



# The levels of evidence in nutrition research

Last Updated : 06 September 2023

Nutrition and health are inextricably linked. Nutrition researchers try to unravel these connections in order to arrive at reliable nutritional advice. However, not all types of research can be used to draw equally firm conclusions. Understanding the different types of study designs is important for distinguishing between reliable and less robust findings. This article explores the various study designs commonly used in nutrition research, their purpose, how strong their evidence is and discusses the strengths and limitations of each design.

## Systematic reviews and meta-analysis

A single study is not enough to make a general statement with certainty about a certain link between nutrition and health. That is where systematic reviews and meta-analysis come in: in this type of research, researchers gather all relevant studies on a particular topic and analyse these collectively. As a result, the risk between a certain exposure/factor (e.g., overweight) and outcome/disease (e.g., cancer) can be estimated reasonably well.<sup>[1](#)</sup>

In a meta-analysis, results of multiple studies are pooled, following a rigorous protocol to find all the original research studies done on a question, and weighted with statistical methods into a single summary estimate. Large and well-conducted studies with high-quality evidence are given more 'weight' than small or poorly conducted studies with low quality data. A meta-analysis can only be carried out if the studies look at the same research question and use similar methods to measure relevant variables.

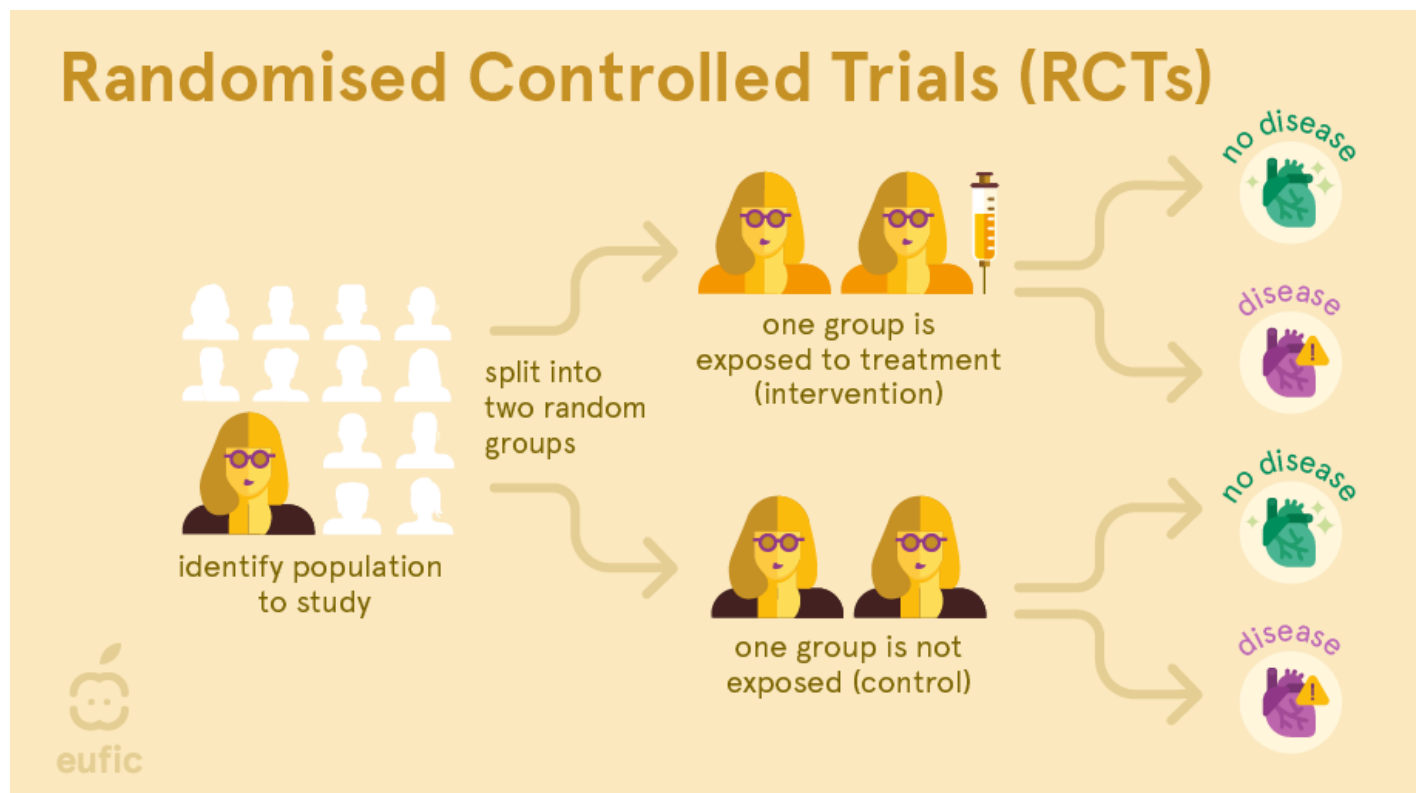
Systematic reviews are similar to meta-analysis, but without the use of a statistical analysis. Although systematic reviews and meta-analysis can reduce bias by pooling data from all relevant studies investigating a particular topic, they are only as good as the studies they include. It is important to check whether data from flawed studies are included or if there are data from studies that use different methods to measure variables – resulting in a comparison of 'apples and oranges.' To reduce the risk of

