T+ + F- = (1.96) 2 \times \{ 0.43 (1 - 0.43)\} / 0.052$$

$$= 3.841 \times 0.2451 / 0.0025 = 376.57$$

So $n_{Sn} = 376.57 / 1 - 0.10 = 418.41$ or 419 patients

Since we are trying to find out the diagnostic accuracy, we take the higher number of required patients. In the above calculations, we will take 1131 patients as the required sample size for determining the diagnostic accuracy of elevated TLC count for predicting acute appendicitis.

iii. Most relevant tests for undertaking statistical analysis:

Descriptive and inferential statistics shall be employed for data analysis. The numerical data shall be expressed as mean \pm SD while the categorical data shall be expressed as frequency and percentages. **Chi-square test** shall be used to compare percentages and a p-value of less than 0.05 shall be considered statistically significant.

The following **2 x 2 table** shall be employed to determine sensitivity, specificity, positive predictive value and negative predictive value of elevated TLC counts in predicting acute appendicitis.

Table: Leukocyte counts and operative diagnosis among the patients.			
Operative diagnosis	Leukocyte counts		Total
	> 10,000/mm³	< 10,000/mm³	
Appendicitis	A	B	A+B
Normal appendix	C	D	C+D
	A+C	B+D	-

- Sensitivity = $A / A+C \times 100 = X \%$
- Specificity = $D / B+D \times 100 = X \%$
- Positive predictive value (PPV) = $A / A+B \times 100 = X \%$

- Negative predictive value (NPV) = $D / C+D \times 100 = X \%$

3. CORRELATIONAL STUDIES ¹⁷⁻²⁰:

i. Research question:

Are the low entry-test scores of students related to their failure in the First professional MBBS examination at the Medical College?

ii. Sample size Calculation:

Since in this study we are dealing with dichotomous variables (i.e., proportion of low-scorers in the entry tests and proportion of failures in the First professional MBBS exams), the required sample size shall be calculated by using the formula:

$$n = 4 \times Z\alpha^2 \times p (1-p) / w^2$$

Where:

Z α is the confidence level. (i.e., 1.96)

p is the pre-study estimate of the proportion to be measured.

w is the width of the confidence interval (i.e., ± 10)

Taking 95% confidence level, the confidence interval as ± 10 and the p as 20%, the calculations shall be:

$$n = 4 \times (1.96)^2 \times 0.20(1-0.20) / 0.20^2 = 4 \times 3.841 \times 0.16 / 0.04 = 61.456 \text{ students}$$

Hence, for conducting this correlational study, a minimum of 62 students should be recruited.

iii. Most relevant tests for undertaking statistical analysis:

Correlation statistical test (Correlation coefficient):

The correlation matrix of the Pearson coefficients shall be presented. It shall describe and measure the degree of association or relationship between the two variables (or sets of scores for them) under investigation. (i.e., low entry-test scores and failures in the First professional MBBS examination). The direction, form, and strength of

the associations between the scores shall be indicated.

Displays of scores (i.e., scatterplots and matrices):

Displays of scores correlated for participants shall be generated. The scatterplots shall graphically represent the data whereas the correlation matrices shall show the correlation among the variables in tabular form. In the interpretation of correlations, the positive or negative direction of the correlation of scores shall be highlighted. The plot of distribution of the scores shall indicate if they are normally or non-normally distributed. This shall also indicate the degree and strength of association between the scores.

Multiple variable analysis (i.e., partial correlations and multiple regression):

For correlating more than two variables (such as poor educational background, low motivation level, low entry test scores), the effects of the third variable shall be controlled to examine a prediction equation of multiple variables that could explain the outcome.

Regression analysis:

4. EXPERIMENTAL STUDIES^{9,21-23}:

i. Research question:

What is the effect of using the FCR-set (Flexor carpi radialis) versus FCU-set (Flexor carpi ulnaris) of tendon transfers on restoring the lost extension of the wrist and digits in cases of high radial nerve palsy?

ii. Sample size Calculation:

The sample size shall be calculated using the formula

$$n = Z^2 p q / e^2$$

Where:

Z is the standard normal deviate for the 95% CI. It is constant as 1.96 error of deviation for 95% CI.

p is the prevalence of high radial nerve palsy needing tendon transfers, as reported in the published literature. (i.e., 1.42% or 0.0142).

q is 1-p. Here it is 1-0.0142 (i.e., 0.9858)

e is the margin of error. It is set at 5% (i.e., 0.05).

