**Proper Reporting of Methodological Aspects of Sample Size Determination in Health Research: An Umbrella Review**

Md. Shihab Mostafa1, Jakia Sultana Mim1, Sidratul Muntaha1, Sabbir Hossain1, Mahfuzer Rohman1, Md. Fakrul Islam1, Prosenjit Basak Arka1, Md Jamal Uddin, PhD1,2\*

Md. Shihab Mostafa

[shihabmostafa475@gmail.com](mailto:shihabmostafa475@gmail.com)

Jakia Sultana Mim

[jakiasultanamim99@gmail.com](mailto:jakiasultanamim99@gmail.com)

Sidratul Muntaha

[Sidratul99muntaha@gmail.com](mailto:Sidratul99muntaha@gmail.com)

Md Sabbir Hossain

[bmsabbirhossainsakib@gmail.com](mailto:bmsabbirhossainsakib@gmail.com)

Mahfuzer Rohman

[mahfuzzihad@gmail.com](mailto:mahfuzzihad@gmail.com)

Md. Fakrul Islam

[fokhrulislumneshat@gmail.com](mailto:fokhrulislumneshat@gmail.com)

Prosenjit Basak Arka

[prosenjitbasakarka@gmail.com](mailto:prosenjitbasakarka@gmail.com)

Md Jamal Uddin

[jamal-sta@sust.edu](mailto:jamal-sta@sust.edu)

1. Department of Statistics, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh
2. Faculty of Graduate Studies, Daffodil International University (DIU), Dhaka.

\*Corresponding author: Md Jamal Uddin, Professor, Department of Statistics, Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh. Email: [jamal-sta@sust.edu](mailto:jamal-sta@sust.edu)

**Abstract**

**Background**

As health research continues to expand, the need for a reliable procedure to determine a sample size that accurately represents a given population has become increasingly important. Although numerous systematic reviews have addressed this topic, a clear and consistent method for estimating sample size remains elusive. This umbrella review seeks to bring together and synthesize findings from all relevant systematic reviews on sample size determination in health studies into a single comprehensive analysis.

**Method**

The study searched for systematic reviews (SLRs) on sample size determination in medical or health studies from January 1, 2000, to December 31, 2023, using PubMed, ResearchGate, ScienceDirect, and Google Scholar evaluated references of pertinent articles. Systematic reviews of health-related studies that explored the sample size determination process were included. We also assessed the methodological quality of the featured systematic reviews using the “Assessing the Methodological Quality of Systematic Reviews 2” checklist. The following key factors were considered during the review of papers: study design, population size, study hypothesis, power calculations, significance level, effect size, sample size formula, sampling techniques, statements about sample size, and standard deviation.

**Results**

Out of 1,42,276 screened publications, a total of 13 systematic reviews (SLRs) were included. Most of the SLRs focused on recommending methods for proper sample size calculation. However, the majority of the systematic reviews did not adequately address many critical factors or provide necessary explanations for calculating sample size. The adoption rates of these factors in the SLRs were as follows: study design (30.8%), population size (7.7%), study hypothesis (23.1%), power calculations (46.2%), significance level (38.5%), effect size (38.5%), sample size formula (7.7%), sampling techniques (30.8%), statements about sample size (69.2%), and standard deviation (7.7%).

**Conclusion**

The umbrella review highlights significant gaps in existing SLRs regarding the inclusion and explanation of critical factors for sample size determination, with most failing to provide comprehensive guidance. To address this, future SLRs should prioritize incorporating essential factors—such as power calculations, effect size, and sampling techniques—and offer detailed explanations to enhance the validity and practical application of sample size determination methods. Thorough explanations and clear steps are particularly crucial in medical research, where accurate measures are vital for consistent and reliable results.

**Introduction**

An effective way to determine sample size accurately to reflect a particular population is important due to the growing demand for research [1]. Sample size estimates are required for journal publication demands, ethical committee clearance, funding body approval, research project approval, and most importantly to support the validity of study results [2]. Sample size determination involves mathematics to determine the ideal number of samples to include in a study before it starts. The issue of determining the sample size will arise for the student or researcher regardless of the form of research design employed for the study [3]. In medical studies, it's critical to establish a sample size that will allow for trustworthy results.

A sample size calculation, also known as a power calculation, provides a response to the question, "How many individuals or observations are required to be incorporated in a study?" at the beginning of research study preparation [4]. The answer is standard error will be lower, and the research's statistical power and precision will be high if a study is properly designed and has the appropriate sample size [5]. However, because of the needless engagement of additional subjects and the resulting higher expenditures, a study with an excessive sample size could be considered unethical [6]. Conversely, a study with an inadequate sample size will not be able to identify effects that are clinically significant. Because of this, the study may be inappropriate in the use of subjects and additional resources and be ineffective from a scientific standpoint [7]. On the other hand, sampling techniques are frequently employed in research studies to provide more accurate estimates in less time and at a lower cost. In applied statistics research challenges, choosing the appropriate sampling techniques and estimating the sample size are critical steps in reaching the right conclusions. Techniques for calculating sample size and carrying out power analysis primarily rely on the study's design and primary measure. There are numerous ways to estimate sample size for various study types and various results measures [8]. Optimizing the number of samples and sampling techniques used in a study increases the likelihood that the data can be understood and reduces research waste [8, 9].

In most research projects, informal suggestions for sample size are based on the expertise of the researchers and may be sufficient, as in the case of pilot studies [10, 11]. However, a growing number of scholarly publications have mandated proof of sample size computation or other specifications to be given in the methodology portion of a work, and the computation may be a component of a checklist prior to submission to a publication [12, 13]. Still, sample size is rarely considered by healthcare researchers, who instead select the most practical numbers (30, 50, 100, etc.) or durations (one month, six months, one year, etc.) for their investigations [14]. Even with major advances in research, determining the right sample sizes for medical and health-related studies is still challenging and often fragmented areas. Previous research has mostly focused on specific study designs and provided limited techniques for determining sample sizes [15–17]. Because of this restricted focus, there have been many systematic literature reviews (SLRs) that focus on certain areas of determining sample size. Although these SLRs provide insightful information, they often lack a comprehensive perspective, which presents difficulties for researchers and practitioners looking for a common understanding of sampling techniques and how to accurately use them in a variety of study scenarios.

Umbrella reviews have become a powerful tool in recent years for combining data from different review-level studies, providing a broader and more integrated understanding of complex research topics. [18]. In contrast to traditional systematic literature reviews, umbrella reviews facilitate the aggregation and critical assessment of findings from a variety of reviews, which helps to identify trends, gaps, and areas for improvement. The umbrella review approach is becoming more popular in public health research and practice as a way to maximize resource allocation and inform evidence-based policymaking [19]. This umbrella review aimed to evaluate all SLRs that addressed sample size determination in medical or health-related studies, assessing whether these reviews included the necessary critical factors and provided adequate explanations for accurate and reliable sample size determination.

**Methods**

**Study protocol and data sources**

This umbrella review was conducted according to the suggested Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) which was provided in 2020 [20]. The International Prospective Register of Systematic Reviews had this protocol registered (PROSPERO; registration number CRD42024600718). Moreover, “PubMed”, “ResearchGate”, “ScienceDirect”, and “Google Scholer” were utilized for the required published articles from January 1, 2000 to December 31, 2023.

