ECDC **TECHNICAL GUIDANCE**

Sample size guidance for surveillance data

www.ecdc.europa.eu

ECDC TECHNICAL GUIDANCE

Sample size guidance for surveillance data

EUROPEAN CENTRE FOR
DISEASE PREVENTION
AND CONTROL

This report was commissioned by the European Centre for Disease Prevention and Control (ECDC), coordinated by Joana Gomes Dias, and produced by the Data Science Institute and Centre for Statistics, Hasselt University.

Contributing authors (in an alphabetical order)*:*

- Liesbeth Bruckers, Christel Faes, Zoë Pieters and Bryan Sumalinab: Data Science Institute and Centre for Statistics, Hasselt University;
- Joana Gomes Dias and Gaetano Marrone: ECDC.

Suggested citation: European Centre for Disease Prevention and Control. Sample size guidance for surveillance data. Stockholm: ECDC; 2023.

Stockholm, July 2023

ISBN 978-92-9498-645-0
doi: 10.2900/037252
Catalogue number TQ-05-23-276-EN-N

Contents

Figures

Tables

Executive summary

Sample size calculation is an essential step in ensuring that analysis of the data which are collected through surveillance systems can support the desired public health objective(s). The main aim of sample size calculations is to determine the sample size needed from the population under study to possibly detect a statistically significant result.

This document provides guidance to support disease experts in doing sample size calculations according to different surveillance objectives. Its main aims are to:

- introduce the statistical terminology and key concepts used in relation to sample size and power calculations;
- provide guidance for doing sample size calculations for typical situations that involve the methods most frequently used in surveillance analyses;
- illustrate these calculations with examples and easy-to-use R scripts.

This document is organised as follows: Section 1 presents an overall introduction to sample size calculations for surveillance data and a list of objectives for this guidance. Section 2 enumerates the key statistical concepts and definitions used in sample size calculation and provides illustrative examples drawn from different sample size requirement situations, which are mostly relevant in the context of ECDC. Section 3 presents sample size calculations for surveys using simple random sampling, where formulae are provided according to the different situations being considered. Each outcome is illustrated with a scenario, along with the provision and explanation of R codes to obtain the required sample size. Most computations are done with available R packages.

1 Introduction

A series of scientific guidances on statistical methods is published by ECDC to address the needs of surveillance experts and increase biostatistical capacity. Sample size calculation is an essential step in ensuring that analysis of the data which are collected through surveillance systems can support the desired public health objective(s). The main aim of sample size calculations is to determine the sample size needed from the population under study to possibly detect a statistically significant result.

This document provides guidance to support disease experts in doing sample size calculations according to different surveillance objectives. It provides a practical overview of sample size calculation methods for different statistical analyses, illustrated with examples from infectious disease surveillance settings. It also provides a hands-on step-by-step guidance on how sample size calculations are obtained, with code snippets throughout the document.

1.1 Objectives

The objectives of this guidance for sample size calculations for surveillance data are:

- to introduce the statistical terminology and key concepts used in relation to sample size and power calculations;
- to provide guidance for doing sample size calculations for typical situations that involve the methods most frequently used in surveillance analyses;
- to illustrate these calculations with examples and easy-to-use R scripts (a free software environment for statistical computing and graphics).

1.2 Structure of the document

This document is organised as follows: Section 1 presents an overall introduction to sample size calculations for surveillance data and a list of objectives for this guidance. Section 2 enumerates the key statistical concepts and definitions used in sample size calculation, such as sampling methods, hypothesis testing, confidence intervals, types of outcome variables, representativeness, power, and level of significance. It also provides illustrative examples drawn from different sample size requirement situations, which are mostly relevant in the context of ECDC. This includes typical sample size calculations when the outcome under study is either a binary outcome, a non-binary categorical outcome, or a continuous outcome.

Sample size calculations for surveys using simple random sampling are presented in Section 3, where formulae are provided according to the different situations being considered. The situations range from obtaining the sample size for the estimation of a population proportion with a prespecified level of precision, to the sample size for the comparison of two population means. In addition, sample size calculations for categorical non-binary outcome and continuous outcome are also discussed in Section 3. Each outcome is illustrated with a scenario, along with the provision and explanation of R codes to obtain the required sample size.

The same structure is followed for each section:

- First, the scenario used for the specific sample size calculation is introduced ('Scenarios 1, 2, 3 and 4');
- Then, the theoretical background and the formula behind the calculations are provided ('Theoretical background');
- The scenario and the theoretical background are used to illustrate the sample size calculation in R software ('Implementation of the scenario in R');
- Lastly, a conclusion is formulated based on the R output, and some remarks are provided for the sample size calculation ('Conclusion').

Most computations are done with available R packages. A step-by-step guide to download and install R can be found here. To install R packages not included with the R installation, the `'install.packages'` function can be used. The code in the following page will install and load the necessary packages and source files.

R code for installing necessary R packages

```
## Installing R packages - done only once.
install.packages("epiR")
install.packages("pwr")
install.packages("TrialSize")


## Loading R packages - each time a new session in R is started
library(epiR)
library(pwr)
library(TrialSize)

# function to correct for finite sample size (see Section 8)
source("finite_correct.R")

# function to compute sample size for testing
# difference in two dependent proportions (see Section 8)
source("mcnemar.R")
```

Key concepts for cluster sampling and clusters of equal and unequal sizes are discussed in Section 4. Section 5 covers a few other issues such as types of bias, missingness and weighting.

2 Concepts and definitions used in the guidance document

2.1 General background

Sampling is the process of selecting a specific part (called 'sample') of a given target population. The ultimate goal is to obtain a sample that closely represents an approximation of the characteristics of the target population, in order to draw valid inferences. Such a sample size is defined as 'representative' of the target population.

The sample should be selected using appropriate methods. In addition to choosing a suitable sampling method, the sample size is also vital as it directly influences both the level of precision of the findings and the cost of the study.

The sample size should be determined based on several considerations, such as the statistical test that will be used, the variability in the parameter, the study design, the precision of the estimates, etc. A sample size that is too small will result in a lack of precision and is more likely to have random errors, leading to misleading results. A sample size that is too large on the other hand, will be too time- and resource-consuming. Moreover, a very large sample size will demonstrate statistical significance even for very small differences which may not be of importance to public health.

The appropriate sample size depends on different factors, such as the surveillance goals, the type of outcome being investigated, and the expected results. Therefore, three aspects need to be considered before determining the size of a sample.

The first consideration is the goal of the surveillance activity, which needs to be clearly specified before a sample size calculation can be conducted. Surveillance activities that need a sample size calculation include:

- the estimation of a parameter with a desired precision (e.g. the estimation of disease prevalence or the estimation of diagnostic accuracy);
- the comparison of a parameter versus a hypothesised value (e.g. testing if the disease proportion is significantly different than the hypothesised value);
- the comparison of two population parameters, i.e. hypothesis testing (e.g. comparison of the accuracy of diagnostic tests, or comparison of disease prevalence at different time points).

The second consideration is the type of outcome being investigated, since this affects the statistical data analysis and equivalently the sample size calculation method.

There are three common categories of outcomes:

- continuous (e.g. a test result taking many possible values, such as blood pressure);
- binary (e.g. disease or vaccination status with two possible values: 'yes' or 'no');
- categorical variables, which are of two types:
  - ordinal categorical variable (e.g. the disease stage of ovarian cancer, which takes multiple possible values with an intrinsic ordering – stage I up to IV);
  - nominal categorical variable (e.g. genotype (AA, Aa, aa), where there is no intrinsic ordering to the categories).

The third consideration is about the expected results. Sample size calculations for binary outcomes require information about the expected prevalence of the outcome. For continuous outcomes, the expected average value and variability of the outcome need to be considered, while the minimum relevant detectable difference must be decided for the comparison of outcomes. For other aspects related to the design of surveillance systems, an estimate of the intra-cluster correlation coefficient, etc might be needed. The investigator needs to make assumptions on these parameters before conducting the survey. Therefore, these pieces of information have to be obtained from expert knowledge, prior to available information, or from a pilot study.

The following sections introduce some of the most commonly used concepts and definitions of relevance to sample size calculations, such as sampling size, sampling methods, hypothesis testing, confidence intervals, types of outcome variable, and effect size.

2.1.1 Population and sample

A sample from the target population is often taken by selecting units from the sampling frame using a specific sampling process.

In any surveillance study, the target and study population need to be defined. The **target population** is the population being investigated. The **study population** is the population from which data can actually be collected [1].

Ideally, the study and target population are the same, but often they are not. For example, consider a national health interview survey where the target population is the non-institutionalised civilian population residing in the country at the time of the interview. This includes residents of households and non-institutional group quarters (e.g. homeless shelters, rooming houses, and group homes). For the interview, the residents will be contacted via their home address. As a consequence of the contacting process, the study population will only include individuals who have a fixed household address.

A population is said to be **finite** if the number of units can be counted. On the other hand, if the total population is unknown and/or large, it is called an **infinite** population. The sample size is usually larger when sampling units are drawn from an infinite population, as compared to a finite population. This will be demonstrated in the sample size calculations in Section 3.

The **sampling units** are the building blocks of the population. These are the units that can be identified prior to the drawing of the sample, and subsequently can be selected for a sample. The sampling units can be the individuals in the study population, clusters of individuals such as households, or other pre-defined building blocks. For instance, individuals might be the intended focus group, but if only home addresses are accessible and a list of all individuals is not available, one can use the households as the (primary) sampling units, instead of the individuals. So, eventually, sampling units can be individuals (e.g. patients suffering from a disease), places (e.g. geographical areas such as countries), clusters (e.g. laboratories), or even objects (e.g. medical records) [2].

A **sampling frame** is then obtained from the study population. It is a 'list' containing all the units that have a non-zero probability of being selected. For example, a city residential directory is used as a sampling frame for a study or survey in which the sampling units are the residents of that city [3].

2.1.2 Sampling methods

A sampling method refers to the method in which sampling units are selected from the study population. Some common sampling methods in surveillance studies include simple random sampling, stratified sampling, clustered sampling, and multi-stage sampling.

Simple random sampling is a sampling method in which each unit in the study population has an equal, non-zero probability of being selected to participate in the study or survey. The method provides unbiased estimates if the population is homogeneous.

On the other hand, if the population is heterogeneous, i.e. the population consists of different subgroups within which the outcome is believed to be different, **stratified sampling** is a useful sampling method. In stratified sampling, the population is first divided into subgroups, also called strata. Then the units are randomly sampled from within each of these strata according to a chosen probability sampling method (e.g. simple random sampling within each stratum). The sample size can be assumed to vary or be the same in each stratum. When the sample size in each stratum varies according to the relative importance of the respective stratum in the population, it is called **proportionate sampling**. If the sample size in each stratum is not proportional to the size of that stratum in the population, it is called **disproportionate sampling**. For example, a region-based stratified sampling process can be used to estimate the prevalence of hepatitis C virus (HCV) in a given country. In this case, the strata are the regions of the country, and the samples are taken randomly from within each region. Using stratified sampling, every region is ensured to be represented in the sample. Consequently, estimates will be more precise.

In **cluster sampling**, units are grouped into larger units with homogeneous characteristics. These larger units are called clusters, and a random sample of these clusters is selected. All the units in the selected clusters are included in the sample. Clusters are often natural groupings in the population, such as households, cities, or hospitals, within which the outcome is often more homogeneous as compared to the population as a whole. For example, suppose that, in order to estimate the proportion of individuals with an influenza infection in the past, a sample of clinics is selected randomly. All patients of the sample clinics are included in the survey. An advantage of cluster sampling as compared to simple random sampling and stratified sampling is that it is often more time- and cost-efficient. However, note that samples from the same cluster (clinic, in the previous example) tend to be correlated. The correlation between pairs of elements in the same clusters is called **intra-cluster correlation (ICC)**.

Rather than collecting all units from the selected clusters, **multi-stage sampling** can be used to randomly select units from the selected clusters. Multi-stage sampling is a sampling process done in more than one step (or stage), with a different sampling frame used in each step. For instance, to estimate the proportion of individuals with an influenza infection in the past, a two-stage sampling process can be used. The clinics are selected in the first stage, and sample patients are chosen (ideally in a random manner) from each selected clinic in the second stage. The units from the first and second stages of a multi-stage sampling are called **primary** and **secondary sampling units**, respectively.

2.1.3 Hypothesis testing

Several of the sample size calculations correspond to situations where the aim is hypothesis testing. Therefore, brief descriptions of some key concepts and frequently used terminologies pertinent to hypothesis testing are provided here. Most of the terminologies and concepts are discussed in more details in the later sections.

Hypothesis testing is a statistical tool proposed by Neyman and Pearson [4]. It is used to decide whether or not the data support a particular hypothesis. First, the **null hypothesis** (denoted as H_0) is formulated. For a two-sample situation, this hypothesis implies no difference between a specified parameter of the two populations (for example, two means or two population proportions). Then, the **alternative hypothesis** (denoted as H_1) is defined, which is the opposite of the null hypothesis (in this case, it would be that the two means or the two population proportions are different).

In hypothesis testing, the objective is to specifically determine whether or not there is enough evidence to reject the null hypothesis. If the null hypothesis is rejected, the conclusion is that the alternative hypothesis is true. If the null hypothesis cannot be rejected, it may be true or there might not be enough data to be able to detect that the null hypothesis is false.

Consider this as a first example: an investigator states as null hypothesis H_0 that the global prevalence of hepatitis C virus (HCV) is 2.5%. The alternative hypothesis H_1 can be that the investigator expects that the global prevalence of HCV is different from 2.5%. After setting up these hypotheses, the aim of hypothesis testing is to see if there is enough evidence to reject the null hypothesis. This is an example of a one-sample test, namely the hypothesis test of a population proportion against a pre-defined value.

Now consider this as a second example: an investigator states as null hypothesis H_0 that the global prevalence of HCV is the same in injecting drug users and non-drug users. The alternative hypothesis H_1 can be that the investigator expects that the global prevalence of HCV is greater among injecting drug users compared to non-drug users. This is an example of a two-sample test, where the comparison of two groups (here, injecting drug users and non-drug users) is of interest.

There is a distinction between **two-sided** (or **two-tailed**) and **one-sided** (or **one-tailed**) hypothesis tests. Two-sided hypothesis tests are the most common. These have as the alternative hypothesis, that the parameter of interest (either a proportion or a mean) is not equal to the pre-specified value in the null hypothesis (for a one-sample test), or that the two parameters are different from one another (for a two-sample test). Hence, the interest does not lie in one direction specifically.

The first example presented above is an example of a two-sided test, i.e. the alternative hypothesis is that the global prevalence of HCV is different from 2.5%. However, a one-sided hypothesis test is used if the interest lies in observing an effect in a specific direction: for example, a parameter that is greater (or smaller) than a pre-defined value in the null hypothesis (for a one-sample test), or that the parameter is greater (or smaller) in one population compared to the second population (for a two-sample test)–. The second example presented above is an example of a one-sided test in which the alternative hypothesis is that the global prevalence of HCV is greater among injecting drug users than non-drug users.

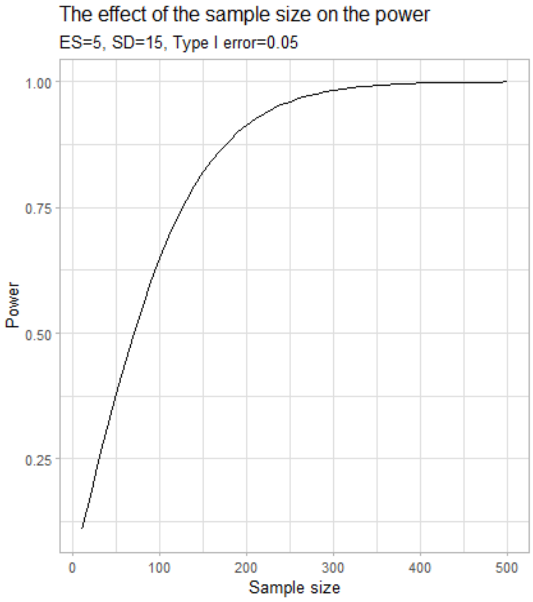
As hypothesis testing is a tool to choose between the null and alternative hypotheses based on the available information, there is a risk that an incorrect decision is made. Two fundamental errors can be made while choosing the hypothesis. **Type I error** (or the significance level), denoted by α, represents the probability of rejecting the null hypothesis when the null hypothesis is true. It is also known as the false positive conclusion. **Type II error**, denoted by β, represents the probability of accepting the null hypothesis when the alternative hypothesis is actually true. These are illustrated in Table 1.

Table 1. Illustration of type I and type II error for hypothesis testing

		Truth	
		H_0 is true	H_0 is false
Study findings	H_0 is true (fail to reject H_0)	Correct acceptance	Type II error (β)
	H_0 is false (reject H_0)	Type I error (α)	Correct rejection Power (1-β)

A related concept is the **power**, denoted as (1-β), which is the probability of rejecting the null hypothesis when the alternative is true, i.e. the probability of making the correct conclusion when the alternative hypothesis is true. The power is equal to one minus the type II error (β). It should be noted that the power of a study increases with an increase in the sample size. Figure 1 illustrates this relationship. In addition, there is an inverse relationship between the type I (α) and type II (β) errors, such that if one decreases, the other increases.

Figure 1. Illustration of the effect of the sample size on the power



ES: effect size; SD: standard deviation

When performing a sample size calculation to test a statistical hypothesis, the power (1-β) and significance level (α) have to be specified. The most common choices for the significance level are 5% or even lesser, such as 1%, corresponding to a small probability of making a false positive conclusion. The most common choices for the power are 80% (β=0.20) or 90% (β=0.10), corresponding to a large probability to correctly detect a true effect. Specifying a lower significance level and a higher power will lead to a required sample size which is larger. Note that while most sample size calculation methods control the significance level, they do not always control the power.

