EXCLI Journal 2026;25:1636-1673 – ISSN 1611-2156
Received: September 07, 2026, accepted: October 02, 2026, published: November 02, 2026

Scientific Forum:

MASQUERADING EXPLORATION AS CONFIRMATION: OPENING UP PANDORA'S BOX OF QUESTIONABLE RESEARCH PRACTICES

Asghar Ghasemi*

Endocrine Physiology Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, No. 24, Sahid-Erabi St., Yemen St., Chamran Exp., Tehran, Iran

* **Corresponding author:** Asghar Ghasemi, Endocrine Physiology Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, No. 24, Sahid-Erabi St., Yemen St., Chamran Exp., Tehran, Iran;
P.O. Box: 19395-4763; Phone: +98 (21) 224 32 500;
Fax: +98 (21) 224 16 264, 224 02 463; E-mail: Ghasemi@sbmu.ac.ir

https://dx.doi.org/10.17179/excli2026-9787

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/).

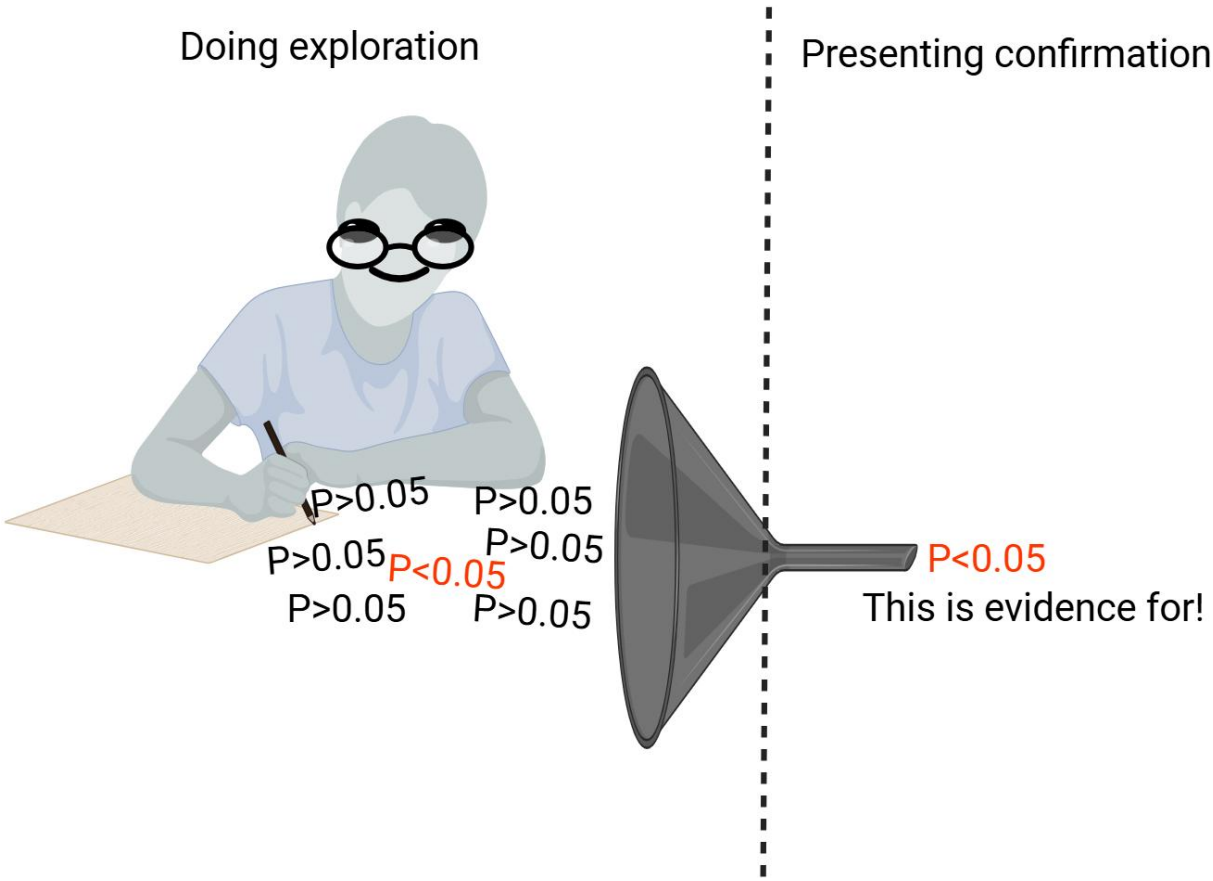

Figure 1: Graphical abstract

EXCLI Journal 2026;25:1636-1673 – ISSN 1611-2156
Received: September 07, 2026, accepted: October 02, 2026, published: November 02, 2026

ABSTRACT

Scientific knowledge generation can be conceptually divided into two complementary, distinct, and equally important phases: exploratory, in which hypotheses are generated to represent data patterns, and confirmatory (refutational), in which predictions based on these hypotheses are tested by collecting new data to provide evidence of practical/clinical utility. However, scientific publishing weights confirmatory research more heavily than exploratory research. The devaluation of exploratory research pressures authors to disguise their exploratory work as confirmatory, leading to questionable research practices (QRPs), including p-hacking (massaging data to yield statistically significant results), HARKing (hypothesizing after results are known), cherry-picking (selective outcome reporting), and fishing expeditions (data dredging). These bad research practices increase false-positive rates and jeopardize the credibility of scientific findings, leading to one of science's worst failings: the replication crisis. Radical reform in how science is conducted and published, including preregistration, redefining statistical thresholds, advancing research training, providing more room for exploratory research, and changing the wrong incentives in scientific practice, may partially solve the problem of blending and its associated QRPs. In conclusion, the distinction between exploratory and confirmatory research is crucial for better bench-to-bed translation of preclinical evidence. Tentative exploratory results must be confirmed before establishing clinical utility.

Keywords: Exploration, confirmation, blending, questionable research practices, P-hacking, HARKing, cherry-picking, fishing expedition

1. INTRODUCTION

Scientific research is a problem-oriented (Bahadoran et al., 2025a) and purpose-directed (Ghasemi et al., 2026) activity. Research activity can be divided into two main phases: (1) the hypothesis-generation phase and (2) the hypothesis-testing phase. Scientific inquiry not guided by a core theory is called exploratory research (postdiction) (Waters, 2007), which uses existing observations of nature (data) to generate ideas about how the world works (Nosek et al., 2018). Exploratory research is an integral part of scientific activity (Waters, 2007) and is vital to science (Kriegeskorte et al., 2010), as data exploration is essential and can lead to novel discoveries (Büttner et al., 2020; Kriegeskorte et al., 2010). The ideas generated in exploratory research are presented as hypotheses that can later be tested with independent data in confirmatory research (Jaeger and Halliday, 1998; Kriegeskorte et al., 2009; Szollosi and Donkin, 2021). The style of scientific inquiry that begins with a core theory is called confirmatory research (prediction) (Waters, 2007) and involves having a theory-laden, *a priori* (established beforehand) idea (hypothesis) about how the world works (Gaus et al., 2015; Nosek et al., 2018). In confirmatory research, theory generates expectations (hypothesis-based predictions) about what will be observed (Waters, 2007), and researchers make new observations (collect new data) to test whether that idea is a reasonable explanation (Nosek et al., 2018), a process called hypothesis testing (Szollosi and Donkin, 2021).

Most research in medicine aims to determine the relations between variables (Daly and Bourke, 2000); such a quest may be theory-driven and thus adhere to a prespecified plan (confirmatory research) or may be primarily data-driven (exploratory research) without following a restricted prespecified plan (Höfler et al., 2023). Despite arguments that the classical model of scientific inquiry is hypothesis-driven (i.e., confirmatory research) (Ghasemi et al., 2025; Yanai and Lercher, 2019), a broad-practice-centered approach (i.e., exploratory research) has been introduced as an epistemology of scientific inquiry (Waters, 2019). Exploratory and confirmatory research are two distinct and complementary approaches to knowledge generation and science (Höfler et al., 2022; Kimmelman et al., 2014; Nosek et al., 2018; Schickore, 2016). Exploratory research provides insights for new or improved scientific claims (Höfler et al., 2023), which confirmatory research supports when they make correct predictions on new data (Höfler et al., 2022). Therefore, transparent exploration and genuine confirmation benefit from one

another (Höfler et al., 2022, 2023), and when they coexist in a good balance, science is most productive (Höfler et al., 2022).

Distinguishing between exploratory and confirmatory research is vital to the health of science (Gaus et al., 2015; Höfler et al., 2022) and will improve translation from preclinical to clinical settings (Kimmelman et al., 2014). A sharp boundary between exploration and confirmation is needed to separate established from new scientific claims (Höfler et al., 2022). If exploratory research is intentionally or unintentionally falsely reported as confirmatory (referred to as nominally confirmatory results, pseudo-confirmatory results, or flawed confirmatory research), it misleads scientific inquiry, wastes resources, and harms human health (Anderson et al., 2001; Büttner et al., 2020; Höfler et al., 2022; Wagenmakers et al., 2012). First, it opens the door for questionable research practices (QRPs), including p-hacking (massaging data to yield statistically significant results), HARKing (hypothesizing after results are known), cherry-picking (selective outcome reporting), and fishing expeditions (data dredging) (Büttner et al., 2020; McIntosh, 2017; Wagenmakers et al., 2012). Second, it biases the scientific literature (Agnoli et al., 2017) by selling artifacts in a particular dataset (spurious effects) rather than genuine effects (Anderson et al., 2001). In other words, it can lead to overconfidence in *post hoc* explanations (postdictions) and inflate the likelihood of believing that there is evidence for a finding when there is not (Nosek et al., 2018). Finally, mistaking exploration for confirmation underestimates outcome uncertainty, increases false-positive results (results arising from chance) (Büttner et al., 2020), and thereby decreases reproducibility (Nosek et al., 2018) and contributes to the replication crisis (Colquhoun, 2014; Kennedy, 2026). Indeed, most published research findings have been reported as false-positive (Ioannidis, 2005).

Both direct evidence (researchers' admission of QRP) and indirect evidence (the replication crisis, more negative results in reports than in published articles, and a high number of published studies framed as confirmatory) indicate that blending is common (Höfler et al., 2022). Blending confirmatory and exploratory research may arise from a wrong incentive atmosphere of publishing in which positive results are rewarded more than negative findings (Nosek et al., 2018) and quantity is valued over quality (Allen and Mehler, 2019; Colquhoun, 2014), causing researchers to play the game to maximize their publications (Gigerenzer, 2018), as well as ineffective training in good statistical and methodological practice (Gigerenzer, 2018; Nosek et al., 2018), which in most cases teaches statistical ritual instead of statistical thinking (Gigerenzer, 2018). Because research training can partially address poor scientific practices (Andrade, 2021; Horton, 2015), this review clarifies the concepts of exploratory and confirmatory research, highlights their positions and differences as complementary modes of scientific inquiry, and summarizes recommendations to avoid blending them.