**Search Strategy**

The study used PubMed and Google Scholar to search for Systematic reviews on sample size determination in public health studies. To narrow the papers, this study employed the following keywords: “Sample size determination”, “sample size calculation”, “sampling techniques”, “Sampling”, “Sample size”, “RCTs” or “Randomized Control trials or health studies”, and “Medical”. Two researchers screened the abstracts, titles, and conclusions for the systematic review articles. They are unknown of the author's names and journals of the publication. Additionally, reference lists of the systematic review articles were searched as an additional source for eligible papers.

**Study eligibility criteria (Inclusion and exclusion criteria)**

SRLs which included sample size determination or sample size calculation reports on health or medical fields were eligible for inclusion in the umbrella review. This study only took those reviews that followed a systematic method, published in the English language as full text, and available for free. This study considered studies reported on sample size calculation, sample size formula, sampling techniques, power calculations, significance level, effect size, hypothesis testing, different study designs, mean differences, standard deviation, and treatment effect which are presented in Table 3. However, opinions and reviews of narratives were not considered for this review. Science aims to focus only on health or medical studies and those SRLs that were not based on RCTs or health or medical or biostatistics fields were excluded from the study. A full details of this exclusion criteria will be found in Table 2.

**Data Extraction**

A standardized method for data extraction was utilized by the two reviewers (MJU and MSM) independently for extracting the given information from each of the systematic reviews: author details, citation, objective of the review, literature search details, included study design, number of studies included by each SRLs, and outcome variables (study design, population size, study hypothesis, power calculation, significant level, effect size, sample size formulae, sampling techniques, statement about sample size, mean difference, standard deviation, and treatment effect).

**Quality Assessment and Risk of Bias**

The AMSTAR-2 checklist was used to measure the quality of included SRLs in this study [21]. The most commonly applied tool for quality appraisal of SRLs is AMSTAR-2 which can be used for both randomized and non-randomized studies (may or may not include meta-analysis) [22]. This checklist has in total sixteen questions and has 7 domains that help to determine the quality of a systematic review. Here, the critical domain includes a previously recorded assessment protocol, literature search adequacy, individual studies exclusion justification, assessment of risk of bias, proper statistical techniques for merging results, and identifying publication bias which is for meta-analysis and potential biases consideration for individual studies interpretation [21]. Because the reviews included in this study varied widely, and a lot of previously published studies with extremely low ratings, there are worries that the tool’s discriminative power may be limited [23]. Two individual reviewers (MJU and MSM) performed this assessment. Any differences in opinion were discussed and settled by agreement.

**Sample Size Determination Process**

**Study design**

The most important step for conducting research and selection of sample size is the selection of a proper study design [24]. A good study design can help to determine or identify the outcome of the study or research properly and can run a study successfully [25]. Different types of study design can be applied based on the objective of the research. A cross-sectional study which can be utilized at a specific time period. This design evaluates the level of exposure and disease concurrently in a particular population. In contrast, a case-control study design is utilized for comparing the health status of the two groups (cases and controls). Moreover, this study can also assess the exposure risk associated with each group. On the other hand, the key feature of a cohort study is it can compare the likelihood of the disease in the two disease-free groups where one is exposed, and another is not exposed to the risk factor that followed over a time period [24].

**Sample size Determination Parameters**

**Study Hypothesis**

A research hypothesis is conducted to summarize the research's basic elements such as sampling techniques, design of the study, sample size, independent and dependent variables. The main purpose of presenting the hypothesis is to set the foundation for statistical significance testing. Moreover, if a study is comparative then there is a necessity for study hypothesis. There are two types of hypotheses: null hypothesis (*H0*) and alternative hypothesis (*H1*). Here, the alternative hypothesis in our example would be that there is a difference in the clinical failure rate and the null hypothesis would be that there is no difference in the clinical failure rates between cemented first molar attachments and bonded tube attachments. One cannot test the alternative hypothesis directly. An alternative hypothesis is accepted automatically if a statistical test rejects the null hypothesis. There are two types of alternative hypotheses: one-sided and two-sided. A two-sided alternative hypothesis is there will be a difference that can go either way, while a one-sided alternative focuses on a particular direction of the result (bonded tube first molar attachments will have a higher long-term failure rate) [24].

**Effect size**

The degree of the association in the target population will determine how likely it is that a properly executed investigation would find a relationship between an independent and dependent variable. It will be simple to find in the sample if it is vast. On the other hand, it will be challenging to identify the relationship in the sample if it is little, necessitating a high sample size. This value is the least difference in measurement among the groups being compared that the investigator expects the study will indicate. Selecting a suitable effect size is the most challenging part of choosing sample sizes. To estimate an appropriate effect size, the investigator should first attempt to find data from earlier studies in the relevant field. As an alternative, one can choose the minimum effect size that in their judgment which would have clinical significance. In situations where data are unavailable, a pilot survey could be essential. Moreover, both the technical proficiency of the statistician and the scientific expertise of the investigator are needed for the examination of the tradeoffs between sample size and effect size [24].

**Estimated measurement variability**

The indicated standard deviation in the measurement mode within each contrast group serves as a representation of this parameter. The larger the sample size required to find the minimal difference with the lower the statistical variability. Ideally, preliminary data gathered from a comparable study population should be used to calculate the expected measurement variability. From a literature review, one can find these parameter estimates. The parameter needs to be estimated by subjective experience or needs to be assumed if the data is not available. When comparing a proportional measurement (as opposed to a mean), there is no need for an additional estimation of measurement variability since the standard deviation may be obtained mathematically from the proportion [26].

**Statistical Power**

Statistical power is important for sample size determination. If the power increases, then the sample size also grows. Although high power is always preferred, there is an obvious compromise with the number of subjects that can be studied realistically, considering the time and resources that are often fixed for research or investigative studies. Statistical power in randomized controlled trials is typically set at a value more than or equal to 0.80 while many specialists in clinical trials are increasingly pushing for a power of 0.90 (Wood and Lombert 1999, Writes 2002) [26].

**Good Sample Size Criteria’s**

The level of precision, the level of confidence or the risk, and the degree of variability are typically the three factors that need to be specified to calculate the proper sample size, in addition to the population size and study goal (Miaoulis and Michener, 1976). [26]

**Level of precision or Significance Level**

The diversity in which the population’s true value is expected to be known as the precision level or sampling error. Similar to how media outlets report polling results for political campaigns, this range is sometimes stated as percentage points which is ±5%. With a precision rate of ±5%, a researcher can determine that between 55% and 65% of farmers in the population have implemented the recommended practice if they discover that 60% of farmers in the sample have done so [26].

**The confidence Level**

The idea of confidence level or risk came from the Central Limit Theorem. The Central Limit Theorem's main hypothesis is that, when a population is sampled repeatedly, its mean value of the attribute that those samples yield is equal to the actual value of the population. Furthermore, the values that these samples yield are regularly dispersed about the true value, with some samples yielding values that are higher and others yielding values that are lower than the true population value. About 95% of sample values in a normal distribution fall between two standard deviations from the underlying population value. In statistical hypothesis testing, the risk of error is another name for this confidence interval. Here, this indicates that 95 out of 100 samples will have the actual population value within the designated accuracy range if a 95% confidence level is used. However, the confidence level can be 90 to 99% (Gupta and Kapoor 1970 and Singh and Masuku 2012) [26].