To illustrate the choice of the type I error and the power, consider the following example: when testing a new drug before it is introduced to the market, the manufacturer will perform a statistical test with the null hypothesis that the drug is ineffective, versus the alternative hypothesis that the drug is effective. From the perspective of the drugs administration agency of that market, a smaller type I error is desired since it would reduce the chances of the investigators approving a non-beneficial drug. On the other hand, the drug manufacturer would want to have a small type II error (higher power) since they want to see the impact of the drug when it is indeed effective.

2.1.4 Confidence interval

A **confidence interval** is a range of plausible values for an unknown parameter, which is used to give an indication of the uncertainty about the parameter. Confidence intervals can be calculated for several parameters, such as a proportion, a mean, differences between means or proportions, etc. Different factors affect the width of confidence intervals, including the confidence level, the variability in the data, and the sample size.

The confidence interval is calculated with a pre-specified **confidence level,** usually chosen as 1-α. The confidence level represents the 'long-run' proportion of corresponding confidence intervals that end up containing the true value of the parameter. For example, a confidence interval with a 95% confidence level gives a range of values within which the true value of the parameter in the population lies with 95% probability. Note that the conclusions from a confidence interval and a hypothesis test are in general similar, in the sense that if a 95% confidence interval contains a hypothesised value, then the corresponding hypothesis test with a 0.05 significance level will almost certainly fail to reject the null hypothesis.

Note that the width of the confidence interval gets smaller with increasing sample size. The smaller the confidence interval, the more precisely the parameter of interest can be estimated. The precision of the estimate can also be referred to as 'margin of error', 'maximum tolerable error' or 'half of the width of the confidence interval'. The maximum tolerable error is the difference between the true population parameter and the estimate of the true population parameter derived from sampling [5].

In this guidance, we will also refer to **absolute error** or **relative error**. Absolute error refers to the allowed deviation from the true population parameter. Relative error is the ratio of the absolute error and the true population parameter. It is often expressed as a percentage. For example, if the parameter of interest is a population proportion, with the true value of 0.5, then an absolute error of 0.1 means that we allow the estimate to lie between 0.4 and 0.6. A relative error of 10% means that we allow a deviation of 10% from 0.5 i.e. from (1-0.1)x0.5=0.45 to (1+0.1)x0.5=0.55.

2.2 Types of outcome variable

The sample size calculation depends on the type of outcome variable being investigated. Outcome variables are generally classified into two types: categorical (qualitative) variables, and numerical (quantitative) variables.

Categorical variables can take on a limited number of possible values or categories. Categorical variables may be further subdivided into nominal and ordinal variables. **Nominal variables** are categorical variables that do not have a natural ordering (e.g. EU countries), while **ordinal variables** do (e.g. stages of ovarian cancer, ranging from stage 1 to 4). A **binary or dichotomous variable** is a special case when the nominal variable only has two categories (e.g. sick healthy or vaccinated/unvaccinated).

Numerical variables are used to describe a quantity measurable with numbers. Furthermore, a numerical variable may be classified as either continuous or discrete. A **continuous variable** is a numerical variable that could take an infinite number of values within any chosen range. Examples of continuous variables are height, weight, age, etc. These are often obtained by measurements. A **discrete variable** is a numerical variable the value of which can be counted. For example, the number of individuals exposed to a risk factor, the number of people living in a household, the number of working hours in a week, etc. They can only take a finite number of values within any chosen range.

2.3 Differences that are considered of importance (effect size)

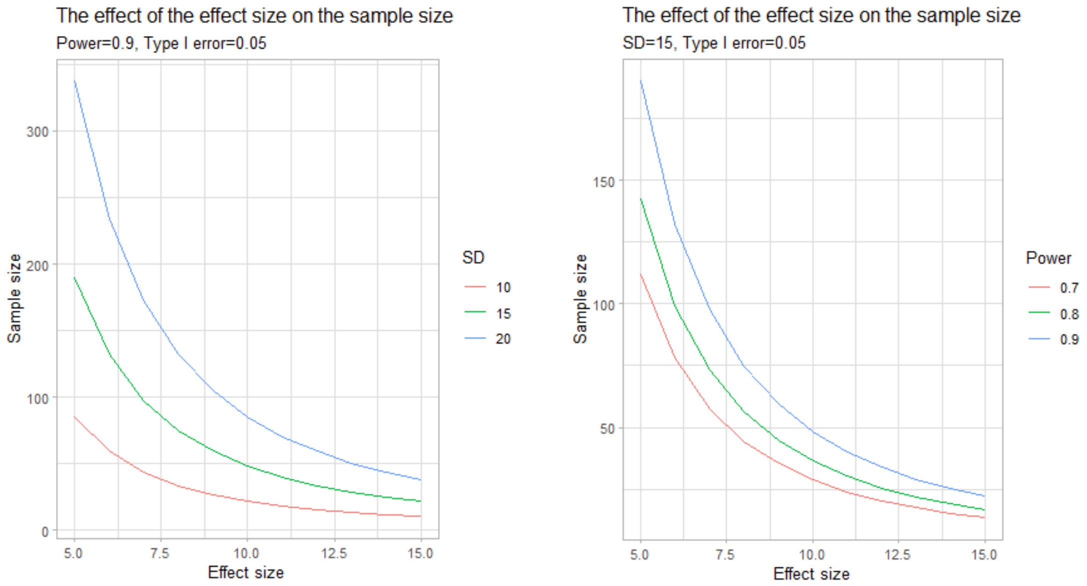
The **effect size** is defined as the minimal difference an investigator wishes to detect. This should be the difference that is epidemiologically relevant and biologically plausible. When comparing two population means, for example, the effect size could refer to the absolute effect size or standardised effect size.

The **absolute effect size** is the minimal relevant difference in the population means/proportions of the two groups. For example, when the outcome of interest is the prevalence of HCV in injecting drug users and non-drug users, the investigator might choose a difference of 0.2% between the injecting drug users and non-drug users as the minimal relevant difference to be detected in the study.

Sometimes the **standardised effect size** is used, often standardising the absolute effect size with respect to the standard deviation. In Sections 3.4.2 and 3.4.3, the standardised effect size is introduced and exemplified.

The effect size needs to be pre-specified by the investigator. It can be based on experts' opinions, historical data, a pilot study, or previous scientific knowledge. It should be noted that the effect size has a large impact on the sample size. There is an inverse relationship between the sample size and the effect size. Thus, a large effect size requires a small sample size, while a small effect size requires a large sample size.

Figure 2. Illustration of the relationship between effect size and sample size: a) Left panel – with power 90%, Type I error 5% and different values for the standard deviation; b) Right panel – with standard deviation 15, Type I error 5% and different values for the power



SD: standard deviation

While conducting a statistical analysis of the study data, it might be the case that no significant difference is found. This result does not imply that the null hypothesis of no (significant) difference is true, but rather indicates that there is insufficient evidence to reject the null hypothesis.

This lack of evidence could have several implications. It could be that the true effect size is smaller than the one considered of interest. Indeed, if the true effect size is smaller than the one considered for the sample size calculation, then there will be insufficient evidence to reject the null hypothesis. It is also possible that the standard deviation of the outcomes is bigger than the one considered, and therefore, the sample size is not enough. Figure 2 (above) illustrates these issues. The left panel shows that if the standard deviation of the outcomes increases, then for a given power and effect size, a bigger sample size is needed in order to be able to reject the null hypothesis. In addition, the right panel illustrates that a higher power will result in a bigger sample size for any given effect size, with differences becoming larger as the effect size becomes smaller. Hence, it could be that a type II error is being made, and there is not enough power to detect it.

2.4 Scenarios

In this section, several scenarios are introduced, which will be used throughout this guidance document to demonstrate sample size calculations in different situations. In scenarios 1 and 2, the outcome of interest is of a binary nature. Scenario 1 describes the prevalence of a bacteria species which is antimicrobial resistant (AMR), and scenario 2 is about the effectiveness of influenza vaccine. Scenario 3 illustrates a categorical outcome, namely the disease severity of a SARS-CoV-2 infection. Lastly, scenario 4 is about a continuous outcome of interest, namely the estimation of the mean cost price of a SARS-CoV-2 polymerase chain reaction (PCR) test.

Introductions to each of the four scenarios are given below. For the specific sample size calculations, extra features will be introduced to illustrate the various sample size calculations for that specific data type.

2.4.1 Scenario 1: Binary outcome (antimicrobial resistance)

Collecting antimicrobial resistance (AMR) surveillance data is important in defining the resistance prevalence and informing an appropriate direction in the use of available antimicrobials. The occurrence of AMR across the EU/EEA varies depending on the bacterial species, antimicrobial groups, and geographical regions. Each year, 30 countries from the EU and EEA report antimicrobial susceptibility testing from invasive isolates (blood or cerebrospinal fluid) collected from local/clinical microbiology laboratories and laboratory networks to the European Antimicrobial Resistance Surveillance Network (EARS-Net). In this scenario, laboratories in each of the 30 countries were randomly selected. Disease experts want to find out how many isolates need to be taken in order to estimate, with a given precision, the proportion of AMR against an antimicrobial or a group of antimicrobials for a bacterial species, such as *Escherichia coli* (*E. coli*).

With this scenario, the sample size calculations for different situations will be illustrated. Firstly, the focus will be on the sample size calculation for the estimation of a population proportion with a desired level of precision, and also for hypothesis testing of a population proportion with a specified prior proportion (Sections 3.1.1 and 3.1.2). Secondly, the focus will be on the comparison of two populations, including sample size calculations for two independent proportions (e.g. the detection of significant differences in the AMR of *E. coli* in two different countries at the same time) (Section 3.2.1), and two dependent proportions (e.g. comparison of a new and a standard diagnostic test on the same isolates) (Section 3.2.4).

This scenario is based on the publication, 'Antimicrobial resistance in the EU/EEA (EARS-Net) – Annual Epidemiological Report for 2019' [6].

2.4.2 Scenario 2: Binary outcome (vaccine effectiveness)

As influenza viruses constantly evolve, vaccines are reformulated every year. Therefore, it is important to calculate annual influenza vaccine effectiveness (VE) estimates at the European level at the beginning of a seasonal influenza epidemic/pandemic, as well as monitor the VE along the course of the influenza season.

To inform these calculations, EU/EEA Member States are required to include both vaccinated and unvaccinated individuals in their surveillance data, who do not have contraindications for the influenza vaccine.

With this scenario, the sample size calculations for the estimation of the relative risk (RR) and odds ratio (OR) of medically attended acute respiratory infections (MAARI) in vaccinated versus unvaccinated children (VE = 1-RR) will be illustrated (Sections 3.2.2 and 3.2.3). In addition, this scenario will be used for the non-inferiority testing of a new vaccine compared to the standard vaccine (Section 3.2.5).

This scenario is based on the publication, 'Protocol for cohort database studies to measure influenza vaccine effectiveness in the European Union and European Economic Area Member States' [7].

2.4.3 Scenario 3: Categorical outcome with more than two categories (disease severity SARS-CoV-2)

In the COVID-19 pandemic, infected individuals showed a wide range of symptoms, which can be classified into three categories: 'mild', 'moderate', or 'severe'. ECDC wants to investigate whether there was a difference in disease severity between men and women. Section 3.3 illustrates a sample size calculation based on the chi-squared test, using multiple categories in the outcome. An alternative approach to come up with a sample size for this setting could be to dichotomise the disease severity (for example, into 'mild'/'moderate' versus 'severe') and then use the methods to compare two proportions.

This scenario is not based on any official reference.

2.4.4 Scenario 4: Continuous outcome (mean cost price of a PCR test)

Across the EU/EEA, PCR tests were conducted to verify infection with the SARS-CoV-2 virus during the pandemic. ECDC wants to conduct a survey among laboratories throughout the EU/EEA to estimate the mean cost price of a PCR test to diagnose an individual with a SARS-CoV-2 infection. For this scenario, it is assumed that the cost price of a PCR test will follow a normal distribution.

With this scenario, the sample size calculations for the estimation of one sample mean with a desired precision, for the hypothesis testing of the sample mean with a specified power and for the comparison of two sample means (Section 3.4) will be illustrated.

This scenario is not based on any official reference.

3 Sample size calculations for one or two samples (simple random sampling)

3.1 Binary outcome for one sample

3.1.1 Sample size calculation to estimate a population proportion with a desired level of precision

Scenario 1 (Section 2.4.1) illustrates the sample size calculation for estimating a population proportion with desired precision. This type of calculation is based on the confidence interval. Sample size calculations for both finite and infinite populations are discussed. The desired precision can be specified either as an absolute or a relative error.

In order to estimate the proportion of antimicrobial resistant (AMR) *Escherichia coli* (*E. Coli*) present in the country X, ECDC would like to organise a survey obtaining isolates from the national reference laboratory. Disease experts at ECDC expect the proportion of AMR *E. coli* to be around 8%. They want to know the number of isolates that need to be collected and tested to be 95% confident that the proportion of AMR *E. coli* is within 15% of the true population proportion. The national reference laboratory in the country X has 15 000 isolates available for testing. A perfect diagnostic test to detect *E. coli* in the isolates is used in the survey (sensitivity and specificity equal to 100%).

How many isolates need to be collected, at random, from the national reference laboratory in the country X?

Theoretical background

Formula-based methods are available to perform sample size calculations for the estimation of a population proportion, under simple random sampling [3]:

$$n = \frac{z^2_{1-\frac{\alpha}{2}} N P_y(1-P_y)}{z^2_{1-\frac{\alpha}{2}} P_y(1-P_y) + (N-1)\epsilon_r^2 P_y^2}$$

The following quantities are used in the equation:

- *n* = the required sample size;
- $z_{1-\frac{\alpha}{2}}$ = the value from the standard normal curve corresponding to the desired confidence level of $1-\alpha$. Use $z_{1-\frac{\alpha}{2}}$ = 1.96 for 95% confidence level;
- *N* = the study population size;
- P_y= the population proportion for outcome *Y* of interest (in the expression above, the proportion is on the 0–1 scale);
- ϵ_r= the relative error.

The formula above generates the sample size, *n*, which is required to estimate the expected proportion with a desired precision. The formula will return the number of isolates needed to be 95% confident that the expected proportion of isolates containing AMR *E. coli* will be anywhere between (1-ϵ_r)*P_y and (1+ϵ_r)*P_y, or ϵ_r%, from the expected population proportion of P_y. In the example, this is between 6.8% and 9.2%.

When planning a study, it is desirable to determine the sample size so that the precision or maximum tolerable error is sufficiently small. The authors of the function '`epi.sssimpleestb`' refer to the desired precision as the maximum tolerable relative error. Note that a relative error of ϵ_r corresponds to an absolute error of $\epsilon_r * P_y$. In the example, the relative error of 15% in a population proportion of 8% thus corresponds to an absolute error of 1.2%.

Choosing the desired precision is not a statistical issue, but rather a practical one. It must be defined on a case-by-case basis. A precision of 15% might be fine in the example above, but in another setting a 10%, 5% or a smaller value is desirable. As a consequence of the precision ϵ_r becoming smaller, the sample size *n* will increase.

The population proportion P_y used in the formula is unknown, and is therefore replaced by an estimate of P_y. In case no good estimate is available, it is important to realise that a smaller value of P_y while fixing the relative error will lead to larger sample sizes. If on the other hand an absolute error is fixed, sample sizes will be the largest when the population proportion P_y is 0.5 (50%).

Implementation of the scenario in R

The formula above is implemented in the function '`epi.sssimpleestb`' of the package '`epiR`' in R software. The function requires the specification of the following arguments:

- **N** corresponds to the study population size. For the scenario discussed above, the study population size is 15 000. If **N** is left unspecified by the user, a default population size of 1 000 000 will be assumed. In the situation of an infinite population, either a large number or the default value of 1 000 000 can be used. **Py** corresponds to the expected population proportion of the outcome of interest. In this scenario, the expected population proportion is 8%, thus **Py** gets the value of 0.08.
- The desired precision is specified by the arguments **error** and **epsilon**. The scenario assumes a relative error of 15%. To indicate that the precision is 'relative', we need to specify 'relative' for the argument **error** in combination with a value of 0.15 in the argument **epsilon**. Alternatively, an absolute error can be defined in the function by specifying 'absolute' for the argument **error** and setting **epsilon** equal to 0.012 (=0.08x0.15).
- For surveys that involve the use of a diagnostic method, the characteristics of the diagnostic method are specified in the arguments **se** and **sp**. The argument **se** corresponds to the diagnostic sensitivity of the method used to detect positive outcomes and is indicated with a value ranging from 0–1; **sp** corresponds to the diagnostic specificity of the method used to detect positive outcomes. In this scenario, it is assumed that the diagnostic test is perfect and therefore both **se** and **sp** will be assigned the value of 1. However, if a diagnostic method is not perfect, then its sensitivity and specificity can be entered at the corresponding arguments.
- For surveys not involving a diagnostic test, or for which **se** and **sp** cannot be calculated or are unavailable, the arguments **se** and **sp** should be specified to be equal to 1.
- Since the investigator of the study is interested in a 95% confidence level, the argument **conf.level** will take the value 0.95, which is also the default value of the argument.

R code for sample size calculation

```
# Setting:
# N= 15 000
# Proportion (Py) = 0.08
# Relative error = 0.15 (or absolute error = 0.012)
# Sensitivity= 1
# Specificity= 1
# Confidence level = 0.95

# The function to calculate the sample size:
epi.sssimpleestb(N = 15 000,       # population size
            Py = 0.08,             # expected population prevalence
            Epsilon = 0.15,        # maximum tolerable error
            error = "relative",    # the error we refer to is a relative error
            se = 1,                # sensitivity
            sp = 1,                # specificity
            conf.level = 0.95      # confidence level
            )
```

R output

```
[1] 1737
```

Conclusion

A total of 1 737 isolates needs to be collected at random from the national reference laboratory in the country X to meet the requirements of the study, i.e. the disease experts are 95% confident that the estimated proportion of AMR *E. coli* isolates is within 15% of the true population proportion. The relative error of 15% translates, for this example, into an absolute error of 1.2%.

