Introduction to Study Design - Where does the data come from?

Inhalt

"We need less research, better research and research done for the right reasons"

"Huge sums of money are spent annually on research that is seriously flawed through the use of inappropriate designs, unrepresentative samples, small samples, incorrect methods of analysis, and faulty interpretation"

Professor Doug Altman, Statistician (The scandal of poor medical research, doi:10.1136/bmj.308.6924.283 pmid:8124111)

Objectives

This Chapter is an introduction to planning a research study. We will highlight the components of a **good research plan** and the formulation of **research questions**. We will also clarify the difference between a research question and **hypothesis testing**, illustrate the correct use of hypothesis testing, and elucidate the difference between **clinical and statistical significance**. Last but not least, you will be introduced to common **study designs** in medical research and the most commonly used tools for **good reporting quality** and evaluating the **internal validity** of a study. A carefully selected research question and rigorous study design are the backbones of good data quality and research.

Planning a Research Project

To formulate a research question worthy of being investigated, you need to keep up with the publications in your broad research field. Participating in research projects to build up your experience and portfolio would also be best. You will also benefit by participating actively in journal clubs, conferences and workshops relevant to your broad research field.

A research idea may be brand new and include an unanswered research question for a specific population, intervention and health condition. Alternatively, your research question may aim to improve the design and analysis of a published study with many limitations that question the credibility of its finding. For the latter, you may consider more study participants, allow more diversity in the characteristics of the participants, or apply recent analysis methods that perform better.

To decide whether the research question should be investigated, you need to ensure that you fulfil a series of minimum requirements:

- Your idea is **truly novel**; you have not found similar published research. You may summarise in three paragraphs what has been done, what is missing and why you focus on the specific knowledge gaps.
- The research question should primarily serve the patient's needs and then clinical curiosity. This is important if you want to investigate a research question that has already been answered; however, the study has a questionable design and conduct.
- You have carefully considered the project timeline and aim to secure a budget (e.g., by writing a research grant). You also plan to consider a team of domain-specific experts.

Having a carefully planned research question will increase the project's success. Consequently, the project will not contribute to the increasing **research waste**, namely, a study with poor design, conduct, and report.

Once you have decided on the topic and the research questions, the next step is to plan the study design and the statistical analysis carefully. The research question guides the design and the statistical analysis – not the other way around. Your goal is to select a study design and statistical analysis that serve the research question. Study design and statistical analysis are the topics of the upcoming weeks.

Ethics Approval

To initiate your research project, you need first to get ethics approval. You apply for ethics approval if you plan a project on humans or animals. A project based on published studies (e.g., systematic reviews) does not require ethical approval.

You must fill out the study synopsis **before** study initiation. When preparing your project synopsis for ethics approval, important material to consider includes the **General Data Protection Regulation** and the site for the Ethics Committee in the MHH:

- **General Data Protection Regulation:** https://gdpr-info.eu/
- **MHH Ethics Committee:** https://www.mhh.de/ethikkommission
- **Application forms:** https://www.mhh.de/ethikkommission/antragsstellung

The PICOT Framework

After deciding on the research questions worthy of being investigated, the next step is to work on the *PICOT elements* of the questions. The abbreviation PICOT stands for **Participant**, **Intervention**, **Comparator**, **Outcome** and **Time**. The PICOT framework is essentially a synopsis of your research study's inclusion and exclusion criteria. In the synopsis for ethics approval, you are asked to indicate the PICOT elements of your study.

The participant population may include demographic, social and clinical characteristics. The intervention or exposure may include the dosage, mode, frequency and duration of delivery. Likewise, for the comparator, if it is relevant to the research question. Make sure to include an efficacy outcome (e.g., improvement of symptoms) and a safety outcome (e.g., adverse events) to account for the risk-benefit trade-off when making recommendations. The time may include the study duration and specific time points in a longitudinal assessment (if applicable).