**Degree of Variability**

In the third criterion, the allocation of attributes in the population is considered, together with the degree of variability in the characteristics under investigation. The lesser the sample size required, the more homogeneous the population is for the variables. The greater the degree of population heterogeneity, the greater the sample size necessary to achieve a particular degree of precision. For instance, a percentage of 50% denotes more unpredictability than either 80% or 20%. This is due to the fact that 80% and 20%, respectively, show that a sizable majority possess the relevant quality. A more conservative sample size may be chosen since a proportion of .5 reflects the greatest variability in a population; that is, the sample size may be greater than if the true variation of the population characteristic were utilized.

**Strategies for Determining Sample Size**

There are several strategies for sample size determination which include using a sample size of a similar study, using a census for a small population, and different for calculating sample size [27].

**Using Sample Size of a Similar Study**

Using a similar sample size from other studies that are comparable to the one the researcher intends is a strategy for determining sample size. However, if the methods used in this research aren't reviewed, a researcher can end up duplicating mistakes that were made while figuring out the sample size for a different study. Nevertheless, advice regarding “typical” sample sizes that are employed can be obtained by reviewing the literature in the researcher's field [27].

**Using a Census for Small Population**

Another strategy is to use whole study populations as a sample size. A census is desirable for small populations, but financial issues prevent it from being done for large ones. A census offers information on each member of the population and removes sampling inaccuracy. Furthermore, many expenses such as creating the sampling frame and designing the questionnaire are “fixed”, meaning that they won’t change whether the sample size is 50 or higher (200). Additionally, in order to reach the desired degree of precision, almost the whole population would need to be sampled in a tiny population.

**Using Formulas to Calculate a Sample Size**

Sample size calculation using an appropriate sample size formula is another approach for determining sample size correctly.

For large populations or unknown population sizes, Cochran (1963) provided an equation that will provide a representative sample according to the population size.

Equation 1.

Here, is the required sample size

The critical value from the standard normal distribution (1.96 for a 5% level of significance)

desired level of precision

is the estimated proportion of characteristics that are presented by the population

= 1-

The resulting sample size using the above formula is given below

For finite or known population size

Equation 2.

Here, n is the required sample size

is the standard sample size

N is the known population size

However, there are different kinds of formulas according to the different requirements. But these are the most common formulas that are used for determining sample size [27].

Other sample size formula is given below [15]:

|  |  |  |
| --- | --- | --- |
| Types of Study | Sample Size Formula | Explanation |
| Case-Control Study |  | * r = control to cases ratio (1 if same numbers of patient in both groups) * = it is the desired power (0.84 for 80% power and 1.28 for 90% power) * = it it’s the standard normal variate with 1,96 at 5% error * = proportion in cases * = proportion in controls * = required sample size |
| Diagnostic tests | For determining Sensitivity:  For determining Specificity:  Where: | * Z = it is conventionally taken as 1.96 in lieu of 90% confidence interval * p = prevalence of rate of disease in study population * W = it is maximum acceptable width of 95% CI, is conventionally taken as 10% * Specificity and sensitivity and p-values are inferred from the previous studies |
| Animal Study | For one-way ANOVA design (for group comparison):  Minimum number of patients or groups:  Maximum number of patients or groups: | * k = number of groups * n = number of patients per group * r = number of repeated measurements * If the study involves sacrificing of animals, then the n should be multiplied by r. |
| One within factor, repeated-measure ANOVA (one group, repeated measures):  Minimum number of patients or groups:  Maximum number of patients or groups: |
| One between, one within factor, repeated measures ANOVA (group comparison, repeated measurements):  Minimum number of patients or groups:  Maximum number of patients or groups: |
| Cohort Study |  | * = it is the desired power (0.84 at 80% power) * = it is the standard normal variate in 1.96 at 5% error * = possibility of event in controls. From previous studies * = possibility of an event in experimental. From previous studies * m = number of control subjects per experimental subject * p = [p1 + (m.p0)]/m+1 |

**Sampling Techniques**

Sampling Techniques are another method for determining a representative sample size for the population. If the sample size does not represent the population, then the result of the research will give wrong information to the nation. Therefore, accurate sampling techniques is very important in sample size determination [28]. Different types of sampling techniques are-

**Simple Random Sampling**

In the simple random sampling procedure, there is an equal chance of inclusion for each unit in the provided sample. This technique provides an unbiased and better estimate of the parameters if the population is homogeneous. Moreover, in case the population is homogeneous, this method offers a more impartial and superior approximation of the parameters [28].

**Stratified Random Sampling**

In cases where the population is heterogeneous, stratified random sampling is a helpful technique for gathering samples. This approach divides the whole diverse population into several homogeneous groupings, or strata, each of which is homogenous within itself. Units are then randomly sampled from each of these strata. The relative relevance of each stratum in the population determines the sample size for each stratum. This procedure is known as a stratified random sampling [28].

**Cluster Sampling**

Dividing the population into groups or clusters and selecting a random sample from every cluster is known as a cluster sampling technique. Within the chosen clusters, each observation is part of the sample. A statistical population's "natural" yet often homogeneous groupings can be identified through the use of cluster sampling. When a researcher has access to a comprehensive list of groups or "clusters" within a population but not all of the units within that population, they will typically utilize cluster sampling. This sampling technique might be more cost-effective and practicable than stratified or simple random sampling [28].

**Systematic Random Sampling**

With systematic random sampling, the initial sample unit is selected at random, and the remaining sample units are chosen in a systematic manner. If there is population size of N and n units need to be selected, the R = N/n (Here, R is the sampling interval). From the remaining R (Sampling Interval) to the previously chosen number, the first number is chosen at random [28].

**Multistage Sampling**

Multiple phases of unit selection are included in multistage random sampling. At each level, the sampling designs could be the same or different. In multistage sampling approach, sometimes known as multistage cluster sampling, uses samples that are somewhat clustered. The main benefit of this sampling procedure is that it allows resources to be focused on a small number of frame units, however, sampling error may rise as a result of this sampling technique [28].

**Results**

**Systematic Literature Review Characteristics**

The study search showed 1,42,246 results. After duplication, 1,42,133 references were matched at the title and abstract level for the inclusion criteria, of those 66 reviews were eligible for full-text eligibility assessment. 13 reviews finally fulfilled the inclusion criteria and were included for data extraction. The list of 53 articles was excluded during the full-text review. The PRISMA flowchart highlights the results of the literature search and study selection (Figure 1).

**113** duplicates removed

**1,42,246** imported for screening

From 2010 to 2020

**Imported**

**1,42,067** publications considered irrelevant

**1,42,133** publications screened at title and abstract

**66** publications assessed for eligibility

**13** included for data extraction

**53** publications excluded

**12** with ineligible publication type (narrative, scoping, conference abstract, etc.)

**4** published prior to January 1st, 2000

**7** with missing key information (text, protocol, outcomes)

**15** not RCT studies or health-related studies

**5** not open access

**10** didn’t explain about the outcome (sample size determination)

Screening

Included

*Figure 1. PRISMA Flowchart for Study Selection*

**Distribution of Sample Size Determination Strategies Included by SLRs**

From **Figure 2**, just 30.8% of systematic review studies included study design. Moreover, 7.7% stated about sample size determination formulae and approximately 30.8% explained different sampling techniques. Lastly, almost two-thirds of the SLRs (69.2%) talked about sample size statements which was what would happen if the sample size were large or small.

**Distribution of Different Parameter Explanation Included by SLRs**

**Figure 3** shows that 46.2% of the studies incorporate power calculation. On the other hand, just 38.5% of SLRs explained the significance level and effect size for SSD. Additionally, only 7.7% talked about standard deviation in their studies.