Note that when, in the code, the argument **error** is changed into error = "absolute" and the argument **epsilon** into epsilon = 0.012, the same result is obtained. This is to be expected as the relative error can always be translated into an absolute error. If there a finite population of isolates was not available, but instead an infinite population was available, we would change the argument **N** into N = 1 000 000, which would lead to a higher sample size of 1 964. Thus, if the population is finite, the sample size can be slightly reduced.

If the diagnostic test used was not perfect, but had a sensitivity of 95% and specificity of 99%, we would change the arguments **se** and **sp** into se = 0.95 and sp = 0.99. This results in a sample size of 2 035. In the case of an imperfect test, the required sample size in general would be larger than with a perfect test.

3.1.2 Sample size calculation for hypothesis testing of a population proportion

Scenario 1 (Section 2.4.1) is used to demonstrate the sample size calculation for a binary parameter that is being compared to a hypothesised value. For this sample size calculation, the function `'pwr.p.test'` from the R package `'pwr'` will be used. Hypothesis tests can be one-sided or two-sided. The sample size calculation for an infinite population size is discussed first, but a small sample correction is presented thereafter for a finite population.

This section continues to build on scenario 1, previously discussed in Section 3.1.1. To estimate the proportion of antimicrobial resistant (AMR) *E. coli* present in the country X, ECDC would like to organise a survey obtaining isolates from the national reference laboratory, assuming an infinite population. Historically, the proportion of AMR *E. coli* among isolates has been 8%. However, disease experts want to investigate whether the current proportion is different from the historical value of 8%. The experts opt for a two-sided test, given that they do not know if the current proportion will be lower or higher than 8%. For the sample size calculations, they assume that the population proportion value currently equals 6.5%. The investigators want to detect a difference with 90% power of the test, assuming that the test is carried out at a significance level of 5%.

Theoretical background

The sample size calculation for a single population proportion being compared to a known value, is often calculated based on the assumed adequacy of the normal approximation to the binomial distribution (formula not shown). That approach is reasonable when both the population proportion and the proportion under the null hypothesis are close to 0.5. If this requirement is not met, it is better to use an arcsine transformation of the proportions, which is more generally applicable [8]. The formula for sample size calculation for hypothesis testing of a population proportion using the arcsine transformation for a two-sided test is given by:

$$n = \left(\frac{z_{1-\beta} + z_{1-\frac{\alpha}{2}}}{2\arcsin(\sqrt{p_1}) - 2\arcsin(\sqrt{p_2})} \right)^2$$

The quantities used in the equation are:

- *n* = the required sample size;
- $z_{1-\frac{\alpha}{2}}$= the value from the standard normal distribution corresponding to the desired significance level (α). Use $z_{1-\frac{\alpha}{2}}$ = 1.96 for a two-sided hypothesis at a 5% significance level;
- $z_{1-\beta}$ = the value from the standard normal distribution corresponding to the desired study power (1-β). Use $z_{1-\beta}$ = 0.84 for 80% power and $z_{1-\beta}$ =1.28 for 90% power;
- p_1= the value of the proportion under the null hypothesis;
- p_2= the true population proportion.

Implementation of the scenario in R

The formula is implemented in the function `'pwr.p.test'` of the package `'pwr'` in R software. The function requires the specification of the following arguments:

- **h** corresponds to the effect size. In order to calculate the effect size, the R package `'pwr'` has a built-in function to calculate the effect size determined by $2arcsin(\sqrt{p_1}) - 2arcsin(\sqrt{p_2})$. The function `'ES.h'` can be used either outside the function to store the value it returns or directly within the `'pwr.p.test'` function. In the R code, the latter will be demonstrated. The function 'ES.h' has two arguments, namely **p1** and **p2**, corresponding to the proportion under the null hypothesis and the true population proportion, i.e. p_1 and p_2 in the formula, respectively. In the scenario, **p1** will be assigned the value of 0.08 and **p2** the value of 0.065;
- The investigator requested a significance level of 5%. Therefore, 0.05 is the value chosen for the argument **sig.level**;
- **power** corresponds to 0.9, as the investigator expressed the desire of being 90% sure of detecting a difference if in reality there is one;
- **alternative** corresponds to the type of alternative hypothesis that is of interest. Depending on the hypothesis test that is under investigation, the argument *alternative* can take several values: 'two.sided', 'greater' or 'less'. A two-sided alternative hypothesis is in place when the investigator is interested in a difference and not so much about the direction of this difference (H_1: $p_2 \neq p_1$), which is the case in the example. An alternative hypothesis can be 'greater' (H_1: $p_2 > p_1$) or 'less' (H_1: $p_2 < p_1$). For instance, in this scenario, the *alternative* is 'two.sided' since the investigator would like to detect a difference from p_1 or 0.08.

Note that the same function can be used to do a power calculation, instead of sample size calculation. One argument in the function, namely **n** gets the value 'NULL', which is the default value when the aim is to calculate the sample size given a preferred power. Instead, if the sample size was already calculated and the investigator would like to know the corresponding power, **n** would get the value of the sample size and the **power** either gets the value 'NULL' or is not mentioned in the function. Consequently, the function will return the power for that specific setting. Note that only one element at a time can get the value 'NULL' or cannot be referenced in the function.

R code for sample size calculation

```
# Setting:
# Proportion under null hypothesis (p1) = 0.08
# Population proportion (p2) = 0.065
# Significance level = 0.05
# Power = 0.90

# The function to calculate the sample size:
pwr.p.test(h=ES.h(p1=0.08, p2=0.065),      # effect size
          sig.level= 0.05,                  # significance level
          power = 0.90,                     # power of the test
          alternative = 'two.sided'         # two-sided hypothesis
           )
```

R output

```
     proportion power calculation for binomial distribution (arcsine transformation)

              h = 0.0579191
              n = 3132.222
      sig.level = 0.05
          power = 0.90
    alternative = two.sided
```

Conclusion

For an infinite population size, a total of 3 133 isolates needs to be collected at random, and tested from the national reference laboratory in the country X to meet the requirements of the study, i.e. disease experts at ECDC will have 90% power of detecting a difference if it exists, assuming that a two-sided test is carried out at a significance level of 5%. Note that the output gives a sample size n = 3132.222 corresponding to an effect size of h = 0.0579191. When the estimated sample size is not a whole number, this number should be rounded upward to the next whole number, resulting in a required sample size of 3 133.

If a two-sided hypothesis is not of interest, but rather a one-sided alternative hypothesis H_1: $p_2 > p_1$, we will change the argument **alternative** into alternative = 'greater'. This results in a required sample size of 2 553. Thus, note that to reliably detect an effect for a two-sided hypothesis, it requires a larger sample than to detect an effect for a one-sided hypothesis at the same significance level.

R code for power calculation

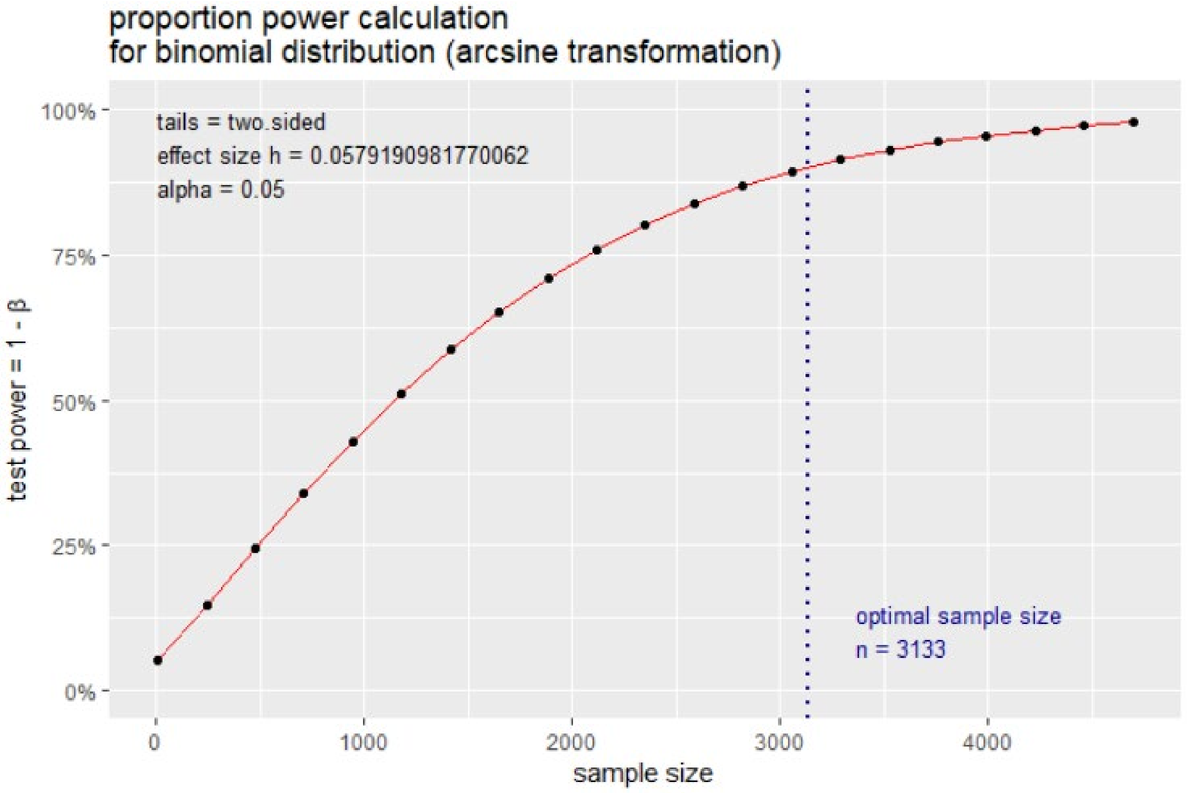
In addition to the numerical output via the '`pwr`' package, the package also provides a graphical overview of changes in power as the sample size changes. In order to obtain the graph, the result from the '`pwr.p.test`' function first has to be saved in R. Then, the results can be obtained using the '`plot`' function.

```
# Save the analysis into an object:
one.prop <- pwr.p.test(h=ES.h(p1=0.08, p2=0.065),
          sig.level= 0.05, power = 0.90, alternative = 'two.sided' )

# Make plot:
plot(one.prop)
```

R output

Figure 3. Power calculations as a function of sample size for one sample proportion



Conclusion

The graph that R returns shows that when the sample size increases, the power increases as well. While the graph is steep when the sample sizes are small, it levels off at higher sample sizes. This means that when the sample size is small, there is a large gain in power when increasing the sample size. However, if the sample size is high, the gain in power by increasing the sample size gets smaller.

R code for finite population correction

If the population size is no longer infinite, but finite, then an extra step needs to be performed after calling the function '`pwr.p.test`' in R. Let us assume that in the example used previously, the only quantity that changes is the population size *N*. The (finite) population size now consists of 15 000 isolates. In order to correct for this finite sample size, the following formula is used:

$$n_{corr} = \frac{n\ x\ N}{n + N}$$

where *n* is the sample size obtained from the function '`pwr.t.test`', i.e. 3132.222 and *N* is the population size, namely 15 000 isolates which can be sampled (at random). In order to correct for this finite sample size, we can use the function '`correct.N`' (see Annex for the source file). The function has two arguments:

- **n**, i.e. the sample size obtained from the appropriate R function.
- **N**, i.e. the population size. For this scenario, **N** would get the value 15 000.

```
# Correction of sample size if population size is finite
correct.N(n=3132.222, N=15 000)
```

R output

```
[1] 2591.151
```

Conclusion

For an infinite population size, a total of 3 133 isolates needs to be collected at random and tested from the national reference laboratory in the country X to meet the requirements of the study. Instead if the population is finite and consists of 15 000 isolates that can be sampled, then it is sufficient to sample 2 592 isolates at random to meet the requirements of the study.

Note again, that it is important to round the number given in the R output upwards to the nearest integer. In the situation of a finite population, a slightly smaller sample size is required as compared to a sample taken from an infinite population, because each sample proportionately contains more information in a finite population than in an infinite population.

3.2 Binary outcomes for two samples

In this section we consider the two-sample problem in which two population proportions need to be compared. There are three common summary measures to compare the two proportions p_1 and p_2:

- using the **absolute difference** of the proportions $p_1 - p_2$;
- using a **relative risk** defined as the ratio of the two proportions p_1/p_2; or
- as an **odds ratio** defined as $\frac{p_1/(1-p_1)}{p_2\,/(1-p_2)}$.

Sample size calculations related to each of these summary measures are presented in this section. As indicated by Julious (2010) [9], although the three approaches seem to be quite different, they are the same – approximately and algebraically. Thereafter, sample size calculations for dependent proportions will be described, as well as sample size calculations for non-inferiority tests.

3.2.1 Sample size estimation for a difference in two independent proportions

Scenario 1 (Section 2.4.1) is now used to illustrate the sample size calculation to estimate a difference between two independent population proportions. To perform the calculation, the function '`pwr.2p.test`' from the R package '`pwr`' is used.

Disease experts from ECDC would like to conduct a survey in two countries, X and Y, and obtain isolates from their national reference laboratories. The aim is to estimate whether there is a difference in the proportion of AMR *E. coli* isolates collected in the same time period between the two countries. The disease experts would like to detect a difference in the proportion of AMR *E. coli* isolates as small as 2.5%. They expect one country to have a proportion of AMR *E. coli* isolates of 10%, while the other is expected to have a proportion of 7.5%. Again, if the difference exists, the investigators wish to be 90% confident of detecting it (= power of the test), assuming that a two-sided test is carried out at a significance level of 5%.

How many isolates should be sampled at random, in the reference laboratories of both countries, so that the disease experts achieve the desired power to test their hypothesis?

Theoretical background

Similar to Section 3.1.2, this section focuses on the sample size calculations that are implemented in the R function '`pwr.2p.test`'. The advantage of the proposed calculation and function is that it does not rely on the normal approximation of the binomial distribution, which requires both proportions to be close to 0.5. Therefore, this section displays the formula for the sample size calculation to test for a difference in two independent populations using the arcsine transformation [8].

$$n = \frac{1}{2}\left(\frac{z_{1-\beta} + z_{1-\frac{\alpha}{2}}}{\arcsin(\sqrt{p_1}) - \arcsin(\sqrt{p_2})}\right)^2$$

The quantities used in the equation are:

- n = the number of subjects in each group;
- $z_{1-\frac{\alpha}{2}}$= the value from the standard normal distribution corresponding to the desired significance level (α). Use $z_{1-\frac{\alpha}{2}}$ = 1.96 for a two-sided hypothesis at a 5% significance level;
- $z_{1-\beta}$ = the value from the standard normal distribution corresponding to the desired study power (1-β). Use $z_{1-\beta}$ = 0.84 for 80% power and $z_{1-\beta}$ =1.28 for 90% power;
- p_1= the expected population proportion in the first group;
- p_2= the expected population proportion in the second group.

Implementation of the scenario in R

The formula is implemented in the function '`pwr.2p.test`' of the package '`pwr`' in R software. Although the name of the function is different from the function '`pwr.p.test`', which is used to calculate the sample size for a proportion with the intent of doing hypothesis testing, the arguments of both functions are the same. Note also, that the effect size is determined in the same way. More information is available in Section 3.1.2.

R code for sample size calculation

```
# Setting:
# N = infinite population size
# Proportion (p1) = 0.10
# Proportion (p2) = 0.075
# Significance level = 0.05
# Power = 0.90

# The function to calculate the sample size:
pwr.2p.test(h = ES.h(p1 = 0.10, p2 = 0.075),  # effect size
           sig.level = 0.05,                  # significance level
           power = 0.90,                      # power of the test
           alternative = 'two.sided'          # two-sided hypothesis
           )
```

R output

```
    Difference of proportion power calculation for binomial distribution (arcsine transformation)

              h = 0.08869008
              n = 2671.628
      sig.level = 0.05
          power = 0.90
    alternative = two.sided

NOTE: same sample sizes per group
```

Conclusion

When the study population under investigation is of an infinite size, a total of 5 344 isolates are needed. This means 2 672 isolates each in the reference laboratories of country X and Y need to be selected at random and tested. Notice that the sample size obtained is the sample size per group*.* As before, when the estimated sample size is not a whole number, this number should be rounded upward to the next integer.

Note that, the smaller the difference is between the population proportions in the two groups ($|p_1 - p_2|$), the larger is the required sample size in both the groups. Keeping the difference between the two population proportions constant, the closer the population proportions are to 0.5, the larger the sample size.

R code for finite population correction

If the study population is finite, rather than infinite as discussed above, then the sample size needs to be corrected. The formula to do that has been described in Section 3.1.2. The function created in R to correct for finite population size `'correct.N'` has also been described in Section 3.1.2. For example, if the number of isolates available per country is 15 000, this would lead to the following code:

```
# Correction of sample size if population size is finite
correct.N(n=2671.628 , N= 15 000)
```

R output

```
[1] 2267.727
```

Conclusion

If the population is finite and consists of 15 000 isolates per country that can be sampled, then it is sufficient to sample 2 268 isolates at random to meet the requirements of the study from each country. This will result in a total sample size of 4 536 isolates.

