I'm not a bot

Researchers should consider many factors for the research to be successful when conducting research. One of the essential steps is identifying the target population for the research study early on while planning a market research study and thinking about goals and objectives. A target population sets a clear direction on the scope and object of the research and data types. This article will explain what the target population is in a research example and Frequently ask questions about the target population with all the details. What is the target population in research?The population that the intervention is intended to study and take conclusions from is known as the target population. A target population, also referred to as a target audience, is a group of people with particular characteristics that may be effectively defined to distinguish them from the general population. The target population aims to comprehend and assess their preferences and behaviors to promote a particular good or service or to research a specific feature that frequently manifests itself in their behavior, such as behavior patterns. It is a notion that has to do with business market segmentation tactics. A target population is typically a group or collection of factors you want to learn more about. The target population is a subset of the general public identified as the targeted market for a given product, advertising, or research. It is a subset of the entire population chosen to serve as the objective audience.How to determine the target population in a research designIdentifying the target population requires careful preparation, precise research questions, and a robust study design. It will assist you in obtaining trustworthy and accurate information that responds to your study issue or concern.The target group is frequently chosen based on characteristics or demographics such as age, gender, employment, income, or health condition. The research's conclusions are then extrapolated to the broader population from whom the target sample was selected. How to choose your target populationNarrowing objectives is one of the most sensible things for businesses to do when identifying a target market so that their goods or services may be promoted effectively. The quality of the service will rise, and controllability will be significantly aided by an attempt to offer a more specific target population. First, you should define the target population you wish to attract with your marketing campaigns to create a target audience. The following steps will assist you in determining the population. Set business or brand goals: This is essential in creating a content strategy or developing a marketing plan. Define your study’s objective: You should clearly state your study's motivation and objective.Determine the population features: Determine the features of your intended audience. Gender, age, educational background, socio-economic situation, and career should be determined.Market research: Market research is the process of collecting and analyzing data about the market in which a business operates to understand the target audience and the potential of a specific product or service. Target population determination is an ongoing process that requires research, analysis, and adjustments. You can identify the specific group of people you want to reach with your marketing efforts and create a strategy to interact with them effectively. Here are 3 target population examples. Determining target population according to specific features: Let's say you are researching school service fees. First, you should consider specific characteristics when determining the target population. As parents whose children go to school and parents who send their children to school by service, you can narrow down your target audience in your research. Determining target population according to characteristic features: Another example is you are launching a new perfume as a cosmetics company. Your further perfume appeals to young women. You should determine the age group of the target population. Then you focus on gender, and maybe you can consider the women using perfume. You can choose your target population in this way. Determining target population according to target audience: The last example is that you are a business manager and want to measure the customer's satisfaction with a product. The target population would be all the customers of a particular business. The target population is a big group of people that researchers are looking at. The target population is very similar to other population types. There are frequently asked questions about the target population.In quantitative research, the term "target population" refers to the group of people or things the researcher wishes to analyze and draw conclusions about based on the data collected.In qualitative research, the term "target population" refers to the group of participants or subjects who are of interest to the researcher to explore, comprehend, and examine subjective experiences, behaviors, attitudes, or occurrences. Audiences often choose The target group based on criteria like gender, age, education, and other features.The difference between the sample population and the target population is that the target population is a more significant subset of the target population chosen for the research project. In contrast, the sample population is the entire group of people or things the researcher is interested in working with.The difference between the target and accessible population is that an accessible population is that portion of the target population that may theoretically be included in the study. In