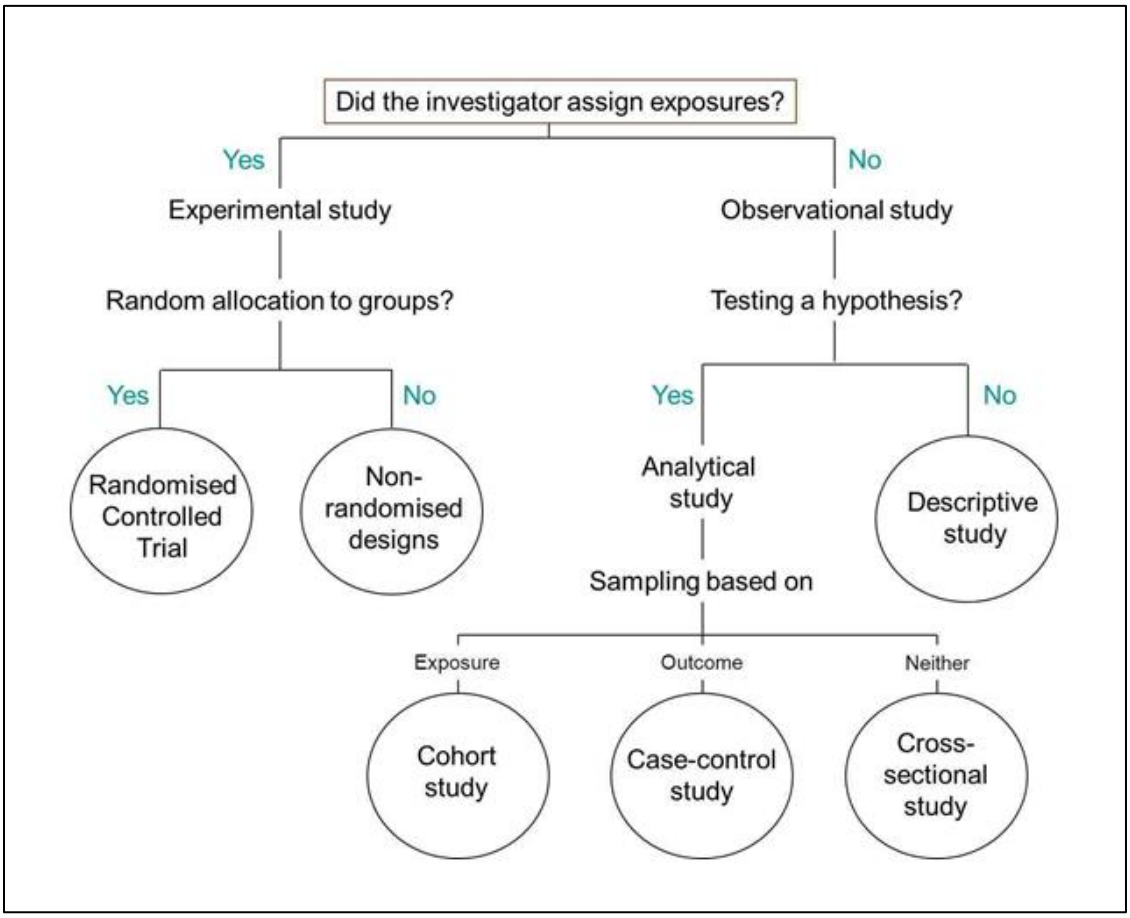
As a simple example, let us consider the following research question:

"We want to identify the demographic, social and clinical characteristics that can predict stroke in adults with diabetes who receive medication for hypertension after a follow-up of 2 years."

The PICOT elements are listed below:

PICOT	Exact elements
Participants	▪ Demographic characteristics (e.g., age≥ 18 years old, gender) ▪ Social characteristics (e.g., education and income) ▪ Clinical characteristics (e.g., blood pressure and insulin level)
Intervention	Medication for hypertension
Comparator	None
Outcome	Risk of stroke
Time	2 years

Study Designs in Medical Research



Source: Figure 5.1 in AFMC Primer on Population Health

Study designs can be based on quantitative or qualitative methods. **Quantitative study designs** include experimental studies (i.e., randomised controlled trials and non-randomised designs) and observational studies (i.e., cohort studies, case-control and cross-sectional

studies). **Qualitative study designs** include grounded theory, phenomenology and hermeneutics, and ethnography. We focus on the quantitative study designs.

The advantages and disadvantages of the different **quantitative** study designs designate a different **strength of evidence**. The term **strength of evidence** implies the extent of confidence we place in the results of a study.

A study associated with the least biases in the design and conduct provides stronger evidence compared to a study with more flaws in the design and conduct.

The ability to critically appraise the quality and the strength of evidence of a study can determine the quality of decisions made in the healthcare sector. The **hierarchy of evidence** offers a *heuristic* sorting of the study designs by **increasing evidence strength**. Therefore, systematic reviews synthesising randomised controlled trials are placed at the top of the evidence hierarchy, followed by randomised controlled trials, cohort, case-control and cross-sectional studies. Even lower in the evidence hierarchy are case reports, followed by ideas/opinions/editorials and animal research. In-vitro studies are at the bottom of the evidence hierarchy.

However, undue reliance on the evidence hierarchy may be misleading.

This can be the case when the design and conduct of a study do *not* reflect the study's position in the evidence hierarchy. For instance, a systematic review with questionable design and conduct offers lower-quality evidence than a randomised controlled trial with high standards in the design and conduct. Similarly, a carefully planned, conducted and analysed observational study offers higher evidence than a randomised controlled trial with fundamental design, conduct and reporting flaws. By the same token, observational studies offer better evidence to evaluate potential harms than randomised controlled trials because the former can be conducted for a longer period.

Systematic Reviews

A systematic review constitutes a **meta-research**. Several *published and unpublished studies* on a pre-specified health condition, population, interventions and design are searched, retrieved, appraised and synthesised. The aim is to summarise the available relevant evidence qualitatively and quantitatively (if appropriate). The anticipated clinical and methodological inconsistencies in the collated studies, alongside the quality of their design, can frame the conclusions for the strength of the available evidence.

Systematic reviews have also been advocated for the **design of future research to minimise research waste**. Systematic reviews were originally developed to increase the likelihood of detecting a statistically significant difference (known as power) in the compared interventions, contrary to an individual randomised controlled trial. **Pairwise meta-analysis** is the statistical tool to synthesise a series of studies on two interventions and has received attention from the wide research community since the '70s.

The systematic review methodology has been substantially advanced and refined in the last three decades. Systematic reviews can now include different study designs, investigate more

than two interventions, and target prevalence, aetiology, prognosis, and diagnosis rather than therapy alone. Systematic reviews on multiple interventions are known as **comparative effectiveness reviews**. The tool to synthesise a network of multiple interventions is known as **network meta-analysis**. Network meta-analysis also constitutes the analysis tool of **living systematic reviews**. A living systematic review is a new generation of meta-research constantly updated when a new relevant study is published to enable an up-to-date evidence ecosystem.