**Distribution of Important Items Included by SLRs**

From **Figure 4**, just 7.7% of the SLRs explained the population size for determining sample size while the percentage is 23.1% for the study hypothesis.

**Characteristics of Systematic Reviews included in the Umbrella Review**

This umbrella review incorporates a total of 13 systematic reviews without meta-analysis [29–41]. These SLRs were published between 2000 and 2023, and the majority (12) were published before the year 2020 (**Table 1**). This included reviews were used approximately 1 to 7 databases for keyword searches (**Table 2**). On the other hand, the approximate number of included studies for systematic reviews ranged from 60 to 324. The majority of the studies that were included in the SLRs were based on Randomized Control trials (RCTs) and their main objective was to overview the sample size determination process in the medical or health field. The 13 systematic reviews included only those studies that explained the sample size calculation or determination and sampling techniques.

**Characteristics of Search Criteria across Included SLRs**

**Table 2** highlighted that most of the paper properly explained their search strategy while only three SLRs did not [29, 30, 34]. On the other hand, 6 studies did not include the PRISMA flowchart whereas 7 did (**Table 2**). The number of reviewers included in the SLR studies ranged from two to four. The majority of the reviews included literature sources information while just one did not [34]. Additionally, 8 studies did not mention the funding sources, and the rest (5) of them were mentioned.

**Methodological Quality**

Almost all the included SLRs failed to fulfill the 7 critical domains of the “Assessing the Methodological Quality of Systematic Reviews 2” checklist (**Table 5**). Of 13 SLRs, 10 studies failed to give review protocol. On the other hand, only one study did not include comprehensive literature search strategies in their study. Moreover, five systematic reviews did not incorporate the list of excluded studies and justification. Unfortunately, most (8) of the SLRs failed to show satisfactory techniques for assessing the risk of bias. Furthermore, 4 studies did not account for the risk of bias when interpreting or discussing the results. For publication bias assessment and discussion, 3 of the SLRs failed to fulfill this critical domain. The remaining critical domain related to meta-analysis was difficult to measure because the majority of the SLRs did not include meta-analysis in their studies. None of the included SLRs fulfill all nine of the non-critical domains of the AMSTAR 2 checklist (**Table 5**).

**SSD Outcomes**

A diverse range of outcomes (12) outcomes for sample size determination were reported by the thirteen systematic reviews (**Table 3**). This includes study design, population size, study hypothesis, power calculation, significance level, effect size, sample size formulas, sampling techniques, statements about sample size, mean difference, standard deviation, treatment effect, and many more.

**Findings across SLRs for Sample Size Determination**

**Study Design**

While only 4 studies stated study design the majority, 9 studies did not mention study design (Table 3). From them, a study talked about parallel-group randomized trials and cluster randomized trial [29]. Moreover, another mentioned parallel and cross-over study design or trial design [30] whereas a study mentioned the cross-sectional, cohort, and open-cohort studies [35]. A systematic literature review (SLR) discussed the percentage of studies that mentioned their design but did not specify the types of study designs used by the included research [38].

**Population Size**

Only a single study mentioned population size for sample size determination [29]. They stated which studies discussed population size prior to sample size calculation.

**Study Hypothesis**

Out of the 13 SLRs, just 3 studies discussed the study hypothesis. However, the study did not explain the hypothesis properly. In other words, the main purpose of the hypothesis for sample size determination was unclear in the study [29]. Others adequately outlined the basic goal of the hypothesis and the necessity of having the SSD [30, 32]. Studies explained that in statistical hypothesis testing, there are key concepts that help us assess the validity of our conclusions: significance level (α), power, and errors (Type I and Type II).

**Power Calculation**

Power calculations for sample size determination were discussed in 6 SLRs whereas 7 studies did not discuss (**Table 3**). The majority of the paper stated that the standard power measure used by different studies for sample size calculation. Here, most of the study used the power 0.8 and some others used 0.90 for sample size determination.

**Significance Level**

Among the 13 systematic literature reviews (SLRs), 5 explicitly discussed the significance level, while the remaining 8 did not mention it. Among the 5 that did, 2 studies explained the standard significance level used in determining sample size for research studies [29, 30]. These reviews also discussed the concepts of Type I error and Type II error, which are important for hypothesis testing. The SLRs that mentioned the significance level generally stated that for sample size determination, a one-tailed test often uses a significance level of 0.025, while a two-tailed test uses a significance level of 0.05. These levels represent the acceptable probability of making a Type I error, which helps guide how large a sample should be to achieve reliable statistical results.

**Effect Size**

Out of the 13 systematic literature reviews (SLRs), only 5 included effect size in their findings related to sample size determination. Some of these studies provided explanations about standard effect sizes used for calculating the appropriate sample size [29, 30, 32]. One of the studies specifically mentioned effect size measures of 0.88 and 0.96 as standard values for their sample size calculations [31]. The other studies that discussed effect size emphasized its importance in determining sample size but did not specify exact values. They highlighted that effect size is a critical factor in designing studies, as it reflects the magnitude of the relationship or difference being studied. Larger effect sizes typically require smaller sample sizes, while smaller effect sizes need larger samples to detect statistically significant results.

**Sample Size Formula**

Sample size formulas for sample size calculation were mentioned in 1 SLR while 12 studies did not include a sample size formula [34]. This study used a standard sample size formula when the population size was unknown which was

Here, is the required sample size

The critical value from the standard normal distribution (1.96 for a 5% level of significance)

desired level of precision

is the estimated proportion of characteristics that are presented by the population

= 1-

Another formula for SSD they used (known population size)

Here, s is the required sample size

is table value of chi-square for 2 degrees of freedom at necessary confidence level (3.841)

N= Population size

P= Population proportion (0.50)

d = degree of accuracy expressed in proportion (0.05).

**Sampling Techniques**

From the 13 systematic literature reviews, just 4 talked about different sampling techniques. Two of them explained both probability sampling (Simple Random sampling, Stratified Sampling, Systematic random sampling, multistage sampling, and Cluster sampling) and non-probability sampling methods (Convenience sampling, snowball sampling, quota sampling, and judgment sampling) [34, 39]. Another one explained only simple random sampling [36] whereas one study highlighted the cluster sampling method [35].

**Statement about Sample Size**

Out of the 13 studies, 9 explicitly discussed sample size in their findings (as summarized in **Table 3)**. These studies explained the implications of both large and small sample sizes. The studies indicated that a larger sample size generally leads to more precise and reliable results. It reduces the margin of error and increases the statistical power of the study, making it easier to detect true effects or differences. Conversely, small sample sizes can lead to less reliable results, higher variability, and a greater chance of producing Type II errors (failing to detect a true effect). Smaller samples might also introduce bias and make generalizing the findings to a broader population more difficult [29, 30, 32, 34, 35, 37, 38, 40, 41].