2. GOAL-ORIENTED CLASSIFICATION OF RESEARCH: EXPLORATORY, DESCRIPTIVE, AND CONFIRMATORY

The research goal may be to refine an unfamiliar problem about which little is known (exploratory research), to characterize a previously known problem (descriptive research), or to tackle problem-solving (confirmatory research) (Ghasemi et al., 2026). Exploration refers to a mode of scientific inquiry involving free and flexible data examination (Ditroilo et al., 2025) to learn about a new and under-researched topic (Leavy, 2023) for which there is no covering theory (Waters, 2019), to refine an unfamiliar problem, or "*to see what is there?*" (Andrew et al., 2011; Jaeger and Halliday, 1998; Onwuegbuzie and Leech, 2006; Swaraj, 2019). Exploration probes the unknown and unforeseen relationships/patterns (Ditroilo et al., 2025; Due, 2022; Fife and Rodgers, 2022) in previously collected data (Nosek et al., 2018) to generate new hypotheses/modify previous hypotheses about why something occurred (Höfler et al., 2022; Nosek et al., 2018). The core benefit of exploration is "the potential to find the new, whenever

it might be hidden, whatever it might look like" (Höfler et al., 2022), which makes exploration vital for discovering possibilities not yet considered (Nosek et al., 2018). Overall, exploration develops theory from data (Stebbins, 2001).

Confirmation refers to a mode of scientific inquiry in which a theory-laden *a priori* hypothesis is formulated by deduction (Ghasemi et al., 2025; Jaeger and Halliday, 1998). Then, according to a prespecified plan, researchers collect data to empirically and statistically test predictions based on this hypothesis, in accordance with an evidential norm (Ditroilo et al., 2025; Höfler et al., 2022; Nosek et al., 2018), to confirm whether evidence supports or refutes the claim (Höfler et al., 2022; Kennedy, 2026) for its practical/clinical utility (Kimmelman et al., 2014). Testing predictions is vital for establishing evidence for explanatory claims (Nosek et al., 2018). The way in which confirmatory research produces knowledge is called the hypothetico-deductive theory of science (Ghasemi et al., 2025). Confirmatory research includes three types: predictive, explanatory, and evaluative research. Prediction uses pre-existing knowledge or theory to forecast future events (Andrew et al., 2011; Onwuegbuzie and Leech, 2006). Explanatory research (also called etiological research when it investigates the causes of disease (Luijken et al., 2022)) seeks to establish the causal relationship between a specific variable (cause or risk factor) and an outcome (effect), using pre-existing knowledge (van Diepen et al., 2017). Evaluation is the operation that associates symbolic entities (i.e., values) with things under evaluation (Mari, 2003).

Descriptive studies represent "*the first scientific toe in the water*" in scientific inquiry (Grimes and Schulz, 2002) that collect data to characterize the problem (Pangaro and McGaghie, 2001) or a population (Andrew et al., 2011). Descriptive studies answer five basic W questions (Grimes and Schulz, 2002): ***who*** has the condition (e.g., age and sex), ***what*** is the condition being studied (e.g., clear case definition), ***why*** the condition arises (providing clues about causes to pursue with other studies), ***when*** the condition is common or rare (information about time), and **where** the condition does or does not arise (e.g., latitude plays a role in vitamin D deficiency). Descriptive studies also answer an implicit question: So what? (that is related to public health effect). Descriptive studies are used for trend analysis, health-care planning, and hypothesis generation (Grimes and Schulz, 2002). As the name implies, descriptive studies do not test hypotheses (Ditroilo et al., 2025; Grimes and Schulz, 2002), and they involve no comparisons (Pangaro and McGaghie, 2001). These studies do not rely on inferential statistics and report descriptive statistics (Ditroilo et al., 2025). Descriptive studies can provide valuable insights (Ditroilo et al., 2025), including the ability to herald new illnesses (e.g., blindness in newborn infants and high oxygen levels in incubators) or to monitor health (Grimes and Schulz, 2002). However, caution is warranted when interpreting descriptive study results. For example, based on misinterpretation of descriptive studies, it was assumed that routine electronic fetal monitoring can decrease the perinatal mortality rate; however, subsequent clinical trials showed that it actually harms women, as it increases operative deliveries (Grimes and Schulz, 2002). Descriptive questions use both quantitative and qualitative approaches and may deal with individuals (case reports, case series, cross-sectional studies, and surveillance) or populations (ecological correlational studies) (Grimes and Schulz, 2002).

3. COMPARING EXPLORATORY AND CONFIRMATORY RESEARCH

3.1. General characteristics of exploratory and confirmatory research

Exploratory and confirmatory research have been described in different ways (Table 1); among these, night science for exploratory research and day science for confirmatory research (coined by Nobel laureate Francois Jacob) may better contrast these two types of scientific inquiry (Yanai and Lercher, 2019). In the logical steps of systematic day science (where ideas are formulated as hypotheses and tested), only ideas related to the current hypothesis can be

reached. However, in more open, creative night sciences (where ideas are born), back-and-forth between ideas is possible (Yanai and Lercher, 2019, 2021). Table 1 compares overall characteristics of exploratory and confirmatory research, including (1) position in the research cycle phases (observation → hypothesis → experimentation → verification (Nosek et al., 2012)); (2) logic of scientific inquiry and theoretical framework of research; and (3) the nature of the study. An example of quantitative exploratory research is genomic research in which thousands of relationships among many molecular markers are assessed (Goeman and Solari, 2011). An example of qualitative exploratory research is examining the construct of patients' reasons for choosing complementary medicine over conventional medicine (Lewith and Chan, 2002). An example of confirmatory research is assessing a drug's efficacy in humans (Schwab and Held, 2020). Indeed, phase III clinical trials in which the research question can be phrased as a straightforward hypothesis and previous knowledge supports an intervention's efficacy and safety are confirmatory research (Dirnagl, 2020; Ioannidis, 2005; Schwab and Held, 2020).

3.1.1. Position in the research cycle phases

In the research cycle (exploration → description → confirmation), exploratory research is in the initial phase (Ditroilo et al., 2025; Goeman and Solari, 2011) to identify regularities and develop new concepts and provide direction for further research (Swaraj, 2019); descriptive research has an intermediate position to identify the nature and composition of a phenomenon (Onwuegbuzie and Leech, 2006; Tully, 2014), and confirmatory research is in the final stage (Goeman and Solari, 2011) to provide evidence for clinical/practical utility (Kimmelman and London, 2015; Kimmelman et al., 2014). For example, in 1980, researchers serendipitously discovered that the endothelium releases a substance that relaxes nearby vascular smooth muscle and named it endothelium-derived relaxing factor (EDRF) (Furchgott and Zawadzki, 1980). Researchers conducted exploratory research to determine the nature of EDRF until, in 1987, they identified it as nitric oxide (NO) (Ghasemi and Kashfi, 2022). The characteristics of this simple molecule were then described. Finally, FDA-approved NO-based drugs were developed and marketed to treat human diseases, including inhaled NO for pulmonary hypertension in infants in 1999 (Steinhorn, 2011) and latanprostene (an NO-donating prostaglandin analog) as an ophthalmic solution to reduce intraocular pressure in glaucoma in 2017 (Kaufman, 2018).

3.1.2. Logic of scientific inquiry and theoretical framework of research

Confirmatory and exploratory research use deduction and induction, respectively (McIntosh, 2017; Stebbins, 2001), as logical paths to knowledge (Rothchild, 2006). Deduction and induction are classical modes of scientific inquiry (Coccia and Benati, 2022; McGregor, 2018; Rothchild, 2006). In terms of direction (moving between theory and observation), deduction is top-down logic that infers from general to specific (i.e., theory-driven), and induction is bottom-up logic that infers from specific to general (i.e., data-driven) (Coccia and Benati, 2022; Emmert-Streib, 2022; McGregor, 2018; Rothchild, 2006). In confirmatory research, a theory-laden *a priori* hypothesis is articulated to address the known study problem and solve it, either partially or completely. This is followed by a rigidly preplanned research design, data collection, and statistical analysis to test the hypothesis both empirically and statistically (Jaeger and Halliday, 1998); finally, the hypothesis may be refuted or not (Ghasemi et al., 2025) for its practical/clinical utility (Kimmelman et al., 2014). This research mode reflects the Popperian philosophy of science (Schickore, 2016), which follows a deductive falsification logic (Stebbins, 2001). According to Popper, disproving hypotheses (falsification), rather than verifying them, generates knowledge (Fudge, 2014), as theories can never be verified, only refuted (Hessels and Hooge, 2021). This preference for falsifiability over verifiability stems from the asymmetry between these two demarcation criteria: a universal statement can never be proven true by particular statements, no matter how numerous they may be; however, it can be proven

false if it contradicts even a single specific case (Ayala, 2009). Thus, while we can demonstrate the falsity of universal statements, proving their truth remains elusive (Ayala, 1994). For this reason, some prefer the term "hypothesis-driven" rather than "confirmatory research" (Sainani, 2009). In this text, I use "confirmatory" in the sense of hypothesis-driven studies, but I retain the term because it is more frequently used. Confirmatory research uses well-developed theories (Due, 2022; Elliott, 2007) or core theories (Waters, 2019) for developing hypotheses and is a theory-directed (Elliott, 2007) or theory-driven (Schickore, 2016) activity.