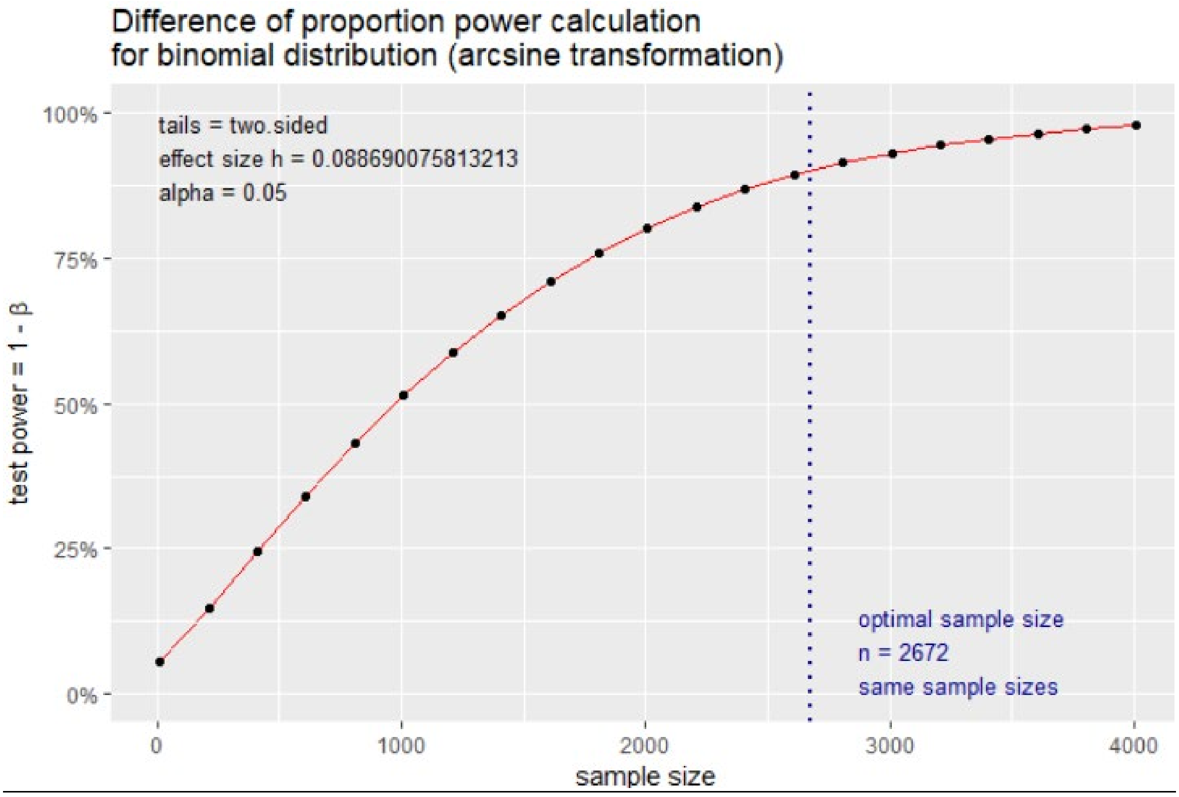
R code for power calculation

Similar to Section 3.1.2, we can visualise the changes in power by using the 'plot' function.

```
# Save the analysis into an object:
two.prop <- pwr.2p.test(h=ES.h(p1=0.10, p2=0.075),
          sig.level= 0.05, power = 0.90, alternative = 'two.sided' )
# Make plot
plot(two.prop)
```

R output

Figure 4. Power calculations as a function of sample size for two independent proportions



Conclusion

The graph shows the increase of power with increasing sample size. The optimal sample size corresponding to a power of 90% are 2 672 isolates in both the groups, or a total of 5 344 isolates.

3.2.2 Sample size estimation for odds ratio

An **odds ratio** is a commonly used statistic to compare the relative odds of the occurrence of an outcome of interest (e.g. disease). The odds ratio represents the odds that an outcome will occur in population 1, compared to the odds of the outcome occurring in population 2. This section explains sample size calculations when testing the null hypothesis that the odds ratio equals 1, versus the alternative hypothesis that the odds ratio is different from 1. Scenario 2 (Section 2.4.2) is used to illustrate the sample size calculation, using the function `'RelativeRisk.Equality'` from the R package `'TrialSize'`. A comparison is made with a sample size calculation based on an absolute risk difference, using the function `'pwr.2p.test'` from the R package `'pwr'` (introduced in Section 3.2.1).

Disease experts from ECDC want to estimate the influenza vaccine effectiveness (VE) at the beginning of a seasonal influenza epidemic among infants and children in the country Z. In the study, medically attended acute respiratory infection (MAARI) caused by the influenza virus is under investigation. The research question is formulated in terms of the odds ratio, comparing the odds of a MAARI for vaccinated children with the odds of a MAARI for unvaccinated children. Previous data suggest that 1.5% of the unvaccinated children have a MAARI during the flu season. Based on an expected vaccine effectiveness of 50%, the odds ratio is expected to be 0.496 (further details are given in the subsequent sections). An equal number of vaccinated and unvaccinated children will be selected at random from the computerised databases of general practitioners (GPs) throughout the year.

How many children need to be selected to have 90% power, for a one-sided 5% significance level? A one-sided hypothesis is used because the aim is to show that the odds ratio is lower than 1.

Theoretical background

The following formula determines the sample size when the difference between two populations is expressed in terms of the odds ratio [10]:

$$n_c = \frac{(z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2}{\log^2(OR)} \left(\frac{1}{kp_T(1-p_T)} + \frac{1}{p_C(1-p_C)} \right)$$

The following quantities are used in the equation:

- n_C = the number of subjects in the second group;
- k = the ratio of sample size in the first versus second group;
- $z_{1-\frac{\alpha}{2}}$= the value from the standard normal distribution corresponding to the desired significance level (α). Use $z_{1-\frac{\alpha}{2}}$ = 1.96 for a two-sided hypothesis at a 5% significance level;
- $z_{1-\beta}$ = the value from the standard normal distribution corresponding to the desired study power (1-β). Use $z_{1-\beta}$ = 0.84 for 80% power and $z_{1-\beta}$ =1.28 for 90% power;
- p_T= the expected population proportion in the first group;
- p_C= the expected population proportion in the second group;
- OR = odds ratio. The odds ratio is the odds of the outcome in the first group relative to the odds of the outcome in the second group. The OR can be calculated with the following formula: $\frac{p_T/(1-p_T)}{p_C\,/(1-p_C)}$.

Implementation of the scenario in R

Firstly, the scenario is implemented using the odds ratio. Secondly, a comparison is made with the absolute risk difference approach.

Implementation of the sample size calculation using odds ratio in R

The formula based on the odds ratio approach is implemented in the function '`RelativeRisk.Equality`' of the package '`TrialSize`' in R software. The function requires the specification of the following arguments:

- **alpha** corresponds to the desired two-sided significance level. In the scenario, as it is in a one-sided test, this argument is assigned the value of 2*0.05, which corresponds to the one-sided significance level of 0.05.
- **beta** corresponds to 1 minus the power (power=1-β). In the scenario, this argument is assigned the value of 0.1.
- **k** corresponds to the ratio of sample size in the two groups. In the scenario, it is assigned the value of 1.
- **pt** corresponds to the expected proportion in the first group. In the scenario, this is the group of vaccinated children. **pt** is assigned the value of 0.0075, i.e. the expected proportion of vaccinated children experiencing MAARI. This comes from the information that the experts expect a 50% vaccine effectiveness, so 0.5 times the proportion of the vaccinated children experiencing MAARI (1.5%).
- **pc** corresponds to the expected proportion in the second group. In the scenario, this is the group of unvaccinated children. **pc** will be assigned the value of 0.015, i.e. the expected proportion of unvaccinated children experiencing MAARI.

R code for sample size calculation

```
# Setting:
# N = infinite population size
# Expected population proportion in group 1 (pt) = 0.0075
# Expected population proportion in group 2 (pc) = 0.015
# Expected odds ratio (OR) = 0.496
# Significance level = 0.05 (one-sided). The function uses a two-sided significance level so alpha= 0.10 (two-sided)
# Power = 0.90 or beta = 0.10
# Equal sample sizes for group 1 and group 2 (k=1)

# Calculate the odds ratio
pt<-0.0075
pc<-0.015
OR <- (pt*(1-pc))/(pc*(1-pt))

# The function to calculate the sample size:
RelativeRisk.Equality(alpha = 0.1,  # significance level
               beta = 0.1,          # type II error
               or = OR,             # expected odds ratio
               k = 1,               # ratio of sample sizes in the two groups
               pt = pt,             # expected proportion in group 1(vaccinated)
               pc = pc              # expected proportion in group 2 (unvaccinated group
               )
```

R output

```
3523.422
```

Conclusion

For a study population of infinite size, a total of 7 048 children are needed, i.e. 3 524 unvaccinated and 3 524 vaccinated children, to have 90% power to show that the odds ratio is less than 1. Note that the sample size given in the R output is the size per group.

Calculations based on the absolute risk difference

The implementation of the absolute risk difference approach is similar to the approach described in Section 3.2.1. The formula is implemented in the function '`pwr.2p.test`' of the package '`pwr`' in R software.

R code for sample size calculation

The expected proportion of unvaccinated children experiencing MAARI is 1.5% (p2), and the expected proportion of vaccinated children experiencing MAARI equals 0.75% (p1).

How many children are needed to have 90% power to detect a difference (in fact a decrease) of proportion between the vaccinated and unvaccinated groups, using a one-sided test with a 5% level of significance?

```
# Setting:
# N = infinite population size
# Expected proportion in group 1 (p1) = 0.0075
# Expected proportion in group 2 (p2) = 0.015
# Significance level (alpha) = 0.05
# power = 0.90

pwr.2p.test(h = ES.h(p1 = 0.0075, p2 = 0.015),      # effect size
            sig.level = 0.05,                        # significance level
            power = 0.90,                            # power of the test
            alternative = 'less'                     # one-sided hypothesis
            )
```

R output

```
Difference of proportion power calculation for binomial distribution (arcsine transformation)

          h = -0.0721432
          n = 3290.85
  sig.level = 0.05
      power = 0.90
  alternative = less

NOTE: same sample sizes per group
```

Conclusion

The sample size provided in the R output is the required sample size per group. A total of 6 582 children, 3 291 unvaccinated and 3 291 vaccinated children, need to be sampled in the country Z to meet the requirements of the study.

Notice that this number is similar to the sample size based on the odds ratio, although slightly lower (about 7%). When the population proportions are close to 0.5, the two methods will be even more similar to each other.

3.2.3 Sample size estimation for relative risk

The relative risk is another commonly used statistic for the comparison of proportions. Relative risk is the ratio of the probability of an outcome in group 1 to the probability of the outcome in group 2. This section discusses how sample size calculations can be obtained when we want to test the null hypothesis that the relative risk equals 1, versus the alternative hypothesis that the relative risk is less than 1.

Disease experts from ECDC want to estimate the influenza vaccine effectiveness (VE) at the beginning of a seasonal influenza epidemic among infants and children in the country Z. In the study, medically attended acute respiratory infection (MAARI) caused by the influenza virus is under investigation. The research question is formulated in terms of the relative risk, comparing the risk of a MAARI for vaccinated children with the risk of a MAARI for unvaccinated children. Previous data suggest that 1.5% of the unvaccinated children have a MAARI during the flu season. The experts anticipate a 50% vaccine effectiveness, corresponding to a relative risk of 0.5. An equal number of vaccinated and unvaccinated children will be selected at random from the computerised databases of general practitioners (GPs) throughout the year.

How many children need to be selected to have 90% power, for a one-sided 5% significance level? A one-sided hypothesis is used because the aim is to show that the relative risk is lower than 1.

Theoretical background

The formula to calculate the sample size for detecting a relative risk is as follows [1]:

$$n = \frac{r+1}{r(\lambda-1)^2\pi^2}\left[z_{1-\frac{\alpha}{2}}\sqrt{(r+1)p_c(1-p_c)} + z_{1-\beta}\sqrt{\lambda\pi(1-\lambda\pi)+r\pi(1-\pi)}\right]^2$$

The following quantities are used in the equation:

- *n* = the total sample size required (in all groups together);
- *r* = the allocation ratio of the sample in group 1 versus group 2, i.e. the ratio of the sample size in group 1 versus the sample size in group 2;
- $z_{1-\frac{\alpha}{2}}$ = the value from the standard normal distribution corresponding to the desired significance level (α). Use $z_{1-\frac{\alpha}{2}}$ = 1.96 for a two-sided hypothesis or $z_{1-\alpha} = 1.64$ for a one-sided hypothesis at a 5% significance level;
- $z_{1-\beta}$ = the value from the standard normal distribution corresponding to the desired study power (1-β). Use $z_{1-\beta}$ = 0.84 for 80% power and $z_{1-\beta}$ =1.28 for 90% power;
- π = the expected population proportion for group 2;
- λ = the relative risk (the ratio of the probability of outcome in group 1 to the probability in group 2);
- p_c = the pooled population proportion calculated by $p_c = \frac{\pi(r\lambda+1)}{r+1}$. Note that this reduces to the average of the population proportions of group 1 and group 2 if equal-sized groups are assumed (i.e. if allocation ratio *r=1*).

Implementation of the scenario in R

The formula is implemented in the function '`epi.sscohortc`' of the package '`epiR`' in R software. The function requires the specification of the following arguments:

- **irexp0** is the expected proportion of the outcome in group 2. In the scenario, this is the unvaccinated group of children. The argument is assigned the value of 0.015.
- **irexp1** is the expected proportion of the outcome in group 1. In the scenario, this is the vaccinated group of children. The value is obtained as $\lambda * \pi = 0.5 * 0.015 = 0.0075.$
- The investigator required a significance level of 5%. Therefore, 0.95 is the value chosen for the confidence level specified in the argument **conf.level**;
- **power** corresponds to 0.90, as the investigator wanted to be 90% sure of detecting a relative risk less than 1;
- **sided.test** corresponds to 1 if the test is one-sided, and 2 if two-sided. Here, it will be a one-sided test since we are testing the alternative hypothesis that the relative risk is less than 1.
- **r** is the ratio of the sample size in group 1 versus in group 2. In this scenario, it is specified as 1.

R code for sample size calculation

```
# Setting:
# N = infinite population size
# Expected proportion in group 1 (irexp1) = 0.0075
# Expected proportion in group 2 (irexp0) = 0.015
# Significance level (alpha) = 0.05
# Power = 0.90

epi.sscohortc(irexp1 = 0.0075,        # proportion of outcome in the vaccinated group
              irexp0 = 0.015,         # proportion of outcome in the unvaccinated group
              power = 0.90,           # power of the test
              r = 1,                  # ratio of the sample sizes in the two groups
              sided.test = 1,         # one-sided hypothesis
              conf.level = 0.95)      # confidence level
```

R output

```
$n.total
[1] 6772

$n.exp1
[1] 3386

$n.exp0
[1] 3386

$power
[1] 0.9

$irr
[1] 0.5

$or
[1] 0.4962217
```

Conclusion

For the specified significance level and power, a total of 6 772 children (3 386 vaccinated and 3 386 unvaccinated) are needed in the study. The relative risk (irr in the R output) is the risk of the outcome in the vaccinated group, divided by the risk of the outcome in the unvaccinated group. It equals 0.5.

The sample size based on the relative risk is similar to the sample size calculations based on the odds ratio or the risk difference. When the population proportions are close to 0.5, the results will be even closer to each other.

3.2.4 Sample size estimation for a difference in two dependent population proportions

Scenario 1 (Section 2.4.1) is used in this section to illustrate the sample size calculation to estimate a difference between two dependent population proportions. Two approaches will be illustrated in this section. To perform the calculations, the function ‘`McNemar.Test`’ from the R package ‘`TrialSize`’ and the function ‘`sampleSizeMcNemar`’ is used (see Annex for the source file).

This section continues to build on scenario 1, previously discussed in Section 3.1.1. A new diagnostic test is developed to detect antimicrobial resistant (AMR) *E. coli*. The new diagnostic test needs to be compared with the old test. ECDC wants to organise a survey obtaining isolates from the national reference laboratory, assuming an infinite population. Each isolate will be analysed with both the old and new diagnostic tests, in order to make a comparison between them. The investigators wish to detect a difference in the proportion of isolates that test positive for AMR using the two diagnostic tests with 90% power of the test, assuming that the test is carried out at a significance level of 5%.

Depending on the information available from literature or pilot data, for the expected difference between the two dependent proportions, a different method for the sample size calculation can be chosen.

For the first method, investigators need to be able to make a guess about the proportion of tests that would shift from a positive outcome (using the old test) to a negative outcome (using the new test), and vice versa. Suppose that the investigators believe that the proportion of tests that are positive using the old test, but negative using the new test is 1%. In addition, investigators believe that the proportion of tests that are negative using the old test, but positive using the new test is 2%.

For the second method, investigators only need to make a guess about the proportion of isolates that would test positive using the old and new tests. For the sample size calculations, investigators assume that the proportion of isolates that test positive for AMR based on the old test is 8%, while the proportion of isolates that test positive for AMR based on the new test is 9%.

How many isolates should be sampled at random, so that the investigators achieve the desired power to test their hypothesis?

Theoretical background for method 1

In contrast to the previous sections, two correlated/dependent proportions are of interest here. McNemar's test is designed for the comparison of two correlated proportions [11]. Studies with a 'before' and 'after' design are examples of dependent proportions. For example, when the binary outcome is collected on the same unit before and after an intervention or treatment. Another example is the comparison of a new and a standard diagnostic test on the same isolates. Here, we are interested in testing the null hypothesis that the proportions for the old test (or before intervention) and the new test (or after intervention) are the same, and the alternative hypothesis is the inequality of the two proportions.