Cochrane Collaboration is at the forefront of planning, conducting, analysing and reporting high-quality systematic reviews. Many founding and co-founding members have been the leading force behind the constantly advanced methodology of systematic reviews.

Advantages

- Systematic reviews are **resourceful and labour-saving** for the reader of the constantly explosive number of medical publications.
- They can also **uncover inconsistencies** in the design, conduct, results, and reporting of the collected studies to frame their implication on the conclusions delivered to the end-users.
- Systematic reviews **elucidate knowledge gaps and research flaws** in the available evidence. Hence, ***systematic reviews are often coined as more reliable than any individual study***.
- Importantly, they form the **backbone of evidence-based practice and decision-making**.

Disadvantages

- Systematic reviews can be **time-consuming** to conduct and publish.
- They become **outdated** if the authors do not plan to update the collected studies every time new relevant evidence emerges.
- Systematic reviews require **years of training** to gain expertise in literature search and quality appraisal.

Randomised Controlled Trials

Randomised controlled trials (RCTs) have been coined as the **gold standard** of experimental research for preventing **selection bias**, reducing bias in the results.

Selection bias occurs when two or more groups of participants differ beyond their assigned intervention or exposure. Consequently, any difference observed in the compared interventions may be attributed to the different characteristics of the participants rather than the interventions alone. For instance, disease severity can affect how an intervention works: the higher the disease severity in a patient, the more likely the patient will benefit from the intervention. Suppose more severely ill patients receive the new intervention, and moderately ill patients receive the placebo. The difference between the new intervention and placebo would be larger than if patients were randomly assigned to the new intervention or placebo. Therefore, the difference is biased upwards; in other words, the difference is overestimated. Randomising the patients to receive either the experimental or the control

arm balances important characteristics in the compared groups that act as **confounders (also known as 'effect modifiers')**. Consequently, any difference observed in the **average** outcome of the compared groups can be explained only by the assigned intervention.

"The main purpose of randomisation is to eliminate selection bias and balance known and unknown confounding factors in order to create a control group that is as similar as possible to the treatment group"

Akobeng AK. Understanding randomised controlled trials. Arch Dis Child. 2005;90(8):840-4. (doi: 10.1136/adc.2004.058222)

A successful design and conduct of an RCT depend on a series of vital actions:

a) the **randomisation method**,

b) the **allocation concealment**,

c) **blinding** the patient, clinician, outcome assessor and statistician from the assigned interventions,

d) attempts to minimise premature discontinuation of the patients, and

e) **avoid** any **protocol violation**.

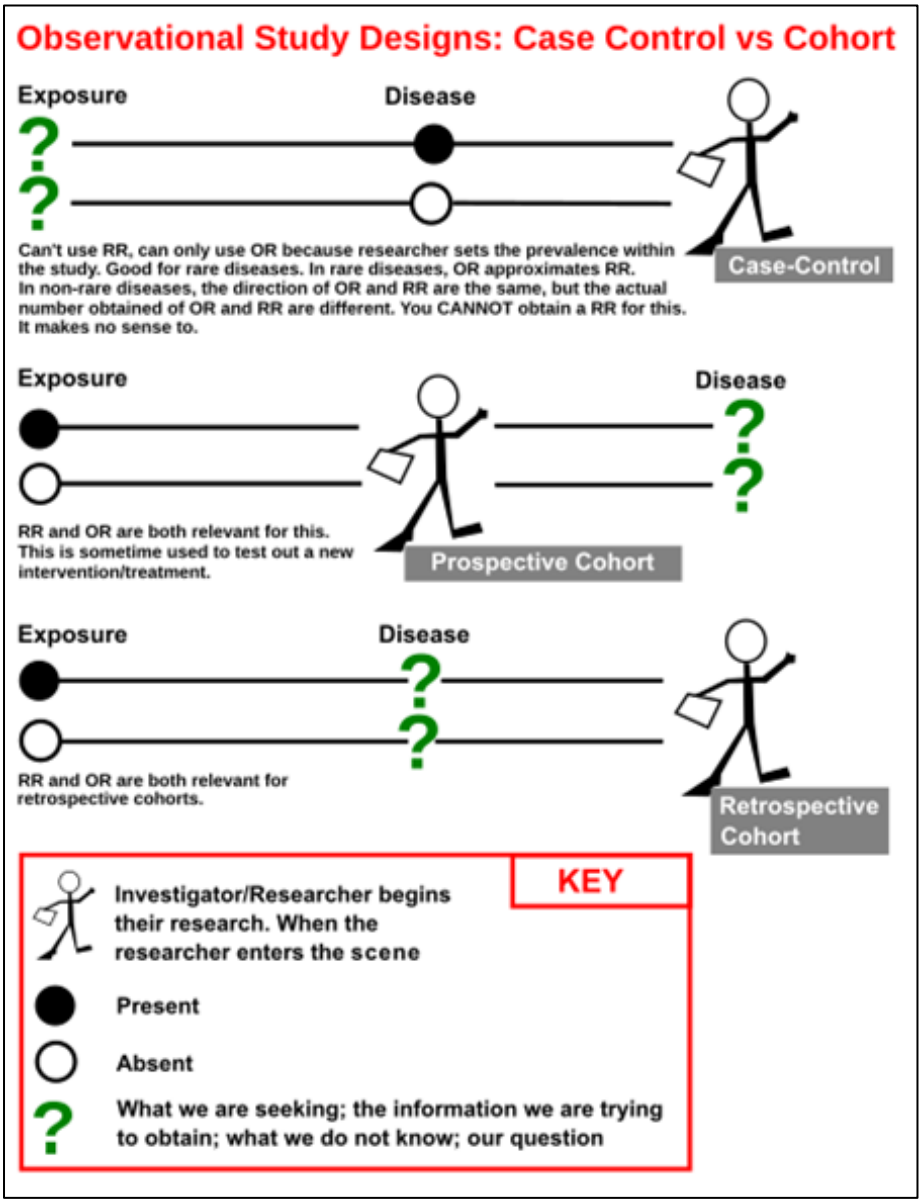
These actions are interrelated to some extent. For instance, an unsuccessful or lack of allocation concealment will compromise the blinding to the assigned intervention and result in premature discontinuation. Consequently, the study's internal (and possibly external) validity will be compromised, leading to questionable results. Akobeng (2005) offers a fruitful discussion on core terminologies related to the design, conduct, and analysis of an RCT.