bias and improve reporting and transparency, using a set of guidelines called the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) is endorsed by most of the high-quality scientific journals.<sup>2</sup>

## Randomised controlled trials (RCTs)

A randomised controlled trial (RCT) is a type of intervention study where the researcher actively intervenes to change any aspect of nutrition to see what the effect this has on a certain health outcome.

In an RCT, a group of human subjects with same condition are identified and then randomly allocated to either receive the treatment (e.g., an omega-3 fatty acid supplement) or to a control group that does not receive the treatment (e.g., a 'placebo' supplement that looks identical to the original supplement but does not contain the substance being studied). After a defined time period, the effects in both groups are measured and compared to each other. As only one factor is deliberately changed between the groups (and other possible factors that influence the relationship are kept the same or as similar as possible), this type of research can allow us to identify [cause-and-effect](#) relationships.<sup>1</sup> These types of study are also used in medical research, for example to test the effects of new drugs or vaccines.



Preferably, RCTs are performed as a double-blind study: both the researcher and the participants are unaware who is in which group. This is important because a participant's response or a researcher's measurement of the outcome could be impacted by knowing who is being treated. For example, the placebo effect is a well-known phenomenon where a person observes an improvement in symptoms or effects in themselves after taking a fake or non-active 'treatment' (e.g., pills that don't contain any active ingredient). On top of that, true randomisation is important. If one group were in some way more ill (or less healthy) than the other group at the start, this might make this group appear to have worse outcomes, even if the administered treatment really had no effect.

RCTs also have limitations. They may not be suitable for answering certain research questions, such as the effect of whole diets (e.g., ketogenic, vegan) on the prevention of chronic diseases like cancer or cardiovascular disease. It would be impossible to control compliance (e.g., how strictly the study participants stick to the prescribed diet), a huge number of subjects would be required to show a significant difference in outcome, and it would cost a huge amount of money and time. Furthermore, interventions which deliberately expose participants to something thought to be harmful (e.g., alcohol, smoking, contaminants) or withholding participants from treatment thought to improve health (e.g., certain antibiotics or chemotherapeutic agents) in order to definitely prove a cause-and-effect relationship raise ethical concerns. Time is also an important factor. The more robust and reliable studies are those lasting longer.

It is also important not to generalise results of RCTs too quickly. RCTs often have strict criteria of whom to include and exclude. If a study was only carried out on a specific group of people (e.g., middle-aged women with diabetes), the study may not be applicable to the wider population. Lastly, RCTs may often run for short time periods since they are expensive to carry out. So, they may not be able to tell us about the long-term effects of dietary patterns and changes.