contrast, a target population is the whole set of instances the researcher desires to analyze.The difference between the target population and the sampling frame is that the sampling frame is a subset of the target population. The sampling frame typically serves as the starting point for the sampling process. The sample is a portion of the population and the individuals or items you have access to. In contrast, the target population is the entire group of people or objects to which you desire to generalize the results of your study. The difference between the population and the sample is the target population is much larger than the sample. The target population is the entire group or items the researcher intends to examine and draw conclusions about. On the other hand, a sample is a subset of the target population that is selected for study.ConclusionIn conclusion, a researcher's target population is the group they are trying to comprehend. When the target audience is analyzed, new information may be found that enables the business to launch various advertising campaigns tailored to the target population's income levels and attitudes. This article has explained the definition of the target population, How to determine the target population in a research design, and examples of the target population. When you read this article, you can learn all details about the target population. Choosing the optimal number of samples while testing your medical device product boils down to a common trade-off: cost vs. benefit. Staying within your testing budget is important but ensuring that your data is robust enough to withstand scrutiny from the regulatory body is paramount and avoids costly delays with re-testing. As a 3rd party contract testing lab, DDL will not make the final determination of sample size for our clients. However, there are common tendencies which we have observed in the sampling that our customers see when submitting device applications. We will outline these tendencies and other observations in the hope that they will be useful for those who are researching how to best structure their testing regimen. Before discussing the common sampling trends DDL has observed, it is relevant to address pre-design verification activities. Feasibility and characterization testing are often conducted during pre-design verification studies using lower sample quantities than Design Verification. When determining the sample size during Pre-DV, it is important to consider multiple samples per unique design input (e.g. mold cavity) as it may help mitigate any false positives and issues down the road. In addition, the risk of the device’s manufacturing process should also be considered when determining the sample size for these early, initial phase studies. During design verification studies, the first factor that needs to be made when determining sample size is whether the test data you require will be attribute or variable. Attribute data, also called binomial data, are qualitative results. Pass/fail or go/no-go are common types of attribute data – for example whether a measured dimension falls within the tolerances on the drawing. Variable data output is shown in values. For example, the seal strength of a heat-sealed pouch or the tensile strength of a poly film is typically shown as variable results. The next factor in determining sample size is evaluating your risk tolerance. An internal regulatory or quality department or a consultant will provide good guidance on risk tolerance. In addition, ISO 14971 is called out by a significant number of ISO standards on how to apply risk management to medical devices which may ultimately be helpful in determining the appropriate sample size. As one can expect, a high-risk product requires more test samples in order to achieve an acceptable confidence interval. Statistically speaking, a higher risk product means that you need to assign a more stringent acceptable quality level (AQL), p0, to your experimental design. The AQL represents the maximum allowable proportion of defective items in a lot. For example, if a maximum of 5% of your parts can be defective, your p0 value would be 0.05. From there, using the cumulative geometric distribution function, you can determine your optimum sample size for an attribute test. Similar principles can be applied to variable testing when it comes to risk but instead of a sample size increasing with a higher risk product, the Cpk or K-value will change. Ultimately, the change in these values will make sure the device is still held to higher levels of risk that will ensure that the device is still adequate for use against the appropriate risk assessment. The most common sample sizes DDL sees for attribute tests are 29 and 59. For example, to obtain a 95% confidence that your product’s passing rate is at least 95% – commonly summarized as “95/95”, 59 samples must be tested and must pass the test. If your product has lower risk and you are able to accept a lower passing rate of 90%, only 29 passing samples are needed to obtain 95% confidence, or “95/90”. These numbers all assume that there will be no failures in any of the samples. That, unfortunately, is not always the case. In the event of an isolated failure, a different equation – the negative binomial distribution, must be used. In order to maintain the same confidence intervals as stated above with one failure, the sample size is 46 for a p value of 0.10 and 93 for a p value of 0.05 and increases with additional failures. At the end of the day, determining sample size for an attribute test is a straightforward task once the statistical requirements are known, but its importance cannot be overstated. Conversely, with