Different methods exist to compute the sample size for testing dependent proportions. These depend on the information that is available about the outcome. The method presented in this section can be used when the proportion of subjects that shift from a positive outcome (e.g. using the first test) to a negative outcome (e.g. using the second test), and vice versa, are available. The formula to calculate the sample size is given below [10]:

$$n = \frac{\left[z_{1-\frac{\alpha}{2}}(\psi+1) + z_{1-\beta}\sqrt{(\psi+1)^2 - (\psi-1)^2\pi_{disc}}\right]^2}{(\psi-1)^2\pi_{disc}}$$

The quantities used in the equation are:

- *n* = the required sample size;
- $z_{1-\frac{\alpha}{2}}$ = the value from the standard normal distribution corresponding to the desired significance level (α). Use $z_{1-\frac{\alpha}{2}}$ = 1.96 for a two-sided hypothesis at a 5% significance level;
- $z_{1-\beta}$ = the value from the standard normal distribution corresponding to the desired study power (1-β). Use $z_{1-\beta}$ = 0.84 for 80% power and $z_{1-\beta}$ =1.28 for 90% power;
- p_{01} = the proportion of subjects that shift **from not having** the outcome based on the first test, **to having** the outcome based on the second test;
- p_{10} = the proportion of subjects that shift **from having** the outcome based on the first test, **to not having** the outcome based on the second test;
- $\psi = \frac{p_{01}}{p_{10}}$ is the ratio of the proportion of subjects that make the shift from a negative to positive outcome with the proportion of subjects that make the shift from a positive to negative outcome;
- $\pi_{disc} = p_{01} + p_{10}$ is the total proportion of subjects that make a shift in the outcome.

Implementation of the scenario in R

The formula is implemented in the function '`McNemar.Test`' from the R package '`TrialSize`'. The function requires the specification of the following arguments:

- **alpha** is the significance level. The investigator requires a significance level of 5%. Therefore, 0.05 is passed to the argument **alpha**.
- **beta** corresponds to the type II error (β) (power = 1 – β). In the scenario, this argument is assigned the value of 0.1.
- **psai** is the ratio of $\frac{p_{01}}{p_{10}}$. This is equal to ψ in the formula and is assigned a value of $\frac{p_{01}}{p_{10}} = \frac{0.02}{0.01} = 2$.
- **paid** is the sum of $p_{01} + p_{10}$. This is equal to π_{disc} and has a value of $p_{01} + p_{10} = 0.02 + 0.01 = 0.03$.

R code for sample size calculation

```
# Setting:
# Proportion shifting from a negative to a positive outcome (p01) = 0.02
# Proportion shifting from a positive to a negative outcome (p10) = 0.01
# Significance level= 0.05
# Power = 0.90, beta = 0.10

p01=0.02
p10=0.01
psai=p01/p10
paid=p01+p10

mcnemar<-McNemar.Test(alpha=0.05,       # significance level
                beta=0.10,              # type II error
                psai=psai,              # ratio of proportions for discordant cell
                paid=paid)              # sum of proportions for discordant cell
mcnemar
```

R output

```
[1]  3148.071
```

Conclusion

The total number of isolates required for the specified significance level and power is 3 149. Note that the output gives a sample size n = 3148.071. This number should be rounded upward to the next integer, resulting in a required sample size of 3 149.

Theoretical background for method 2

While the previous method requires information about the proportion of subjects that shift from a positive to a negative outcome, and vice versa, this information is not always available. Often, only information about the proportion of positive outcomes with the old and new test is available. However, this information is not sufficient to determine an 'exact' sample size. Lachenbruch (1992) [11] proposes the calculation of a range of possible sample sizes. The method only needs information on the proportions in the two dependent groups (e.g. the proportions before and after intervention).

The formula for the sample size calculation is given below:

$$n = \frac{0.25(z_{1-\beta} - z_{1-\frac{\alpha}{2}})^2}{(0.5 - s)^2(|p_2 + p_1 - 2p_{11}|)}$$

The quantities used in the equation are:

- *n* = the required sample size;
- $z_{1-\frac{\alpha}{2}}$= the value from the standard normal distribution corresponding to the desired significance level (α). Use $z_{1-\frac{\alpha}{2}}$ = 1.96 for a two-sided hypothesis at a 5% significance level;
- $z_{1-\beta}$ = the value from the standard normal distribution corresponding to the desired study power (1-β). Use $z_{1-\beta}$ = 0.84 for 80% power, and $z_{1-\beta}$ =1.28 for 90% power;
- p_1= the expected proportion having a positive outcome, for the first test;
- p_2= the expected proportion having a positive outcome, for the second test;
- p_{11}= the expected proportion having a positive outcome for the first and second test. This probability is rarely known in practice, but ranges between min(p_1, p_2) and max(0, p_1 + p_2 - 1);
- s = (p_2 - p_{11})/(p_2 + p_1 - 2p_{11}), which is calculated for the minimal and maximal value of p_{11}. In addition, a value of s midway between the minimal and maximal values are computed. This is called the midpoint.

Implementation of the scenario in R

The formula is implemented in the function '`sampleSizeMcNemar`' (see Annex for the source file). The function requires the specification of the following arguments:

- **p1** and **p2**, correspond respectively to the proportion of positive outcomes for the first test, and the proportion of positive outcomes for the second test. In the scenario, **p1** is assigned the value 0.08, and **p2** the value 0.09.
- **alpha** is the significance level. The investigator required a significance level of 5%. Therefore, 0.05 is passed to the argument **alpha**.
- **power** corresponds to 0.9, as the investigator required 90% surety of detecting a difference.
- specify **plot=TRUE** to make a plot of the possible sample sizes as a function of the proportion of tests that make a switch from a positive to a negative result among the tests that make a shift.

R code for sample size calculation

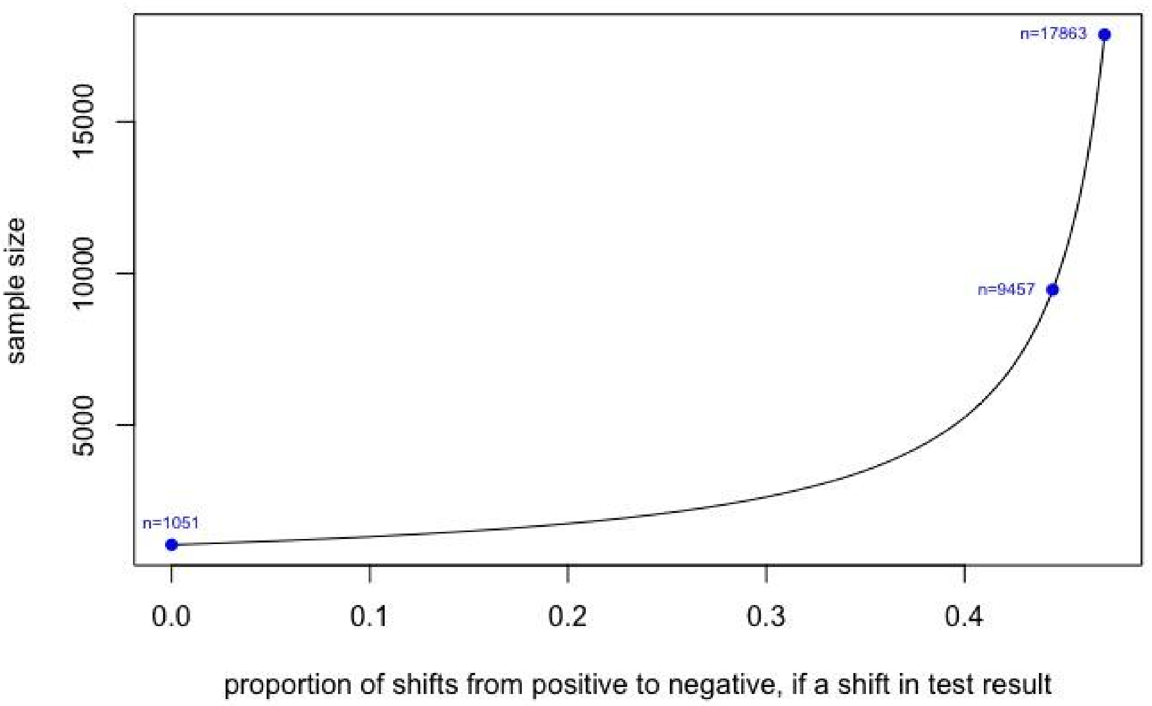
```
# Setting:
# N = infinite population size
# Proportion having positive outcome for first test (p1) = 0.08
# Proportion having positive outcome for second test (p2) = 0.09
# Significance level= 0.05
# Power = 0.90

sampleSizeMcNemar(p1=0.08,        # proportion having positive outcome for test 1
            p2=0.09,              # proportion having positive outcome for test 2
            alpha = 0.05,         # significance level
            power = 0.90)         # power of the test
```

R output

```
N_min N_mid N_max
17863  9457  1051
```

Figure 5. Possible sample sizes as a function of the proportion of tests that make a switch from a positive to a negative result, among the tests that make a shift



Conclusion

Note that this method does not specify an exact sample size, but a range of possible sample sizes. In the R output, three values for the sample size are given: the minimum, midpoint and maximum. These are the sample sizes based on the smallest and the largest possible proportion of tests that make the shift from a positive to negative outcome, among the tests that make a shift (from positive to negative or vice versa), and the midpoint of these. The values given by the minimum and maximum reflect the level of uncertainty on the sample size. From the output, we see that at least 1 051 samples are needed, and at most 17 863. The midpoint value of 9 457 samples is a reasonable compromise. Therefore, the number of isolates needed to be sampled is 9 457.

The plot illustrates the relationship between the sample size and the proportion of shifts from positive to negative, among the tests with a shift in the test result. Note that in the previous method, it was assumed that the proportion of isolates that shift from a positive to a negative result equals 0.01, and the proportion switching from negative to positive equals 0.02. As a result, the proportion of shifts from positive to negative among the isolates with a change in outcome equals 0.01/(0.01+0.02)=0.33. In the plot, the sample size corresponding to this proportion is 3 153, resembling the result in the previous section. The plot illustrates the uncertainty in the sample size due to the missing information about the amount of shifts in the outcome, and gives an idea of the possible range of sample sizes required for the expert to make the best choice based on the available information.

3.2.5 Sample size estimation for a non-inferiority test

The objective of a non-inferiority study is to show that a treatment is not inferior to a standard treatment or control. Non-inferiority can be investigated in terms of the difference between two treatments or in terms of the relative effect (e.g. odds ratio, risk or benefit) of the treatments for the disease under study.

Two-sample proportion test for non-inferiority using absolute difference

Scenario 2 (Section 2.4.2) is used to illustrate the sample size calculation for a non-inferiority test. In order to perform the calculation, the function `‘TwoSampleProportion.NIS’` from the R package `‘TrialSize’` is used.

Disease experts from ECDC want to compare two different vaccines which are used in vaccination centers in the country X. It is of interest to establish non-inferiority of the new vaccine in protecting children against the medically attended acute respiratory infection (MAARI) caused by the influenza virus, as compared to the standard vaccine. A difference in the rate of effectiveness between the two vaccines of less than 10% is considered to be of no clinical importance. Thus, the non-inferiority margin is chosen to be 10% (i.e. δ = -0.10). Previous data suggest a rate of effectiveness of 65% for the standard vaccine. The experts expect a rate of effectiveness of 60% for the new vaccine.

Using a one-sided 5% significance level, how many children should be sampled at random so that the experts can investigate non-inferiority of the new vaccine with 90% power?

Theoretical background

Let ε = $p_1 - p_2$ be the difference between the population response rates of a treatment (p_1) and a control (p_2). For a non-inferiority test, the following hypotheses are used:

H0: $\varepsilon \leq \delta$ versus Ha: $\varepsilon > \delta$,

where δ is the non-inferiority margin. When δ < 0, the rejection of the null hypothesis indicates non-inferiority of the treatment against the control.

Chow et al., (2018) [10] proposed the following formula to determine the sample size required in a non-inferiority study:

$$n_1 = kn_2$$

$$n_2 = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{(\epsilon - \delta)^2}\left[\frac{p_1(1-p_1)}{k} + p_2(1-p_2)\right]$$

The following quantities are used in the equation:

- n_1= the required sample size in the treatment group;
- n_2= the required sample size in the control group;
- $k = \frac{n_1}{n_2}$ = the ratio between the sample size in the treatment and control groups;
- $z_{1-\alpha}$= the value from the standard normal distribution corresponding to the desired significance level (α). Use $z_{1-\alpha} = 1.64$ for one-sided hypothesis at a 5% significance level;
- $z_{1-\beta}$ = the value from the standard normal distribution corresponding to the desired study power (1-β). Use $z_{1-\beta}$ = 0.84 for 80% power, and $z_{1-\beta}$ =1.28 for 90% power;
- p_1 = the population response rates in the treatment group;
- p_2 = the population response rates in the control group;
- ε = the difference between p_1 and p_2 (calculated as p_1- p_2);
- δ = the non-inferiority margin.

Implementation of the scenario in R

The formula above is implemented in the function 'TwoSampleProportion.NIS' of the R package 'TrialSize' in R software. The function requires the specification of the following arguments:

- **alpha** corresponds to the desired one-sided significance level. In the scenario, this argument is assigned the value of 0.05.
- **beta** corresponds to the type II error (β) (power = 1 – β). In the scenario, this argument is assigned the value of 0.1.
- **p1** corresponds to the population response rate for the treatment. In the scenario, it is assigned the value 0.6.
- **p2** corresponds to the population response rate for the control. In the scenario, it is assigned the value 0.65.
- **k** corresponds to the ratio of sample size in the treatment and control group. In the scenario, it is assigned the value of 1 such that the same number of children receive treatment (with the new vaccine) and control (with the standard vaccine).
- **delta** corresponds to the difference between **p1** and **p2** (this is ε in the formula). In the scenario, it is assigned the value of -0.05.
- **margin** corresponds to the non-inferiority margin. In the scenario, this argument is assigned the value of -0.10. The non-inferiority margin is the amount by which the effect of a new treatment is not considered to be unacceptably worse than the effect of the standard treatment. In the scenario, this means that the difference between the effectiveness of the standard vaccine and the new is lower than 10%.

R code for sample size calculation

```
# Setting:
# N = infinite population size
# p1 = 0.6
# p2 = 0.65
# Significance level (alpha) = 0.05 (one-sided)
# Power = 0.90 or beta = 0.10

TwoSampleProportion.NIS(alpha = 0.05,     # significance level
                        beta = 0.10,      # type II error
                        p1 = 0.6,         # rate of effectiveness of the new vaccine
                        p2 = 0.65,        # rate of effectiveness of the standard vaccine
                        k = 1,            # ratio of sample size for vaccinated and unvaccinated groups
                        delta = -0.05,    # difference in the rates of effectiveness of two vaccines
                        margin = -0.10)   # non-inferiority margin
```

R output

```
1601.439
```

Conclusion

The sample size returned is the sample size per group. Hence, a sample of 1 602 vaccinated children with the new vaccine, and 1 602 vaccinated children with the standard vaccine need to be included in the study to have 90% power to demonstrate non-inferiority of the new vaccine. This results in a total sample size of 3 204 children.

Two-sample proportion test for non-inferiority using odds ratio

Often the aim is to investigate the relative effect of the treatments for the disease under study. Odds ratio has been frequently used to assess the association between a binary treatment variable and a binary disease outcome [10].

Scenario 2 is used to illustrate the sample size calculation for a non-inferiority test. In order to perform the calculation, the function `'RelativeRisk.NIS'` from the R package `'TrialSize'` is used.

Disease experts from ECDC want to compare two different vaccines which are used in vaccination centers in the country X. It is of interest to establish non-inferiority, in terms of the odds ratio (OR), of the new vaccine in protecting vaccinated children against the medically attended acute respiratory infection (MAARI) caused by the influenza virus, as compared to the standard vaccine. Previous data suggest the OR to be equal to 1.45 (based on information that 3.5% of children receiving the standard vaccine experienced MAARI, and the experts expect 5% of children receiving the new vaccine to experience MAARI). For the new vaccine to be non-inferior to the standard vaccine, it is specified that an OR not lower than 0.90 is considered as a non-clinically important difference.

Using a one-sided 5% significance level, how many children should be sampled at random such that the experts can investigate non-inferiority of the new vaccine with 90% power?

Theoretical background

For a non-inferiority test, the following hypotheses are used:

$$H_0\text{: OR} \leq \delta' \text{ versus } H_1\text{: OR} > \delta'$$

where δ′ is the non-inferiority margin for the OR. The above hypotheses can be rewritten as:

$$H_0\text{: } \log(\text{OR}) \leq \delta \text{ versus } H_1\text{: } \log(\text{OR}) > \delta$$

where δ is the non-inferiority margin in natural logarithmic scale, and log refers to the natural logarithm. When δ < 0, the rejection of the null hypothesis indicates non-inferiority of the treatment against the control.

Chow et al. (2018) [10] proposed the following formula to determine the sample size required in a non-inferiority study using the odds ratio approach:

$$n_c = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{(\log(OR) - \delta)^2} \left(\frac{1}{kp_T(1-p_T)} + \frac{1}{p_C(1-p_C)} \right)$$

The following quantities are used in the equation:

- n_C = the required sample size in the control group;
- n_T = the required sample size in the treatment group;

- $k = \frac{n_T}{n_C}$ = the ratio between the sample size in the treatment and control groups;
- $z_{1-\alpha}$= the value from the standard normal distribution corresponding to the desired significance level (α). Use $z_{1-\alpha} = 1.64$ for one-sided hypothesis at a 5% significance level;
- $z_{1-\beta}$ = the value from the standard normal distribution corresponding to the desired study power (1-β). Use $z_{1-\beta}$ = 0.84 for 80% power, and $z_{1-\beta}$ =1.28 for 90% power;
- p_T = the population response rates in the treatment group;
- p_C = the population response rates in the control group;
- OR = odds ratio between the treatment and control; calculated as OR = pt(1-pc)/pc(1-pt).
- δ = the non-inferiority margin in log scale.

