

Research Problems and Process

Formulating the Problem · Hypothesis · Research Design · Experimental Design

Every piece of formal research follows a common logical sequence: a broad area of interest is narrowed into a precise research problem, the problem is expressed as research questions and objectives, a testable hypothesis is proposed, and a research design and (where relevant) an experimental design are chosen to gather the evidence needed to test it. These notes work through that sequence step by step.

1. Identification and Formulation of the Research Problem

1.1 What is a research problem?

A **research problem** is a clearly stated area of concern, a gap in existing knowledge, or a difficulty that a researcher wants to resolve through systematic enquiry. It is the starting point of every study: until the problem is identified and stated precisely, it is impossible to decide what data to collect or how. Problems are typically identified from the researcher's own field experience, from unresolved questions or contradictions noticed while reviewing published literature, from theoretical gaps, or from a practical difficulty that needs a solution.

1.2 Formulating the problem

Formulation is the process of translating a broad, loosely defined area of interest into a specific, workable research problem. It generally proceeds from a broad subject area, through a literature survey and discussion with experts to locate a genuine gap, to a precisely defined and delimited problem statement that specifies the population, the variables involved, and the scope of the study. A well-formulated problem is then expressed as a set of research questions and objectives, which in turn lead to a testable hypothesis.

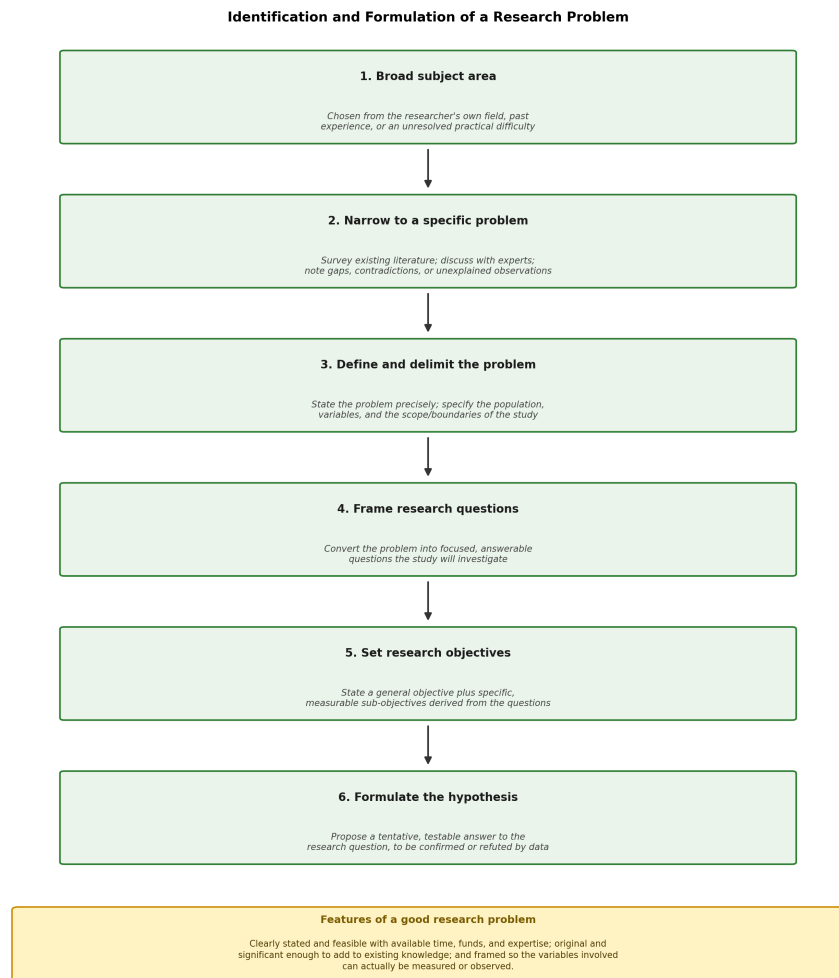


Fig. 1 - The research process from a broad subject area to a formulated hypothesis, and the features that mark out a good research problem.

1.3 Research questions and research objectives

A **research question** is a focused, answerable question derived directly from the research problem; it states precisely what the study intends to find out (for example, "does treatment with compound X reduce fungal infection rate in stored grain?"). A **research objective** restates that question as a goal the study will pursue - typically one broad, general objective capturing the overall aim, supported by several specific, measurable sub-objectives that break the general aim into concrete, achievable steps. Good objectives are specific, measurable, achievable with the resources available, and directly traceable back to the research questions they answer.