Table 1: General characteristics of exploratory and confirmatory research

	Exploratory Research	Confirmatory Research	Ref.
Other names	Night science	Day science	Due, 2022; Hessels and Hooge, 2021; Nosek et al., 2018; Yanai and Lercher, 2019
	Hypothesis-generating	Hypothesis-testing	
	Hypothesis-driven	Data-driven	
	Idea-generating	Idea-testing	
	The context of discovery	The context of justification	
	Data-contingent research	Data-independent research	
	Postdiction	Prediction	
Position in the research cycle	Initial phase	Final phase	Goeman and Solari, 2011
Logic of scientific inquiry	Induction	Deduction	McIntosh, 2017; Stebbins, 2001
Theoretical framework	Theory-informed	Theory-directed/ driven	
Presence of core theory	No	Yes	Anderson et al., 2001; Due, 2022; Elliott, 2007; Schickore, 2016; Waters, 2007; Waters, 2019
Using background and local theories	Yes	Yes	
Theoretical influence	Minimum	Relies on well-developed theories	
Study nature	Unstructured and open-minded	Structured and rigorous	Goeman and Solari, 2011; Höfler et al., 2022; Nosek et al., 2018
Mildness	High	Low	Goeman and Solari, 2011
Flexibility	High	Low	Goeman and Solari, 2011
Protocol	*Post hoc*	Prespecified	Goeman and Solari, 2011
Researchers' degree of freedom	High	Low	Höfler et al., 2022; Waters, 2007
Examples	Genomic research in which thousands of relationships among many molecular markers are assessed	Assessing a drug's efficacy in humans	Goeman and Solari, 2011; Schwab and Held, 2020

In exploratory research, researchers formulate a research question to address an unknown study problem and better understand it. Then, the exploratory research space is defined (Höfler et al., 2023) to examine data, identify regularities among variables, and develop new theoretical ideas, concepts, and hypotheses from the collected data, or refine existing hypotheses, leading

to discoveries (Elliott, 2007; Höfler et al., 2023; Jaeger and Halliday, 1998; Schickore, 2016; Schwab and Held, 2020). This inductive logic from particulars to generals is necessarily probabilistic (Hayes et al., 2010) and should be tested in confirmatory research (Jaeger and Halliday, 1998; Kriegeskorte et al., 2009; Szollosi and Donkin, 2021). Exploration occurs when no well-developed theories exist (Due, 2022) and is therefore not guided by an encompassing/covering theory (Anderson et al., 2001; Waters, 2019). However, exploratory research is not totally theory-free (Elliott, 2007; Schickore, 2016; Waters, 2007) but works with a minimum of theoretical influence and uses local theories (theories about the specific entity (Elliott, 2007)) or background theories (Waters, 2007) (e.g., the nature of the cell cycle in biology (Elliott, 2007)); thus, background theories are a necessary prerequisite for any scientific exploration (Yanai and Lercher, 2021), and it is said that exploratory research is theory-informed (Schickore, 2016).

3.1.3. Study nature

Exploratory research is open-minded (Höfler et al., 2022). Open-mindedness in exploration (Stebbins, 2001) indicates gaps in theory formulation (Höfler et al., 2022) and can be characterized by mildness, flexibility, and *post hoc* characteristics (Goeman and Solari, 2011; Thompson et al., 2020). Exploratory research is mild in the sense that it allows for some false-positive results (Goeman and Solari, 2011). This is reasonable because exploratory research occurs early in the research cycle and must be highly sensitive to detect all strategies that might be useful (Dirnagl, 2019; Kimmelman et al., 2014). Exploratory research is flexible because it does not dictate *a priori* hypotheses and allows freedom to select and choose hypotheses (Goeman and Solari, 2011). Exploratory research is flexible in its definitions, designs, experimental parameters, including the use of multiple instruments and techniques, the range of interventions used, the variables measured, the selection of animal strains, and the statistical analysis (Anderson et al., 2001; Dirnagl, 2019; Elliott, 2007; Höfler et al., 2022, 2023; Ioannidis, 2005; Schickore, 2016; Swaraj, 2019; Waters, 2007). Exploratory research is *post hoc* in the sense that it allows for choices inherent in procedures conducted after seeing the data (Goeman and Solari, 2011), representing exploratory data analysis (EDA).

These characteristics give researchers greater degrees of freedom to focus on the most promising patterns in exploratory research (Waters, 2007). More degrees of freedom mean more exploratory work (Thompson et al., 2020). However, exploratory research is not unbounded and is conducted within boundaries (i.e., the exploratory research space); this space includes all data patterns explored among all patterns that could be explored (Höfler et al., 2023). The choice between a narrow and a broad exploratory research space depends on the availability of resources (Höfler et al., 2023). Although a large search space is desirable, it requires substantial resources and increases the risk of false-positive results (Höfler et al., 2023).

In contrast to the open-mindedness of exploratory research, confirmatory research is characterized by strictness, inflexibility, and prespecified characteristics (Goeman and Solari, 2011). Confirmatory research is in the final stage of the research cycle and is rigorous in that it seeks to minimize false positives (Goeman and Solari, 2011). It is inflexible, as an *a priori* hypothesis dictates all research steps, and all choices regarding the testing procedure must be set in stone before data collection (Goeman and Solari, 2011). In the prespecified confirmation plan, objectives are stated, hypotheses are framed, outcomes are defined, and the statistical procedures (i.e., confirmatory data analysis, CDA) expected to be applied are described (Andrade, 2023). Thus, during confirmation, researchers' degrees of freedom are restricted to provide evidence for (or against) a clearly stated hypothesis (Höfler et al., 2022; Thompson et al., 2020; Waters, 2007). It follows an in-advance research plan before data collection (Goeman and Solari, 2011), and researchers commit themselves to this plan before they see the data (Wagenmakers et al., 2012). Confirmatory research is a structured, preplanned, and rigorous

process that adheres to an evidential norm (Höfler et al., 2022). According to evolutionary theories of causation, humans evolved to detect relationships because detecting causes was survival-valuable (Hollenbeck and Wright, 2017). In this context, an *a priori* standard for prediction safeguards science (Hollenbeck and Wright, 2017).

3.2. Research problem, research purpose, and research question

Table 2 compares research problem, research purposes, and research question/hypothesis between exploratory and confirmatory research. Identifying the research problem, an obstacle that must be overcome to achieve a goal, is the starting point of the research process (Bahadoran et al., 2025a). In exploratory research, researchers pursue an unfamiliar problem (e.g., an unnoticed association between two phenomena (Bahadoran et al., 2025a)) for which no covering theory exists (Andrew et al., 2011; Jaeger and Halliday, 1998; Onwuegbuzie and Leech, 2006; Swaraj, 2019; Waters, 2019). In other words, exploration approaches a new area of research to begin with or "become familiar with something by testing it or experimenting with it" (Stebbins, 2001). In contrast, confirmatory research pursues a known problem from a theory-focused perspective (Waters, 2019).

Table 2: Comparison of research problem, research purpose, and research question/hypothesis in exploratory and confirmatory research

	Exploratory Research	Confirmatory Research	Ref.
Research problem	Unfamiliar	Known	Waters, 2019
Research purpose	Hypothesis generation to lead discovery	Hypothesis testing to provide evidence for practical utility	Elliott, 2007; Höfler et al., 2023; Jaeger and Halliday, 1998; Kimmelman et al., 2014; Lange et al., 2025; Schickore, 2016; Schwab and Held, 2020
Research question/ hypothesis	Broad/no *a priori* hypothesis	Narrow/*a priori* hypothesis	Höfler et al., 2022; Jaeger and Halliday, 1998; Kimmelman et al., 2014; Schwab and Held, 2020; Waters, 2019

One of the main differences between exploratory and confirmatory research is that the former aims to generate (not test) a new hypothesis or refine existing ones, whereas the latter aims to test a predefined hypothesis to ascertain its veracity (Andrew et al., 2011; Due, 2022; Elliott, 2007; Höfler et al., 2023; Jaeger and Halliday, 1998; Kimmelman et al., 2014; Lange et al., 2025; Onwuegbuzie and Leech, 2006; Schwab and Held, 2020; Stebbins, 2001; Swaraj, 2019). The goal of exploratory research is to identify regularities among variables (producing inductively derived generalizations about the situation under study (Stebbins, 2001)), to develop new theoretical ideas, claims, and concepts (Elliott, 2007; Höfler et al., 2023), and to generate hypotheses that justify further research (Dirnagl, 2020; Schwab and Held, 2020). Exploration aims to generate new ideas and weave them into a grounded theory, or a theory that emerges directly from data (Stebbins, 2001). In contrast, confirmation aims to test a specific prediction of a particular hypothesis or theory using new data and to provide evidence for practical/clinical utility (Elliott, 2007; Jaeger and Halliday, 1998; Kimmelman et al., 2014; Schwab and Held, 2020). Overall, exploration aims at discovery, while confirmation aims to provide evidence (Lange et al., 2025).

Formulation of the research question (Bahadoran et al., 2025b), based on the study problem (Bahadoran et al., 2025a), must be aligned with the study's purpose (Maedche et al., 2019; Newman et al., 2003). With an idea for addressing the research problem, the researcher articulates the research question (Tully, 2014). The research question is broad in exploratory research and narrow in confirmatory research (Höfler et al., 2022; Waters, 2019). In confirmatory research, the question should be converted into a testable hypothesis. In exploratory research, we have no *a priori* hypothesis, and hypotheses are generated as findings of the study.

3.3. Research design and data collection

Table 3 compares study design and data collection in exploratory and confirmatory research with respect to the qualitative/quantitative approaches used, common study designs employed, sample sizes, data sources, the number and relevance of measured variables, and the extent to which experimental parameters are allowed to vary. Designs are tools for answering questions in exploratory research (Mayo et al., 2013; McGaghie et al., 2001) and for testing hypotheses in confirmatory research (Lange et al., 2025), and are regarded as the study's architectural backbone (Polit and Beck, 2004). Researchers need to determine which design best answers the question (Hulley et al., 2013). The distinction between exploratory and confirmatory research does not directly appear in the study design (De Groot, 2014).

Table 3: Comparison of study design and data collection between exploratory and confirmatory research

	Exploratory Research	Confirmatory Research	Ref.
Quantitative/ qualitative	Mainly qualitative	Mainly quantitative	Booth, 2006; Eldredge, 2002; Stebbins, 2001; Swaraj, 2019
Study designs used	Literature research, survey, focus group, case studies, toxicological screening and genome-wide association studies, phases I and II clinical trials*	Case-control, cohort, phase III clinical trials	Gaus et al., 2015; Luijken et al., 2022; Swaraj, 2019; Ditroilo et al., 2025; Due, 2022
Sample size	Low	High	Dirnagl, 2019; Ioannidis, 2005; Kimmelman et al., 2014; Mogil and Macleod, 2017
Data source	Previously collected data is used, or data is collected during the study	New data is collected	Anderson et al., 2001; Ghasemi et al., 2025
Variable relevancy	Low	High	Anderson et al., 2001
Number of variables measured	High	Low	Anderson et al., 2001; Dirnagl, 2019; Ditroilo et al., 2025
Allowing variation in experimental parameters	High	Low	Elliott, 2007

* The distinction between exploratory and confirmatory research is not sharp in terms of study design; for example, phase IV clinical trials are both confirmatory and exploratory (see the text for details).

Exploratory questions are more open-ended and often use qualitative research designs (Booth, 2006; Eldredge, 2002; Stebbins, 2001; Swaraj, 2019) to identify variables related to a phenomenon (Ringsted et al., 2011; Swaraj, 2019). However, exploratory research also uses a quantitative approach to identify data patterns that yield novel insights or relationships between variables, for example, to test a new compound in mice (Höfler et al., 2023; Kimmelman et al., 2014; Schwab and Held, 2020). In addition, both quantitative and qualitative data may be gathered during exploration (Stebbins, 2001). Literature reviews, surveys, focus groups, and case studies are designs used to conduct exploratory research (Swaraj, 2019). Toxicological screening (Gaus et al., 2015) and genome-wide association studies (Luijken et al., 2022) are also exploratory.