Implementation of the scenario in R

The formula based on the odds ratio approach is implemented in the function '`RelativeRisk.NIS`' of the R package '`TrialSize`' in R software. The function requires the specification of the following arguments:

- **alpha** corresponds to the desired one-sided significance level. In the scenario, this argument is assigned the value of 0.05.
- **beta** corresponds to the type II error (β) (power = 1 – β). In the scenario, this argument is assigned the value of 0.1.
- **or** corresponds to the odds ratio between the treatment and control groups.
- **k** corresponds to the ratio of sample size in the treatment and control groups. In the scenario, it is assigned the value of 1 such that the same number of children receive treatment (with the new vaccine) and control (with the standard vaccine).
- **pt** corresponds to the population response rate in the treatment group. In the scenario, **pt** is assigned the value of 0.05.
- **pc** corresponds to the population response rate in the control group. In the scenario, **pc** is assigned the value of 0.035.
- **margin** corresponds to the non-inferiority margin in log scale. In the scenario, the non-inferiority margin for the OR was 0.90; this gives a non-inferiority margin for log(OR)=log(0.90)=-0.10. Therefore, this argument is assigned the value of -0.10.

R code for sample size calculation

```
# Setting:
# N = infinite population size
# Proportion (pt) = 0.05
# Proportion (pc) = 0.035

# Significance level (alpha) = 0.05 (one-sided)
# Power = 0.90 or beta = 0.10

# Calculate the odds ratio
pt<-0.05
pc<-0.035
OR <- (pt*(1-pc))/(pc*(1-pt))

RelativeRisk.NIS(alpha = 0.05,      # significance level
                 beta = 0.10,       # type II error
                 or = OR,           # odds ratio
                 k = 1,             # ratio of sample size for vaccinated and unvaccinated groups
                 pt = pt,           # probability of outcome in the vaccinated group
                 pc = pc,           # probability of outcome in the unvaccinated group
                 margin = -0.10)    # non-inferiority margin in log-scale
```

R output

```
1944.579
```

Conclusion

The sample size returned is the sample size per group. Hence, a total of 3 890 vaccinated children, consisting of 1 945 children vaccinated with the new vaccine, and 1 945 vaccinated with the standard vaccine, need to be included in the study to have 90% power to demonstrate non-inferiority of the new vaccine.

3.3 Sample size calculation for two categorical outcomes

Scenario 3 (Section 2.4.3) is used to illustrate the sample size calculation for testing independence between two categorical variables. The function `'pwr.chisq.test'` from the R package `'pwr'` will be used to perform the calculations.

In the COVID-19 pandemic, infected individuals showed a wide range of symptoms which can be classified into three categories: 'mild', 'moderate', or 'severe'. ECDC wants to investigate whether there is a difference in disease severity between men and women.

The data can be organised in a two-way contingency table. A chi-squared test is often employed to investigate independence between the two categorical variables in the contingency table (null hypothesis). It is expected that more men have 'mild' symptoms, whereas women have more 'moderate' and 'severe' symptoms. The cell proportions given in Table 2 below reflect this hypothesis. These proportions are assumed for the alternative hypothesis that there is a dependence between the two categorical variables. The cell proportions are obtained with respect to the total study size, e.g. the proportion of 0.3 in the upper left cell in Table 2 represents the proportion of men with mild symptoms among all the individuals in the study. Both genders comprise 50% of the individuals for the survey.

Using the chi-squared test, the investigator wants to be able to detect the association between gender and disease severity with a power of 80% and a 5% significance level. What is the sample size needed to achieve this?

Table 2. Expected proportions under the alternative hypothesis of non-independence between gender and severity

		Severity			
		Mild	Moderate	Severe	total
Gender	Men	0.3	0.1	0.1	0.5
	Women	0.1	0.2	0.2	0.5
	Total	**0.4**	**0.3**	**0.3**	**1**

Theoretical background

A contingency table is a tabular representation of two categorical variables. Here, we consider a contingency table of two categorical variables with rows *i=1,…,r* and columns *j=1,…,c*.

Each cell in the table represents the observed frequencies for a particular combination of the categories of the two variables. Let o_{ij} be the observed frequencies for cell *ij*.

Table 3. Illustration of an *r x c* contingency table for two categorical variables

		Variable 2				
		1	2	⋯	c	total
Variable 1	1	o_{11}	o_{12}	⋯	o_{1c}	$n_{1\cdot}$
	2	o_{21}	o_{22}	⋯	o_{2c}	$n_{2\cdot}$
	⋮	⋮	⋮	⋯	⋮	⋮
	r	o_{r1}	o_{r2}	⋯	o_{rc}	$n_{r\cdot}$
	Total	$n_{\cdot 1}$	$n_{\cdot 2}$	⋯	$n_{\cdot c}$	n

The corresponding cell proportions, as provided in Table 2, can be calculated as $p_{ij} = o_{ij}/n$ with row and sum proportions written as $p_{i.}$ and $p_{.j}$, respectively. Under the null hypothesis, when the two categorical variables are independent, the expected frequency e_{ij} is equal to:

$$e_{ij} = \frac{n_{i.} \times n_{.j}}{n}$$

where $n_{i.}$ and $n_{.j}$ are the totals for the *ith* row and *jth* column, respectively.

There is no closed form sample size formula for this setting. The calculation involves the estimation of a non-centrality parameter, for which there is no analytical formula, but it can be estimated using software. To obtain the sample size, the software requires the following:

- Cell proportions P_{0ij}under the null hypothesis, which can be calculated as:
  $$P_{0ij} = e_{ij}/n = p_{i.} \times p_{.j}$$
- Cell proportions P_{1ij}under the alternative hypothesis.
- The effect size *w,* defined as:
  $$w = \sqrt{\sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(P_{1ij} - P_{0ij})^2}{P_{0ij}}}$$
- the degrees of freedom (r-1)(c-1)

Implementation of the scenario in R

For the scenario, the cell proportions (P_{1ij}'s) in Table 2 are used for the alternative hypothesis. The expected proportions under the null hypothesis of independence of disease severity and gender (P_{0ij}'s) are calculated via the above expression for the expected frequencies. These are provided in Table 4. For example, the expected proportion for men with mild symptoms is equal to $0.5 \times 0.4 = 0.2$.

Table 4. Expected proportions under the null hypothesis independence

		Severity			
		Mild	Moderate	Severe	total
Gender	Men	0.2	0.15	0.15	0.5
	Women	0.2	0.15	0.15	0.5
	Total	**0.4**	**0.3**	**0.3**	**1**

Given these proportions under the null and alternative hypotheses, the effect size *w* equals 0.17. Using the chi-squared test, the investigator wants to detect this effect size of 0.17 with a power of 80% and a 5% significance level. What is the sample size needed to achieve this?

The formula above is implemented in the function ‘`pwr.chisq.test`’ of the package ‘`pwr`’ in R. The function requires the specification of the following arguments:

- **w** corresponds to the effect size (ES). In the scenario, the effect size is 0.17.
- **sig.level** is the specified significance level of $\alpha = 5\%$.
- **power** is the power (1- β) to detect an association with a specified effect size. In the scenario, the investigators specified a power of 0.8.
- **df** corresponds to the degrees of freedom (r-1)(c-1) where r is the number of rows and c the number of columns of the contingency table. In the scenario, the data results in a 2x3 contingency table, hence the degree of freedom is (2-1)*(3-1)=2.

R code for sample size calculation

```
# Setting:
# Effect size = 0.17
# Significance level= 0.05
# Power = 0.80
# Degrees of freedom = 2

# The function to calculate the sample size:
pwr.chisq.test(w=0.17,               # effect size
               sig.level= 0.05,      # significance level
               power = 0.80,         # power of the test
               df = 2                # degrees of freedom
)
```

R output

```
   Chi-squared power calculation

         w = 0.17
         N = 333.3802
        df = 2
  sig.level = 0.05
      power = 0.8

NOTE: N is the number of observations
```

Conclusion

With an 80% power, we need a sample size of 334 – 167 men and 167 women – to detect a difference in disease severity between the genders. Note that in the R output of this function, N is the total sample size and not the population size.

3.4 Continuous outcome

3.4.1 Sample size calculation to estimate a population mean with a desired precision

Scenario (Section 2.4.4) is used to illustrate the sample size determination for a continuous population parameter with a specified level of confidence and precision using simple random sampling, and demonstrate the use of the function ‘`epi.sssimpleestc`’.

ECDC wants to organise a survey among 450 laboratories to estimate the average cost price of a PCR test across the EU/EEA, to diagnose individuals with a SARS-CoV-2 infection. ECDC anticipates the average cost price to be around EUR 70 and the standard deviation of the cost price to be EUR 15. They wish to construct a two-sided 95% confidence interval for the mean, so that they are 95% confident that the expected cost price is within 10% of the true cost price.

From how many laboratories, sampled at random, should ECDC collect their cost price for a PCR test to be 95% confident that the estimate is within 10% of the true cost?

Theoretical background

The sample size calculations for the estimation of a population mean with desired precision is given by the formula below [3]:

$$n = \frac{z^2_{1-\frac{\alpha}{2}} N V_x^2}{z^2_{1-\frac{\alpha}{2}} V_x^2 + (N-1)\epsilon_r^2}$$

The following quantities are used in the equation:

- *n* = the required sample size;
- $z_{1-\frac{\alpha}{2}}$= the value from the standard normal curve corresponding to the desired confidence level (1-α). Use $z_{1-\frac{\alpha}{2}}$= 1.96 for a 95% confidence level;
- N = the population size;
- V_x^2= the relative variance for the variable calculated as the squared standard deviation (σ_x^2) divided by the squared population mean ($\underline{x}^2$), i.e. $V_x^2 = \sigma_x^2/\underline{x}^2$;
- ε_r= the relative error.

The formula above generates the sample size, *n*, to estimate the average with a desired precision. In Section 2.1.4, a detailed explanation of the terminology used to indicate the precision and the error has been given. Applied to this scenario, a relative error of 10% means that we allow a deviation of 10% from EUR 70, thus from (1-0.1)xEUR 70=EUR 63 to (1+0.1)xEUR 70=EUR 77. The formula above will then return the number of laboratories needed to be 95% confident that the expected cost price from the study will be anywhere between EUR 63 and EUR 77, or within 10% from the population mean.

Implementation of the scenario in R

The formula above is implemented in the function ‘`epi.sssimpleestc`’ of the package ‘`epiR`’ in R software. The function requires the specification of the following arguments:

- **N,** i.e. the population size. The function assumes a default population size of 1 000 000. The default value can be used if the assumption of an infinite population is justifiable. If the study population is smaller, then the accurate number of the study population under investigation can be entered here. For the scenario, this would imply a study population of 450 laboratories in the EU/EEA.
- **xbar** corresponds to the expected mean of the continuous variable. In the scenario, the average cost price of a PCR test is estimated to be EUR 70.
- **sigma** corresponds to the expected standard deviation of the data. The expected standard deviation of the cost price is EUR 15 and is entered for this argument. The arguments **xbar** and **sigma** are used to calculate the relative variance V_x^2 .
- The precision is specified by the arguments **error** and **epsilon**. In the scenario, the precision is expressed as a relative error of 10%. This is indicated by specifying ‘relative’ for the **error** in combination with a value of 0.1 in the argument **epsilon.** Note that in the scenario, a relative error of 10% corresponds to an absolute error of EUR 7(70*0.1 = 7). Thus, the sample size can also be obtained by specifying **error** equal to ‘absolute’ and **epsilon** equal to 7.
- Since the investigator of the study is interested in a 95% confidence level, the argument **conf.level** will take the value 0.95.

R code for sample size calculation

```
# Setting:
# N= 450
# Mean = 70
# Standard deviation = 15
# Relative error = 0.10 (or absolute error = 7)
# Confidence level = 0.95

# The function to calculate the sample size:
epi.sssimpleestc(N=450,              # population size
           xbar=70,                  # the average cost price of a PCR test
           sigma= 15,                # standard deviation of the cost price of a PCR test
           epsilon= 0.10,            # tolerable error
           error = "relative",       # the error we refer to is a relative error
           conf.level= 0.95          # confidence level
)
```

R output

```
[1] 18
```

Conclusion

A total of 18 laboratories needs to be sampled (at random) from the 450 available laboratories to meet the requirements of the survey, i.e. ECDC is 95% confident that the estimated mean cost price of a SARS-CoV-2 PCR test is within 10% of the true population mean.

3.4.2 Sample size calculation for one population mean hypothesis testing

Scenario 4 is used to illustrate the sample size determination for a continuous parameter that is compared to a hypothesised value using simple random sampling. In this section, the use of the function 'pwr.t.test' is demonstrated.

Similar to the previous section, ECDC wants to conduct a survey among laboratories throughout the EU/EEA, to estimate the average cost price of a PCR test to diagnose individuals with a SARS-CoV-2 infection. ECDC wishes to detect a difference of EUR 5 between the study population mean and the hypothesised value of EUR 70. The population standard deviation of the cost price is EUR 15.

In this setting, ECDC is interested in a deviation of EUR 5 in either direction of the population mean (= two-sided test). The investigator wishes to be 90% sure of detecting a difference (= power of the test), assuming that a two-sided test is carried out at a significance level of 5%.

From how many laboratories, sampled at random, should ECDC collect their cost price for a PCR test to achieve the desired power to test this hypothesis?

Theoretical background

Here, the hypotheses of interest are the null hypothesis (H_0: $\mu = \mu_0$) that the population mean (μ) equals the hypothesised value (μ_0), and the alternative hypothesis (H_1: $\mu \neq \mu_0$) that the population mean differs from the hypothesised value. Alternatively, the interest of the investigator could also be 'one-sided' i.e. the mean is smaller than the hypothesised value (H_1: $\mu < \mu_0$), or larger than the hypothesised value (H_1: $\mu > \mu_0$). The formula for determining the sample size for a one-sample mean used in hypothesis testing is given below [12]:

$$n = \left(\frac{\sigma}{d}\right)^2 \left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2$$

The following quantities are used in the equation:

- n = the number of units in the sample;
- $z_{1-\frac{\alpha}{2}}$= the value from the standard normal distribution corresponding to the desired significance level (α). Use $z_{1-\frac{\alpha}{2}}$ = 1.96 for a two-sided hypothesis at a 5% significance level;
- $z_{1-\beta}$ = the value from the standard normal distribution corresponding to the desired study power (1-β). Use $z_{1-\beta}$ = 0.84 for 80% power, and $z_{1-\beta}$ =1.28 for 90% power;
- σ= the assumed population standard deviation;
- d= the difference between the population mean and the hypothesised value (μ - μ_0).

Implementation of the scenario in R

The formula above is implemented in the function `'pwr.t.test'` of the package `'pwr'` in R software. The function requires the specification of the following arguments:

- **d** corresponds to the effect size (ES). The effect size needs to be computed by the user as the standardised difference between the mean and the hypothesised value, i.e. as $\frac{\mu-\mu_0}{\sigma}$ where μ is the population mean, μ_0 is the hypothesised value, and σ is the standard deviation of the outcome. The calculated effect size, in the R code below, is saved in an object called **ES** and is then passed to the argument **d**.
- The investigator required a significance level of 5%. Therefore, 0.05 is passed to the argument **sig.level**.
- **power** corresponds to 0.9, as the investigator wanted to be 90% sure of detecting a difference.
- **type** corresponds to the type of test (one- or two-sample test). Since information on the price will only be gathered from laboratories across the EU/EEA, this is a one-sample test and **type** gets assigned the value `'one.sample'`. If the investigator wants to evaluate whether there is a difference in the average cost price of PCR tests in laboratories of two different countries, then the argument **type** gets assigned the value `'two.sample'` (see the Section 3.4.3).
- **alternative** corresponds to the alternative hypothesis that is of interest. Depending on the hypothesis test that is under investigation, the argument **alternative** can take several values: `'two.sided'`, `'greater'` or `'less'`. A two-sided alternative hypothesis is in place when the investigator is interested in a difference and not in the direction of this difference (H_1: $\mu \neq \mu_0$), which is the case in the scenario. An alternative hypothesis can be 'greater' (H_1: $\mu > \mu_0$) or 'less' (H_1: $\mu < \mu_0$).

R code for sample size calculation

```
# Setting:
# Infinite population size - the function does not use N and assumes an infinite population (see later for finite
population)
# Mean difference = 5 # Difference between population mean (μ) and hypothesised value (μ0)
# Standard deviation = 15
# Power = 0.90
# Significance level = 0.05
# Two-sided test


# Calculate the Effect Size (ES)
# Formula ES = Mean difference/Standard deviation
ES <- 5/15


# The function to calculate the sample size
pwr.t.test(d= ES,                      # effect size
           sig.level = 0.05,           # significance level
           power = 0.90,               # power of the test
           type = "one.sample",        # type of test (one sample)
           alternative = "two.sided"   # two-sided hypothesis
)
```

R output

```
     One-sample t-test power calculation

              n = 96.50801
              d = 0.3333333
      sig.level = 0.05
          power = 0.90
    alternative = two.sided
```

Conclusion

When the population size is infinite, a total of 97 laboratories needs to be sampled (at random) to meet the requirements of the survey, i.e. ECDC is 90% confident of detecting a difference, assuming that a two-sided test is carried out at a significance level of 5%. Note that if a fractional value is proved in the R output, then it is important to round that number upward. In the R code, the sample size provided is 96.50801, which is rounded upward to the required sample size of 97 laboratories.