variable testing, there are many product dependent factors to consider before arriving at a sample size quantity. Determining your sample size using either statistical method will not only ensure that regulatory requirements are met, it also provides evidence that the quality of the product is high increasing patient safety. Research population and sample serve as the cornerstones of any scientific inquiry. They hold the power to unlock the mysteries hidden within data. Understanding the dynamics between the research population and sample is crucial for researchers. It ensures the validity, reliability, and generalizability of their findings. In this article, we uncover the profound role of the research population and sample, unveiling their differences and importance that reshapes our understanding of complex phenomena. Ultimately, this empowers researchers to make informed conclusions and drive meaningful advancements in our respective fields. What Is Population? The research population, also known as the target population, refers to the entire group or set of individuals, objects, or events that possess specific characteristics and are of interest to the researcher. It represents the larger population from which a sample is drawn. The research population is defined based on the research objectives and the specific parameters or attributes under investigation. For example, in a study on the effects of a new drug, the research population would encompass all individuals who could potentially benefit from or be affected by the medication. When Is Data Collection From a Population Preferred? In certain scenarios where a comprehensive understanding of the entire group is required, it becomes necessary to collect data from a population. Here are a few situations when one prefers to collect data from a population: 1. Small or Accessible Population When the research population is small or easily accessible, it may be feasible to collect data from the entire population. This is often the case in studies conducted within specific organizations, small communities, or well-defined groups where the population size is manageable. 2. Census or Complete Enumeration In some cases, such as government surveys or official statistics, a census or complete enumeration of the population is necessary. This approach aims to gather data from every individual or entity within the population. This is typically done to ensure accurate representation and eliminate sampling errors. 3. Unique or Critical Characteristics If the research focuses on a specific characteristic or trait that is rare and critical to the study, collecting data from the entire population may be necessary. This could be the case in studies related to rare diseases, endangered species, or specific genetic markers. 4. Legal or Regulatory Requirements Certain legal or regulatory frameworks may require data collection from the entire population. For instance, government agencies might need comprehensive data on income levels, demographic characteristics, or healthcare utilization for policy-making or resource allocation purposes. 5. Precision or Accuracy Requirements In situations where a high level of precision or accuracy is necessary, researchers may opt for population-level data collection. By doing so, they mitigate the potential for sampling error and obtain more reliable estimates of population parameters. What Is a Sample? A sample is a subset of the research population that is carefully selected to represent its characteristics. Researchers study this smaller, manageable group to draw inferences that they can generalize to the larger population. The selection of the sample must be conducted in a manner that ensures it accurately reflects the diversity and pertinent attributes of the research population. By studying a sample, researchers can gather data more efficiently and cost-effectively compared to studying the entire population. The findings from the sample are then extrapolated to make conclusions about the larger research population. What Is Sampling and Why Is It Important? Sampling refers to the process of selecting a sample from a larger group or population of interest in order to gather data and make inferences. The goal of sampling is to obtain a sample that is representative of the population, meaning that the sample accurately reflects the key attributes, variations, and proportions present in the population. By studying the sample, researchers can draw conclusions or make predictions about the larger population with a certain level of confidence. Collecting data from a sample, rather than the entire population, offers several advantages and is often necessary due to practical constraints. Here are some reasons to collect data from a sample: 1. Cost and Resource Efficiency Collecting data from an entire population can be expensive and time-consuming. Sampling allows researchers to gather information from a smaller subset of the population, reducing costs and resource requirements. It is often more practical and feasible to collect data from a sample, especially when the population size is large or geographically dispersed. 2. Time Constraints Conducting research with a sample allows for quicker data collection and analysis compared to studying the entire population. It saves time by focusing efforts on a smaller group, enabling researchers to obtain results more efficiently. This is particularly beneficial in time-sensitive research projects or situations that necessitate prompt decision-making. 