## Observational research

Observational research involves simply observing the habits or behaviours in large groups of people to investigate the relationship between lifestyle factors and health outcomes. The researcher does not intervene in any way but compares the health outcomes of people who make different diet or lifestyle choices. These studies are used to identify [correlations](#) and develop hypotheses for further testing.<sup>1</sup>

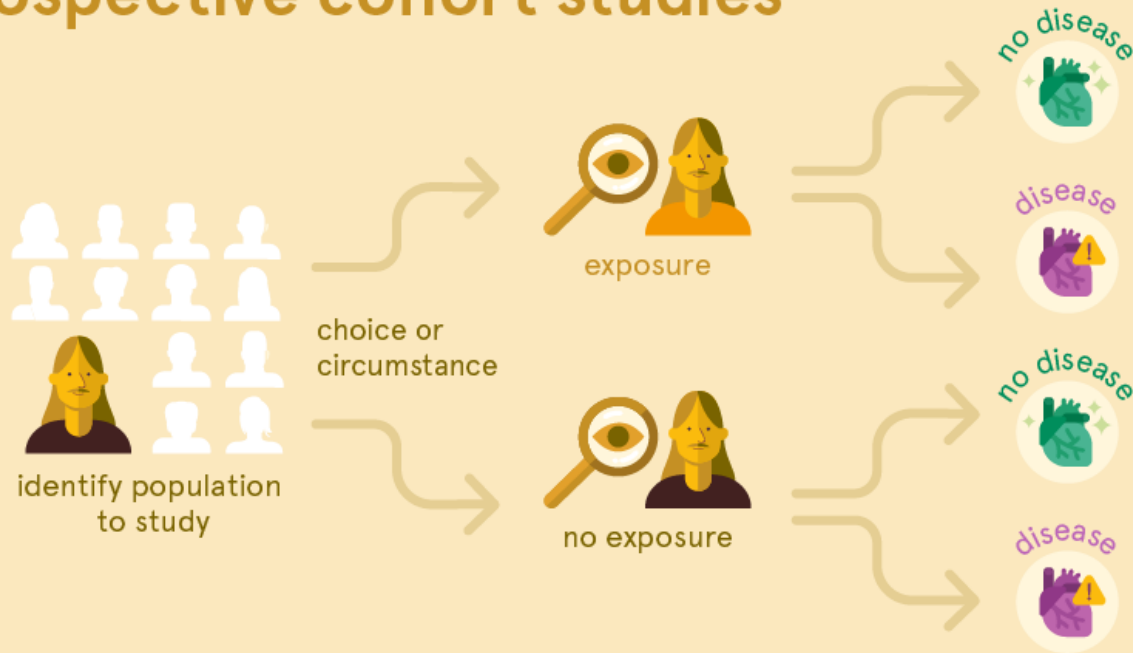
For example, researchers observe that people who drink alcohol are more likely to develop lung cancer than those who don't. However, it may also be that people who drink alcohol also tend to smoke more often and that researchers fail to include this factor into their analysis (e.g., it was not measured or not thought to influence the relationship). Here, smoking is a so-called confounding factor: a factor associated with both the exposure (drinking alcohol) as well as the outcome (lung cancer) and therefore could distort the data. In an RCTs these confounding factors are evenly distributed across the study groups (assuming randomisation is done correctly), but this is unlikely to be the case in observational research. Because of the presence of confounding factors, **observational research cannot prove cause-and-effect**. The interest of observational studies lies in revealing these relationships in order to address future research.

The next sections describe three common observational study designs: prospective cohort studies, case-control studies, and cross-sectional studies.

## Prospective cohort studies

A prospective cohort study is a study that follows a group of people over time. At the start of the study, researchers ask participants to complete a questionnaire (e.g., about their dietary habits, physical activity levels, etc.) and may also take measurements of weight, height, blood pressure, blood profile or other biological factors. Years later, the researchers look whether the participants have developed disease and study whether exposure from the questionnaire or biological measurements are associated with the disease.<sup>1</sup>

# prospective cohort studies



One of the main advantages of this design is that researchers observe participants' natural exposures and behaviours without intervention, providing insights into real-life scenarios. It also allows researchers to examine the long-term effects of nutrition or other lifestyle-related exposures on real diseases. Since chronic diseases, such as heart disease and osteoporosis, often take many decades to develop, a cohort may be more suitable compared to a RCT where often intermediate markers for these diseases are measured (e.g., narrowing of the arteries or bone density). These markers don't always develop into the disease.