Key points

- The sequence runs broad area → specific problem → research questions → research objectives → hypothesis - each step narrows and sharpens the one before it.
- A good research problem is feasible within the available time, funds and expertise; original and significant enough to add to existing knowledge; and stated so its variables can actually be measured or observed.
- Objectives should be specific and measurable - vague objectives ('to study X') cannot later be checked off as achieved or not achieved.

2. Hypothesis: Meaning, Role, and Types

2.1 Meaning and role

A **hypothesis** is a tentative, testable statement that predicts the relationship expected between two or more variables, formulated before data are collected. It plays a central role in research: it converts a general research question into a specific, checkable claim; it guides the researcher in deciding what data to collect, from whom, and by what method; and it provides the standard against which the collected evidence is judged - the hypothesis is ultimately either supported or rejected by the data, but strictly speaking is never itself finally "proved" true.

2.2 Types of hypothesis

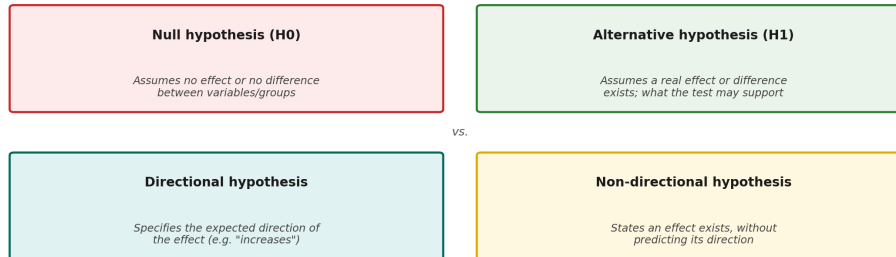
By statistical role, every hypothesis test involves a pair of complementary statements: the **null hypothesis (H0)**, which assumes no real effect or no difference exists, and the **alternative hypothesis (H1)**, which assumes a real effect or difference does exist and is what a significant test result is taken to support. By the specificity of the prediction, a hypothesis may be **directional**, stating the expected direction of an effect (e.g. "increases" or "is greater than"), or **non-directional**, merely asserting that some effect or difference exists without predicting which way it runs.

What a Hypothesis Is, and What It Does

A hypothesis is a tentative, testable statement predicting the relationship between two or more variables, put forward before data collection begins.

Role: it sharpens a vague research problem into a specific, checkable claim, guides which data are collected and how, and is ultimately either supported or rejected by the evidence - it is never itself "proved" true.

Types of Hypothesis, by Statistical Role and by Direction



A Priori vs A Posteriori Hypothesis Testing

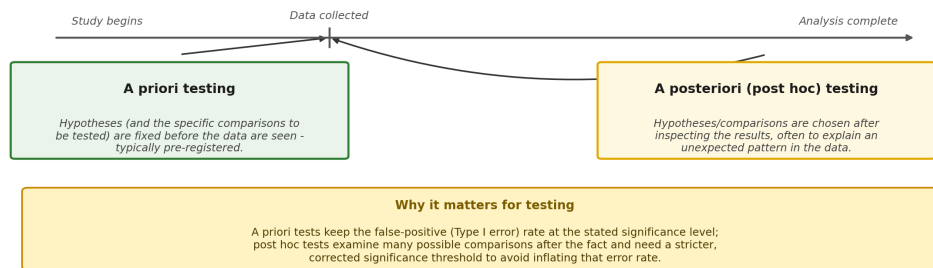


Fig. 2 - What a hypothesis is and does; its main types by statistical role and by direction; and the distinction between a priori and a posteriori hypothesis testing.

2.3 A priori and a posteriori hypothesis testing

This distinction concerns *when*, relative to seeing the data, a hypothesis or comparison is decided on. In **a priori (planned) testing**, the specific hypotheses and comparisons to be tested are fixed before the data are collected or examined - often formally pre-registered - based on theory or prior evidence. In **a posteriori (post hoc) testing**, hypotheses or comparisons are chosen only after the results have been inspected, typically to explain a pattern that was not predicted in advance (for example, running pairwise comparisons between several groups only after an overall ANOVA turns out significant).

The distinction matters for the validity of the statistical conclusion. An a priori test keeps the chance of a false positive (Type I error) at the stated significance level, because only one or a few pre-specified comparisons are made. A posteriori tests, by contrast, are typically drawn from many possible comparisons that could have been made once the data are in hand, so the probability of at least one false positive rises sharply unless a stricter, corrected significance threshold (such as a Bonferroni or Tukey correction) is applied.