RCTs that are designed and conducted with high standards can infer **causation**. Namely, they can provide firm evidence about the efficacy or safety of the compared interventions for a (pre-defined, patient-important) outcome. Using regulatory terms, RCTs are confirmatory trials; namely,

"[...] confirmatory trials [are necessary to] provide firm evidence of efficacy or safety. In such trials the key hypothesis of interest follows directly from the trial's primary objective, is always pre-defined, and is the hypothesis that is subsequently tested when the trial is complete. In a confirmatory trial it is equally important to estimate with due precision the size of the effects attributable to the treatment of interest and to relate these effects to their clinical significance"

European Medicines Agency. NOTE FOR GUIDANCE ON STATISTICAL PRINCIPLES FOR CLINICAL TRIALS. ICH Topic E 9 Statistical Principles for Clinical Trials. September 1998 (CPMP/ICH/363/96).

Observational Study Designs



Source: Wikipedia: Cohort statistics

A. Cohort Study

A cohort study is an observational study design conducted prospectively or retrospectively. In a **prospective cohort study**, a group of people with pre-defined characteristics are recruited. Their pre-defined outcome is measured over a long period (usually longer than the duration of an RCT). For instance, adults over 60 with type II diabetes are recruited and followed up for three years. Demographic, social and clinical characteristics are measured at the beginning of the study; the clinical characteristics are also measured every six months. The outcome of interest is experiencing any of the following cardiovascular diseases (CVDs) by the end of the study: stroke, heart attack or heart failure. A cohort study with repeated measurement of the outcome during the study is also known as a **longitudinal cohort study**.

A **prospective cohort study** can also be **comparative** when more than one group/exposure is considered. In the previous example, we could have considered a group of adults over 60 with and without type II diabetes. We are interested in identifying characteristics that differentiate the risk of CVD in these two groups.

In a **retrospective cohort study**, the researchers resort to repositories and patient records in the hospital to select participants by exposure status. They further collect information on the outcome and pre-defined characteristics. Alternatively, the researchers survey participants who have the exposure of interest to retrieve information from the past regarding their outcome and other important characteristics. Hence, the outcome and exposure precede the study initiation in the retrospective cohort study. On the contrary, exposure exists at study initiation in the prospective cohort study, and the outcome occurs during the study. As the terms indicate, with a prospective cohort study, the researchers follow one or more groups of participants *over time*. In contrast, the researchers look back in time with a retrospective study.

Advantages

- An advantage of the prospective cohort study is that it determines the **temporal relationship** between the exposure and the outcome; namely, the exposure did happen before the outcome.
- By following the participants over time, the **risk of recall bias** is minimised, contrary to the retrospective cohort study.
- The **long duration** of the prospective cohort study allows collecting sufficient information of important characteristics.

Disadvantages

- **Prospective cohort studies** can be **costly** and associated with a **high risk of follow-up losses**.
- Since no randomisation is performed by design, the results are highly sensitive to **confounding bias**. This is a characteristic shared by *all* observational studies. Consequently, **observational studies cannot identify any cause-effect relationships**.
- Cohort studies **do not administer any intervention**; they are largely about **exposure** and characteristics that may predict the risk of the disease.

B. Case-control Study

In a **retrospective cohort study**, the researchers separate the participants by exposure and then examine the outcome status. In contrast, in a **case-control study**, the researchers separate the participants by outcome status and then look backwards at the exposure. The **case group** refers to participants *with the outcome* (e.g., with disease), and the **control group** refers to participants *without the outcome* (e.g., healthy). Therefore, a case-control study is a **comparative observational study** by definition. In practice, the researchers identify the cases and match them with the controls based on specific characteristics, such as age and gender.

Advantages

- Contrary to a prospective cohort study, a case-control study is **relatively inexpensive** and has a **shorter duration**.
- This design is often used to **investigate rare diseases** as there is little or no prior knowledge about the association between exposure and outcome.