**Standard Deviation**

Only one study talked about standard deviation for sample size calculation [29]. According to this study, the majority of the included studies in the SLR assumed incorrect standard deviation when calculating statistical power for SSD (either too large or too small). However, underestimating the standard deviation can affect the trial’s ability to detect important treatment differences in randomized control trials while overestimating the standard deviation can lead to an inflated sample size, requiring more participants than necessary for the study, which can waste resources and time

Table 1. *Characteristics of Included SLRs*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Author,  Year | Included study Design | Review Objective | Study Selection Criteria | Evidence Type | Studies Included |
| [29]  (Copsey et al., 2018) | Report of RCT of two treatment arms in a hip or knee osteoarthritis population | summarize current practice in calculating the sample size for trials of hip and knee osteoarthritis, including the sample size, target difference, and justification for the chosen inputs | Sample size calculation method, effect size, target difference, standard deviation | Human | 116 |
| [30]  (McKeown et al., 2015) | Primary publications of randomized, double-blind clinical trials of pharmacologic and interventional pain treatments | describe the extent to which sample size calculations are reported comprehensively in analgesic RCTs published in the pain literature | Sample size calculation or power analysis, statistical test | Human | 172 |
| [31]  (Guo et al., 2014) | Functional Magnetic Resonance Imaging studies, RCT | assess the reporting of effect sizes and variance components in observational fMRI studies involving clinical human participants | Study design, sample size calculations. | Human | 100 |
| [32]  (Abdulatif et al., 2015) | Randomized controlled trials, controlled clinical | evaluated the pitfalls in reporting sample size calculation in parallel-group superiority RCTs | Sample size calculation, effect size | Human | 194 |
| [42]  (Alam et al., 2014) | RCTs, Dermatologic Surgery | To assess the quality of reporting of randomization, blinding, sample size, and power analysis in RCTs | Randomization, sample size, power analysis | Human | 324 |
| [34]  (Rahi, 2017) | Paradigms, sampling issues | contribute a detailed systematic review of research paradigms, sampling, and instrument  development issues in the field of business research | Sampling techniques, sample size | Human | N/A |
| [35]  (Martin et al., n.d.) | Stepped-wedge cluster, randomized trails | Determine adherence to reporting each of the sample sizes recommended, identify the power methodology used, determine whether appropriate methodology is being used | Sample size calculation, significance level, power, treatment effect, cluster size | Human | 60 |
| [36]  (Williamson, 2003) | Nursing | discusses the theoretical limitations of the use of random sampling and probability theory in the production of a significance level (or P-value) in nursing research | Sampling, study design, quantitative research, | Human | 131 |
| [37]  (Malekinejad et al., 2008) | HIV biological or behavioral surveillance | operational and analytical characteristics of RDS studies and discuss factors that may affect recruitment | Chain-referral sampling, respondent-driven sampling, sample size | Human | 123 |
| [38]  (Vasileiou et al., 2018) | Interview based studies | aims to further contribute to the dialogue of sample size in qualitative research by offering empirical evidence around justify cation practices associated with sample size | Sample size justifications, sampling techniques | Human | 214 |
| [39]  (Barros et al., 2015) | Men who have Sex with Men, Sex Work, Sex Workers | assess if there were any relations between the study populations and the sampling methods used to recruit them. | Sampling, Sampling Techniques | Human | 268 |
| [40]  (Abdul Latif et al., 2011) | Physical Medicine, Rehabilitation, RCTs | To assess systematically the reporting of sample size calculation in randomized controlled trials (RCTs) in 5 leading journals in the field of physical medicine and rehabilitation | Study design, Sample size calculation, Results of trial, type of control, parameter for a priori calculation | Human | 111 |
| [41]  (Schuster et al., 2021) | Psychological depression treatment, RCTs | investigates the practice of SSP (Sample size Planning) in current trials for depression | Sample size, sample size planning | Human | 78 |

\*\*\* N/A = Not Available

Table 2. *Characteristics of Search Criteria across Included SLRs*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Author | Availability of search Strategy | PRISMA flowchart availability | Literature Sources Information | Number of databases searched | Number of reviewers involved | Tabulated description of individual studies | Tabulated quality assessment of individual studies | Funding reporting |
| [29]  (Copsey et al., 2018) | N | Y | Y | 7 | 4 | N | Y | Y |
| [30]  (McKeown et al., 2015) | N | N | Y | 3 | 3 | Y | N | N |
| [31]  (Guo et al., 2014) | Y | Y | Y | N/A | 2 | Y | Y | N |
| [32]  (Abdulatif et al., 2015) | Y | Y | Y | 1 | 2 | Y | N | Y |
| [42]  (Alam et al., 2014) | Y | N | Y | 1 | 2 | N | N | N |
| [34]  (Rahi, 2017) | N | N | N | N/A | N/A | N | N | N |
| [35]  (Martin et al., n.d.) | Y | Y | Y | 3 | 2 | N | Y | Y |
| [36]  (Williamson, 2003) | Y | N | Y | 1 | N/A | Y | N | N |
| [37]  (Malekinejad et al., 2008) | Y | N | Y | 3 | N/A | Y | Y | N |
| [38]  (Vasileiou et al., 2018) | Y | Y | Y | N/A | N/A | N | N | Y |
| [39]  (Barros et al., 2015) | Y | Y | Y | 11 | 3 | Y | N | N |
| [40]  (Abdul Latif et al., 2011) | Y | Y | Y | 1 | 4 | Y | N | N |
| [41]  (Schuster et al., 2021) | Y | N | Y | 4 | 2 | N | Y | Y |

\*\*\*Y = Yes

\*\*\*N = No

\*\*\* N/A = Not Available

Table 3. *Reported Outcomes in Included SLRs*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SLRs | Study Design | Population Size | Study Hypothesis | Power Calculation | Significance Level (One side or Two side Test) | Effect Size | Sample Size Formula | Sampling Techniques | Statement about Sample Size | Standard Deviation | Other Outcomes |
| [29]  (Copsey et al., 2018) | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ |  |  | ✔ | ✔ | Attrition, Noninferiority margin, mean difference |
| [30]  (McKeown et al., 2015) | ✔ |  | ✔ | ✔ | ✔ | ✔ |  |  | ✔ |  | Trail size, IQR, Variability, Treatment effect |
| [31]  (Guo et al., 2014) |  |  |  |  |  | ✔ |  |  |  |  | Variability, autocorrelation variance-covariance matrix, t-test, z-test, and F-statistics |
| [32]  (Abdulatif et al., 2015) |  |  | ✔ | ✔ |  | ✔ |  |  | ✔ |  | Variance in studies |
| [42]  (Alam et al., 2014) |  |  |  | ✔ | ✔ | ✔ |  |  |  |  | Post hoc power analysis, Jadad score |
| [34]  (Rahi, 2017) |  |  |  |  |  |  | ✔ | ✔ | ✔ |  | N/A |
| [35]  (Martin et al., n.d.) | ✔ |  |  |  | ✔ |  |  | ✔ | ✔ |  | Variation on design, Inter Quartile Range (ICC), Treatment effect |
| [36]  (Williamson, 2003) |  |  |  |  |  |  |  | ✔ |  |  | N/A |
| [37]  (Malekinejad et al., 2008) |  |  |  |  |  |  |  |  | ✔ |  | Respondent-Driven sampling Method |
| [38]  (Vasileiou et al., 2018) | ✔ |  |  |  |  |  |  |  | ✔ |  | N/A |
| [39]  (Barros et al., 2015) |  |  |  |  |  |  |  | ✔ |  |  | N/A |
| [40]  (Abdul Latif et al., 2011) |  |  |  | ✔ |  |  |  |  | ✔ |  | Trails control |
| [41]  (Schuster et al., 2021) |  |  |  | ✔ | ✔ |  |  |  | ✔ |  | Sample size with replacement |