3. Manageable Data Collection Working with a sample makes data collection more manageable. Researchers can concentrate their efforts on a smaller group, allowing for more detailed and thorough data collection methods. Furthermore, it is more convenient and reliable to store and conduct statistical analyses on smaller datasets. This also facilitates in-depth insights and a more comprehensive understanding of the research topic. 4. Statistical Inference Collecting data from a well-selected and representative sample enables valid statistical inference. By using appropriate statistical techniques, researchers can generalize the findings from the sample to the larger population. This allows for meaningful inferences, predictions, and estimation of population parameters, thus providing insights beyond the specific individuals or elements in the sample. 5. Ethical Considerations In certain cases, collecting data from an entire population may pose ethical challenges, such as invasion of privacy or burdening participants. Sampling helps protect the privacy and well-being of individuals by reducing the burden of data collection. It allows researchers to obtain valuable information while ensuring ethical standards are maintained. Key Steps Involved in the Sampling Process Sampling is a valuable tool in research; however, it is important to carefully consider the sampling method, sample size, and potential biases to ensure that the findings accurately represent the larger population and are valid for making conclusions and generalizations. While the specific steps may vary depending on the research context, here is a general outline of the sampling process: 1. Define the Population Clearly define the target population for your research study. The population should encompass the group of individuals, elements, or units that you want to draw conclusions about. 2. Define the Sampling Frame Create a sampling frame, which is a list or representation of the individuals or elements in the target population. The sampling frame should be comprehensive and accurately reflect the population you want to study. 3. Determine the Sampling Method Select an appropriate sampling method based on your research objectives, available resources, and the characteristics of the population. You can perform sampling by either utilizing probability-based or non-probability-based techniques. Common sampling methods include random sampling, stratified sampling, cluster sampling, and convenience sampling. 4. Determine Sample Size Determine the desired sample size based on statistical considerations, such as the level of precision required, desired confidence level, and expected variability within the population. Larger sample sizes generally reduce sampling error but may be constrained by practical limitations. 5. Collect Data Once the sample is selected using the appropriate technique, collect the necessary data according to the research design and data collection methods. Ensure that you use standardized and consistent data collection process that is also appropriate for your research objectives. 6. Analyze the Data Perform the necessary statistical analyses on the collected data to derive meaningful insights. Use appropriate statistical techniques to make inferences, estimate population parameters, test hypotheses, or identify patterns and relationships within the data. Population vs Sample — Differences and examples While the population provides a comprehensive overview of the entire group under study, the sample, on the other hand, allows researchers to draw inferences and make generalizations about the population. Researchers should employ careful sampling techniques to ensure that the sample is representative and accurately reflects the characteristics and variability of the population. Research Study: Investigating the prevalence of stress among high school students in a specific city and its impact on academic performance. Population: All high school students in a particular city Sampling Frame: The sampling frame would involve obtaining a comprehensive list of all high schools in the specific city. A random selection of schools would be made from this list to ensure representation from different areas and demographics of the city. Sample: Randomly selected 500 high school students from different schools in the city The sample represents a subset of the entire population of high school students in the city. Example 2: Research Study: Assessing the effectiveness of a new medication in managing symptoms and improving quality of life in patients with the specific medical condition. Population: Patients diagnosed with a specific medical condition Sampling Frame: The sampling frame for this study would involve accessing medical records or databases that include information on patients diagnosed with the specific medical condition. Researchers would select a convenient sample of patients who meet the inclusion criteria from the sampling frame. Sample: Convenient sample of 100 patients from a local clinic who meet the inclusion criteria for the study The sample consists of patients from the larger population of individuals diagnosed with the medical condition. Example 3: Research Study: Investigating community perceptions of safety and satisfaction with local amenities in the neighborhood. Population: Residents of a specific neighborhood Sampling Frame: The sampling frame for this study would involve obtaining a list of residential addresses within the specific neighborhood. Various sources such as census data, voter registration records, or community databases offer the means to obtain this information. From the sampling frame, researchers would randomly select a cluster sample of households to ensure representation from different areas within the neighborhood. Sample: Cluster sample of 50 households randomly selected