Disadvantages

- Like with any observational study, the risk of **selection bias** is high as the characteristics of cases and controls may be systematically different from the target population. This systematic inconsistency is largely attributed to the **sampling method** applied. **Non-probability sampling**, such as *convenience sampling*, is often applied for recruitment. For instance, the researchers resort to the hospital they work for easy access (convenience). It is unlikely for a hospital to represent the target population entirely. Hence, selection bias is imminent.
- Like with the retrospective cohort study, a case-control study is not immune to **recall bias**.
- The lack of randomisation by design is likely to introduce **confounding bias** in the results. However, matching the cases with the controls and proceeding with a proper regression analysis may mitigate (but not eliminate) confounding bias.

C. Cross-sectional Study

A cross-sectional study belongs to observational studies. Most **surveys** are cross-sectional studies. For instance, randomly sampling households from a national telephone directory and contacting the people via telephone for an interview constitutes a cross-sectional survey. The researchers **record the outcome, exposure, and other important characteristics** of the interviewee on the spot. However, the recruitment of the sample to interview can take some time. Thus, **a cross-sectional study cannot be longitudinal** by design.

The cross-sectional studies can **infer an association between an outcome and exposure** but **not a cause-effect relationship** – the latter is plausible only in a high-quality RCT. It is impossible to determine whether the exposure preceded the outcome by design. Cross-sectional studies can **inform the prevalence of a disease**, that is, the proportion of the population with the disease. However, they **cannot measure the incidence of the disease** as it would require following the participants within a specific period (longitudinally). A cross-sectional study can be used to **inform the hypotheses to be investigated with a cohort study**.

Advantages

- Like with the case-control studies, cross-sectional studies are **relatively inexpensive**, **easy to perform**, and **short in duration**.
- Since participants are interviewed only once, cross-sectional studies **do not suffer from losses to follow-up**, contrary to prospective cohort studies.

Disadvantages

- While the cross-sectional studies attempt to represent the target population, they are **not immune to non-response bias**, thus, distorting the intended sample. For instance, when the randomly selected households who answer the telephone have systematically different characteristics from the households that do not answer the telephone, the collected sample will not be representative of the target population.
- Like with the case-control and retrospective cohort studies, cross-sectional studies are also **prone to recall bias**.
- Contrary to case-control studies, cross-sectional studies **cannot be used to study rare diseases**.

Optional Reading List 🕮

Research questions and PICOT framework

1. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. **The well-built clinical question: a key to evidence-based decisions**. ACP J Club. 1995;123(3):A12-3. PMID: 7582737

2. [1]Guyatt GH, Osoba D, Wu AW, et al. **Methods to explain the clinical significance of health status measures**. Mayo Clin Proc. 2002;77(4):371-83. doi: 10.4065/77.4.371

3. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. **Utilization of the PICO framework to improve searching PubMed for clinical questions**. BMC Med Inform Decis Mak. 2007;7:16. doi: 10.1186/1472-6947-7-16

4. Creswell J. **Chapter 7: Research Questions and Hypotheses**. In: Research design: Qualitative, quantitative, and mixed methods approaches (3rd ed.). Los Angeles: Sage; 2009. Online access

5. Riva JJ, Malik KM, Burnie SJ, Endicott AR, Busse JW. **What is your research question? An introduction to the PICOT format for clinicians**. J Can Chiropr Assoc. 2012;56(3):167-71. PMID: 22997465

6. Kersting C, Kneer M, Barzel A. **Patient-relevant outcomes: what are we talking about? A scoping review to improve conceptual clarity**. BMC Health Serv Res. 2020;20(1):596. doi: 10.1186/s12913-020-05442-9

7. Zapf A, Rauch G, Kieser M. **Why do you need a biostatistician?** BMC Med Res Methodol. 2020;20(1):23. doi: 10.1186/s12874-020-0916-4

Study designs in medical research

1. Kendall JM. **Designing a research project: randomised controlled trials and their principles**. Emerg Med J. 2003;20(2):164-8. doi: 10.1136/emj.20.2.164
2. Mann CJ. **Observational research methods. Research design II: cohort, cross sectional, and case-control studies**. Emerg Med J. 2003;20(1):54-60. doi: 10.1136/emj.20.1.54
3. Akobeng AK. **Understanding randomised controlled trials**. Arch Dis Child. 2005;90(8):840-4. doi: 10.1136/adc.2004.058222
4. Sutton AJ, Cooper NJ, Jones DR. Evidence synthesis as the key to more coherent and efficient research. BMC Med Res Methodol. 2009;9:29. doi: 10.1186/1471-2288-9-29
5. Sedgwick P. **Cross sectional studies: advantages and disadvantages**. BMJ. 2014;348:g2276. doi: 10.1136/bmj.g2276
6. Sedgwick P. **Bias in observational study designs: case-control studies**. BMJ. 2015;350:h560. doi: 10.1136/bmj.h560
7. Sedgwick P. **Bias in observational study designs: cross sectional studies**. BMJ. 2015;350:h1286. doi: 10.1136/bmj.h1286
8. Murad MH, Asi N, Alsawas M, Alahdab F. **New evidence pyramid**. Evid Based Med. 2016;21(4):125-7. doi: 10.1136/ebmed-2016-110401
9. [2]Krauss A. **Why all randomised controlled trials produce biased results**. Ann Med. 2018;50(4):312-322. doi: 10.1080/07853890.2018.1453233
10. Eichler HG, Pignatti F, Schwarzer-Daum B, et al. **Randomized Controlled Trials Versus Real World Evidence: Neither Magic Nor Myth**. Clin Pharmacol Ther. 2021;109(5):1212-1218. doi: 10.1002/cpt.2083

[1] We will not delve into the framework of minimally clinically important differences.