Case-control and cohort studies fall between purely exploratory and purely confirmatory studies (Gaus et al., 2015), with a tendency toward confirmation. In clinical trials, phases I and II are exploratory (Ditroilo et al., 2025), and phase III performs a confirmatory test of the treatment (Ditroilo et al., 2025). However, even in phase III, assessing secondary outcome variables and sub-group analyses are exploratory (Gaus et al., 2015). Phase IV clinical trials, the first real-world test of a new drug, are both exploratory (monitoring for unknown and unpredictable phenomena such as unpredicted side effects) and confirmatory (testing the hypothesis that efficacy established in phase III works in the real world or establishing the effectiveness) (Due, 2022). This case has been nicely presented in a paper by Due, in which he provides details of the confirmatory (testing the hypothesis that imatinib is effective in pediatrics, based on its efficacy established in adults; it has been confirmed) and exploratory (discovering potential side effects of imatinib; growth-stunting was found) nature of the phase IV trial of imatinib (an oncoprotein inhibitor) for managing chronic myelogenous leukemia in pediatric patients (Due, 2022). If the studies included in a meta-analysis are exploratory, the results will be exploratory; however, studies included in a meta-analysis are usually confirmatory, and meta-analyses tend to be confirmatory (Gaus et al., 2015).

Confirmation requires a larger sample size (Dirnagl, 2019), which is appropriate for providing evidence that a hypothesis is false or not (Kennedy, 2026). In addition, *a priori* sample size calculation is obligatory for confirmation (Dirnagl, 2019), but not required for exploration (Fife and Rodgers, 2022), although it is recommended (Dirnagl, 2019).

Regarding data sources, in exploratory research, data may be collected during the study, or the researcher may use previously collected data (archival datasets) to explore interesting patterns (Anderson et al., 2001; Ditroilo et al., 2025; Fife and Rodgers, 2022; Lange et al., 2025). In confirmatory research, after formulating a hypothesis, the researcher conducts empirical observation (collection of fresh data) guided by the prespecified hypothesis to see whether logically derived, hypothesis-based predictions align with empirical observations (Aczel, 2024; Fife and Rodgers, 2022; Ghasemi et al., 2025; Lange et al., 2025). In prediction (confirmation), data are collected to test ideas about what will occur (Nosek et al., 2018).

Without a biological theory to guide the study, exploratory research typically measures many available or easily measured variables (Anderson et al., 2001; Dirnagl, 2019; Jacobucci, 2022). Given the limited existing knowledge, having multiple dependent variables in exploratory research is a more efficient use of resources (Nosek et al., 2012). However, measuring many variables increases the likelihood of false positives (Nosek et al., 2012); it is like a wild-goose chase, but this is the price that must be paid for advancing knowledge (Thomas et al., 1985). Exploratory research does not test a specific prediction of a particular theory and therefore has a looser structure and involves extensive variation of experimental parameters (Elliott, 2007). For example, it tests various substances (Elliott, 2007) and employs multiple instruments and techniques to determine the properties of a phenomenon (Schickore, 2016). In contrast, confirmatory research has clear research objectives; it collects only data on variables relevant

to testing a hypothesis (Anderson et al., 2001; Ditroilo et al., 2025), and justifies each variable (Jacobucci, 2022).

3.4. Statistical data analyses

The distinction between CDA, which is like putting a hypothesis on trial, and EDA, which is like detective work, is central to statistical practice (Fife, 2020; Gaus et al., 2015; Goeman and Solari, 2011; Wasserstein et al., 2019). In this section, I first overview statistical inference and then compare CDA and EDA.

3.4.1. Statistical inference: An overview

Inference (to carry forward (Raymond and Darsaut, 2025)) refers to "the process of using facts we know to learn about facts we do not know" (Epstein and King, 2002). Deductive inference (from logical links between premises to conclusions) belongs to pure logic, but inductive inference (from observations to generalization or theory) bridges the gap between experience and knowledge (Raymond and Darsaut, 2025) using evidence (sense data (Djulbegovic et al., 2009)). Statistical inference, a type of inductive inference, tries to address the problem of induction: moving from past observations on known cases (sample) to future expectations for unknown cases (population), and thus helps bridge the gap between experience and knowledge (Raymond and Darsaut, 2025). Evidence is a certain kind of information, related to a specific scientific claim (Macnaughton, 1998), that serves as "a guide to truth" (Djulbegovic et al., 2009) and as a basis for deriving inferences (Goodman and Royall, 1988). Evidence allows humans to believe something rationally (Martini, 2020), and the nature of inference is to constitute a claim as science (McMullin, 2013). Scientific knowledge emerges from evidence, and the quality of evidence depends on methodology (Marshall, 1997).

Since there is always the possibility of error (for example, due to measurement or sampling), data never directly entails claims about phenomena, and the transformation of data to potential falsifiers is done through statistical inference in which statistical hypothesis testing is used as a methodological decision procedure to distinguish signals from noise at least in the long run (Uygun Tunç et al., 2023). A statistical hypothesis test is based on collecting data to provide evidence about a statistical hypothesis that in turn provides evidence about a theoretical hypothesis (Kennedy, 2026). If an observation is rare under a hypothesis, then it can be regarded as evidence against that hypothesis (Goodman and Royall, 1988; Salsburg, 2014). The Neyman-Pearson approach to hypothesis testing is the most widely used statistical decision procedure for making dichotomous decisions in science (Salsburg, 2001; Turner et al., 2020; Uygun Tunç et al., 2023). This approach provides a methodological rule for distinguishing between the null hypothesis (H0) and a specified alternative hypothesis (H1) while controlling error rates in the long run (Uygun Tunç et al., 2023). H0 (the hypothesis under test) of no relation between variables/no difference between groups is assumed true unless observations lead to its rejection, providing evidence in favor of H1 (the study objective) of a relation between variables/difference between groups (Chen et al., 2010; Rajić, 2026). The observed value of a test statistic, a quantity derived from the sample, is compared with the null distribution, the probability distribution of the test statistic when the null hypothesis is true, to compute the p-value (Chen et al., 2010; Rajić, 2026), which is the probability, under H0 and the assumption of the statistical model, of obtaining data at least as extreme as those observed (Fife and Rodgers, 2022; Kennedy-Shaffer, 2019). The p-value indicates the degree of conformity between the test statistic obtained from the observed data and the pattern predicted by the null distribution and other test assumptions (Fife and Rodgers, 2022; Greenland et al., 2016). The p-value is compared with a prespecified cut-off, named type I error (α), which is the long-run probability of rejecting a true H0 under a specified testing procedure. α, commonly set at 0.05 (Kennedy-Shaffer, 2019), is fixed in advance and is not affected by data (Greenland et al., 2016); If p-

value < α, H0 is rejected, and if p-value > α, H0 is not rejected, which provides a dichotomous claim based on continuous test statistics (Uygun Tunç et al., 2023). Rejection of H0 is a claim about the likelihood that the data are as extreme as the observed data would have been if the null hypothesis were true (Nosek et al., 2018). Thus, "p-values provide a translational mechanism from theory to data" (Fife, 2020).

Correction for multiple testing is another issue related to error control in statistical inference. Upon conducting a single significant test, the probability of type I error is called the individual error rate (IER) (Bender and Lange, 2001). If k independent tests are performed, the probability of not rejecting all k null hypotheses when in fact all are true (the universal null hypothesis (Perneger, 1998)) is called the family-wise error rate (FWER) (Bender and Lange, 2001), which is the probability of rejecting at least one true H0 among all hypotheses tested (Chen et al., 2010). The probability of making FWER with multiple comparisons, when k independent statistical tests are performed, can be calculated using the equation below (Bender and Lange, 2001; Srinivas et al., 2015): $FWER=1-(1-IER)^{k}$. Poor control for multiple comparisons can lead to false-positive results (Button, 2019), and to avoid an inflated risk of type I error, multiple comparison procedures (MCP) control the family-wise type I error rate to not exceed a predefined significance level (conventionally 0.05) (Ludbrook, 1998). One statistical approach to address the multiplicity problem is to adjust the prespecified α level by the number of tests performed (Chen et al., 2010; Sainani, 2009), for which Bonferroni correction is the best-known method (Colquhoun, 2014; Srinivas et al., 2015). The Bonferroni-adjusted significance-level threshold for k significant tests is calculated from the equation below (Perneger, 1998): $\text{Bonferroni-adjusted significance level threshold}=1-(1-\alpha)^{\frac{1}{k}}$; this equation can be approximated by α/k (Bender and Lange, 2001; Perneger, 1998; Sainani, 2009). For example, when FWER is fixed at 5 % and k = 10, the IER would be $1-(1-0.05)^{\frac{1}{10}}=0.0051$ or approximately $0.05/10=0.005$; this means a p-value < 0.005 is needed to reach statistical significance (Perneger, 1998; Srinivas et al., 2015).

Another error in statistical hypothesis testing is the type II error (β), which is the pre-study long-run probability of failing to reject a false H0 under a specified alternative, effect size, and testing procedure (Greenland et al., 2016). Statistical power (1-β) of a test is the pre-study probability of rejecting H0 under a specified alternative, effect size, and testing procedure (Greenland et al., 2016). To investigate whether an effect exists, an experiment should be designed with a reasonable chance of detecting it; in statistical terms, it should have sufficient power to detect a correct H1 (Gigerenzer, 2018; Greenland et al., 2016). Power depends on the α level, the standard deviation of the underlying population, the sample size, and the size of the effect we hope to detect (Colquhoun, 2014; Curran-Everett, 2017) and thus can be increased by increasing sample size and reducing measurement error (Sterne and Smith, 2001). Underpowered studies increase false-negative rates (Dirnagl, 2016) and overestimate the effect sizes in significant findings (Colquhoun, 2014; Dirnagl, 2016; Gigerenzer, 2018), a phenomenon known as the inflation effect (Colquhoun, 2014), thereby potentially masking real effects (Sainani, 2009). Furthermore, in underpowered studies, the chance of replicating a significant finding is low; thus, power matters for the probability of successful replication (Gigerenzer, 2018).

3.4.2. Comparison between CDA and EDA

Table 4 indicates that CDA and EDA differ in four main aspects: (1) evidential status and supporting inference; (2) function of statistical tests used; (3) error control; and (4) preplanning for statistical analysis.