So $e^2 = (0.05)^2 = 0.0025$

So the sample size calculation will be

$$n = Z^2 p q / e^2$$

$$= (1.96)^2 \times 0.0142 \times 0.9858 / 0.0025$$

$$= 3.841 \times 0.0142 \times 0.9858 / 0.0025$$

$$= 21.20 \text{ or } 22 \text{ patients.}$$

Hence, for conducting this randomized control trial, a minimum of 22 patients should be recruited in each group.

iii-Most relevant tests for undertaking statistical analysis:

Descriptive analyses of the participants: Frequencies and percentages shall be used to express the categorical data. The numerical data will be presented as Mean \pm SD.

Group comparison statistics: These shall be employed for comparing the groups in terms of the outcomes. (i.e., restoration of the lost extension of the wrist and digits in the affected upper limb)

To answer the hypotheses, **inferential statistics** shall be used. The percentages of various variables shall be compared by employing the χ^2 test. $P < 0.05$ shall be considered statistically significant.

The statistic tests shall include the **t-test** or the family of parametric analysis of variance statistics (e.g., ANOVA, ANCOVA (analysis of covariance))

5. CASE CONTROL STUDIES^{12,19}:

i. Research question:

Does the use of sedative hypnotics predispose to Colle's fractures of the distal radius among elderly patients?

ii. Sample size Calculation:

Keeping the CI at 95% and 80% power of the study, the sample size will be calculated using the formula:

$$n = \frac{1+r}{r} \times p(1-p) \frac{(Z_{1-\beta} + Z_{1-\alpha/2})^2}{(p_1 - p_2)^2}$$

Where:

n = The required sample size

r = Controls to cases ratio (1 when the numbers are equal in both groups)

p = Proportion of population = (P1+P2)/2.

In this study it is $0.4+0.3/2 = 0.35$

$Z_{1-\beta}$ = It is the desired statistical power. It is 0.84 for 80% power and 1.28 for 90% power). In this study we take it as 0.84.

$Z_{1-\alpha/2}$ = It is the standard value for the corresponding level of confidence.

At 95% CI, it is 1.96 and at 99% CI, it is 2.58. In this study we take it as 1.96.

Taking the expected proportion of use of sedative hypnotics in cases as 40% and in controls as 30%, the

P1 = Proportion in cases. (40%=0.4)

P2 = Proportion in controls. (30%=0.3)

So the calculations:

$$n = \frac{1+1/1}{1} \times 0.35 (1-0.35) \frac{(0.84+ 1.96)^2}{(0.40-0.30)^2}$$

$$= 2 \times 178.36$$

$$= 356.72 \text{ or } 357$$

Considering 10% dropout rate of the study participants, the sample size will become:

$$n=357+36=393 \text{ patients}$$

Hence, for conducting this case control study, a minimum of 393 patients are required in each group of cases and controls.

iii. Most relevant tests for undertaking statistical analysis:

Odds ratio (OR) calculation: The OR shall identify if there is any association between use of sedative hypnotics and Colle's fractures.

All elderly patients who will present with Colle's fractures will be considered as cases. The other patients who will present with other illnesses will be taken as controls. Both groups shall be evaluated using retrospective questioning regarding use of sedative hypnotics.

The Odds ratio calculation will be carried out as:

Exposure (i.e., sedative hypnotics)	No. of Cases (i.e., Colle's fractures)	No. of Controls (other illnesses)
Yes	a	b
No	c	d

Odds of being exposed among the cases = a/c

Odds of being exposed among the controls = b/d

Exposure odds ratio = (a/c)/(b/d) = (a × d)/(b × c)
(Cross-product ratio)

6. COHORT STUDIES ^{12,24,25}:

i. Research question:

Does texting on mobile devices with thumb for over two hours daily cause deQuervain's tenosynovitis among diabetic patients?

ii. Sample size Calculation:

The required sample size shall be determined using the formula:

$$n = \frac{p_1(1-p_1) + p_2(1-p_2) \times C}{(p_1 - p_2)^2}$$

Where:

n = The required sample size

p_1 and p_2 = Proportion of the two groups. In this study p_1 is 40%=0.4.