[2] The title should **not** be taken literally. This article serves to remind us that a study's quality should not be automatically equated to its position in the evidence pyramid. **You should always critically appraise the study at hand**.

Useful Links

European Medicines Agency

- Clinical trials in human medicine
- Clinical Trial Regulation

European Union Drug Regulating Authorities Clinical Trials Database (EudraCT)

- Trial registration

EU Clinical Trials Register

- Searching protocols and results

DRKS - German Clinical Trials Register

- Registration for clinical trials conducted in Germany

Cochrane Collaboration (Systematic Reviews)

- Cochrane Library
- Cochrane Handbook

PROSPERO

- Registration of systematic reviews protocols

Introduction to Hypothesis Testing

Research questions and hypotheses are often used interchangeably. However, they are two distinct components: a research question is a synopsis of the PICOT. For instance, the following text is a **research question** and can be investigated using a proper study design:

> "What are the demographic, social and clinical predictors of stroke in adults with diabetes who receive medication for hypertension after a follow-up of 2 years?"

On the other hand, the following text refers to a **hypothesis**:

> "We test the null hypothesis (none of the characteristics predict stroke) against the alternative hypothesis (at least one characteristic predicts stroke) via a multivariable logistic regression model and a significance level of 5%."

A hypothesis (that there is at least one predictor of stroke) can be investigated using a specific statistical analysis (a multivariable logistic regression model) and a minimally accepted error of 5%.

Note that we specify **one** hypothesis testing for each research question separately – not for each PICOT element of the research question.

Type of Variables

When we formulate a hypothesis, we must distinguish between two types of variables, the dependent and independent variables:

- The **dependent variable** refers to the measured outcome (e.g., having a stroke, yes or no). We do *not* know a priori the values of the dependent variable. They are a **'surprise'** for the investigator.
- The **independent variables** are more than one and refer to the Participant, Intervention, and Comparator elements of the PICOT framework. The investigators know *a priori* the measurements of the independent variables through the inclusion and exclusion criteria.

As a simple example, let us consider the following research question again:

> "What are the demographic, social and clinical predictors of stroke in adults with diabetes who receive medication for hypertension after a follow-up of 2 years?"

The following table illustrates the dependent variable (one) and independent variables:

Dependent variable (one)	Independent variables (many)
▪ Stroke (yes/no)	▪ Demographic (e.g., age and gender) ▪ Social (e.g., education and marital status) ▪ Clinical (e.g., cardiovascular- and diabetes-related) ▪ Medication

Hypothesis Testing: Null versus Alternative Hypothesis

Let us use another easy example to demystify the elements of the hypothesis:

"We want to investigate which dietary plan can improve fasting insulin substantially in people with insulin resistance: eating more meat than vegetables or the other way around."

Disclaimer: let us assume for the moment that the design of this study is ethically approved, and more details on the design are out of the scope of this lecture.

Fasting insulin is the **dependent variable** and refers to the outcome. We do **not** know each participant's fasting insulin level at the start of the study. The dietary plan is the (unique) **independent variable** and refers to exposure. We know the dietary plan of each participant at the start of the study because the participants chose the dietary plan.

Now, we will formulate the research question into hypothesis testing. Hypothesis testing comprises the null hypothesis and the alternative hypothesis. The **null hypothesis** is the *devil's advocate* and implies no association between the dependent and independent variables.

Using our simple example, the null hypothesis would be:

"Both dietary plans yield the same fasting insulin."

which is equivalent to:

"Fasting insulin is ***not*** associated with any of the dietary plans."

Statistically speaking, the null hypothesis is defined as follows:

"The **true** average fasting insulin in dietary plan A ***equals*** the **true** average fasting insulin in dietary plan B."

$$\mu_A = \mu_B$$

which is equivalent to:

"The difference in the true average fasting insulin of the compared dietary plans ***equals*** 0."

$$\mu_A - \mu_B = 0$$

The **alternative hypothesis** implies the opposite 'belief', that is,

"The compared dietary plans yield ***different*** fasting insulin"

which is equivalent to:

"Fasting insulin ***is*** associated with the dietary plan."

Statistically speaking, we do **not** give a specific direction to the alternative hypothesis because we have *no prior* knowledge on which dietary plan may result in higher or lower average fasting insulin:

"The **true** average fasting insulin in dietary plan A ***differs*** from the **true** average fasting insulin in dietary plan B."

$$\mu_A \neq \mu_B$$

which is equivalent to:

"The difference in the true average fasting insulin of the compared dietary plans ***differs from*** 0."

$$\mu_A - \mu_B \neq 0$$

Subsequently, we measure the fasting insulin in each participant in our sample. Then, we calculate the average fasting insulin in each group.

Disclaimer: We assume that the sample size is enough to represent the target population.

The last step is to apply a **statistical test** to decide between two scenarios:

Scenario 1

"We **reject** the null hypothesis, and therefore, we **accept** the alternative hypothesis."

Correct: $\mu_A \neq \mu_B$

Scenario 2

"We **fail to reject** the null hypothesis, and therefore, we **do not accept** the alternative hypothesis."