Table 4: Comparison between statistical analysis in exploratory and confirmatory research

Type	Exploratory Data Analysis (EDA)	Confirmatory Data Analysis (CDA)	Ref.
Evidential status	Lower	Higher	Allen and Mehler, 2019; Höfler et al., 2022; Kriegeskorte et al., 2010; Rasmussen et al., 2026
Robustness of supporting inference	Less	More	Allen and Mehler, 2019; Kriegeskorte et al., 2010
Function of statistical analysis	Hypothesis-filtering	Hypothesis-testing	Höfler et al., 2023
Error control	Low	High	Ditroilo et al., 2025; Jaeger and Halliday, 1998
Planning	Unplanned (*post hoc*)	Planned	Andrade, 2023; Goeman and Solari, 2011; Höfler et al., 2022
Number of statistical tests	Limited number of pre-specified tests	Large number of un-planned statistical tests	Sainani, 2009
Freedom in analysis	Allowed	Not allowed	Höfler et al., 2022
Fine-tuning of analysis according to data	High	Low	Wagenmakers et al., 2012

3.4.2.1. Evidential status and supporting inference

In confirmatory research, inferential statistics are heavily relied on (Stebbins, 2001), and hypothesis-based predictions are tested with new data to provide evidence for or against the study claim (Nosek et al., 2018). Unlike confirmation, which adheres to an evidential norm (Höfler et al., 2022), exploration opens the door to novelty at the cost of falling short of that norm (Höfler et al., 2022). Statistical inference in confirmatory research rests on the assumption that analytic choices are independent of the data used to evaluate the hypothesis, an assumption that does not hold in exploratory research (Rasmussen et al., 2026). Thus, statistical hypothesis testing, designed mainly for confirmation, cannot be used with the same confidence for exploration (Dirnagl, 2019; McIntosh, 2017; Nosek et al., 2018). Following inspection of data, the tendency to see patterns in random data (apophenia), confirmation bias (the tendency to focus on evidence that is in line with our expectations), and hindsight bias (the tendency to assume that an event having been predictable after its occurrence) can affect our conclusions (Munafò et al., 2017). This means statistical inference is less robust after inspecting data (Allen and Mehler, 2019); therefore, confirmatory analyses are likely to have higher evidential value than exploratory ones (Allen and Mehler, 2019; Kriegeskorte et al., 2010). In other words, the evidential weight of hypothesis testing is higher in confirmatory inference than in exploratory inference because of temporal priority and constrained analytical flexibility in confirmatory research (Rasmussen et al., 2026).

Accordingly, explorative and confirmatory interpretations of p-values and similar inferential statistics differ (Gaus, 2014; Gigerenzer, 2018; McIntosh, 2017). They should be distinguished, as the former typically establishes a new hypothesis, whereas the latter provides statistical evidence for an *a priori* hypothesis (Gaus, 2014; Gaus et al., 2015). Exploratory research relies much more on proper descriptive statistics (Dirnagl, 2019; Ditroilo et al., 2025; Gaus et

al., 2015; Lange et al., 2025) and visualizations (Fife and Rodgers, 2022). However, data reduction techniques are used in exploratory research, including principal component analysis to remove highly correlated variables, exploratory factor analysis to detect pattern similarity, and cluster analysis to identify similar observations (De Groot, 2014; Ditroilo et al., 2025). As long as researchers are honest about the exploratory nature of their work, many common analytic procedures used for CDA can also be used in EDA (Fife and Rodgers, 2022).

3.4.2.2. Function of statistical tests used

Statistical analysis in confirmatory research supports *hypothesis testing*; that is, a methodological procedure that turns observed data into error-controlled scientific evidence (Ditroilo et al., 2025). In contrast, statistical tests in EDA are used for *hypothesis filtering* rather than hypothesis testing (Höfler et al., 2023). Data exploration pursues two contrasting goals: comprehensiveness and efficiency (Höfler et al., 2023). Comprehensiveness is the ability to discover all possible patterns in the dataset through thorough research across all variables and potential analyses, without missing anything that might be hidden in the data (Höfler et al., 2023). For example, when identifying unknown risk factors for a disease, all possible variables, analyses, and observational levels should be considered (Höfler et al., 2023). Comprehensive data exploration carries the risk of false-positive results (Höfler et al., 2023). In exploratory research, we want to select several promising hypotheses for further probing in a subsequent validation phase (Goeman and Solari, 2011). Exploration aims to advance science with new insights without polluting the literature with many claims, as sub-sequent non-confirmation would waste other investigators' resources (Höfler et al., 2023). To avoid many false-positive results, exploration should be efficient (i.e., be true and relevant) (Höfler et al., 2023). Therefore, rigorous filtering is performed to identify patterns of interest, from which new hypotheses are derived (Höfler et al., 2023). To be true, an explored data pattern should not be due to chance; because the dominant filtering method is statistical testing, researchers look for statistically significant patterns in the data set to eliminate random patterns (Höfler et al., 2023; Thomas et al., 1985). Thus, statistical tests in exploratory research serve as a tool to "update subjective beliefs" about generated hypotheses that warrant future attention (Ditroilo et al., 2025).

3.4.2.3. Error control

In CDA, strict type I error control is applied during hypothesis testing to decrease false-positive results, as the defined threshold norm (usually $p < \alpha$) should not be affected by the data (Höfler et al., 2022); this is done by determining a predefined significance level threshold (confirmatory alpha) before the study and making corrections for multiple comparisons during statistical analyses (Gaus et al., 2015; Goeman and Solari, 2011; Jaeger and Halliday, 1998; Kennedy, 2026). In EDA (e.g., early studies in drug development), researchers are less strict about committing to a type I error rate than in CDA, and may adopt values of 0.1 and 0.2 (Durham et al., 2012).

Adjusting for multiple testing has generally been accepted in confirmatory research (Bender and Lange, 2001; O'brien, 1983; Sainani, 2009), and most multiple testing procedures have been designed for CDA (Goeman and Solari, 2011). Indeed, in strict CDA, "only one chance to identify a significant result is allowed, i.e., only one test is permitted (Gaus et al., 2015)," and when testing multiple hypotheses, adjustment for the α level is required (Fife and Rodgers, 2022). With multiple testing correction, p-values retain their diagnosticity (Nosek et al., 2018). Multiplicity control in EDA is challenging as the choice and number of tests are data-dependent (Bender and Lange, 2001) and applying multiple correction is difficult (Bender and Lange, 2001; Gaus et al., 2015). In addition, because there are many statistical tests, multiple correction may reduce the probability of detecting true associations to virtually undetectable levels (Thomas et al., 1985), and false-negative rates will be high (Sainani, 2009; Srinivas et al.,

2015). On the other hand, typical frequentist statistics suggest that when multiple focal comparisons are performed, the statistical threshold should be adjusted (Benning et al., 2019). In addition, appropriate adjustment for multiple comparisons, including controlling the false discovery rate (FDR), has been recommended for EDA (Menyhart et al., 2021).

Statistical power is important in any experimental study, and underpowered studies lead to high rates of false negatives and false positives (Dirnagl, 2019). For CDA, it has been proposed that power levels of 80 %, 90 %, and 95 % are considered insufficient, good, and optimal, respectively (Kennedy, 2026). Exploratory research does not control the error rate as rigorously as confirmatory research (Ditroilo et al., 2025). However, it aims to minimize type II error (maximize power) to reduce the likelihood of false-negative findings (Jaeger and Halliday, 1998; Srinivas et al., 2015); thus, high statistical power is desirable (Dirnagl, 2019). However, controlling the type II error rate is also difficult in exploratory research because sample size is small and not determined by an *a priori* power analysis (Ditroilo et al., 2025). Except for very large datasets and when type I error is high, exploratory research is underpowered (Ditroilo et al., 2025).

3.4.2.4. Preplanning for statistical analysis

CDA is preplanned: the types and numbers of statistical tests and the significance-level threshold (confirmatory alpha) are predefined before the study begins (before seeing the data) (Andrade, 2023; Höfler et al., 2022; Kennedy, 2026). By specifying the analytic pipeline before observing the data, researchers' degrees of freedom for statistical analysis are limited, the number of statistical tests is reduced, and the problem of forking paths is avoided (Nosek et al., 2018). Preplanned statistical analysis is an important feature of confirmation: testing with a high risk of failure if a claim is wrong (severe testing), and this risk must not be reduced by analytical decisions (Höfler et al., 2022). A test is severe when it can demonstrate that a claim is false (Ditroilo et al., 2025; Lakens, 2019), and theories that resist severe testing are accorded higher confidence within the scientific community (Ditroilo et al., 2025).

EDA is a *post hoc* (Latin = after this) analysis (Goeman and Solari, 2011), referring to unplanned statistical tests (Ditroilo et al., 2025) used to examine data for findings not specified *a priori* (before the study is conducted) but analyzed after seeing the data (Andrade, 2023; Goeman and Solari, 2011; Höfler et al., 2022; Srinivas et al., 2015). Exploration allows free inquiry into the behavior of the data set (Höfler et al., 2022, 2023). Because there are many variables (Dirnagl, 2019) and many choices for analyzing the data (Nosek et al., 2018), the number of statistical tests increases (Dirnagl, 2019; Höfler et al., 2023) while their severity decreases (Ditroilo et al., 2025; Höfler et al., 2023). In a dataset with v variables, g groups, and t time points, the total number of possible pairwise comparisons is: $1/2vgt(g+t-2)$ (Gaus et al., 2015). For example, with four variables, three groups, and two time points, the total number of possible pairwise comparisons is 72. In addition, observing the data can influence the selection of tests to conduct (i.e., statistical decisions are influenced by the observed data); in the "Garden of Forking Paths," observing the data may make some paths more likely, and hindsight bias may influence the findings, i.e., P-values leave with unknown diagnosticity and become uninterpretable (Nosek et al., 2018), and thus there is no need for their reporting (Mogil and Macleod, 2017). This means that knowledge generated by exploration is less robust than statistical significance would suggest, and that using test statistics in exploration is less helpful than in confirmation (Dirnagl, 2019). Thus, exploratory analysis is susceptible to spurious fluctuations in the data (Kriegeskorte et al., 2010). Unplanned statistical analysis is an important feature of exploration, which looks for novel patterns in data (Andrade, 2023; Höfler et al., 2023; Höfler et al., 2022; Srinivas et al., 2015) to generate new hypotheses (Höfler et al., 2023).

To sum up, statistically, a precise *a priori* hypothesis, establishing the significance level threshold beforehand and correction for multiple testing, and not deviating from preplanned

analysis have been proposed as three prerequisites for providing confirmatory statistical evidence, and studies that do not fulfill one or more of these prerequisites are exploratory (Gaus et al., 2015; Fife and Rodgers, 2022).