The p_2 is 20%=0.2.

C = It is the standard value for the corresponding level of α and β selected for the study.

With 95% CI and 80% power for the study, and taking the proportions of the two groups as

40% and 20%, respectively (from the published literature), the sample size will be calculated as

$$n = 0.4(1-0.4) + 0.2(1-0.2) \times 7.85 / (0.4-0.2)^2$$

$$= 10 \times 7.85$$

$$= 78.5$$

Considering 10% dropout of study participants, the sample size will become:

$$n = 78.5 + 7 = 85.5 \text{ or } 86 \text{ patients}$$

Hence, for conducting this cohort study, a minimum of 86 patients are required in each group.

- iii. Most relevant tests for undertaking statistical analysis:

Relative risk (RR) of exposure:

We are attempting to determine the association between an exposure (i.e. texting on mobile devices with thumb for over two hours daily) and an outcome (i.e. deQuervain's tenosynovitis). The incidence of outcome shall be determined in the exposed group and the non-exposed group and then these will be compared. For calculating the incidence in each group, we will divide the number of subjects that have developed the outcome by the total number of subjects in the group. To determine the RR of exposure, we shall divide the incidence in the exposed group by the incidence in the non-exposed (control) group. This shall tell us how much higher or lower the risk of obtaining the outcome is for a person who is exposed to the factor than for a person who is not.

Relative risk (RR) = Risk in the exposed group / Risk in the Un-exposed group

- An RR of >1 means positive association between the exposure and outcome. It is weak if it is 1.01 – 1.50, moderate if it is 1.51 – 3.00 and very strong if it is >3.00.
- An RR of <1 means a negative association.

An RR of 1 means no association between exposure and outcome.

The following table shall be employed to determine an association between the exposure (texting) and Outcome (de-Quervain's disease) among diabetic patients:

Exposure status	Disease	No disease	Total Number
Exposed	A	B	C
No-exposed	D	E	C

There will be equal number of subjects in each of the exposed and non-exposed groups.

Incidence in the exposed group = Number with outcome / total Exposed = A / C = F %

Incidence in the Non-exposed group = Number with outcome / total Non-exposed = D / E = G %

RR = Incidence in exposed group / Incidence in Non-exposed group = F / G = H (The H will be the RR)

7. Validating and Developing a Questionnaire

- i. Research question:

Are the first year MBBS students satisfied with their anatomy lectures delivered by the junior faculty?

- ii. Steps for the development and validation:

Perform review of the relevant literature: A thorough literature review will help elucidate the questionnaire and also identify any pre-existing related constructs that will guide the current one. Thus, the construct is perfectly aligned to the existing theoretical works. Items from these pre-existing research works may be adapted with permission from original authors.²⁶⁻²⁹

Carry out interviews and/or focus groups: Once the items of the questionnaire have been developed, it will be initially administered to individuals who have close resemblance to the actual target population of the study. This will help to ensure that that the understanding of the

questionnaire items matches how the actual respondents of the study would comprehend them.²⁹

Synthesize the literature review and interviews/focus groups: At this stage, the information obtained so far from the literature review and the interviews will be merged together. This will ensure that construct is not only theoretically well-founded but also easily understandable by the target population sample.

Develop items: The items of the questionnaire will be carefully developed so that they truly reflect the intended construct. Also, the items will be written in clearly understandable words of the target respondents.^{26,29-32}

Conduct expert validation: Once the items have been developed, validity evidence will be collected in the form of expert validation. A panel of 10 experts will be identified. A content validation form will be created that will define the construct and give experts the opportunity to provide appropriate feedback. The experts will particularly assess the representativeness, clarity, relevance and distribution of the item. They will also provide free text comments.^{33,34}

Employ cognitive interviews to ensure response process validity. A minimum of 10 participants shall be interviewed. The cognitive interviews shall be based on the following principles of psychology: Comprehension of an item stem and answer choices; retrieval of appropriate information from long-term memory; judgment based on comprehension of the item and their memory; and selection of a response. The cognitive interviews shall help to refine the items and shall ensure that the respondents understand the items as intended.³⁵⁻³⁷