\*\*\* ✔ = outcome present

\*\*\*N/A = Not Available

Table 4. *Summary of Evidence from Included SRLs*

|  |  |  |
| --- | --- | --- |
| **Phenomenon of Interest** | **Author,**  **year** | **Synthesized Findings** |
| Sample size determination | [29] | Study design, Population size, study hypothesis, power calculation, significance level, effect size, statement about samples size |
| Mean difference, standard deviation attrition, Noninferiority margin |
| [30] | Study design, Study hypothesis, Power calculation, Significance level. Statement about sample size, Variability |
| Treatment effect, Trail size, IQR |
| [31] | Effect size, Variability |
| autocorrelation variance-covariance matrix, t-test, z-test, and F-statistics |
| [32] | Study design, Study hypothesis, power calculation, statement about sample size |
| Variance in studies |
| [42] | Power calculation, significance level, effect size |
| Post hoc power analysis, Jadad score |
| [34] | Sample size formula, sampling techniques, statement about sample size |
| N/A |
| [35] | Study design, Significance level, sampling techniques, statement about sample size |
| Treatment effect, Variation on design, Inter Quartile Range (ICC) |
| [36] | Sampling techniques |
| N/A |
| [37] | statement about sample size |
| Respondent-Driven sampling Method |
| [38] | Study design, sampling techniques, statement about sample size |
| N/A |
| [39] | sampling techniques |
| N/A |
| [40] | Power calculation, significance level, statement about sample size |
| Trails control |
| [41] | Power calculation, significance level, statement about sample size, sample size with replacement |
| N/A |

\*\*\*Green = Effective and important

\*\*\*Red = less effective and less important

Table 5. *Quality Appraisal of the Includes Systematic Reviews using AMSTAR-2 Tool*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SLRs | Components of PICO | Review Protocol | Explanation of study design | Comprehensive literature search strategy | Study selection in duplicate | Data extraction in duplicate | List of excluded studies and justification | Study characteristics | Satisfactory technique for assessing the risk of bias | Source of funding | Appropriate methods | Assess potential impact of risk of bias on the results | Account for risk of bias when interpreting or discussing | Satisfactory explanation and discussion of any heterogeneity | Publication bias assessed and discussed | Potential sources of conflict of interest | Quality Assessment |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |  |
| [29] | Y | P | Y | Y | Y | Y | Y | Y | N | Y | N/A | N/A | Y | Y | Y | Y | Low |
| [30] | Y | N | Y | Y | Y | Y | N | Y | N | N | N/A | N/A | N | N | Y | N | Critically Low |
| [31] | Y | N | Y | Y | Y | Y | Y | Y | Y | N | N/A | N/A | Y | Y | Y | N | Low |
| [32] | Y | N | Y | Y | Y | Y | P | N | N | Y | N/A | N/A | Y | Y | N | Y | Critically Low |
| [42] | Y | N | Y | Y | Y | Y | N | N | N | N | N/A | N/A | N | Y | Y | N | Critically Low |
| [34] | Y | P | Y | N | N | N | N | Y | N | N | N/A | N/A | N | N | N | N | Critically Low |
| [35] | Y | N | Y | Y | Y | Y | Y | N | Y | Y | N/A | N/A | Y | Y | Y | Y | Low |
| [36] | Y | N | Y | Y | N | N | Y | N | N | N | N/A | N/A | N | Y | Y | N | Critically Low |
| [37] | Y | N | Y | Y | Y | Y | N | N | Y | N | N/A | N/A | Y | Y | Y | N | Critically Low |
| [38] | Y | N | Y | Y | Y | Y | Y | Y | Y | Y | N/A | N/A | Y | Y | Y | Y | Low |
| [39] | Y | P | Y | Y | Y | Y | Y | N | N | N | N/A | N/A | Y | Y | Y | N | Critically Low |
| [40] | Y | N | Y | Y | Y | Y | Y | N | N | N | N/A | N/A | Y | Y | Y | N | Critically Low |
| [41] | Y | N | Y | Y | Y | Y | N | N | Y | Y | N/A | N/A | Y | Y | N | Y | Critically Low |

\*\*\*Y = Yes, N = No; P = Partial yes, PICO = Population or participants, and conditions of interest; N/A = Not applicable

Figure 2: *Sample Size Determination Strategies Included by SLRs*

|  |  |
| --- | --- |
|  |  |
|  |  |

Figure 3: *Difference Parameter Explanation Included by SLRs*

|  |  |
| --- | --- |
|  |  |
|  |  |

Figure 4: *Important Items of SSD included by SLRs*

|  |  |
| --- | --- |
|  |  |

**Discussion**

This umbrella review systematically summarized the sample size determination method in the health and medical field. Through this review, the target is to provide information about sample size determination on medical or health-related studies from all SLRs in a single study and identify whether previous studies provided necessary information about sample size determination.

We included 13 SLRs, some of them including only sample size calculation methods for RCTs, and 5 other studies included medical or health-related studies to show the required process for sample size determination. In cross-sectional studies, the main aim of sample size calculation is to estimate the prevalence of unknown parameters from the target population using a random sample. So, it is necessary to identify the adequate sample size to measure the prevalence with better precision. A case-control study is another type of epidemiological observational study. The target is to identify the risk factors which are correlated with specific diseases [43]. When calculating the sample size for unmatched case-control studies, researchers must make certain assumptions. These include the number of cases and controls who are assumed to have encountered risk variables from similar studies or pilot studies (as well as the assumed odds ratio, or OR), the level of confidence (nearly 95%), and the proposed power of the study (which would be from 80%) [44]. In clinical trials, inadequate sample size may fail to identify the research hypothesis or may fail to estimate the impact or correlation. A randomized controlled trial's minimal sample size calculation requires knowledge of the following: power, significance level, population's underlying event rate, and size of desired treatment impact. Aside from this, other variables such as anticipated compliance rates and, less frequently, an unbalanced allocation ratio should be taken into account when calculating the sample size [45].

Independently of the heterogeneity of included SLRs in terms of search, inclusion, or exclusion criteria, and analytical methods, we found that the majority of studies did not include important factors for sample size determination. We found that almost all the studies except one, did not include a single sample size formula for accurately determining sample size which is very important for cross-sectional, case-control, clinical trials, and cohort studies. On the other hand, representative samples from the population were another crucial part of proper sample size determination. For a representative sample size investigator needs different sampling techniques [28]. However, most of the studies did not incorporate sampling techniques for sample size calculation. Moreover, power calculation is another important part of SSD. If the power increases, then the sample size also grows [28]. However, most of the studies had no information about statistical power calculation. Furthermore, those studies that included power calculation did not include satisfactory information about it for sample size determination. Researchers would not get proper knowledge of which value they would use for sample size determination. Some studies, on the other hand, did not explain about study design. Sample size depends on the specific study design which is another drawback for the included SLRs [24]. If one did not know about the population size he/she could not identify the adequate sample size [44]. Although, there were ways to determine sample size using different methods or formulas [34]. Information about population size could help to accurately measure study sample size which was missed by most of the SLR studies. The majority of the studies had no information about standard deviation, significance level, and study hypothesis. A research hypothesis is conducted to summarize the research's basic elements such as sampling techniques, design of the study, sample size, independent and dependent variables [24]. On the other hand, the diversity in which the population’s true value is expected to be known as the significance level [28]. In contrast, underestimating the standard deviation could affect the trial’s ability to detect important treatment differences in randomized control trials while overestimating the standard deviation could lead to an inflated sample size, requiring more participants than necessary for the study, which could waste resources and time [29]. However, some studies properly explained small or large sample sizes and effect sizes which would be a strong outcome from those studies from sample size determination. The lack of methodological consistency in sample size calculation across the reviewed SLRs suggested that researchers often did not prioritize or had clear frameworks for determining appropriate sample sizes. This gap undermined the robustness and reproducibility of findings, particularly in meta-analyses and quantitative SLRs where sample size was important. Researchers conducting Systematic Literature Reviews (SLRs) should be required to follow standardized guidelines for calculating and reporting sample sizes to ensure methodological rigor, enhance reproducibility, and improve the reliability of findings.