3.5. Study findings

Table 5 compares differences between study findings in exploratory and confirmatory research in terms of study output, false-positive rates, biases, sensitivity/specificity, and validity. A data-driven analysis can only produce exploratory findings (Bender and Lange, 2001), that is, hypotheses (Gaus, 2014), which, even when conducted at the highest quality, are inherently tentative and should be confirmed (Dirnagl, 2019; Ditroilo et al., 2025). A hypothesis generated from a dataset should be tested on data other than the one that generated it (Thomas et al., 1985). This means that hypotheses generated in exploratory research do not provide evidence for practice unless they pass the confirmatory phase. Unconfirmed exploratory results should be accompanied by the caveat that they are yet to be confirmed and, without independent validation, should be treated as well-reasoned speculations (Kriegeskorte et al., 2010) and as information that guides the latter stages of clinical translation (Kennedy, 2026; Kimmelman and London, 2015). Confirmatory studies, by contrast, have higher evidential value and provide evidence for a scientific proposition (Gaus et al., 2015; Worrall, 2010) and, subsequently, for clinical practice (Due, 2022; Kimmelman et al., 2014). To turn data into evidence, they must be relevant to a specific scientific claim, or hypothesis (Macnaughton, 1998).

Table 5: Study findings in exploratory and confirmatory research

Research Findings	Exploratory Results	Confirmatory Results	Ref.
Study output	Hypotheses/theories	Evidence for practice	Due, 2022; Kimmelman et al., 2014
False-positive results	High	Low	Button, 2019
Overall chance of bias	High	Low	Dirnagl, 2020; Swaraj, 2019
Sensitivity/ specificity	High sensitivity is desired (minimizing the risk of false negatives) and detecting all strategies that might be useful	High specificity is desired (minimizing the risk of false positives) and excluding all strategies that will prove useless in clinical trials	Dirnagl, 2019; Kimmelman et al., 2014; Lange et al., 2025; Schwab and Held, 2020
Internal validity	Low	High	Dirnagl, 2016; Dirnagl, 2019; Lange et al., 2025
External validity	Low	High	

Exploratory research emphasizes high sensitivity and is thus associated with higher false-positive rates and lower false-negative rates (Dirnagl, 2016, 2019; Lange et al., 2025). Exploration is inherently associated with false-positive findings (Dirnagl, 2020; Goeman and Solari, 2011; Kimmelman et al., 2014; Schwab and Held, 2020) and is more likely to lead to spurious results (Anderson et al., 2001). False-positive results can be statistically significant when chance alone is at play (Colquhoun, 2014) and can provide apparent evidence (Nosek et al., 2018). High false-positive rates in exploratory findings stem from small sample sizes, failure to account for the type I error rate, and running many tests to explore the data without multiple-comparison correction (Anderson et al., 2001; Ioannidis, 2005; Nosek et al., 2018). The more

flexible the research, the less likely the findings are to be true (Ioannidis, 2005). Spurious results are artifacts or apparent evidence (not real effects) that exist only in a particular data set under examination (Anderson et al., 2001; Andrade, 2023; Nosek et al., 2018), reflecting chance occurrences and idiosyncrasies of the data set (Cook and Ranstam, 2017). *Post hoc* strategies, lack of theoretical justification, and the flexibility of conducting undisclosed data analysis often lead to spurious results in exploratory research (Anderson et al., 2001; Höfler et al., 2022; Srinivas et al., 2015) that disguise incidental findings as evidence for a hypothesis (Höfler et al., 2022). Confirmatory research aims for high specificity (excluding strategies that would prove useless in clinical trials) (Kimmelman et al., 2014). Confirmatory research must be highly specific and therefore has fewer false positives and more false negatives (Dirnagl, 2016, 2019; Lange et al., 2025). Adherence to a prespecified plan, along with assessing plausible relations with a low number of prespecified statistical tests in confirmatory studies, prevents fine-tuning of analysis according to data (Wagenmakers et al., 2012), reduces biases (Altman et al., 2017; Cook and Ranstam, 2017), reduces false positive and spurious findings (Cook and Ranstam, 2017; Kriegeskorte et al., 2010), and gives credibility (trustworthiness) to the results (Altman et al., 2017; Cook and Ranstam, 2017).

Both exploratory and confirmatory research are subject to biases, factors that produce findings when they should not be produced (Ioannidis, 2005). However, data acquired through exploratory research are more likely to be subjected to uncontrolled bias (Dirnagl, 2020; Kennedy, 2026; Swaraj, 2019). An exploratory approach is more prone to selection bias, as greater exploration increases the likelihood of finding something in the noise (Kriegeskorte et al., 2010). A confirmatory approach is more prone to model-misspecification bias, since the more we assume, the greater the chance that our conclusions rest on sand (Kriegeskorte et al., 2010). In addition, confirmation bias, the tendency toward evidence that supports our previous beliefs (Giray, 2026), can occur in confirmatory research (Ghasemi et al., 2025). Controlling bias is equally important in both exploration and confirmation (Dirnagl, 2016, 2019; Lange et al., 2025) and is achieved through measures such as randomization, blinding, and prespecified inclusion/exclusion criteria (Dirnagl, 2016, 2019; Lange et al., 2025; Munafò et al., 2017).

Quality criteria of quantitative research are (Frambach et al., 2013): (1) validity (internal and external), (2) reliability, and (3) objectivity. Research validity is the extent to which research represents what it actually intends or claims to measure (Cooper and Schindler, 2014; Roberts and Priest, 2006). Internal validity (integrity) refers to the degree to which study findings support correct conclusions, the soundness of the study in answering the research question, or the extent to which observed effects can be attributed to the independent variable (Brian Haynes, 2006; Frambach et al., 2013; Hulley et al., 2013). Internal validity matters for both exploratory and confirmatory research, with greater emphasis on the latter (Lange et al., 2025), and is achieved by controlling bias. External validity (generalizability) refers to the degree to which the study's conclusions can be applied with confidence beyond the study or from the research sample to the population (Cooper and Schindler, 2014; Frambach et al., 2013; Hulley et al., 2013; Roberts and Priest, 2006). It is an issue of confirmation rather than exploration (Dirnagl, 2016, 2019; Lange et al., 2025). *Post hoc* strategies are inherently vulnerable to limited external validity (generalizability) (Srinivas et al., 2015). Reliability is the extent to which the results are consistent if the study were replicated (Frambach et al., 2013). Replication studies are a special case of confirmatory research (Lange et al., 2025). Objectivity means removing biases and gathering value-free information (Frambach et al., 2013). Researcher neutrality is more crucial in confirmatory than exploratory research (Lange et al., 2025).

To sum up, exploratory findings are more prone to bias (Dirnagl, 2020; Swaraj, 2019) and error (Szollosi and Donkin, 2021), less likely to be true (Ioannidis, 2005), and cannot serve as a basis for claims or future recommendations, such as treating a disease (Dirnagl, 2019; Ditroilo

et al., 2025). Because exploratory research has insufficient error control, these studies cannot support claims (Ditroilo et al., 2025). However, exploratory research is at the frontiers of current knowledge (Dirnagl, 2020; Schwab and Held, 2020), and subsequent confirmatory studies should eliminate false-positive results (Kimmelman et al., 2014). Findings from confirmatory research are more likely to be true than those from exploratory research (Head et al., 2015; Ioannidis, 2005). This is reasonable, as confirmatory research aims to test clinical utility (Kimmelman et al., 2014). For example, clinical trials have a vital role in the evolution of evidence underpinning clinical practice (Altman et al., 2017).

4. PURE EXPLORATION AND CONFIRMATION: POLES OF A SINGLE CONTINUUM

Pure exploratory and pure confirmatory research are two poles on a continuum (Fife and Rodgers, 2022; Gaus et al., 2015; Jacobucci, 2022), which has not yet been fully described except in a few cases (Fife and Rodgers, 2022; Jacobucci, 2022). Indeed, exploratory and confirmatory research are not diametrically opposed (Due, 2022), and their distinction is not necessarily sharp (Waters, 2007). Some argue that dichotomizing research as exploratory or confirmatory is too superficial (Waters, 2007) and too simplistic (Hessels and Hooge, 2021; Jacobucci, 2022). Some have criticized the use of the labels confirmatory and exploratory for research as it applies a single-dimension solution to multi-dimensional problems (Jacobucci, 2022). Much empirical research has a mixed character (De Groot, 2014) and operates somewhere between pure exploratory ("where the hypothesis is found in the data") and pure confirmatory ("where the entire analysis plan has been explicated before the first participant is tested") research (Gaus et al., 2015; Lange et al., 2025; Wagenmakers et al., 2012). By independently replicating experiments, a single study can include both exploration and confirmation (Kriegeskorte et al., 2010). The key is to distinguish confirmatory and exploratory parts transparently and explicitly (De Groot, 2014; Fife, 2020). This means the distinction between EDA and CDA is teleological in that it depends on analysts' intentions: looking for new patterns (EDA) or evidence in favor of previously recognized patterns (CDA) (Fife and Rodgers, 2022), and it can only be assessed through reporting (Jacobucci, 2022).

Table 6 compares exploratory and confirmatory research from the dimensions each prioritizes. Both types of research equally emphasize study rationale, randomization and blinding (to prevent bias), validated data-collection protocols, computational reproducibility, and transparent reporting (Lange et al., 2025). Exploratory research prioritizes hypothesis generation and looking for high sensitivity. Confirmatory research emphasizes preplanning, *a priori* hypotheses, sample size calculation before the study, predefined inclusion and exclusion criteria, error-based hypothesis testing, higher level of evidence, internal and external validity of study findings, and preregistration.

Table 6: Similarities and differences between exploratory and confirmatory research (Aczel, 2024; Allen and Mehler, 2019; Bender and Lange, 2001; Benning et al., 2019; Büttner et al., 2020; Dirnagl, 2016, 2019; Ditroilo et al., 2025; Fife and Rodgers, 2022; Kennedy, 2026; Lange et al., 2025; Sainani, 2009; Thomas et al., 1985)

		Exploration			Confirmation		
Plan and design	Rationale for the study	●	●	●	●	●	●
	Preplanning	●			●	●	●
	Theory-laden *a priori* hypothesis				●	●	●
	A priori sample size calculation	●			●	●	●
	Predefined inclusion/exclusion criteria	●	●		●	●	●
	Randomization	●	●	●	●	●	●
	Blinding	●	●	●	●	●	●
	Primary endpoint				●	●	
	Using validated protocols for data collection	●	●	●	●	●	●
	Researcher's degree of freedom	●	●	●			
Statistical analysis	Hypothesis testing				●	●	●
	Severity of statistical tests	●			●	●	●
	Diagnosticity of P-values	●			●	●	●
	Error control	●			●	●	●
	Correction for multiple testing	●			●	●	●
	Evidential value	●			●	●	●
Study findings	Providing evidence for practice				●	●	●
	Hypothesis generating	●	●	●	●		
	Internal validity	●	●		●	●	●
	External validity				●	●	
	Researcher neutrality	●			●	●	●
	Sensitivity	●	●	●	●		
	Specificity	●			●	●	●
	Replicability	●			●	●	●
Reporting	Preregistration	●			●	●	●
	Transparent reporting	●	●	●	●	●	●
	Computational reproducibility	●	●	●	●	●	●

5. BLENDING CONFIRMATORY AND EXPLORATORY RESEARCH

Blending is defined as reporting exploratory results as confirmatory (intentionally or unintentionally), or misusing exploratory research for confirmatory purposes (Höfler et al., 2022). Blending (also called intransparent exploration (Höfler et al., 2022)) results from three interrelated factors: authors (researchers), editors/reviewers (as gatekeepers of scientific publishing), and organizations (universities, research institutes, grant agencies), which collectively shape the research or scientific culture.