Conduct pilot testing: This shall check for adequate item variance, reliability analysis and convergent/discriminant validity with respect to other measures. The questionnaire shall be pilot tested on a certain number of the target

population. The data obtained shall then reviewed to evaluate item range and variance, assess score reliability of the whole scale and review item and composite score correlations. Descriptive statistics (e.g., means and standard deviations) and histograms shall also be reviewed to know the distribution of responses by item. This analysis shall help identify items that may not be functioning in the way the designer intended.²⁹

Conclusion

The current review is expected to serve as a comprehensive guide for future researchers in planning and executing their quantitative research more thoroughly and robustly.

References

1. Ringsted C, Hodges B, Scherpbier A. 'The Research Compass': An Introduction to Research in Medical Education: AMEE Guide No. 56. *Med Teach*. 2011;33(9):695-709. doi: 10.3109/0142159X.2011.595436.
2. Ratelle JT, Sawatsky AP, Beckman TJ. Quantitative Research Methods in Medical Education. *Anesthesiology*. 2019;131(1):23-35. doi: 10.1097/ALN.0000000000002727.
3. Saaq M, Ashraf B. Modifying "PICO" Question Into "PICOS" Model for More Robust and Reproducible Presentation of the Methodology Employed in a Scientific Study. *World J Plast Surg*. 2017;6(3):390-392. PMID: 29218294; PMCID: PMC5714990.
4. Browner SW, Newman BT, Hully BS. Estimating Sample Size and Power: Applications and Examples. In B.S, Hully, R.S, Cummings, S.W, Browner, G.D, Grady, B.T, Newman, (Eds.). *Designing Clinical Research* (4th ed.). Philadelphia: Lippincott Williams & Wilkins; 2013. p. 55-83.
5. Banerjee A, Chitnis UB, Jadhav SL, Bhawalkar JS, Chaudhury S. Hypothesis Testing, Type I and Type II Errors. *Industrial Psychiatry J*. 2009;18(2):127-31. doi: 10.4103/0972-6748.62274.
6. McConnell MM, Monteiro S, Bryson GL. Sample Size Calculations for Educational Interventions: Principles and Methods. *Can J Anaesth*. 2019;66(8):864-873. doi: 10.1007/s12630-019-01405-9.
7. Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ. Sample Size Calculations: Basic Principles and Common Pitfalls. *Nephrol Dial Transplant*. 2010;25(5):1388-93. doi: 10.1093/ndt/gfp732. Erratum in: *Nephrol Dial Transplant*. 2010;25(10):3461-2.
8. Gordis L. *Epidemiology*. 4th ed. Philadelphia: Elsevier/Saunders; 2010. p. 1-375.