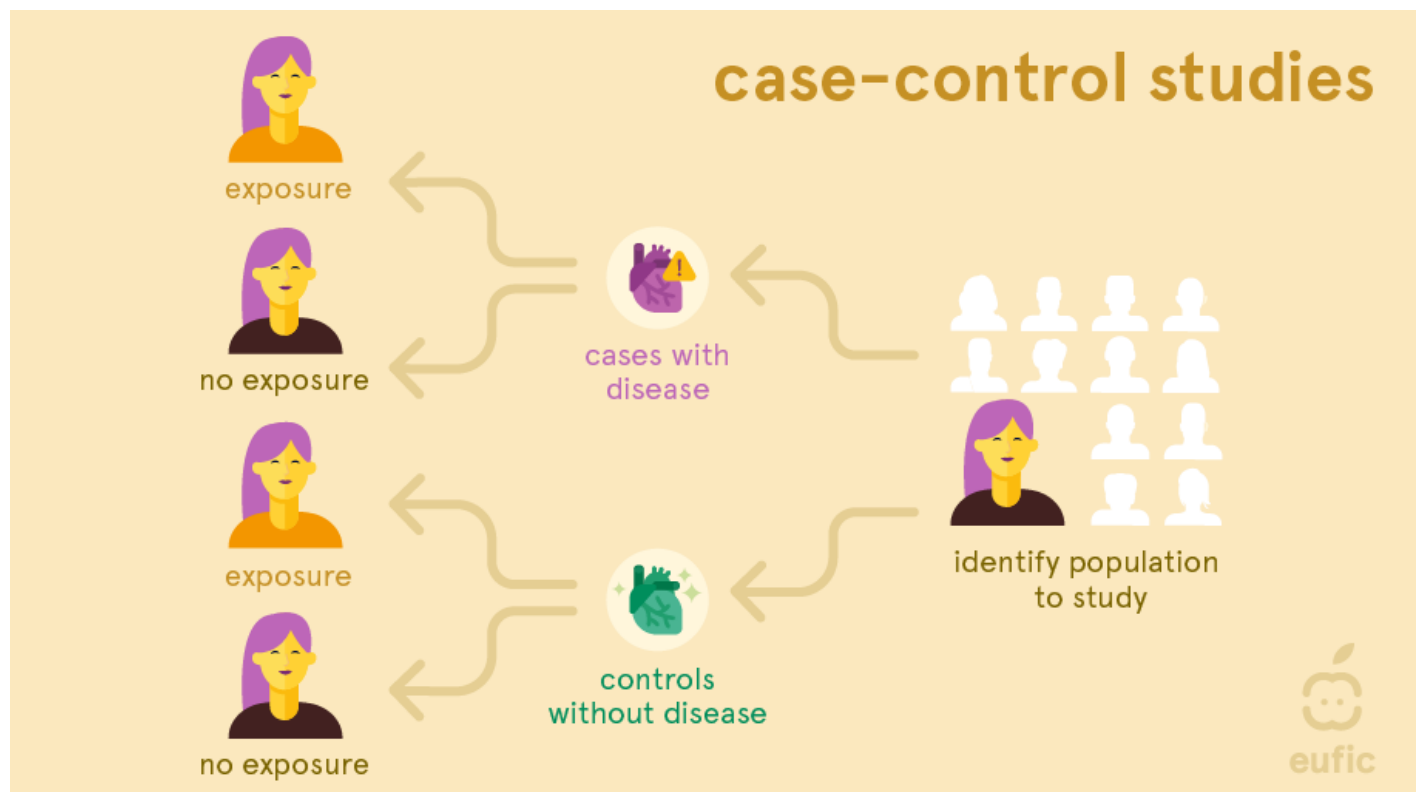
One of the main limitations of cohort studies is that they cannot establish causation definitively due to potential confounding factors. It is also important to consider how participants' food intake is measured. Cohorts often use what's called a Food Frequency Questionnaires (FFQs) which measures a person's average dietary intake over time. While FFQs are one of the best methods available to assess dietary intake, it can be hard to accurately estimate typical intake, portion size, and preparation methods. As data is self-reported, it may include subjective interpretations (e.g., participants may underreport, overreport or simply forgot their past habits: a problem called recall bias). FFQs also don't represent lifetime behaviour patterns: people may have changed their behaviour over the intervening years (e.g., smokers may quit smoking or meat eaters may become vegetarian), resulting in misclassification of participants and potential bias. When a validated FFQ is used, generally a part of bias is limited.