from different blocks within the neighborhood The sample represents a subset of the entire population of residents living in the neighborhood. Summary To summarize, sampling allows for cost-effective data collection, easier statistical analysis, and increased practicality compared to studying the entire population. However, despite these advantages, sampling is subject to various challenges. These challenges include sampling bias, non-response bias, and the potential for sampling errors. To minimize bias and enhance the validity of research findings, researchers should employ appropriate sampling techniques, clearly define the population, establish a comprehensive sampling frame, and monitor the sampling process for potential biases. Validating findings by comparing them to known population characteristics can also help evaluate the generalizability of the results. Properly understanding and implementing sampling techniques ensure that research findings are accurate, reliable, and representative of the larger population. By carefully considering the choice of population and sample, researchers can draw meaningful conclusions and, consequently, make valuable contributions to their respective fields of study. Now, it’s your turn! Take a moment to think about a research question that interests you. Consider the population that would be relevant to your inquiry. Who would you include in your sample? How would you go about selecting them? Reflecting on these aspects will help you appreciate the intricacies involved in designing a research study. Let us know about it in the comment section below or reach out to us using #AskEnago and tag @EnagoAcademy on Twitter, Facebook, and Quora. As a library, NLM provides access to scientific literature. Inclusion in an NLM database does not imply endorsement of, or agreement with, the contents by NLM or the National Institutes of Health. Learn more: PMC Disclaimer | PMC Copyright Notice The importance of estimating sample sizes is rarely understood by researchers, when planning a study. This paper aims to highlight the centrality of sample size estimations in health research. Examples that help in understanding the basic concepts involved in their calculation are presented. The scenarios covered are based more on the epidemiological reasoning and less on mathematical formulae. Proper calculation of the number of participants in a study diminishes the likelihood of errors, which are often associated with adverse consequences in terms of economic, ethical and health aspects. Keywords: Cross-sectional studies, Dermatology, Epidemiology, Prevalence, Risk factors, Sampling studies Investigations in the health field are oriented by research problems or questions, which should be clearly defined in the study project. Sample size calculation is an essential item to be included in the project to reduce the probability of error, respect ethical standards, define the logistics of the study and, last but not least, improve its success rates, when evaluated by funding agencies. Let us imagine that a group of investigators decides to study the frequency of sunscreen use and how the use of this product is distributed in the "population". In order to carry out this task, the authors define two research questions, each of which involving a distinct sample size calculation: 1) What is the proportion of people that use sunscreen in the population?; and, 2) Are there differences in the use of sunscreen between men and women, or between individuals that are white or of another skin color group, or between the wealthiest and the poorest, or between people with more and less years of schooling? Before doing the calculations, it will be necessary to review a few fundamental concepts and identify which are the required parameters to determine them. First of all, we must define what is a population. Population is the group of individuals restricted to a geographical region (neighborhood, city, state, country, continent etc.), or certain institutions (hospitals, schools, health centers etc.), that is, a set of individuals that have at least one characteristic in common. The target population corresponds to a portion of the previously mentioned population, about which one intends to draw conclusions, that is to say, it is a part of the population whose characteristics are an object of interest of the investigator. Finally, study population is that which will actually be part of the study, which will be evaluated and will allow conclusions to be drawn about the target population, as long as it is representative of the latter. Figure 1 demonstrates how these concepts are interrelated. Graphic representation of the concepts of population, target population and study populationWe will now separately consider the required parameters for sample size calculation in studies that aim at estimating the frequency of events (prevalence of health outcomes or behaviors, for example), to test associations between risk/protective factors and dichotomous health conditions (yes/no), as well as with health outcomes measured in numerical scales.1 The formulas used for these calculations may be obtained from different sources - we recommend using the free online software OpenEpi (www.openepi.com).2 When approaching the first research question defined at the beginning of this article (What is the proportion of people that use sunscreen?), the investigators need to conduct a prevalence study. In order to do this, some parameters must be defined to calculate the sample size, as demonstrated in chart 1. Description of different parameters to be considered in the calculation of sample size for a study aiming at estimating the frequency of health