Worth noting

- "A priori" and "a posteriori" are also classical philosophical terms for knowledge that is independent of experience versus knowledge derived from experience; in research methodology and statistics the terms instead describe the timing of a decision relative to the data, as above.
- Post hoc comparisons are a legitimate and common part of exploratory data analysis - the caution is not against using them, but against reporting them, uncorrected, as if they were planned confirmatory tests.

Key points

- H_0 assumes no effect; H_1 assumes a real effect - a significance test never proves H_0 true, it only fails to reject it.
- Directional hypotheses commit to a predicted direction of effect and are tested with a one-tailed test; non-directional hypotheses are tested with a two-tailed test.
- A priori tests are planned before the data are seen; a posteriori (post hoc) tests are chosen after seeing the data and require a stricter, corrected significance threshold.

3. Research Design: Meaning, Significance, and Types

3.1 Meaning and significance

A **research design** is the overall plan, structure, or blueprint that specifies how data relevant to the research problem will be collected, measured, and analysed. It is significant because it keeps a study focused and feasible within its available resources; it minimises bias and error by fixing the procedure in advance; it ensures the data actually collected are capable of answering the research question that was asked; and it makes efficient use of the researcher's time, funding, and effort by avoiding wasted or irrelevant data collection.

3.2 Types of research design

Research designs are commonly grouped into four broad types, distinguished chiefly by how much is already known about the problem and whether variables are actively manipulated. **Exploratory** designs are used when little prior research exists; they aim to generate insight, ideas, and testable hypotheses rather than final answers. **Descriptive** designs systematically describe the characteristics of a population or phenomenon as it naturally occurs, answering who, what, where and when, without manipulating any variable. **Analytical (diagnostic)** designs use facts or data already available to diagnose the causes of a problem and critically analyse relationships between variables. **Experimental** designs go further still: the researcher actively manipulates one or more independent variables under controlled conditions specifically to test a cause-and-effect hypothesis.

Research Design: Meaning and Significance

A research design is the overall plan or blueprint for collecting, measuring, and analysing data to answer the research questions.

Significance: it keeps the study focused and feasible, minimises bias and error, ensures the data collected can actually answer the research question, and makes the most efficient use of the researcher's time, money, and effort.

Four Basic Types of Research Design

Exploratory

Used when little is known about a problem; generates insight, ideas, and testable hypotheses for later study.

Descriptive

Systematically describes the characteristics of a population, situation, or phenomenon (who, what, where, when).

Analytical (diagnostic)

Uses facts or existing information already available to diagnose causes and critically analyse relationships between variables.

Experimental

The researcher actively manipulates one or more variables under controlled conditions to test cause-and-effect hypotheses.

Fig. 3 - The meaning and significance of research design, and its four basic types.

Key points

- Research design is chosen before data collection begins and should follow directly from the research objectives and hypothesis already formulated.
- Exploratory and descriptive designs do not manipulate variables; analytical designs work from existing data; only experimental designs test cause-and-effect directly through manipulation.
- The choice of design shapes every later decision - sampling, data collection instruments, and the statistical analysis that will be valid on the resulting data.

4. Experimental Design and the Concept of Replicates

4.1 Informal experimental designs

Informal experimental designs rely on simple comparison of magnitudes rather than formal statistical procedures, and offer comparatively little control over extraneous variables. Three common forms are the **before-and-after design without control**, in which a single group is measured before and after a treatment and the difference is taken as the treatment effect (though other changes over the intervening time cannot be ruled out); the **after-only design with control**, in which a treated group is compared, at a single point in time, against an untreated control group; and the **before-and-after design with control**, which measures both groups both before and after treatment and takes the treatment effect as the difference between the two groups' changes - the strongest of the three informal designs, because it controls for both the passage of time and pre-existing differences between groups.

4.2 Formal experimental designs

Formal experimental designs build replication and randomization into the design itself and are analysed with precise statistical procedures such as analysis of variance (ANOVA). In the **completely randomized design (CRD)**, treatments are assigned to experimental units entirely at random, and it is the simplest formal design, but it is best suited to otherwise homogeneous experimental material. The **randomized block design (RBD)** adds "local control": units are first grouped into blocks that are internally as uniform as possible, and every treatment appears exactly once in every block, which removes block-to-block variation from the estimate of treatment effect. The **Latin square design (LSD)** extends this idea to control two sources of extraneous variation at once, arranging treatments in a grid so that each treatment occurs exactly once in every row and exactly once in every column.