R code for plot of power as function of sample size

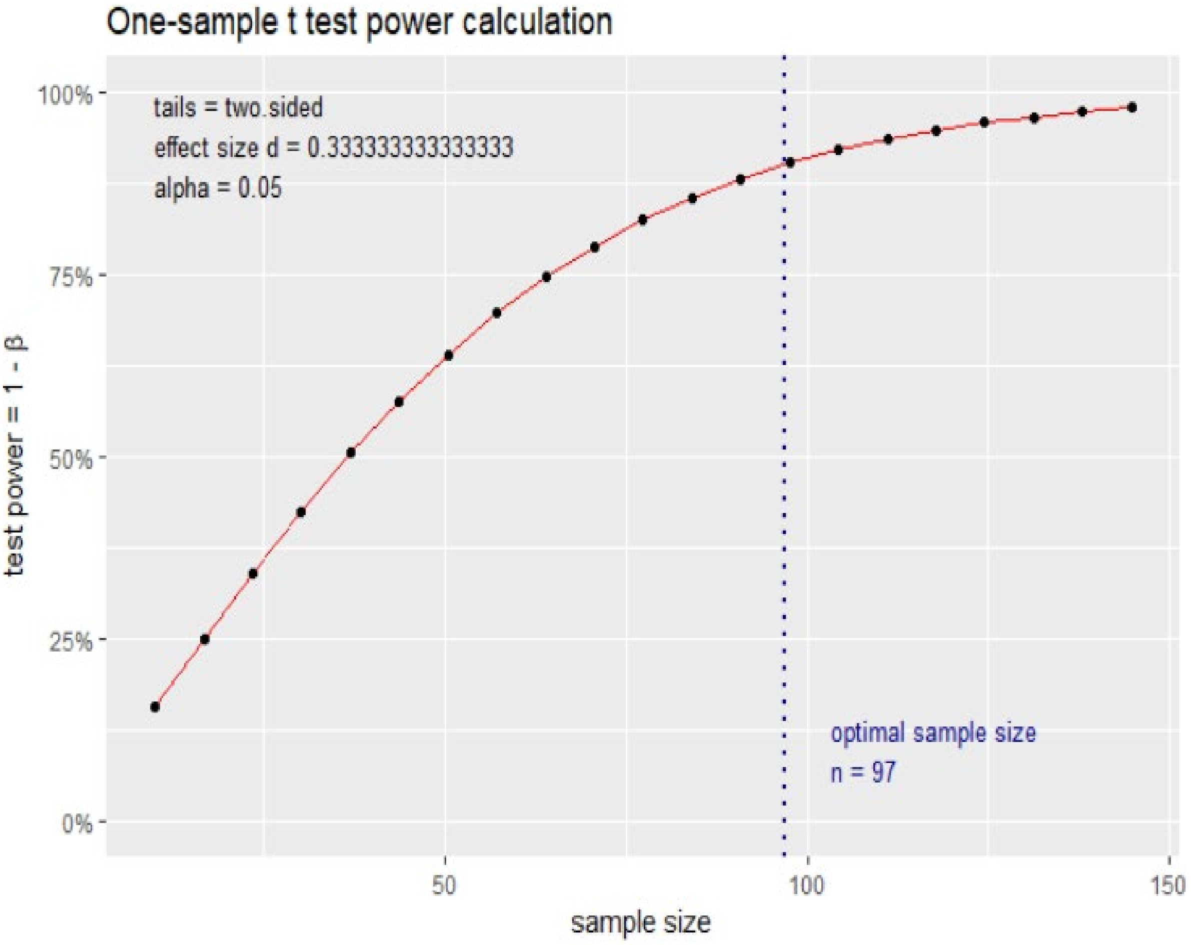
We can visualise the changes in power using the `‘plot’` function.

```
# Save the analysis into an object:
one.mean <- pwr.t.test(d= ES, sig.level = 0.05, power = 0.90,
                      type = "one.sample", alternative = "two.sided")

plot(one.mean)
```

R output

Figure 6. Power calculations as a function of sample size for one population mean



Conclusion

The power increases with increasing sample size. From the plot, it can be observed that there is a steep increase in the power for increases in the sample size, if the sample size is smaller than 97. The gain in power by increasing the sample size beyond the optimal sample size of 97 is rather limited.

R code for finite population size

If the population size is not infinite but rather finite, then an extra step needs to be performed after calling the function `‘pwr.t.test’` in R. Let us assume that in the scenario used previously, the only quantity that changes is the population size *N*. The (finite) population size now consists of 450 laboratories. In order to correct for this finite sample size, use the function `‘correct.N’` (see Annex for the source file). The function has two arguments:

- *n*, i.e. the sample size obtained from the appropriate R function. For this scenario, *n* would get the value 96.50801.
- *N*, i.e. the population size. For this scenario, *N* would get the value 450.

For details on the formula used to correct for a finite sample size, see Section 3.1.2.

```
# If the population size is no longer infinite but rather finite,
# A third step needs to be undertaken in order to obtain the sample size

# Finite population size correction
correct.N(n=96.50801, # Sample size obtained from the function, in this case pwr.t.test
              N=450      # Population size
              )
```

R output

```
[1] 79.46563
```

Conclusion

If the population is infinite, a total of 97 laboratories are needed to meet the requirements of the survey. If the population only consists of 450 laboratories which can be sampled, then it is sufficient to sample 80 laboratories at random.

3.4.3 Sample size calculation for two population mean hypothesis testing

Scenario 4 is used to illustrate the sample size determination for testing differences in the means of two independent samples, using simple random sampling. It relies on the use of the function '`pwr.t.test`' (see Section 3.4.2 for a detailed explanation of the function).

This scenario is similar to the one discussed in Section 3.4.2. The distinction now is that ECDC wants to conduct a survey to estimate the difference between the average cost price of a PCR test in the country Y (μ_1) and the average cost price of a PCR test in country Z (μ_2) to diagnose an individual with a SARS-CoV-2 infection, assuming an infinite population. Again, ECDC wants to detect a mean difference of EUR 5 between the two populations. The population standard deviation of the cost price in both the countries is assumed to be the same (EUR 15). ECDC believes that the difference in the average cost between the two populations is EUR 5 (= two-sided test). ECDC wants to be 90% sure of detecting a difference (= power of the test), assuming that a two-sided test is carried out at a significance level of 5%.

From how many laboratories in the country Y and the country Z should ECDC collect the cost price for a PCR test to achieve the desired power to test this hypothesis?

Theoretical background

In Sections 3.4.1 and 3.4.2, the procedure for estimating the required sample size for a given precision and hypothesis test was discussed. The same concepts and procedures will be used in estimating the sample size in order to estimate differences between the two population means. Consider the null hypothesis in which the two population means are equal (H_0: $\mu_1 = \mu_2$). The alternative hypothesis can either be one-sided (H_1: $\mu_1 - \mu_2 > 0$ or H_1: $\mu_1 - \mu_2 < 0$) or two-sided (H_1: $\mu_1 - \mu_2 \neq 0$). The formula to determine the required sample size for estimating a difference between two (independent) populations means is as follows [12]:

$$n = 2\left(\frac{\sigma}{d}\right)^2 \left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2$$

The following quantities are used in the equation:

- n = the number of samples for each population;
- $z_{1-\frac{\alpha}{2}}$= the value from the standard normal distribution corresponding to the desired significance level (α). Use $z_{1-\frac{\alpha}{2}}$ = 1.96 for a two-sided hypothesis at a 5% significance level;
- $z_{1-\beta}$ = the value from the standard normal distribution corresponding to the desired study power (1-β). Use $z_{1-\beta}$ = 0.84 for 80% power, and $z_{1-\beta}$ =1.28 for 90% power;
- σ= the population standard deviation. It is assumed that the population standard deviations (σ) are the same in both populations and that σ is known. An estimate for σ can be obtained from literature, previous surveys or a pilot study;
- d= the minimum difference that is considered to be significant between the two population means (μ_1 - μ_2).

Implementation of the scenario in R

The formula above is implemented in the function 'pwr.t.test' of the package 'pwr' in R software. Details on the arguments of the function are provided in Section 3.4.2. The scenario considered in this section is a 'two-sample' problem. Only one argument in the function 'pwr.t.test' will change its value as compared to the previous section:

- ***type*** corresponds to the type of test that is being performed (one- or two-sample test). Since the price in two countries are compared, this is a two-sample problem and ***type*** gets assigned the value 'two.sample'. This is in contrast with the value 'one.sample' that was assigned to ***type*** in the scenario in Section 3.4.2.
- Moreover, the effect size **d** that is used here is the standardised effect size, which is the difference between the two means divided by the standard deviation.

R code for sample size calculation

```
# Setting:
# Infinite population size - the function does not use N and assumes an infinite population (see later for finite
population)
# Mean difference = 5 # Difference between population mean of the country Y (μ1) and the country Z (μ2)
# Standard deviation = 15
# Power = 0.9
# Significance level = 0.05
# Two-sided test


# Calculate the Effect Size (ES)
# Formula ES = (μ1-μ2)/Standard deviation
ES <- 5/15


# The function to calculate the sample size
pwr.t.test(d= ES,                    # effect size
        sig.level = 0.05,            # significance level
        power = 0.90,                # power of the test
        type = "two.sample",         # type of test (two sample)
        alternative = "two.sided"    # two-sided hypothesis
)
```

R output

```
     Two-sample t-test power calculation

              n = 190.0991
              d = 0.3333333
      sig.level = 0.05
          power = 0.90
    alternative = two.sided

NOTE: n is number in each group
```

Conclusion

For an infinite population size, a total of 382 laboratories, i.e. 191 laboratories in the country Y and 191 laboratories in the country Z, need to be sampled at random, so that ECDC is 90% confident of detecting a difference, given the study design. It is important to note that the function 'pwr.t.test' will return the sample size per group when the type of test is a two-sample one.

R code to make plot of power as function of sample size

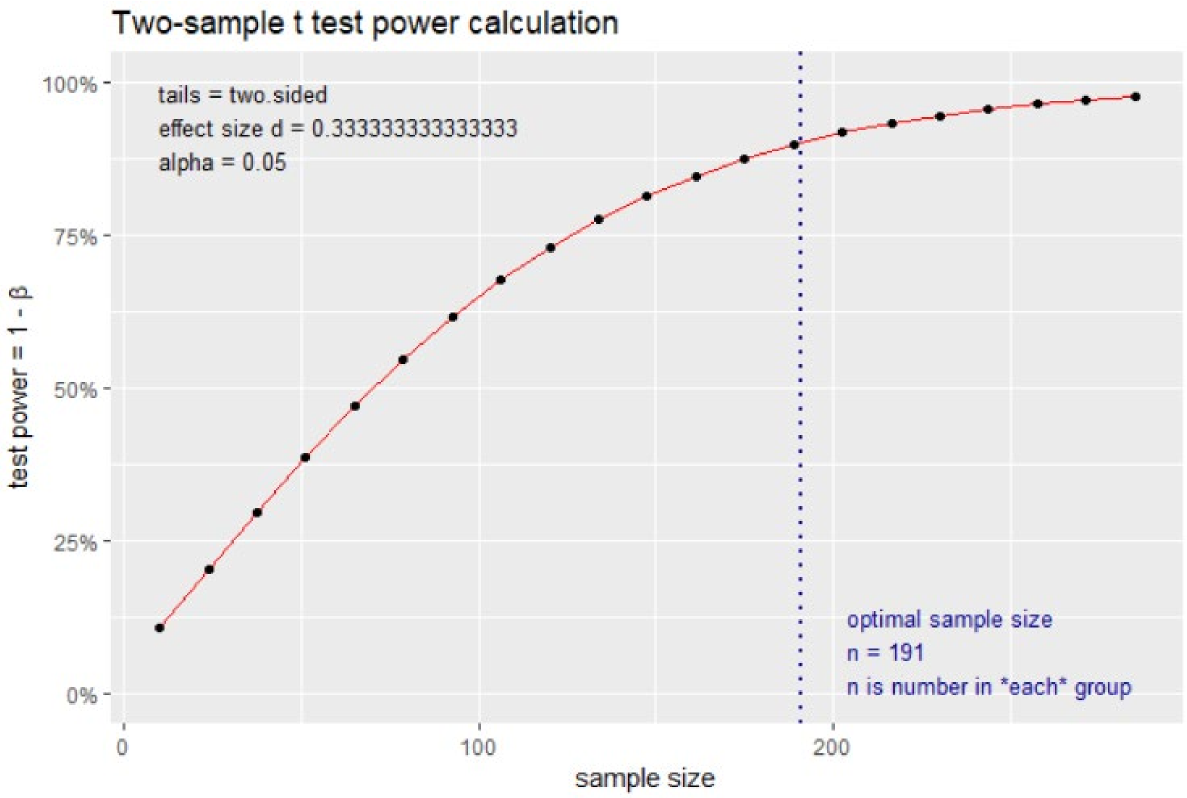
We can visualise the changes in power using the 'plot' function.

```
# Save the analysis into an object:
two.means <- pwr.t.test(d= ES, sig.level = 0.05, power = 0.90,
                  type = "two.sample", alternative = "two.sided" )

plot(two.means)
```

R output

Figure 7. Power calculations as a function of sample size for two population mean



Conclusion

The power increases with increasing sample size. From the plot, it can be observed that there is a steep increase in the power for increases in the sample size, if the sample size per group is smaller than 191. The gain in power by increasing the sample size beyond the optimal sample size of 191 per group is rather limited.

R code for finite sample size correction

If the population is finite, then an extra step needs to be performed after calling the function ‘`pwr.t.test`’ in R. Suppose that the (finite) population size consists of 100 laboratories. In order to correct for this finite sample size, use the function contained in the R script, called ‘`correct.N`’. The function has two arguments:

- *n*, i.e. the sample size obtained from the appropriate R function. For this scenario, *n* gets the value 382.
- *N*, i.e. the population size. For this scenario, *N* gets the value 100.

For details on the formula used to correct for a finite sample size, see Section 3.1.2.

```
# If the population size is no longer infinite but rather finite,
# A third step needs to be undertaken in order to obtain the sample size

# Finite population size
correct.N(n=382, # Sample size obtained from the function, in this case pwr.t.test
        N=100 # Population size
)
```

R output

```
[1] 79.25311
```

Conclusion

This value is the required total sample size, i.e. for the two groups together. For a finite population size of 450 laboratories, it is sufficient to sample 208 laboratories in total, i.e. 104 laboratories in each country, at random to meet the survey requirements. It is important to round the corrected sample size to the next highest even number, yielding a total sample size of 80, implying 40 units/laboratories to be sampled in the country Y and 40 in the country Z.

4 Sample size calculations for cluster sampling

4.1 Key concepts for cluster sampling

This section focuses on sample size calculations for cluster sampling. Firstly, the definition of cluster sampling will be introduced. Secondly, the sample size calculation for a one-stage cluster sampling will be illustrated with the example of the cost price of a PCR test (Scenario 4) and influenza vaccines for MAARI (Scenario 2).

Cluster sampling is a sampling method that can be used when, for any reason, the simple random sampling method is not applicable. To sample individuals from a country's population, one can select regions (clusters) and then select people from these regions. So, for cluster sampling, the population is divided into multiple groups (clusters), and clusters are selected at random for the study.

We will address one-stage and two-stage cluster sampling. In one (or single)-stage cluster sampling, sampling is done just once. In two-stage cluster sampling, a second stage is added; i.e. a number of units of the sampled clusters are selected for the study. For example, a survey among schoolchildren will most likely not select schoolchildren via a simple random sampling method. This would be inefficient as many schools will be involved. Cluster sampling, where schools are the clusters, will be easier to implement and cost less time and money. A **one-stage** cluster sampling will randomly select a number of schools from a list of all schools, and then include all the students from these selected schools. A **two-stage** cluster sampling adds one more stage of sampling. Within every school selected for the study, a number of children are selected. When using cluster sampling, we need to consider that, compared to simple random sampling, children in the same school/cluster are likely to be somewhat similar to one another on account of several different characteristics (such as, socio-economic backgrounds). This implies that adding another child from the same school does not add new information. However, adding another child using simple random sampling (so a child from a different school) might add new information.

Since units in the same cluster are more alike, the variability within clusters is smaller. Information obtained by a cluster sample is typically less than in a random sample of the same size. As a result, there is a loss of efficiency.

A way of summarising the amount of information in the data is via the effective sample size, defined as $neff = n/deff$, with n being the total sample size and $deff$ the design effect. The **design effect** quantifies the loss of efficiency. For example, a design effect of 3 means that the sample size needed using a cluster sampling is three times higher than the sample size that would be needed with simple random sampling. An alternative interpretation is that only ⅓ of the sampled units would be needed to measure a given statistic (proportion or mean) with the same precision, if a simple random sample was used instead of the cluster sample with a design effect of 3. The design effect is a function of the intracluster correlation (ρ) and the cluster sizes (b is the size of the clusters for a one-stage cluster design):

$$deff = 1 + (b - 1)\rho.$$

The design effect increases with increasing cluster sizes and increasing intraclass correlation. The intracluster correlation coefficient is the ratio of the variance between clusters and the total variance (i.e. the sum of the variance within and between clusters). In other words, it is the proportion of the variance that can be attributed to the variability between clusters:

$$\rho = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}$$

where

$\sigma_B{}^2$= variance between clusters;
$\sigma_W{}^2$= variance within clusters.

Let us consider two special cases. On the one hand, if all the values within a cluster are identical, the variance within clusters ($\sigma_W{}^2$) will be 0, and therefore the intracluster correlation ρ will be 1. This implies that the design effect will be as high as the number of sampling units per cluster ($deff = b$). Therefore, the effective sample size will be equal to the number of clusters selected ($neff = n/b$). On the other hand, if the variance between clusters is much smaller than the variance within clusters, i.e. the intracluster correlation is close to 0, then the design effect will be close to 1. If the design effect from clustering is very large, then increasing the number of sampled clusters with a smaller sample size should be considered.

The design effect must be considered in the sample size calculation. There are two 'problems' with this. Firstly, design effects vary from survey to survey and can even vary within a survey. For example, if the sampling unit is a school, then the design effect will be higher for variables that are common for all the students at the school, and lower for variables that are different for different students. Secondly, the design effect is often not known prior to the study, unless previous studies have been conducted on the same outcome. Literature and small pilot studies can be used to come up with a reasonable estimate of the design effect to be used in the sample size calculation.

To illustrate how to use the design effect in sample size calculations, consider Scenario 4. The goal of the study is to estimate the average cost price of a PCR test across the EU/EEA. The average cost price is anticipated to be around EUR 70 with a standard deviation of EUR 15. The sample size calculation for this example, when using simple random sampling (n_{srs}), is discussed in Section 3.4.1. A total of 18 laboratories needs to be sampled (at random) to be 95% confident that the estimated mean cost price of a SARS-CoV-2 PCR test is within 10% of the true population mean.