ouctomes, behaviors or conditions Parameter Description Remark Population size Total population size from which the sample will be drawn and about which researchers will draw conclusions (target population) Information regarding population size may be obtained based on secondary data from hospitals, health centers, census surveys (population, schools etc.). The smaller the target population (for example, less than 100 individuals), the larger the sample size will proportionally be. Expected prevalence of outcome or event of interest The study outcome must be a percentage, that is, a number that varies from 0% to 100%. Information regarding expected prevalence rates should be obtained from the literature or by carrying out a pilot-study. When this information is not available in the literature or a pilot-study cannot be carried out, the value that maximizes sample size is used (50% for a fixed value of sample error). Sample error for estimate The value we are willing to accept as error in the estimate obtained by the study. The smaller the sample error, the larger the sample size and the greater the precision. In health studies, values between two and five percentage points are usually recommended. Significance level It is the probability that the expected prevalence will be within the error margin being established. The higher the confidence level (greater expected precision), the larger will be the sample size. This parameter is usually fixed as 95%. Design effect It is necessary when the study participants are chosen by cluster selection procedures. This means that, instead of the participants being individually selected (simple, systematic or stratified sampling), they are first divided and randomly selected in groups (census tracts, neighborhood, households, days of the week, etc.) and later the individuals are selected within these groups. Thus, greater similarity is expected among the respondents within a group than in the general population. This generates loss of precision, which needs to be compensated by a sample size adjustment (increase). The principle is that the total estimated variance may have been reduced as a consequence of cluster selection. The value of the design effect may be obtained from the literature. When not available, a value between 1.5 and 2.0 may be determined and the investigators should evaluate, after the study is completed, the actual design effect and report it in their publications. The greater the homogeneity within each group (the more similar the respondents are within each cluster), the greater the design effect will be and the larger the sample size required to increase precision. In studies that do not use cluster selection procedures (simple, systematic or stratified sampling), the design effect is considered as null or 1.0. Chart 2 presents some sample size simulations, according to the outcome prevalence, sample error and the type of target population investigated. The same basic question was used in this table (prevalence of sunscreen use), but considering three different situations (at work, while doing sports or at the beach), as in the study by Duquia et al. conducted in the city of Pelotas, state of Rio Grande do Sul, in 2005.3 Sample size calculation to estimate the frequency (prevalence) of sunscreen use in the population, considering different scenarios but keeping the significance level (95%) and the design effect (1.0) constant Target population Prevalence (p) of outcome Sunscreen use at work p=10% Sunscreen use in sports p=35% Sunscreen use at the beach p=50% Acceptable Error 2 p.p. Acceptable Error 5 p.p. Acceptable Error 2 p.p. Acceptable Error 5 p.p. Acceptable Error 2 p.p. Acceptable Error 5 p.p. Sample Health center users investigated in a single day (population = 100) 90 59 96 78 97 80 All users in the area covered by a health center (population size = 1,000) 464 122 687 260 707 278 All users from the areas covered by all health centers in a city (population size = 10,000) 796 137 1794 338 1937 370 The entire city population (N = 40.000) 847 138 2072 347 2265 381 The calculations show that, by holding the sample error and the significance level constant, the higher the expected prevalence, the larger will be the required sample size. However, when the expected prevalence surpasses 50%, the required sample size progressively diminishes - the sample size for an expected prevalence of 10% is the same as that for an expected prevalence of 90%. The investigator should also define beforehand the precision level to be accepted for the investigated event (sample error) and the confidence level of this result (usually 95%). Chart 2 demonstrates that, holding the expected prevalence constant, the higher the precision (smaller sample error) and the higher the confidence level (in this case, 95% was considered for all calculations), the larger also will be the required sample size. Chart 2 also demonstrates that there is a direct relationship between the target population size and the number of individuals to be included in the sample. Nevertheless, when the target population size is sufficiently large, that is, surpasses an arbitrary value (for example, one million individuals), the resulting sample size tends to stabilize. The smaller the target population, the larger the sample will be; in some cases, the sample may even correspond to the total number of individuals from the target population - in these cases, it may be more convenient to study the entire target population, carrying out a census survey, rather than a study based on a sample of the population. When the study objective is to investigate whether there are differences in sunscreen use according to sociodemographic