In a cohort study it is also important that participants are followed up for a long time to accumulate enough data to give robust results. This means they can generally only be used to study diseases that are relatively common. Another concern in cohort studies is selection bias: those selected to be in the study differ from those not selected in some systematic way. Recruitment of participants may be done, for example, through newspapers, phone dialling, the workplace or volunteering, impacting who takes part in the study and how generalisable the results are (i.e., newspapers are often only read by older populations, phone dialling excludes those without phones, volunteering recruits more health-conscious participants, etc.). Another concern occurs if many participants are 'lost to follow-up' (i.e., drop out of

the study) in one exposure group than another, particularly if loss is also related to the outcome being studied.

## Case-control studies

In a case-control study researchers look to the past of people with a disease (called the 'cases') and compare it to people without the disease (the 'controls'). These studies are most often used to study the link between an exposure and a rare outcome.<sup>1</sup> They usually have a smaller sample size than cohort studies and do not require follow-up.



As in cohort studies, recall bias is a problem in case-control studies. This is to even larger extent as people already have the disease of interest when the exposure information is collected or measured and so they might recall their exposure differently from people without disease. Selection bias (cases and/or controls may not be representative for the general population), confounding, and reverse causation can be limitations. Reverse causation occurs when it is challenging to identify if the outcome or exposure came first. For example, if an association was found between the consumption of non-sugar sweeteners and obesity, is this truly because of the higher consumption of non-sugar sweeteners or is it that people with obesity more frequently consumed products containing non-sugar sweeteners to help manage their weight? The choice of an appropriate control group is also one of the main difficulties of this type of study. Control groups should be carefully selected to be similar to the cases (those with the condition of interest) in all aspects except the exposure being studied (e.g., diet). This ensures that any observed difference in outcomes can be attributed to the exposure rather than confounding factors.