9. Creswell JW. Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research. 4th ed. Boston, MA: Pearson Education, Inc.; 2012.
10. Fraenkel JR, Wallen NE, Hyun HH. Survey Research. In: How to Design and Evaluate Research in Education. 8th ed. New York: The McGraw-Hill Companies, Inc.; 2012.
11. Raosoft® Sample Size Calculator. (2023, October 7). Retrieved from <http://www.raosoft.com/samplesize.html>
12. Sharma SK, Mudgal SK, Thakur K, Gaur R. How to Calculate Sample Size for Observational and Experimental Nursing Research Studies? National Journal of Physiology, Pharmacy and Pharmacology. 2020;10(1):1-8. doi: 10.5455/njppp.2020.10.0930717102019.
13. Artino AR Jr, Gehlbach H, Durning SJ. AM Last Page: Avoiding Five Common Pitfalls of Survey Design. Acad Med. 2011;86(10):1327. doi: 10.1097/ACM.0b013e31822f77cc.
14. Habib A, Johargy A, Mahmood K, Humma. Design and Determination of the Sample Size in Medical Research. IOSR Journal of Dental and Medical Sciences. 2014;13(5):21-31. <https://www.iosrjournals.org/iosr-jdms/papers/Vol13-issue5/Version-6/F013562131.pdf>
15. Saaq M, Din NU, Jalil A, Zubair M, Shah SA. Diagnostic Accuracy of Leukocytosis in Prediction of Acute Appendicitis. Journal of the College of Physicians and Surgeons Pakistan. 2014;24(1):67-69.
16. Charan J, Biswas T. How to Calculate Sample Size for Different Study Designs in Medical Research? Indian J Psychol Med. 2013;35(2):121-6. doi: 10.4103/0253-7176.116232.
17. Fraenkel JR, Wallen NE, Hyun HH. Correlational Research. In: How to Design and Evaluate Research in Education. 8th ed. New York: The McGraw-Hill Companies, Inc.; 2012. p. 330-364.
18. Gravetter FJ, Wallnau LB. Statistics for the Behavioral Sciences. 10th ed. Boston, MA: Cengage Learning; 2017.
19. Patra P. Sample Size in Clinical Research, the Number We Need. International Journal of Medical Science and Public Health. 2012;1(1):5-10.
20. Pedhazur EJ. Multiple Regression in Behavioral Research: Explanation and Prediction. 3rd ed. Fort Worth: Harcourt Brace College Publishers; 1997.
21. Lipsey MW, Hurley SM. Design Sensitivity: Statistical Power for Applied Experimental Research. In: The SAGE Handbook of Applied Social Research Methods. SAGE Publications, Inc.; 2009. p. 44-76.
22. Saaq M. Presentation and Management Outcome of High Radial Nerve Palsy With Tendon Transfers. Archives of Bone and Joint Surgery. 2021;9(5):578-584. doi: 10.22038/abjs.2020.51938.2563.
23. Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin; 2002.
24. Gordis L. Cohort Studies. In: Epidemiology. 5th ed. Philadelphia, PA: Elsevier/Saunders; 2014. p. 149-158.
25. Hulley SB, Cummings SR, Newman TB. Designing Cross-Sectional and Cohort Studies. In: Designing Clinical Research. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2013. p. 85-96.
26. Artino AR Jr, LaRochelle JSD, Dezee KJ, Gehlbach H. Developing Questionnaires for Educational Research: AMEE Guide No. 87. Med Teach. 2014;36(6):463-474. doi: 10.3109/0142159X.2014.889814.
27. Cook DA, Beckman TJ. Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application. The American Journal of Medicine. 2006;119(2):166.e7-166.e16. doi: 10.1016/j.amjmed.2005.10.036.
28. Gehlbach H, Artino AR Jr, Durning S. AM Last Page: Survey Development Guidance for Medical Education Researchers. Acad Med. 2010;85(5):925. doi: 10.1097/ACM.0b013e3181dd3e88.
29. Gehlbach H, Brinkworth ME. Measure Twice, Cut Down Error: A Process for Enhancing the Validity of Survey Scales. Rev Gen Psychol. 2011;15(4):380-387. doi: 10.1037/a0025704.
30. DeVellis RF. Scale Development: Theory and Applications. 4th ed. Newbury Park, CA: Sage Publications Inc.; 2016.
31. Dillman DA, Smyth JD, Christian LM. Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method. 4th ed. Hoboken, NJ: Wiley; 2014.
32. Fowler FJ. Survey Research Methods. 4th ed. Los Angeles: SAGE Publications; 2009.
33. Polit DF, Beck CT. The Content Validity Index: Are You Sure You Know What's Being Reported? Critique and Recommendations. Res Nurs Health. 2006;29(5):489-497. doi: 10.1002/nur.20147.
34. Rubio DM, Berg-Weger M, Tebb SS, Lee ES, Rauch S. Objectifying Content Validity: Conducting a Content Validity Study in Social Work Research. Social Work Res. 2003;27(2):94-104. doi: 10.1093/swr/27.2.94.
35. Karabenick SA, Woolley ME, Friedel JM, Ammon BV, Blazeviski J, Bonney CR, et al. Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist*. 2007;42(3):139-151. <https://doi.org/10.1080/00461520701416231>
36. Nápoles-Springer AM, Santoyo-Olsson J, O'Brien H, Stewart AL. Using cognitive interviews to develop surveys in diverse populations. *Medical Care*. 2006;44(11 Suppl 3):S30. <https://doi.org/10.1097/01.mlr.0000245425.65905.1d>
37. Grady DG, GN, editors. *Designing Clinical Research*. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2013. p. 44-55. Available from: https://www.academia.edu/36931058/Designing_Clinical_Research. [Last accessed on 2023 Oct 11].