characteristics (such as, for example, between men and women), the existence of association between explanatory variables (exposure or independent variables, in this case sociodemographic variables) and a dependent or outcome variable (use of sunscreen) is what is under consideration. In these cases, we need first to understand what the hypotheses are, as well as the types of error that may result from their acceptance or refutation. A hypothesis is a "supposition arrived at from observation or reflection, that leads to refutable predictions".4 In other words, it is a statement that may be questioned or tested and that may be falsified in scientific studies. In scientific studies, there are two types of hypothesis: the null hypothesis (H0) or original supposition that we assume to be true for a given situation, and the alternative hypothesis (HA) or additional explanation for the same situation, which we believe may replace the original supposition. In the health field, H0 is frequently defined as the equality or absence of difference in the outcome of interest between the studied groups (for example, sunscreen use is equal in men and women). On the other hand, HA assumes the existence of difference between groups. HA is called two-tailed when it is expected that the difference between the groups will occur in any direction (men using more sunscreen than women or vice-versa). However, if the investigator expects to find that a specific group uses more sunscreen than the other, he will be testing a one-tailed HA. In the sample investigated by Duquia et al., the frequency of sunscreen use at the beach was greater in men (32.7%) than in women (26.2%).3 Although this what was observed in the sample, that is, men do wear more sunscreen than women, the investigators must decide whether they refute or accept H0 in the target population (which contends that there is no difference in sunscreen use according to sex). Given that the entire target population is hardly ever investigated to confirm or refute the difference observed in the sample, the authors have to be aware that, independently from their decision (accepting or refuting H0), their conclusion may be wrong, as can be seen in figure 2. Types of possible results when performing a hypothesis testIn case the investigators conclude that both in the target population and in the sample sunscreen use is also different between men and women (rejecting H0), they may be making a type I or Alpha error, which is the probability of rejecting H0 based on sample results when, in the target population, H0 is true (the difference between men and women regarding sunscreen use found in the sample is not observed in the target population). If the authors conclude that there are no differences between the groups (accepting H0), the investigators may be making a type II or Beta error, which is the probability of accepting H0 when, in the target population, H0 is false (that is, HA is true) or, in other words, the probability of stating that the frequency of sunscreen use is equal between the sexes, when it is different in the same groups of the target population. In order to accept or refute H0 , the investigators need to previously define which is the maximum probability of type I and II errors that they are willing to incorporate into their results. In general, the type I error is fixed at a maximum value of 5% (0.05 or confidence level of 95%), since the consequences originated from this type of error are considered more harmful. For example, to state that an exposure/intervention affects a health condition, when this does not happen in the target population may bring about behaviors or actions (therapeutic changes, implementation of intervention programs etc.) with adverse consequences in ethical, economic and health terms. In the study conducted by Duquia et al., when the authors contend that the use of sunscreen was different according to sex, the p value presented (99.9%).3 Although the type II or Beta error is less harmful, it should also be avoided, since if a study contends that a given exposure/intervention does not affect the outcome, when this effect actually exists in the target population, the consequence may be that a new medication with better therapeutic effects is not administered or that some aspects related to the etiology of the damage are not considered. This is the reason why the value of the type II error is usually fixed at a maximum value of 20% (or 0.20). In publications, this value tends to be mentioned as the power of the study, which is the ability of the test to detect a difference, when in fact it exists in the target population (usually fixed at 80%, as a result of the 1-Beta calculation). In cases where the exposure variables are dichotomous (intervention/control, man/woman, rich/poor etc.) and so is the outcome (negative/positive outcome, to use sunscreen or not), the required parameters to calculate sample size are those described in chart 3. According to the previously mentioned example, it would be interesting to know whether sex, skin color, schooling level and income are associated with the use of sunscreen at work, while doing sports and at the beach. Thus, when the four exposure variables are crossed with the three outcomes, there would be 12 different questions to be answered and consequently an equal number of sample size calculations to be performed. Using the information in the article by Duquia et al.3 for the prevalence of exposures and outcomes, a simulation of sample size calculations was used for each one of these situations (Chart 4). Description of different parameters to be considered in the calculation of sample size for a study aiming