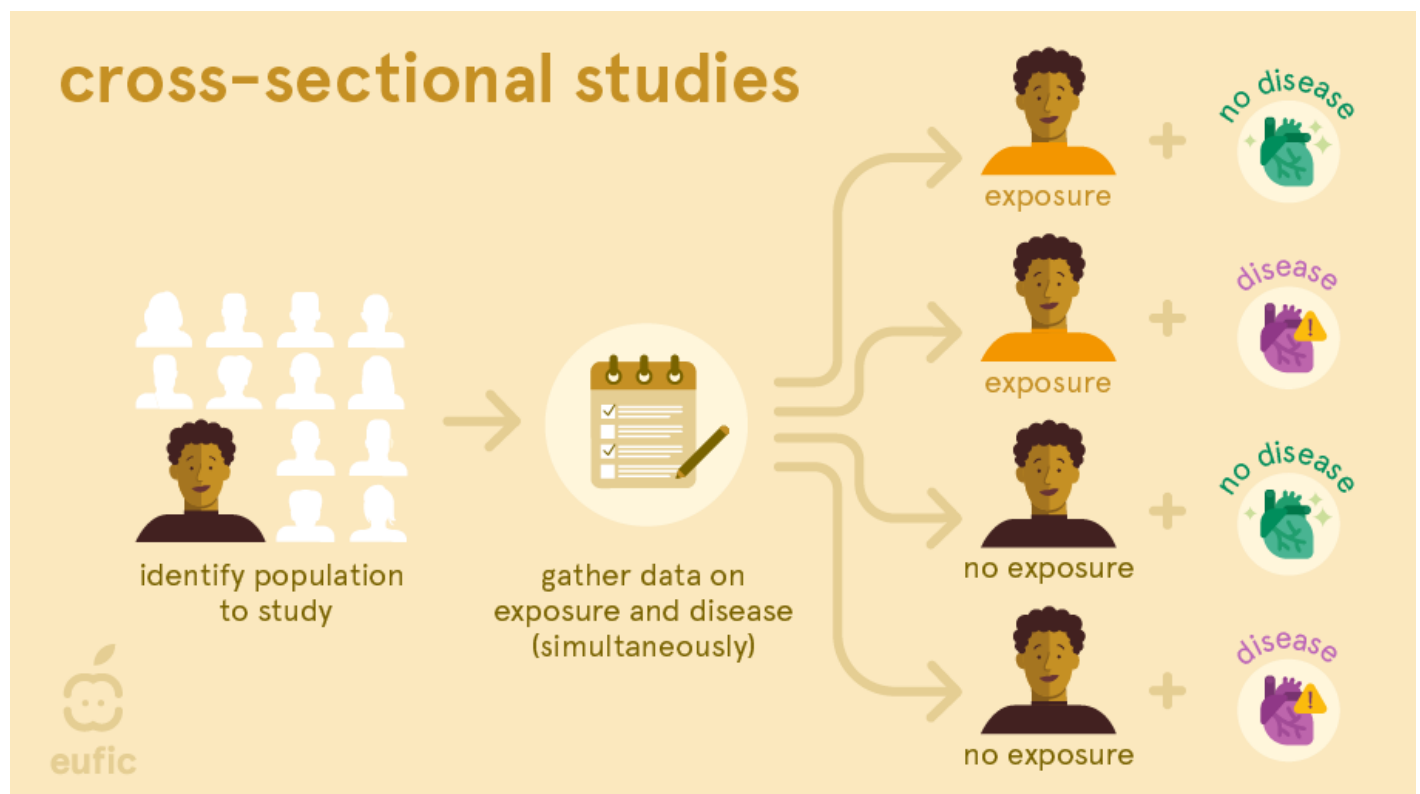
## Cross-sectional studies

A cross-sectional study is a survey or cross-section of a random sample of the population where

information about potential exposures and outcomes is collected at the same time. For example, researchers measure blood pressure and ask questions about, for example, the amount of processed meat each person eats per day. This lets them find out whether there is a link between blood pressure and the amount of processed meat consumed per day.<sup>1</sup>

With cross-sectional studies reverse causality is again a problem: you cannot be sure whether eating processed meat affects blood pressure or vice versa, because the information was obtained at the same time. Like cohort studies, they can also be prone to selection and recall bias. Recall bias may be a particular problem, since participant's knowledge of their health status may influence their reporting of dietary habits (e.g., a person with type 2 diabetes may recall they ate more sweets and sodas than a person without the disease).

Instead, this simple study design can be useful to investigate the possible causes of ill-health at an early stage, examine exposure that do not change over time (e.g., sex, genetic factors) or that occurred many years previously, or estimate the prevalence of dietary habits and health outcomes in a population at a specific time point. They can provide a starting point for further investigation about associations between dietary factors and health outcomes in, for example, a cohort study or RCT.



## Animal and cell studies

Animal and cell (or sometimes called in vitro) studies may provide an indication of the likely effect, however, they cannot be directly applied to humans. Research with animals is an important tool in determining how humans may react when exposed to particular substances. However, because of differences in physiology and the fact that animals are routinely exposed to far higher levels of compounds than typical dietary intakes, for example, results cannot be directly applied to humans.<sup>3</sup> Similarly, isolated cells in a laboratory behave differently than cells in our body. For example, if a test tube shows that substance X causes a cell to burn fat faster, that does not mean that substance X will

help to lose weight in humans. The human body is much more complex than can be imitated in a test tube.

For research into toxic substances, this type of research is the norm. Testing harmful or possibly toxic compounds on humans is dangerous and unethical. Animal testing is therefore used to establish safety guidelines for chemical compounds such as pesticides and environmental contaminants. Because results cannot be extended to humans and people also differ from each other, wide safety margins are used. However, the use of laboratory animals is being substantially reduced following international protocols such as those by the Organization for Economic Cooperation and Development (OECD).<sup>4</sup>

Animal and cell research can complement evidence from observational and experimental research: they can show if there is a mechanism that explains these results. For example, observational research shows that smoking is associated with cancer, while cell studies highlight the specific harmful substances present in tobacco that contributes to the development of cancer. The certainty that a result is accurate increases when there is such a logical explanation.

## **Anecdotes and case studies**

Anecdotes, case reports (on 1 patient) and case series (on several patients) describe a detailed report of individual patient(s) with a specific outcome and/or exposure.<sup>5</sup> They are important for the early identification of health problems and can generate hypotheses about potential causes. However, since they involve a limited number of people, they cannot be generalised to broader populations. A single person's experience or opinion does not provide an objective picture. Therefore, anecdotes and case studies are regarded as low-quality evidence.