The umbrella review's individual SLR selections generally received low scores and varied widely in quality, primarily because different considerations (such as study design, lists of excluded studies, funding details, and an explanation of the impact of heterogeneity in the data analysis plan) were not adequately justified or provided.

**Recommendations**

Researchers conducting systematic literature reviews (SLRs) should be required to report sample size calculation methods as a separate and detailed component of their methodology. Specifically, researchers must provide explicit details of the formulas or statistical software used for sample size calculation. This includes the parameters considered, such as population size, expected effect size, and error margins. Clarity of the sampling process, including the rationale for the chosen sample size, the assumptions made (e.g., anticipated effect sizes), desired power levels (commonly 80% or higher), and significance thresholds (e.g., α = 0.05). This transparency will not only ensure scientific rigor but also allow for easier replication and evaluation of study robustness. Incorporate power analysis into the sample size determination process. Power analysis should be conducted prior to data collection to ensure that the study is adequately powered to detect meaningful effects. This involves specifying the minimum detectable effect size, the acceptable risk of Type I (α) and Type II (β) errors, and the expected variability within the data. These recommendations aim to address the common oversight in SLRs, where insufficient sample sizes lead to underpowered studies and unreliable results. By adopting standardized reporting and rigorous methodologies, researchers can improve the validity, reliability, and generalizability of their findings.

**Strength and Limitations**

There were a number of noteworthy strengths in this review. After a comprehensive and exhaustive search for relevant literature across international databases using replicable criteria, papers were selected for inclusion in this review. On the other hand, this review identified key features of included systematic reviews, which offer a structured synthesis that highlights gaps and inconsistencies. Moreover, the important appraisal of methodologies and the identification of trends in sample size determination provide practical insights that extend the review's utility across clinical trials, public health, and epidemiology. However, there are also certain drawbacks to this study. Three five provided the required papers, which we retrieved. More papers on this topic can be found on other websites. Studies that could have provided updated or contradicted findings pertinent to the evaluation were not included because they did not meet specific temporal requirements when they were included in the first place. It is possible that the search was restricted to studies published in particular languages, which could have resulted in language bias and excluded pertinent papers published in other languages.

**Conclusion**

One of the critical aspects of research is the accurate estimation of sample size, which plays a pivotal role in ensuring the validity and reliability of study findings. This review underscores a significant gap in current systematic literature reviews which is the lack of attention to appropriate sample size determination (SSD). The absence of clear guidelines and rigorous application of SSD methods compromises the reliability of the conclusions drawn from these reviews for sample size determination. When researchers overlook SSD from these SLRs, it can lead to misleading outcomes and hinder the replication of results. Proper study design, understanding of population parameters, sampling techniques, hypothesis testing, and knowledge of sample size calculations are essential to ensure robust research, particularly in medical and health-related fields. Unfortunately, many studies in this review neglected these critical elements, leading to potential inaccuracies in their findings. Therefore, a comprehensive study of SSD is vital, especially in medical ore health-related research, where precise estimates are crucial for reliable outcomes.

**References**

1. Daniel WW, Cross CL. Biostatistics: a foundation for analysis in the health sciences. Wiley; 2018.

2. Pagano M, Gauvreau K, Mattie H. Principles of biostatistics. Chapman and Hall/CRC; 2022.

3. Altman DG. Practical statistics for medical research. Chapman and Hall/CRC; 1990.

4. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: Why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013;14:365–76.

5. S BN, S AE. Relevance of Sample Size Determination in. 2010.

6. Rosner BA. Fundamentals of biostatistics. Thomson-Brooks/Cole Belmont, CA; 2006.

7. Feinstein AR. Principles of medical statistics. Chapman and Hall/CRC; 2001.

8. Singh AS, Masuku MB. International Journal of Economics, Commerce and Management SAMPLING TECHNIQUES & DETERMINATION OF SAMPLE SIZE IN APPLIED STATISTICS RESEARCH: AN OVERVIEW.

9. Freund RJ, Wilson WJ. Statistical methods. Elsevier; 2003.

10. Wilson Vanvoorhis CR, Morgan BL. Understanding Power and Rules of Thumb for Determining Sample Sizes. 2007.

11. Serdar CC, Cihan M, Yücel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. Biochem Med (Zagreb). 2021;31:27–53.

12. Association AP. Publication manual of the American psychological association. American Psychological Association; 2010.

13. Curtis MJ, Bond RA, Spina D, Ahluwalia A, Alexander SPA, Giembycz MA, et al. Experimental design and analysis and their reporting: new guidance for publication in BJP LINKED EDITORIALS. 2015. https://doi.org/10.1111/bph.

14. Freiman JA, Chalmers TC, Smith HA, Kuebler RR. The importance of beta, the type II error, and sample size in the design and interpretation of the randomized controlled trial: survey of two sets of “negative” trials. In: Medical uses of statistics. CRC Press; 2019. p. 357–89.

15. Charan J, Kaur R, Bhardwaj P, Singh K, Ambwani SR, Misra S. Sample Size Calculation in Medical Research: A Primer. Ann Natl Acad Med Sci. 2021;57:74–80.

16. Bujang MA. A step-by-step process on sample size determination for medical research. Malaysian Journal of Medical Sciences. 2021;28:15–27.

17. Bujang MA, Baharum N. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. 2017.

18. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. JBI Evid Implement. 2015;13:132–40.

19. Cairns J, Warren J, Garthwaite K, Greig G, Bambra C. Go slow: An umbrella review of the effects of 20 mph zones and limits on health and health inequalities. Journal of Public Health (United Kingdom). 2015;37:515–20.

20. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. bmj. 2021;372.

21. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. bmj. 2017;358.

22. Ma L-L, Wang Y-Y, Yang Z-H, Huang D, Weng H, Zeng X-T. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? Mil Med Res. 2020;7:1–11.

23. De Santis KK, Lorenz RC, Lakeberg M, Matthias K. The application of AMSTAR2 in 32 overviews of systematic reviews of interventions for mental and behavioural disorders: A cross‐sectional study. Res Synth Methods. 2022;13:424–33.

24. Rao UK. Concepts in sample size determination. Indian Journal of Dental Research. 2012;23:660–4.

25. Arivazhahan A, Senthilvelan J. Study Designs. In: Lakshmanan M, Shewade DG, Raj GM, editors. Introduction to Basics of Pharmacology and Toxicology: Volume 3 : Experimental Pharmacology : Research Methodology and Biostatistics. Singapore: Springer Nature Singapore; 2022. p. 799–817.

26. Singh AS, Masuku MB. International Journal of Economics, Commerce and Management SAMPLING TECHNIQUES & DETERMINATION OF SAMPLE SIZE IN APPLIED STATISTICS RESEARCH: AN OVERVIEW.