Informal Experimental Designs

Before-and-after (no control)

One group, measured before and after treatment. Effect = After - Before. Cannot rule out other changes over time.

After-only, with control

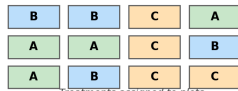
Treatment applied to one group only; compared once against an untreated control group measured at the same time.

Before-and-after, with control

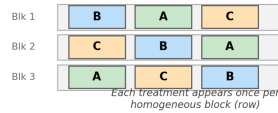
Both groups measured before and after; treatment applied to one only. Effect = (Test change) - (Control change).

Formal Experimental Designs

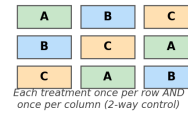
Completely randomized (CRD)



Randomized block (RBD)



Latin square (LSD)



The Concept of Replicates

A replicate is one of several independent experimental units that receive the same treatment, allowing natural variation among units to be measured.

Replication (not repeated measurement of the same unit) estimates "experimental error" - the background variability against which a treatment effect is judged statistically significant - and increases the precision and reliability of the estimated treatment effect.

Treatment A, applied to:



Informal vs formal: the key difference

Informal designs compare group magnitudes directly and offer little control over extraneous variables. Formal designs build in replication, randomization, and (in RBD/LSD) local control of known sources of variation, and are analysed with precise statistical procedures such as ANOVA.

Fig. 4 - Informal experimental designs; formal designs shown schematically (CRD, RBD, and Latin square); and the concept of replicates.

4.3 The concept of replicates

A **replicate** is one of several independent experimental units - separate plants, animals, plots, culture dishes, or samples - that receive exactly the same treatment. Replication is not the same as simply measuring one unit repeatedly: it means applying a treatment independently to multiple units so that the natural biological variation between them can be captured. This variation among replicates gives an estimate of **experimental error** - the background, uncontrolled variability against which a treatment difference is judged, statistically, to be significant or not. Without adequate replication, a real treatment effect cannot be distinguished from ordinary biological variability, and the study's statistical power to detect a genuine effect is low.

In biological research specifically, replicates are essential precisely because living material is inherently variable - no two plants, animals, or cell cultures respond identically even under identical conditions. Applying a treatment to several independent replicates (rather than to a single individual, however carefully measured) is what allows a researcher to report a treatment mean together with a valid measure of its variability (such as a standard error), and to test formally whether an observed difference between treatments exceeds what chance variation alone would produce.

Worth noting

- Biological replicates (independent organisms, plots, or cultures) and technical replicates (repeated measurements of the same biological sample) answer different questions - only biological replicates allow conclusions about the population the samples were drawn from.
- More replicates generally increase statistical power, but the number needed also depends on the size of the effect being looked for and the amount of natural variability in the material - this trade-off is usually formalised through a power analysis before the experiment is run.

Key points

- Informal designs (before-after, after-only-with-control, before-after-with-control) compare magnitudes directly; formal designs (CRD, RBD, Latin square) build in replication, randomization, and statistical control of known variation sources.
- RBD controls one source of extraneous variation via blocking; the Latin square design controls two sources at once, at the cost of requiring equal numbers of rows, columns, and treatments.
- Replication estimates experimental error and is what makes a statistical test of significance possible in the first place - it is distinct from simply repeating a measurement on one unit.

5. Quick Revision Summary

Concept	Essence
Research problem	A stated gap, difficulty, or unresolved question that research sets out to address
Formulation	Narrowing a broad area, via literature review, into a precise, delimited problem statement
Research question / objective	Focused question the study will answer; objective restates it as a measurable goal
Hypothesis	Tentative, testable prediction about a relationship between variables
Null / alternative hypothesis	H0: no effect; H1: a real effect exists - H0 is rejected or not, never "proved"
A priori testing	Hypotheses/comparisons fixed before the data are seen; standard significance level
A posteriori (post hoc) testing	Comparisons chosen after seeing the data; needs a corrected significance threshold
Research design	Overall blueprint for collecting, measuring and analysing data to answer the question
Exploratory / descriptive / analytical / experimental	Generate insight; describe as-is; diagnose from existing data; manipulate variables directly
Informal experimental designs	Before-after; after-only-with-control; before-after-with-control - simple comparisons
Formal experimental designs	CRD, randomized block (RBD), Latin square - built-in replication, randomization, control
Replicate	One of several independent units given the same treatment, to estimate experimental error

Notes prepared for study purposes. Consult current primary literature for details required in formal citation or research use.