Scientific journals form the backbone of scientific communication (Ghasemi et al., 2023). Some editors/reviewers favor significant over nonsignificant findings for publication (Nosek et al., 2018; Stefan and Schönbrodt, 2023; Sterne and Smith, 2001). This demand for positive results encourages authors to mispresent postdictions as predictions, leading to publication bias (Colquhoun, 2014; Gigerenzer, 2018; Höfler et al., 2022; Nosek et al., 2012; Sterne and Smith, 2001). Publication bias, in which published papers are unrepresentative of completed studies (Munafò et al., 2017; Stefan and Schönbrodt, 2023), arises from a strong tendency to accentuate positive results (Sterne and Smith, 2001) and journals' reluctance to publish negative results

(Gigerenzer, 2018), thereby biasing the literature toward positive findings (Colquhoun, 2014). Publication bias may also be researcher-driven (Stefan and Schönbrodt, 2023); selection bias (file-drawer effect, lower publication rates for nonsignificant results) and inflation bias are two major types of researcher-driven publication bias (Head et al., 2015). Selectively reporting positive results more frequently than negative ones increases the likelihood of publishing attractive false-positive results at the expense of accuracy (Nosek et al., 2018), leading to a scientific literature that does not reflect reality (Lakens, 2019). False positives are pervasive in the literature and hinder scientific progress (Head et al., 2015).

In addition to journals, authors are also faced with organizations that emphasize publication as the mechanism of achievement (Nosek et al., 2012), apply wrong economic and reputational incentives, including a "publish positively or perish" culture that values quantity over quality, and evaluate researchers by their number of papers, journal impact factor, h-index, and altimetrics (Büttner et al., 2020; Colquhoun, 2014; Head et al., 2015; Stefan and Schönbrodt, 2023). Mismeasurement of science (flawed metrics used to evaluate scientific research) by silly bibliometric methods such as altmetrics harms science (Lawrence, 2007), and ranking scientists accordingly shifts researchers' focus from scientific problems toward the processes of submission, review, and publication (Nosek et al., 2012).

These actions provide an atmosphere in which quantity is valued over quality, positive (significant) findings are regarded as more valuable than negative ones (Büttner et al., 2020; Head et al., 2015) ("positive results = win" mode of thinking (Yamada, 2018)), exploratory research is considered weak (Hessels and Hooge, 2021), "lower-league research" (Aczel, 2024), less scientific (Collins et al., 2021), or even unscientific (Hessels and Hooge, 2021), and is not generally incentivized (Thompson et al., 2020). However, confirmatory research is considered the holy grail (Hessels and Hooge, 2021) and is more likely to be published (publication bias) (Höfler et al., 2022). In this wrong-incentive research atmosphere (Höfler et al., 2022; Nosek et al., 2018), where only publication matters (Smaldino and McElreath, 2016), some researchers commodify research by playing games to maximize publications (the *strategic-game hypothesis* of the replicability crisis (Gigerenzer, 2018)) because they cannot change the rules (Büttner et al., 2020). They encourage selecting, rationalizing, and reporting tests that maximize reward over accuracy (Nosek et al., 2018) and present postdictions as predictions because the standards for current scientific papers are a confirmatory narrative (Thompson et al., 2020), and exploratory research is not valued (Aczel, 2024); this mischaracterization increases the attractiveness and publishability of findings (Nosek et al., 2018). This leads to hyper-prolific authors who publish a paper every five days (Ioannidis et al., 2018).

In addition to a wrong research atmosphere, limits of human reasoning and hindsight bias (Lange et al., 2025; Nosek et al., 2018), as well as adherence to statistical rituals (the *statistical-ritual hypothesis* of the replicability crisis (Gigerenzer, 2018)), also contribute to the misrepresentation of exploration as confirmation (Nosek et al., 2018; Sterne and Smith, 2001). Hindsight bias, the belief that an outcome is inevitable after it occurs (Büttner et al., 2020), is a normal product of human cognition (Nosek et al., 2012). The degree of freedom in statistical analysis creates a breeding ground for bias, and researchers become subjectively convinced that their arguments are justifiable (Stefan and Schönbrodt, 2023). According to the statistical-ritual hypothesis, the primary goal of statistical analysis is rejecting the null hypothesis or achieving statistical significance (Gigerenzer, 2018), and good standards of scientific practice are replaced by a single convenient surrogate (the p-value) (Gigerenzer, 2018). This type of statistical analysis, which is intrinsically unfalsifiable, violates the principle of falsifiability and produces biased findings (Kennedy, 2026).

When we find a statistically significant association in the data, we can hypothesize an association that subsequent studies can confirm (Höfler et al., 2023). However, in blending, this

association is erroneously treated as a real-world association (Höfler et al., 2023). When an exploratory finding is mischaracterized as confirmatory, it is unclear how likely a statistically significant finding is to be due to chance (Büttner et al., 2020). Mispresenting exploratory results as confirmatory introduces findings that are not sufficiently substantiated by evidence (Höfler et al., 2022) or "distorts scientific evidence" (Kennedy, 2026).

Mispresenting postdictions as predictions (blending) produces pseudo-confirmatory findings (Lange et al., 2025) and falsely reduces uncertainty and decreases reproducibility (Nosek et al., 2018), which is the hallmark of good science (Edwards, 1998). This leads to one of science's worst failings, the replication crisis (Horton, 2015), meaning many published results cannot be reproduced by others (Colquhoun, 2014). Replication is a central principle of the scientific method and may serve as a demarcation criterion between science and nonscience (Nosek et al., 2012).

6. QRPs ASSOCIATED WITH BLENDING

QRPs are activities that occur during the design, analysis, and reporting of a scientific project and provide biased evidence in favor of a claim (Banks et al., 2016). "QRPs are the steroids of scientific completion, artificially enhancing the performance" (John et al., 2012). QRPs have been tagged as the Chrysalis effect, "after the metamorphosis process whereby an ugly caterpillar (initial results) turns into a beautiful butterfly (journal article)" (O'Boyle Jr et al., 2017). Researchers use QRPs to maximize output in an incentive-driven environment (Fraser et al., 2018), and ORPs are often a by-product of misreporting exploratory research as confirmatory (i.e., blending) (Höfler et al., 2022). Allowing freedom to analyze in confirmatory research provides access to the Pandora's Box of tricks, including p-hacking (massaging data to yield statistically significant results), HARKing (hypothesizing after the results are known), cherry-picking (selective outcome reporting), and fishing expeditions ("randomly casting about for new generalizations" (Stebbins, 2001)) that can be used to achieve any desired results (Wagenmakers et al., 2012). Transparent exploration to find new patterns in data should be distinguished from ORPs, which occur after conducting exploration and then presenting confirmation (i.e., blending) (Fife and Rodgers, 2022).

QRPs lie in the grey area between good research practice and scientific misconduct (fabrication, falsification, and plagiarism) (Figure 2) (Stefan and Schönbrodt, 2023). In addition to direct surveys, other indirect measures also indicate that QRPs are widespread (Fraser et al., 2018; Head et al., 2015). For example, the fraction of supported hypotheses is higher in published articles than in their corresponding dissertations (O'Boyle Jr et al., 2017). In addition, approximately 90 % of biomedical research reports findings that support the study hypothesis, raising suspicion about the field-wide presence of QRPs (Büttner et al., 2020). QRPs violate adherence to the evidential norm as a prerequisite for providing evidence in confirmation (Höfler et al., 2022, 2023), reduce the severity of tests used to demonstrate that a claim is false (Lakens, 2019), produce biased study results (Büttner et al., 2020), jeopardize the credibility of scientific findings (Stefan and Schönbrodt, 2023), and are future indicators of irreproducibility in science (Fraser et al., 2018).

EXCLI Journal 2026;25:1636-1673 – ISSN 1611-2156
Received: September 07, 2026, accepted: October 02, 2026, published: November 02, 2026