Now suppose that for practical reasons, a one-stage cluster sampling method has to be used. To limit the travel cost to the laboratories, they are grouped in fives based on their geographical location (region). First, some regions (clusters) are selected, and for each of the selected regions the five laboratories are included. Based on previous studies, a realistic estimate for the intracluster correlation coefficient (ρ) is 0.2. The design effect for the one-stage cluster sampling equals $deff = 1 + (b - 1)\rho$ = 1 + (5-1)0.2 = 1.8. The sample size needed when using the one-stage cluster sampling method therefore equals:

$$n_{clust\ design} = deff \times n_{srs} = 1.8 \times 18 = 33.$$

So, a total of 33 laboratories needs to be sampled, with five laboratories per cluster. This implies seven regions (33/5 = 6.6) have to be sampled for the study. The larger sample size is required to compensate for the loss of precision as a result of the positive correlation among the laboratories in a cluster (ρ=0.2). Some more detailed calculations of the design effect for one-stage and two-stage cluster designs, and clusters of equal and unequal size, are provided below:

- **One-stage cluster design**
  - If all clusters have the same size (*b*), the design effect is obtained as $deff = 1 + (b - 1)\rho$.
  - If the clusters are of unequal sizes, the design effect is obtained as $deff = 1 + ((CV^2 + 1)\bar{b} - 1)\rho$, with $\bar{b}$ as the average cluster size and CV as the coefficient of variation for the number of units per cluster.
- **Two-stage cluster design**

  The population has *A* clusters, and every cluster has *B* units. In the first stage, the clusters are selected, and in the second stage *B* units are sampled from every selected cluster. The sampling fractions for the first and second stage are thus $f_1 = \frac{a}{A}$ and $f_2 = \frac{b}{B}$. The design effect is given by $deff = (1 - f_1)\, f_2[\,1 + \,(B - 1)\;\rho] \,+\, (1 - f_2)\,\rho$ [13]. In practice, the sampling fraction for the first stage is often small ($f_1 \approx 0$) and the expression then simplifies to $deff = 1 \,+\, (b - 1)\,\rho$.

For the **two-sample setting**, where the aim is to test the difference between the means or proportions of the two subgroups, two scenarios can be in place:

- The two samples/subgroups are in different sets of clusters (independent samples). This will be the case when the aim is to compare two regions or two countries.
- The two samples/subgroups 'cut across' the sample and tend to be found in all or many of the clusters. This will be the case when for example, the subgroups are men and women. When households are sampled, the household/cluster will have men and women.

Under the assumption that the variances of the outcome are the same in the two subgroups, the following inequality holds for the design effect for the difference between two means [14].

$$1 < Deff(\bar{y}_1 - \bar{y}_2) \leq \frac{n_2 Deff(\bar{y}_1) + n_1 Deff(\bar{y}_2)}{n_1 + n_2}$$

where, $n_1$and $n_2$are the sample sizes of subgroup 1 and subgroup 2, $Deff(\underline{y}_1 - \underline{y}_2)$is the design effect for the difference between the two means, $Deff\left(\underline{y}_1\right)$, and $Deff(\underline{y}_2)$ are the design effects of the means of the two subgroups. So, the design effect for the difference between two subgroup means or proportions is greater than 1, but less than that obtained when the two subgroups are treated as independent.

Design effects from clustering are typically smaller for differences in means/proportions than for overall means/proportions. When the subgroups are in different sets of the clusters (e.g. when comparing regions), the upper bound applies. If the design effect and sample sizes are similar for the two means/proportions, the design effect for the difference will also be similar to the design effect of the means/proportions.

For other settings and more complex sampling designs, the procedure of multiplying the sample size obtained using simple random sampling by the design effect can still be used. But it is important to reflect the complexity of the design and use the design effect for the statistic of interest.

4.2 Example of cluster sampling

In a one-sample setting with either equal or unequal cluster sizes, the following sections present the formulas and illustrate how to perform the sample size calculations for a binary outcome in R, without first calculating the sample size under simple random sampling. Scenario 2 is used to illustrate the sample size determination and demonstrate the use of the R function `‘epi.ssclus1estb’`.

To estimate the prevalence of influenza in the EU/EEA, a cross-sectional study is planned by ECDC where the computerised databases of general practitioners (GP) across the EU will be sampled and all the patients from these GP databases will be surveyed to determine whether they contracted the virus in the previous year. A pilot study of the prevalence of influenza in five offices showed that 46% of the patients contracted the disease. The intracluster correlation, ρ, was also estimated based on this pilot study and estimated to be equal to 0.2. If ECDC wants to be 90% confident that the survey estimate of influenza prevalence is within 10% of the true population value, how many GP databases would need to be sampled?

Two possible scenarios in terms of the cluster sizes (GP databases) will be considered: a) the GP databases have approximately the same size (75 patients), and b) the number of patients per GP database varies (with an average of 75 patients per database and a standard deviation of 35).

Theoretical background

For the sample size calculations, in case of one-stage cluster sampling, the formula below gives the number of clusters to be sampled (n_c), to estimate a population proportion with desired precision [5]:

$$n_c \geq \frac{z^2_{1-\left(\frac{\alpha}{2}\right)} P_y(1-P_y) \times deff}{(P_y \epsilon_r)^2 \bar{b}}$$

The following parameters are used in the equation:

- *deff* = the design effect. Let ρ be the intracluster correlation coefficient. If the number of units per cluster is approximately the same and equal to *b, the design effect is* $deff = 1 + (b-1)\rho$. If the number of units per cluster varies, the design effect is $deff = 1 + ((CV^2+1)\bar{b} - 1)\rho$ with CV = the coefficient of variation for the number of units per cluster; and $\bar{b}$ = the average number of units per cluster;
- $z_{1-\left(\frac{\alpha}{2}\right)}$ = value from the standard normal curve corresponding to a significance level α. Use $z_{1-\left(\frac{\alpha}{2}\right)}$= 1.96 for 95% confidence level;
- P_y = the estimated population prevalence;
- ε_r = the relative error;
- $\bar{b}$ = the average number of units per cluster.

The formula above generates the number of clusters to be sampled, n_c , to estimate the proportion with a desired precision.

Implementation of the scenario in R

The formula above is implemented in the function `‘epi.ssclus1estb’` of the package `‘epiR’` in R software. The function requires the specification of the following arguments:

- **b** is the average number of units in each cluster and its standard deviation (if the cluster sizes are not approximately equal).
- **Py** is an estimate of the unknown population proportion.
- The precision is specified by the arguments **error** and **epsilon**. In the scenario, the precision is expressed as a relative error of 10%. This is indicated by specifying ‘relative’ for the error in combination with a value of 0.1 in the argument **epsilon**.
- **rho** is the intracluster correlation.
- **nfractional** indicates if the function should return an integer.
- Since the investigator of the study is interested in a 90% confidence level, the argument **conf.level** will take the value 0.90.

4.2.1 Clusters with approximately the same size

When the size of all clusters is approximately 75, the design effect equals 1 +(75-1)*0.2= 15.8. This means that the sample size for the one-stage cluster design needs to be 15.8 times larger than in a simple random sample, keeping all the other conditions the same. The formula returns the number of GP databases needed to be 90% confident that the prevalence rate of influenza in the study will be within 10% from the true population proportion (expected to be 46%).

R code for sample size calculation

```
# Setting:
# sampling units per cluster = 75
# Estimated prevalence rate = 0.46
# Relative error = 0.1
# Intracluster correlation coefficient = 0.2
# Confidence level = 0.90

 epi.ssclus1estb(b = 75,            # (approximate) number of patients per GP database
           Py= 0.46,                # estimated prevalence rate
           epsilon= 0.1,            # tolerable error
           error = "relative",      # the error we refer to is a relative error
           rho = 0.2,               # intracluster correlation
           nfractional = FALSE,     # it returns the rounded (upwards) sample size
            conf.level= 0.90)       # confidence level
```

R output

In the output below:

- n.psu is n_c , the number of clusters to be sampled;
- n.ssu is the total number of units in the study ($n_c * b$).

```
$n.psu
[1] 66.90927
$n.ssu
[1] 5018.195
$DEF
[1] 15.8
$rho
[1] 0.2
```

Conclusion

To be 90% confident that the prevalence rate of influenza in the study will be within 10% from the true population proportion (expected to be 46%), a total of 67 GP databases (with a total of approximately 67*75=5 025 patients) need to be sampled.

4.2.2 Clusters with unequal size

Assuming that the number of patients per GP database is on an average 75, but with a standard deviation of 35, the number of GP databases (clusters) to be sampled is obtained with the following code.

R code for sample size calculation

```
epi.ssclus1estb(b = c(75,35),       # average number of patients per GP database and standard deviation
           Py = 0.46,               # estimated prevalence rate
           epsilon = 0.1,           # tolerable error
           error = "relative",      # the error we refer to is a relative error
           rho = 0.2,               # intracluster correlation
           conf.level = 0.90)       # confidence level
```

R output

In the output below:

- n.psu is n_c , the number of clusters to be sampled;
- n.ssu is the total number of patients in the study ($\bar{b}\ n_c$).

```
$n.psu
[1] 80.74283
$n.ssu
[1] 6055.712
$DEF
[1] 19.06667
$rho
[1] 0.2
```

Conclusion

If the number of patients per GP database is variable (average of 75, with a standard deviation of 35), then the number of GP databases to be sampled should be 81. It is expected that a total of 6 056 patients will be included in the survey. So due to unequal cluster sizes, the number of GP databases to survey will increase from 67 to 81; the number of patients will increase from 5 025 to 6 056.

5 Bias and missingness

Even if the sample size is computed accurately, there are other factors such as bias and missingness that affect the validity of the conclusions of the study. These key concepts will be discussed in this section.

5.1 Source/types of bias

Bias, also called 'systematic error', refers to the systematic distortion that affects the internal validity of the study results [15]. It can also be defined as a deviation from the population value. In this section, three types of bias are considered: a) selection bias, b) response bias, and c) estimation bias.

Selection bias occurs as a result of any error in sample selection. It could happen when some parts of the population do not have the chance of being selected as a sample. For example, when using non-probability sampling methods such as convenience sampling. Using this method, the units included in the sample are those who are easy to select or reach and hence do not represent the entire population. Also, when using probability sampling methods, if some of the units are not included in the sampling frame, it would possibly introduce bias.

A second type of bias is **response bias**, which occurs when not all the chosen units/respondents respond to the survey: non-respondents may have systematic differences from respondents, and this should be taken into account in the analysis. To account for these characteristics, models that deal with missing data can be used. Failing to include them could affect the validity of the results.

Another type of bias is **estimation bias** which can be defined as the difference between the true population value of the parameter being estimated and the value of the estimator observed in the sample. If, on average, the value of the estimator is equal to the population value, then it is called an **unbiased estimator**. Unbiasedness is one of the criteria of a good estimator.

5.2 Missing data and weighting

Representativeness is one of the most important aspects of sampling when conclusions based on the samples collected need to be generalisable to the entire population. Missing data can reduce the representativeness of the sample and potentially introduce bias into the results. Some forms of missing data could be caused by non-response (refusal to participate) of the unit (such as individual or laboratory). It could also be caused by not receiving a reply to some items in a questionnaire (for instance a subject refuses to complete information on questions related to income).

It is recommended to design a survey that maximises the survey response rate. Incentives and appropriate modes of contact (personal contact, or face-to-face interviews) are two of the well-known strategies to increase response rates [16]. Increasing the sample size will not necessarily lead to the elimination of bias because the sample will continue being non-representative of the target population. If missingness/non-response cannot be avoided entirely, post-survey adjustment methods may be used, such as weighting, missing data imputation, and other models for missing data at the analysis stage.

An important concept related to sampling that has not been discussed in the document is weighting. This is particularly useful when using a multi-stage sampling design (not covered in this document) as weights can be used to adjust estimates with the aim of decreasing the bias. Furthermore, the design effect of a multi-stage sampling design can also be obtained as a function of weights.

References

1. Woodward, M. Epidemiology: Study Design and Data Analysis (Third edition). New York: Chapman and Hall/CRC; 27 Dec 2013.
2. Fink, A. How to Sample in Surveys (Second edition). Newbury Park: Sage Publications, Inc; 1 Jan 2011 (online publication).
3. Levy, PS, Lemeshow, S. Sampling of Populations: Methods and Applications (Fourth edition). Hoboken: John Wiley & Sons, Inc; Aug 2008.
4. Neyman, J, Pearson, ES. On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. Biometrika, 20A: pp.175–240. Oxford: Oxford University Press (on behalf of Biometrika Trust); Jul 1928. Available at: https://academic.oup.com/biomet/article-abstract/20A/1-2/175/204353?redirectedFrom=fulltext
5. Stevenson, MA. Sample Size Estimation in Veterinary Epidemiologic Research. Front Vet Sci. 2020;7:539573. Available at: https://www.frontiersin.org/articles/10.3389/fvets.2020.539573/full
6. Antimicrobial resistance in the EU/EEA (EARS-Net) – Annual Epidemiological Report for 2019. Stockholm: ECDC; 2020. Available at: https://www.ecdc.europa.eu/en/publications-data/surveillance-antimicrobial-resistance-europe-2019#:~:text=EARS%2DNet%20data%20for%202019,aureus%20(20.6%25)%2C%20K.
7. Protocol for cohort database studies to measure influenza vaccine effectiveness in the European Union and European Economic Area Member States. Stockholm: ECDC: 2009. Available at: https://www.ecdc.europa.eu/en/publications-data/protocols-case-control-studies-measure-influenza-vaccine-effectiveness-eu-and-eea
8. Ryan, TP. Sample Size Determination and Power. Hoboken: John Wiley & Sons, Inc; 27 Jun 2013.
9. Julious, SA. Sample Sizes for Clinical Trials. New York: Chapman and Hall/CRC; 20 Aug 2009.
10. Chow, S-C, Shao, J, Wang, H, Lokhnygina, Y. Sample Size Calculations in Clinical Research (Third edition). New York: Chapman and Hall/CRC; 24 Aug 2017.
11. Lachenbruch, PA. On the sample size for studies based upon McNemar's test. Stat Med. 1992 Aug;11(11):1521–1525. Available at: https://onlinelibrary.wiley.com/doi/10.1002/sim.4780111110
12. Verma, JP, Verma, P. Determining Sample Size and Power in Research Studies. Singapore: Springer Nature Singapore; 2020.
13. Aliaga, A, Ren, R. Optimal Sample Sizes for Two-stage Cluster Sampling in Demographic and Health Surveys. Demographic and Health Surveys (DHS Program) Working Papers No. 30. Calverton,: ORC Macro; Jul 2006. Available at: https://www.dhsprogram.com/publications/publication-wp30-working-papers.cfm
14. Kalton, G, Brick, JM. Estimating components of design effects for use in sample design. Available at: https://www.semanticscholar.org/paper/Estimating-components-of-design-effects-for-use-in-Kalton-Brick/6541311231bcc7b257c75163ffb55cb8be39e81d
15. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Selection bias and information bias in clinical research. Nephron Clin Pract. 2010;115(2):c94–c99. Available at: https://karger.com/nec/article/115/2/c94/830450/Selection-Bias-and-Information-Bias-in-Clinical
16. Toepoel V, Schonlau M. Dealing with nonresponse: Strategies to increase participation and methods for postsurvey adjustments. Mathematical Population Studies. 2017;24(2):79–83. Available at: https://www.tandfonline.com/doi/citedby/10.1080/08898480.2017.1299988?scroll=top&needAccess=true&role=tab

Annex

```
#####(A)
#function to correct for finite sample size
correct.N <- function(n, N){
  # n = sample size obtained from the function (e.g. pwr.p.test)
  # N = total population size
  # Returns the corrected sample size
  (n*N)/(n+N) }

#####(B)
# function to compute sample size for testing
# difference in two dependent proportions
sampleSizeMcNemar<-function (p1, p2, alpha = 0.05, power = 0.8, plot=T)
{
  ## INPUT:
  # p1: proportion of positive for test 1
  # p2: proportion of positive for test 2
  # alpha: significance level for two-sided test
  # power: power of the test

  ## CALCULATIONS:
  # proportion of positive for test 1 and 2
  p11 <- sort(seq(min(p2, p1), max(0,p2 + p1 - 1), by = -10^-4))
  # proportion of shifts from positive to negative, if a shift in test result
  s <- (p1 - p11)/(p1 + p2 - 2 * p11)
  # sample sizes
  nl <- ceiling(0.25 * (qnorm(alpha/2) + qnorm(1 - power))^2/(0.5 -s)^2/(abs(p1 + p2
- 2 * p11)))
  # min/midpoint/max
  N <- nl[c(1, median(1:length(nl)), length(nl))]

  if (plot==TRUE){
    plot(s,nl,type="l",xlab="proportion of shifts from positive to negative, if a
shift in test result",
         ylab="sample size")
    points(s[c(1, median(1:length(nl)), length(nl))],N,col="blue",pch=16)
    text(s[1],nl[1],paste0("n=", nl[1]),col="blue",pos=2,cex=0.6)
    text(s[median(1:length(nl))],nl[median(1:length(nl))],paste0("n=",
nl[median(1:length(nl))]),col="blue",pos=2,cex=0.6)
    text(s[length(nl)],nl[length(nl)],paste0("n=",
nl[length(nl)]),col="blue",pos=3,cex=0.6)
  }
  names(N) <- c("N_min", "N_mid", "N_max")
  return(N)
}
```

**European Centre for Disease
Prevention and Control (ECDC)**

Gustav III:s Boulevard 40, 16973 Solna, Sweden

Tel. +46 858601000
Fax +46 858601001
www.ecdc.europa.eu

An agency of the European Union
www.europa.eu

Subscribe to our publications
www.ecdc.europa.eu/en/publications

Follow us on Twitter
@ECDC_EU

Like our Facebook page
www.facebook.com/ECDC.EU

Publications Office
of the European Union