## **What is considered the 'best' evidence?**

Generally, the different types of research are organised from single person's experiences and anecdotes having the weakest certainty of evidence to systematic reviews and meta-analysis having the strongest level of evidence. These levels of evidence can be used as a guideline to judge what can be concluded from a particular study. However, they do not substitute for critical appraisal.<sup>6</sup> For example, a strong cohort study may be more useful than a flawed systematic review. Besides, which type of research scientists choose to conduct depends on, among other things, the research question, the amount of time available and the amount of money. Therefore, evidence may be better sorted by its usefulness for investigating a specific research question than by type of study design.

Different types of study designs should be viewed as complementary. For example, observational research can still be meaningful and illuminating when numerous studies consistently show patterns on a large scale.



# How strong is the scientific evidence?



RCTs are often regarded as the 'gold standard' for conducting research and their findings are believed to be more accurate compared to observational research as they can establish cause-and-effect relationships. However, this assumption is not always valid because the intervention/exposure being studied in RCTs may differ from those in observational studies.<sup>6</sup> For example, dietary intakes in observational studies are not interchangeable to some exposures used in RCTs (e.g., intake of omega-3 fatty acids by eating fish is different to omega-3 fatty acids consumed in isolated supplemental form). As a result, it is not surprising that sometimes contradictory results are found between observational and experimental research.<sup>7</sup> When testing findings from observational research further in RCTs, it is therefore important to carefully consider the population being studied, the way the intervention (dietary change) is applied, the comparison group, and the outcome(s) measured. Even small differences in how the study is conducted can lead to varying results.

## Conclusion

Nutrition research is expensive and complex to develop. It is therefore difficult to reach to reliable results that support evidence. A single approach is not sufficient. There are a variety of study designs used in nutrition research which are used to study a variety of different exposures and outcomes. How all these studies can lead to a conclusion depends on the certainty of evidence. A link between an exposure and outcome is more certain if:<sup>8</sup>

- Many long-term studies that follow people over time consistently find a link between exposure to A and the risk of B.
- The studies are generally large, last for a long time, and are well-designed.
- Only a few studies show different or opposite results.
- When possible, controlled experiments (like randomised trials) have also been done to test the findings.



- There's a clear biological explanation for how A could cause B.

In contrast, there is insufficient evidence if:

- There are only a small number of studies suggesting that there is a link between exposure to A (cause) and the risk of B (effect);
- The link found is weak;
- No or insufficient experimental and observational studies have been done and therefore more research is needed.



**Co-funded by  
the European Union**

## References

1. Webb P, Bain C & Page A (2017) Essential epidemiology: an introduction for students and health professionals. Cambridge University Press.
2. [PRISMA. \(2023\). Transparent reporting of systematic reviews and meta-analyses. Retrieved from http://www.prisma-statement.org/?AspxAutoDetectCookieSupport=1](http://www.prisma-statement.org/?AspxAutoDetectCookieSupport=1) (Accessed 05/09/2023)
3. Van der Worp HB et al. (2010) Can animal models of disease reliably inform human studies? PLoS Medicine, 7(3):e1000245.
4. OECD. (2023). Animal Welfare. Retrieved from <https://www.oecd.org/chemicalsafety/testing/animal-welfare.htm> (Accessed 05/09/2023)
5. Mathes T & Pieper D (2017) Clarifying the distinction between case series and cohort studies in systematic reviews of comparative studies: potential impact on body of evidence and workload. BMC medical research methodology, 17:1-6.
6. Flanagan A et al. (2023). Need for a nutrition-specific scientific paradigm for research quality improvement. BMJ Nutrition, Prevention & Health e000650
7. Schwingshackl L et al. (2021) Evaluating agreement between bodies of evidence from randomised controlled trials and cohort studies in nutrition research: meta-epidemiological study. British Medical Journal 374:n1864.
8. [World Health Organization. \(2014\). WHO Handbook for Guideline Development. 2nd edition. Retrieved from https://www.who.int/publications/i/item/9789241548960](https://www.who.int/publications/i/item/9789241548960) (Accessed 02/08/2023)