Wrong: $\mu_A = \mu_B$

Scenario 2 does *not* imply that the compared dietary plans do not differ regarding fasting insulin. Such a statement is seriously misleading and appears in many published articles.

Absence of Evidence is not Evidence of Absence

In his popular article, Professor Doug Altman highlighted that absence of evidence (i.e., failure to reject the null hypothesis) is not evidence of absence (i.e., the two dietary plans have the same effect).

There are several typical reasons to fail to reject the null hypothesis:

- The study has **an inadequate sample** of participants. This can result from having **too many missing data** in the variables or, even worse, **no**t having performed any **sample size calculation**.
- **No sample size calculation, no ethics approval!**
- Not including all **important independent variables** in the analysis model. Such as strategy will yield **substantial unexplained variability** in the outcome, and hence, will **increase our uncertainty** about which dietary plan is better for the fasting glucose.
- Both aforementioned cases imply **poor design of the project** and contribute to the **research waste**.
- The investigated **outcome is rare**. In this case, the association between the dependent and independent variables will be estimated with **great imprecision**, and hence, will **increase our uncertainty** about which dietary plan is better for the fasting glucose.
- It can be the **rare case** where the compared groups are truly similar regarding the measured outcome so that $\mu_A = \mu_B$.

Statistical Significance and Clinical Significance

Another popular misinterpretation in hypothesis testing pertains to the statistical and clinical significance of the test results. Statistical significance and clinical significance should ***not*** be used interchangeably because they are ***not*** the same.

Statistical significance refers to rejecting the null hypothesis and, thus, accepting the alternative hypothesis. **Clinical significance** refers to observing a *clinically important difference* in the compared groups, and this difference is meaningful for the patient. You may have heard the term **minimal clinically important difference**. Guyatt et al. (2002) offer a gentle review of methods to determine the minimal clinically important difference and hence, clinical significance in the context of quality of life measures.

"An association may be statistically significant but clinically non-significant".

This is typical in large trials. They can detect very small differences in the compared groups that are clinically non-significant.

"An association may be statistically non-significant but clinically significant".

This is typical in small trials. They can detect large and clinically significant effects. However, they may be accompanied by a large variance and hence, failure to reject the null hypothesis.

Research Quality Tools: Separating Wheat from Chaff

The daily publication rate of *new* research studies has surged over the last three decades. Pioneering achievements in Computer Science and Technology have led to fundamental advances in conducting and analysing complex research studies. Inevitably, many multinational and multidisciplinary collaborations have been fostered across the wide medical field over the years. Hence, the **publication explosion in medical research**.

With the rise in the published literature, several critical views concerning the quality of explosive published research have also emerged. The critical views concentrate on the **constitution of the research study**, namely, the **planning**, **design**, **conduct** and **reporting**. Any inadequacies in the constitution of a research study may seriously compromise the **quality of the results delivered to the end-users**, such as patients, clinicians and decision-makers. Published studies with questionable constitutions contribute to the **ever-growing body of research waste**.

Tools for Quality Reporting

Transparent communication of the study objectives, design, conduct, analysis and results require **good reporting practices**. The publication explosion in medical research and the numerous cases of inadequacies in reporting the published research led to international initiatives to create a series of minimum, evidence-based, consolidated recommendations. The EQUATOR (Enhancing the QUAlity and Transparency Of health Research) Network is such an initiative.

> **"The EQUATOR Network is an international initiative that seeks to improve the reliability and value of published health research literature by promoting transparent and accurate reporting and wider use of robust reporting guidelines. It is the first coordinated attempt to tackle the problems of inadequate reporting systematically and on a global scale; it advances the work done by individual groups over the last 15 years."**
>
> The EQUATOR Network (https://www.equator-network.org/about-us/)

These recommendations aim to promote **good reporting practices**. Conforming to these recommendations is a requirement from all actors of the research and publication community. Most **(non-predatory) journals** have endorsed the consolidated recommendations: the authors **cannot** submit their work to the Journal if they do not conform to official reporting standards that align with the study design considered.

The **EQUATOR Network** is a large suite of different reporting guidelines for the different study types. Currently, there are **471 reporting guidelines** in total. In this Chapter, we have listed the reporting guidelines for **the most popular study designs in health research**:

a) the *PRISMA 2020 statement* for a systematic review,

b) the *CONSORT 2010 Statement* for the randomised controlled trials, and

c) the *Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement* for observational studies.

Several **reporting guidelines** have been refined to include further reporting items or amend current ones suitable **for the different clinical areas**. The EQUATOR Network also includes **reporting guidelines** consolidated for **study protocols**. Examples include:

a) the *SPIRIT 2013 Statement* for the randomised controlled trials, and

b) the *Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement* for the systematic reviews.

Tools for Internal Validity

A **high-quality research study** stands out for its high standards in the design, conduct, analysis and reporting and the direct applicability of its results to the target population.

The quality of a study is inextricably linked to its internal and external validity, among others.

Internal validity dictates the extent to which a study is free from biases, namely, actions that conceal the true association between the outcome and the interventions or exposures. A well-designed and conducted study has high internal validity; thus, it can establish a cause-effect (randomised controlled trial) or outcome-exposure association (observational study). The **external validity** dictates whether the study results are generalisable and can be applied to the intended population.