27. Israel GD. Determining Sample Size 1 The Level of Precision. 1992.

28. Singh AS, Masuku MB. International Journal of Economics, Commerce and Management SAMPLING TECHNIQUES & DETERMINATION OF SAMPLE SIZE IN APPLIED STATISTICS RESEARCH: AN OVERVIEW.

29. Copsey B, Thompson JY, Vadher K, Ali U, Dutton SJ, Fitzpatrick R, et al. Sample size calculations are poorly conducted and reported in many randomized trials of hip and knee osteoarthritis: results of a systematic review. Journal of Clinical Epidemiology. 2018;104:52–61.

30. McKeown A, Gewandter JS, McDermott MP, Pawlowski JR, Poli JJ, Rothstein D, et al. Reporting of sample size calculations in analgesic clinical trials: ACTTION systematic review. Journal of Pain. 2015;16:199-206.e7.

31. Guo Q, Thabane L, Hall G, McKinnon M, Goeree R, Pullenayegum E. A systematic review of the reporting of sample size calculations and corresponding data components in observational functional magnetic resonance imaging studies. Neuroimage. 2014;86:172–81.

32. Abdulatif M, Mukhtar A, Obayah G. Pitfalls in reporting sample size calculation in randomized controlled trials published in leading anaesthesia journals: A systematic review. Br J Anaesth. 2015;115:699–707.

33. Alam M, Rauf M, Ali S, Nodzenski M, Minkis K. A systematic review of reporting in randomized controlled trials in Dermatologic Surgery: Jadad scores, power analysis, and sample size determination. Dermatologic Surgery. 2014;40:1299–305.

34. Rahi S. Research Design and Methods: A Systematic Review of Research Paradigms, Sampling Issues and Instruments Development. International Journal of Economics & Management Sciences. 2017;06.

35. Martin J, Taljaard M, Girling A, Hemming K. Systematic review finds major deficiencies in sample size methodology and reporting for stepped-wedge cluster randomised trials. https://doi.org/10.1136/bmjopen-2015.

36. Williamson GR. Misrepresenting random sampling? A systematic review of research papers in the Journal of Advanced Nursing. Journal of Advanced Nursing. 2003;44:278–88.

37. Malekinejad M, Johnston LG, Kendall C, Kerr LRFS, Rifkin MR, Rutherford GW. Using respondent-driven sampling methodology for HIV biological and behavioral surveillance in international settings: A systematic review. AIDS Behav. 2008;12 SUPPL. 1.

38. Vasileiou K, Barnett J, Thorpe S, Young T. Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period. BMC Med Res Methodol. 2018;18.

39. Barros AB, Dias SF, Martins MRO. Hard-to-reach populations of men who have sex with men and sex workers: A systematic review on sampling methods. Syst Rev. 2015;4.

40. Abdul Latif L, Daud Amadera JE, Pimentel D, Pimentel T, Fregni F. Sample size calculation in physical medicine and rehabilitation: A systematic review of reporting, characteristics, and results in randomized controlled trials. Archives of Physical Medicine and Rehabilitation. 2011;92:306–15.

41. Schuster R, Kaiser T, Terhorst Y, Messner EM, Strohmeier LM, Laireiter AR. Sample size, sample size planning, and the impact of study context: Systematic review and recommendations by the example of psychological depression treatment. Psychological Medicine. 2021;51:902–8.

42. Alam M, Rauf M, Ali S, Nodzenski M, Minkis K. A systematic review of reporting in randomized controlled trials in Dermatologic Surgery: Jadad scores, power analysis, and sample size determination. Dermatologic Surgery. 2014;40:1299–305.

43. Naing L, Winn T, Rusli BN. Practical issues in calculating the sample size for prevalence studies. Archives of orofacial Sciences. 2006;1:9–14.

44. Lwanga SK, Lemeshow S, Organization WH. Sample size determination in health studies: a practical manual. World Health Organization; 1991.

45. Pourhoseingholi MA, Vahedi M, Rahimzadeh M. Sample size calculation in medical studies. 2013.

Table 6. *PRISMA 2020 Checklist*

| **Section and Topic** | **Item #** | **Checklist item** | **Location where item is reported** |
| --- | --- | --- | --- |
| **TITLE** | | |  |
| Title | 1 | Identify the report as a systematic review. | 1 |
| **ABSTRACT** | | |  |
| Abstract | 2 | See the PRISMA 2020 for Abstracts checklist. | 2 |
| **INTRODUCTION** | | |  |
| Rationale | 3 | Describe the rationale for the review in the context of existing knowledge. | 3 |
| Objectives | 4 | Provide an explicit statement of the objective(s) or question(s) the review addresses. | 4 |
| **METHODS** | | |  |
| Eligibility criteria | 5 | Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses. | 5 |
| Information sources | 6 | Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted. | 4 |
| Search strategy | 7 | Present the full search strategies for all databases, registers and websites, including any filters and limits used. | 5 |
| Selection process | 8 | Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process. | 5 |
| Data collection process | 9 | Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process. | 13 |
| Data items | 10a | List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect. | 6-12 |
| 10b | List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information. | N/A |
| Study risk of bias assessment | 11 | Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process. | 5-6 |
| Effect measures | 12 | Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results. | N/A |
| Synthesis methods | 13a | Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)). | 19-25, 27-30 |
| 13b | Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions. | 19-25, 27-30 |
| 13c | Describe any methods used to tabulate or visually display results of individual studies and syntheses. | 19-25, 27-30 |
| 13d | Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used. | 19-25, 27-30 |
| 13e | Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression). | 19-25, 27-30 |
| 13f | Describe any sensitivity analyses conducted to assess robustness of the synthesized results. | 19-25, 27-30 |
| Reporting bias assessment | 14 | Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases). | N/A |
| Certainty assessment | 15 | Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. | 14-15 |
| **RESULTS** | | |  |
| Study selection | 16a | Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram. | 13 |
| 16b | Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded. |  |
| Study characteristics | 17 | Cite each included study and present its characteristics. | 14 |
| Risk of bias in studies | 18 | Present assessments of risk of bias for each included study. | 25-26 |
| Results of individual studies | 19 | For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots. | 13-14 |
| Results of syntheses | 20a | For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. | 15-18 |
| 20b | Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect. | 15-18 |
| 20c | Present results of all investigations of possible causes of heterogeneity among study results. | 15-18 |
| 20d | Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results. | 15-18 |
| Reporting biases | 21 | Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed. | N/A |
| Certainty of evidence | 22 | Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed. | N/A |
| **DISCUSSION** | | |  |
| Discussion | 23a | Provide a general interpretation of the results in the context of other evidence. | 31-32 |
| 23b | Discuss any limitations of the evidence included in the review. | 32 |
| 23c | Discuss any limitations of the review processes used. | 33 |
| 23d | Discuss implications of the results for practice, policy, and future research. | 33 |
| **OTHER INFORMATION** | | |  |
| Registration and protocol | 24a | Provide registration information for the review, including register name and registration number, or state that the review was not registered. | 4 |
| 24b | Indicate where the review protocol can be accessed, or state that a protocol was not prepared. | N/A |
| 24c | Describe and explain any amendments to information provided at registration or in the protocol. | N/A |
| Support | 25 | Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review. | N/A |
| Competing interests | 26 | Declare any competing interests of review authors. | N/A |
| Availability of data, code and other materials | 27 | Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review. | N/A |