at estimating the frequency of health ouctomes, behaviors or conditions Parameter Description Remark Type I or Alpha error It is the probability of rejecting H0, when H0 is false in the target population. Usually fixed as 5%. It is expressed by the p value. It is usually 5% (p4 anos: 75%(NE) 90% n=1795 n=490 n=654 n=175 n=184 ND r: 3.00 Per capita income: Power PONE: 11.0% PONE: 24.2% PONE: 48.6% ≤133: 50%(E) 80% n=1228 n=360 n=458 n=124 n=128 n=28 >133: 50%(NE) 90% n=1644 n=480 n=612 n=166 n=170 n=36 r: 1.00 Estimates show that studies with more power or that intend to find a difference of a lower magnitude in the frequency of the outcome (in this case, the prevalence rates) between exposed and non-exposed groups require larger sample sizes. For these reasons, in sample size calculations, an effect measure between 1.5 and 2.0 (for risk factors) or between 0.50 and 0.75 (for protective factors), and an 80% power are frequently used. Considering the values in each column of chart 3, we may conclude also that, when the nonexposed/exposed relationship moves away from one (similar proportions of exposed and non-exposed individuals in the sample), the sample size increases. For this reason, intervention studies usually work with the same proportion of individuals in the intervention and control groups. Upon analysis of the values on each line, it can be concluded that there is an inverse relationship between the prevalence of the outcome and the required sample size. Based on these estimates, assuming that the authors intended to test all of these associations, it would be necessary to choose the largest estimated sample size (2,630 subjects). In case the required sample size is larger than the target population, the investigators may decide to perform a multicenter study, lengthen the period for data collection, modify the research question or face the possibility of not having sufficient power to draw valid conclusions. Additional aspects need to be considered in the previous estimates to arrive at the final sample size, which may include the possibility of refusals and/or losses in the study (an additional 10-15%), the need for adjustments for confounding factors (an additional 10-20%, applicable to observational studies), the possibility of effect modification (which implies an analysis of subgroups and the need to duplicate or triplicate the sample size), as well as the existence of design effects (multiplication of sample size by 1.5 to 2.0) in case of cluster sampling. Suppose that the investigators intend to evaluate whether the daily quantity of sunscreen used (in grams), the time of daily exposure to sunlight (in minutes) or a laboratory parameter (such as vitamin D levels) differ according to the socio-demographic variables mentioned. In all of these cases, the outcomes are numerical variables (discrete or continuous)1 , and the objective is to answer whether the mean outcome in the exposed/intervention group is different from the non-exposed/control group. In this case, the first three parameters from chart 4 (alpha error, power of the study and relationship between non-exposed/exposed groups) are required, and the conclusions about their influences on the final sample size are also applicable. In addition to defining the expected outcome means in each group or the expected mean difference between nonexposed/exposed groups (usually at least 15% of the mean value in non-exposed group), they also need to define the standard deviation value for each group. There is a direct relationship between the standard deviation value and the sample size, the reason why in case of asymmetric variables the sample size would be overestimated. In such cases, the option may be to estimate sample sizes based on specific calculations for asymmetric variables, or the investigators may choose to use a percentage of the median value (for example, 25%) as a substitute for the standard deviation. There are also specific calculations for some other quantitative studies, such as those aiming to assess correlations (exposure and outcome are numerical variables), time until the event (death, cure, relapse etc.) or the validity of diagnostic tests, but they are not described in this article, given that they were discussed elsewhere.5 Sample size calculation is always an essential step during the planning of scientific studies. An insufficient or small sample size may not be able to demonstrate the desired difference, or estimate the frequency of the event of interest with acceptable precision. A very large sample may add to the complexity of the study, and its associated costs, rendering it unfeasible. Both situations are ethically unacceptable and should be avoided by the investigator.

1.Duquia RP, Bastos JL, Bonamigo RR, González-Chica DA, Martínez-Mesa J. Presenting data in tables and charts. An Bras Dermatol. 2014;89:280–285. doi: 10.1590/abd1806-4841.20143388. [DOI] [PMC free article] [PubMed] [Google Scholar] 2.OpenEpi.com. Dean AG, Sullivan KM, Soe MM. OpenEpi: Open Source Epidemiologic Statistics for Public Health, Version. [updated 2013 Apr 6]. [2014 Mar 22]. [Internet] www.OpenEpi.com. [Google Scholar] 3.Duquia RP, Baptista Menezes AM, Reichert FF, de Almeida HL., Jr Prevalence and associated factors with sunscreen use in Southern Brazil: A population-based study. J Am Acad Dermatol. 2007;57:73–80. doi: 10.1016/j.jaad.2007.03.012. [DOI] [PubMed] [Google Scholar] 4.Porta M, editor. A dictionary of epidemiology. 5th. ed. New York: Oxford University Press; 2008. [Google Scholar] 5.Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. Emerg Med J. 2003;20:453–458. doi: 10.1136/emj.20.5.453. [DOI] [PMC free article] [PubMed] [Google Scholar]