Developing a rigorous protocol to **increase the internal validity** *may* **compromise the external validity**, as the results may not have immediate applicability outside the strict environment of the study.

This Chapter focuses on **internal validity** and lists (under the **Optional reading list**) the tools established to evaluate internal validity in the most popular study designs in health research:

a) the *Risk of Bias 2 (RoB 2) tool* for randomised controlled trials,

b) the *Risk Of Bias In Non-randomised Studies - of Interventions (ROBINS-I) tool* for non-randomised studies,

c) the *Risk of bias in observational studies of exposures (ROBINS-E) tool* for observational studies, and

d) the *AMSTAR 2 (A MeaSurement Tool to Assess systematic Reviews)* for systematic reviews with randomised or non-randomised controlled trials.

The RoB 2, ROBINS-I, and ROBINS-E tools are commonly applied at the systematic review level. AMSTAR 2 is a critical appraisal tool developed for end-users of systematic reviews. A less popular yet useful approach may be to use these tools when writing the report of the corresponding study to frame its strengths and limitations.

Optional Reading List 🕮

Research Project: the good, the bad, and the ugly

1. Altman DG. **The scandal of poor medical research**. BMJ. 1994;308(6924):283–4. doi: 10.1136/bmj.308.6924.283

2. Altman DG, Bland JM. **Absence of evidence is not evidence of absence**. BMJ. 1995;311(7003):485. doi: 10.1136/bmj.311.7003.485

3. Ioannidis JP. **Why most published research findings are false**. PLoS Med. 2005;2(8):e124. doi: 10.1371/journal.pmed.0020124

4. Bastian H, Glasziou P, Chalmers I. **Seventy-five trials and eleven systematic reviews a day: how will we ever keep up?** PLoS Med. 2010;7(9):e1000326. doi: 10.1371/journal.pmed.1000326

5. Elliott JH, Turner T, Clavisi O, et al. **Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap**. PLoS Med. 2014;11(2):e1001603. doi: 10.1371/journal.pmed.1001603

6. Ioannidis JP. **Why Most Clinical Research Is Not Useful**. PLoS Med. 2016;13(6):e1002049. doi: 10.1371/journal.pmed.1002049

7. Glasziou P, Chalmers I. **Research waste is still a scandal—an essay by Paul Glasziou and Iain Chalmers**. BMJ. 2018;363. doi: 10.1136/bmj.k4645

Research Quality Tools

1. von Elm E, Altman DG, Egger M, et al. **The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies**. Ann Intern Med. 2007;147(8):573–7. doi: 10.7326/0003-4819-147-8-200710160-00010

2. Schulz KF, Altman DG, Moher D; CONSORT Group. **CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials*. Ann Intern Med. 2010;152(11):726–32. doi: 10.7326/0003-4819-152-11-201006010-00232

3. Chan AW, Tetzlaff JM, Altman DG, et al. **SPIRIT 2013 statement: defining standard protocol items for clinical trials**. Ann Intern Med. 2013;158(3):200–7. doi: 10.7326/0003-4819-158-3-201302050-00583

4. Sterne JA, Hernán MA, Reeves BC, et al. **ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions**. BMJ. 2016;355:i4919. doi: 10.1136/bmj.i4919

5. Shea BJ, Reeves BC, Wells G, et al. **AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both**. BMJ. 2017;358:j4008. doi: 10.1136/bmj.j4008

6. Bero L, Chartres N, Diong J, et al. **The risk of bias in observational studies of exposures (ROBINS-E) tool: concerns arising from application to observational studies of exposures**. Syst Rev. 2018;7(1):242. doi: 10.1186/s13643-018-0915-2

7. Sterne JAC, Savović J, Page MJ, et al. **RoB 2: a revised tool for assessing risk of bias in randomised trials**. BMJ. 2019;366:l4898. doi: 10.1136/bmj.l4898

8. Page MJ, McKenzie JE, Bossuyt PM, et al. **The PRISMA 2020 statement: an updated guideline for reporting systematic reviews**. BMJ. 2021;372:n71. doi: 10.1136/bmj.n71

Summary

… and resources!

… plan carefully, instead!

When you plan a research project (your thesis included), you must consider the **research topic** carefully: what would you investigate and why. For instance, you may be interested in a health condition and participant population that has already been investigated (the trial report is published). However, the trial was small and inconclusive, had a questionable design and statistical analysis, and the authors misinterpreted the results. In this case, you aim to conduct a trial using high standards in the design, analysis and reporting so that your research results can benefit the interested end-user, such as patients, clinicians, policy-makers and guideline developers. Alternatively, a knowledge gap exists on a specific health condition and therapy. For instance, this therapy has been used for that condition as an off-label for years, and you aim to elucidate the risk and benefits of that therapy.

To frame a research topic, the investigators (you) need to **determine research questions** and **hypotheses** specific to the population and condition under investigation. The next step is to decide on the **study design** that aligns with the investigated research questions. The appropriateness of the study design will determine the **quality of the collected data** and, by extent, the quality of the **statistical analysis** and conclusions. Lastly, the successful execution of every research project strongly depends on the diversity of knowledge and expertise to deliver quality results. Therefore, recruiting a **multidisciplinary team of experts** is the key to a successful research project. **Clinicians, patients and patient representatives, informaticians, statisticians, sponsors, and regulatory scientists** are among the necessary project members. The research project may be a clinical trial, observational study, or systematic review.