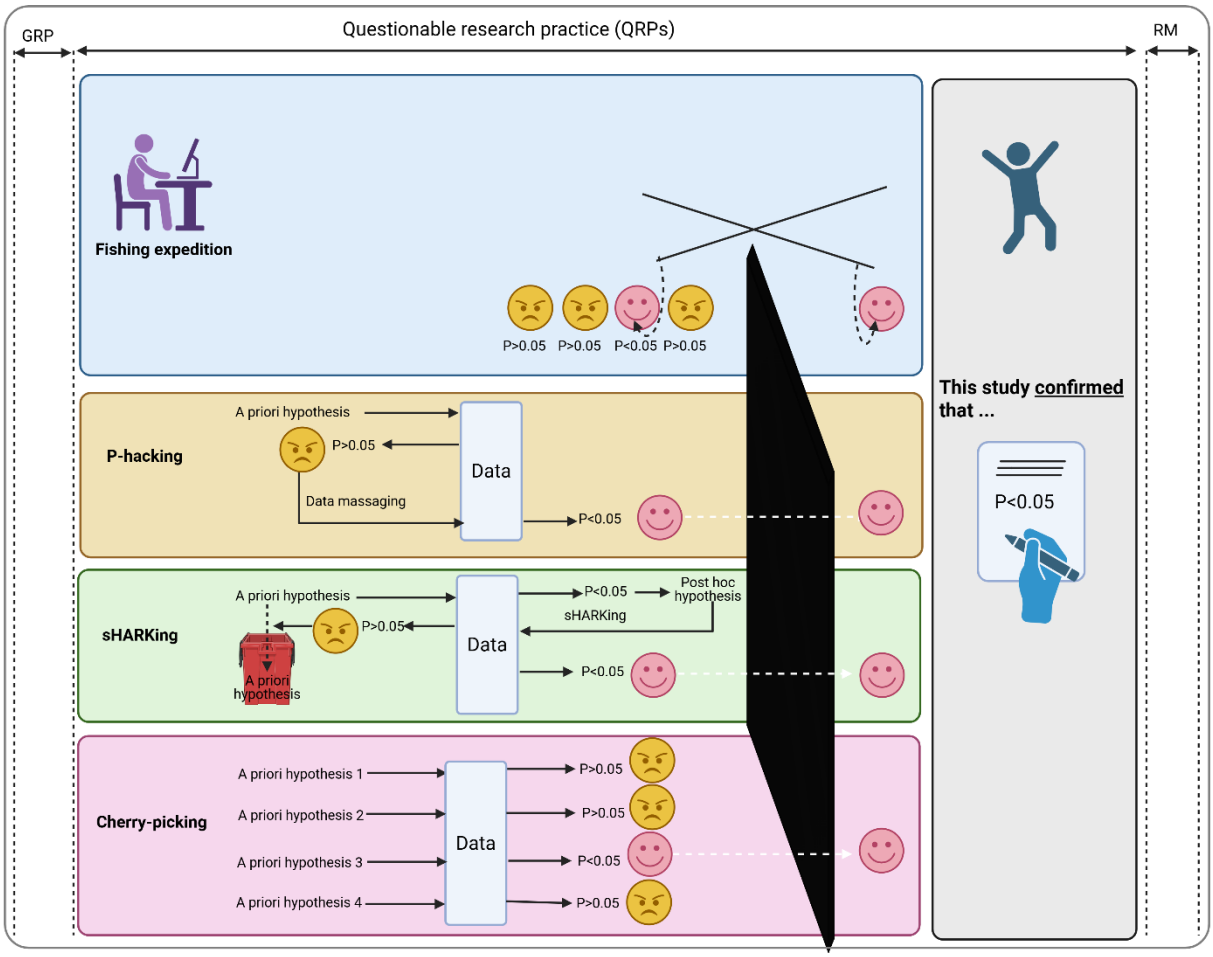

Figure 2: Questionable research practices (QRPs) lie in the grey area between good research practice (GRP) and research misconduct (RM) (fabrication, falsification, and plagiarism). sHARKing, secretly hypothesizing after results are known. See text for details. Created in https://BioRender.com

6.1. P-hacking

Broadly, p-hacking refers to untruthful manipulations a researcher uses to turn a nonsignificant result into a significant one in a single data set (Andrade, 2021; Büttner et al., 2020; Dirnagl, 2019; Head et al., 2015; Stefan and Schönbrodt, 2023; Yamada, 2018). P-hacking refers to exploiting analytic flexibility to probe the data for statistical significance related to one or more planned hypotheses (Andrade, 2021; McIntosh, 2017), and reporting the findings as if from a planned analysis (McIntosh, 2017). Indeed, it is a procrustean practice in which a researcher tries to fit data with the target hypothesis (Stefan and Schönbrodt, 2023). Here, the purpose is not hypothesis testing; rather, it is to obtain a significant result for a hypothesis (Andrade, 2021). P-hacking strategies include (1) using different statistical approaches in a trial-and-error fashion and (2) using different data eligibility specifications (including or excluding outliers, changing the number of variables in the model, covariate selection, using different cut-off values, splitting groups, variable transformation, discretizing variables) (Andrade, 2021; Büttner et al., 2020; Head et al., 2015; Stefan and Schönbrodt, 2023). P-hacking is widespread across multiple scientific disciplines (Head et al., 2015) and produces biased, inflated findings, including potentially false positives (Büttner et al., 2020). P-hacking strategies may be combined to create a P-hacking workflow (See (Stefan and Schönbrodt, 2023) for a comprehensive discussion on P-hacking strategies).

P-hacking is associated with multiple testing (see above) and selective reporting (Stefan and Schönbrodt, 2023), as it exploits multiple testing by conducting multiple tests without correcting for α-error accumulation and selectively reporting only significant results (De Groot, 2014; Stefan and Schönbrodt, 2023). Among the p-hacking strategies listed by Stefan and Schonbrodt, selective reporting ranks highest, with selective reporting of the dependent variable more prevalent than selective reporting of the independent variable (Stefan and Schönbrodt, 2023). Although sometimes p-hacking and selective reporting are considered synonymous under the umbrella of inflation bias (Head et al., 2015) or selective reporting is considered as one strategy for p-hacking (Stefan and Schönbrodt, 2023), the subtle difference between these two is a matter of time in the sense that p-hacking activity is manipulation of data collection and analysis to produce significant results, and selective reporting refers to reporting either statistically significant or nonsignificant results in the published paper (Chuard et al., 2019). In reverse p-hacking, researchers look for nonsignificant results (e.g., group differences in confounding variables) and therefore selectively report nonsignificant ones (Chuard et al., 2019).

P-hacking, a QRP of misusing statistical analysis to turn a nonsignificant result into a significant one and present it as confirmatory, should be distinguished from genuine and transparent exploration in which multiple statistical tests are conducted in search of new patterns and all of them, along with associated procedures, are transparently reported (Menyhart et al., 2021). Conducting multiple tests for hypothesis filtering is a legitimate methodological activity in transparent exploration, but in confirmatory analysis, only prespecified statistical tests are allowed, and any deviation should be rationalized and reported.

6.2. HARKing (hypothesizing after the results are known)

HARKing refers to mispresenting a hypothesis suggested by data (a *post hoc* hypothesis) as if it were *a priori* (McIntosh, 2017; Rubin, 2017; Stefan and Schönbrodt, 2023). HARKing violates the boundary between confirmation and exploration (Höfler et al., 2022) as it is a *post hoc* theorizing (accommodational hypothesizing) in which a hypothesis is generated from observed data (retrofitted to exploratory findings (Button, 2019)). The same data then evaluate the hypothesis (Nosek et al., 2018), presenting an unanticipated finding as if it were an *a priori* prediction (Hollenbeck and Wright, 2017). HARKing has been described as an opportunistic malpractice in which a researcher is indifferent regarding the hypotheses tested and accommodates the paper to fit the data (Stefan and Schönbrodt, 2023).

The problem with HARKing is double-dipping; "the same data cannot be used to generate and test a hypothesis", as it is a logical impossibility (Dirnagl, 2019). Data used to generate a hypothesis should not be used to test it (Anderson et al., 2001; Gaus et al., 2015) because it distorts descriptive statistics, invalidates statistical inference, inflates effect estimates, and increases the chance of positive results (Button, 2019; Höfler et al., 2022; Kriegeskorte et al., 2009, 2010). Building a hypothesis from the data (*post hoc*) is called accommodation (Lipton, 2005), in contrast to formulating a hypothesis before observation (i.e., prediction) (Lipton, 2005). HARKing reverses the typical hypothesis-testing procedure, and the test cannot show that the claim is false (Kennedy, 2026; Lakens, 2019), since *post hoc* hypotheses cannot be shown to be wrong (Lipton, 2005), leading to invalid evidence (Lakens, 2019). In addition, it beautifies results, because fine-tuned analyses are more compelling with the data, are likely to be overly optimistic, and feed our confirmation bias (Button, 2019; Lange et al., 2025).

Hypothesis generation is a normal product of exploration, and subgroup analyses in confirmatory studies can also generate new hypotheses (Srinivas et al., 2015); these should not be confused with HARKing, a QRP of mispresenting exploration as confirmation. To achieve this distinction, HARKing may be classified as secretly HARKing (sHARKing), an unethical QRP, and transparently HARKing (tHARKing), a good scientific practice (Hollenbeck and Wright,

2017). In sHARKing, a researcher analyzes data and, after finding a statistically significant result (often a chance finding or false positive), constructs a hypothesis around it and misrepresents the results as if the study were designed, conducted, and analyzed to test that hypothesis (Andrade, 2021). QRP in HARKing is the misrepresentation of not being transparent about the lack of an *a priori* hypothesis (Andrade, 2021). sHARKing produces many type I errors and harms science (Hollenbeck and Wright, 2017).

On the other hand, tHARKing benefits scientific progress by generating new hypotheses in exploratory research and is encouraged in the discussion section of a confirmatory study report, provided it is clearly marked as exploratory (Hollenbeck and Wright, 2017). In sHARKing, the researcher hypothesizes after examining the data at hand. In tHARKing, he/she presents new hypotheses derived from *post hoc* analyses of all extant studies, to be tested with the data at hand (Hollenbeck and Wright, 2017). tHARKing (e.g., sex-specific analysis) is especially encouraged in high-stakes contexts, where data are costly to obtain (Hollenbeck and Wright, 2017).

6.3. Fishing expedition

A fishing expedition is the indiscriminate examination of associations among combinations of variables in the hope of finding something statistically significant, which may be followed by sHARKing (Andrade, 2021). Unlike p-hacking, in which the same hypothesis is tested in numerous ways (Fife and Rodgers, 2022), data dredging (a fishing expedition taken to an extreme) seeks significant results without hypotheses in mind (Andrade, 2021). Data dredging is associated with high false-positive rates due to the large number of statistical tests conducted (Andrade, 2021). Data mining is a form of fishing expedition that is ethically done in big data (Andrade, 2021).

6.4. Cherry-picking

Cherry-picking refers to testing multiple hypotheses using multiple dependent or independent variables in a research study but reporting only significant findings. At the same time, non-significant results are treated as if they were never studied (Andrade, 2021; Büttner et al., 2020; Fraser et al., 2018; Wagenmakers et al., 2012). In cherry-picking, the orchard of statistically significant outcomes is pruned of unpalatable fruits, and the reader is deceived into seeing a picture more favorable than it truly is (Andrade, 2021; Büttner et al., 2020). Cherry-picking can lead to results that overfit (Goeman and Solari, 2011) or are overoptimistic (Altman et al., 2017), in the sense that patterns selected by the researcher may have arisen by chance and turn out to be false positives in follow-up experiments (Goeman and Solari, 2011). For example, selective outcome reporting increases the probability that a statistically significant finding is due to chance; the probability that a single outcome will yield a statistically significant result ($P < 0.05$) by chance, when H0 is true, is 5 %. However, the probability that a single outcome variable among 10 tested variables will achieve statistical significance when conducting independent tests is 40 % (Büttner et al., 2020; De Groot, 2014; Fife, 2020). In cherry-picked results (selective reporting), probabilities are invalidated, and the probability that the data support a prediction is increased while the probability that it is wrong decreases (Fife and Rodgers, 2022; Lakens, 2019).

7. HOW TO PREVENT PSEUDO-CONFIRMATORY RESEARCH?

Several strategies have been suggested for preventing the blending and its associated QRPs that can be classified into two main categories: (1) Changes in research practice, including (a) adopting preregistration for confirmatory studies (Andrade, 2021; Benjamin et al., 2018; Button, 2019; Fraser et al., 2018; Gigerenzer, 2018; Höfler et al., 2022; Horton, 2015; Nosek

et al., 2018), (b) redefining statistical thresholds, and (c) advancing research training and mentorship. (2) Changes in publication policies, including (a) providing more room for exploratory research, and (b) changing the wrong incentives in science practice.

7.1. Changes in research practice

7.1.1. Preregistration of confirmatory studies

Registration refers to the recording of a study's components in an independent registry system (e.g., https://clinicaltrials.gov, https://osf.io, https://protocols.io, and https://irct.behdasht.gov.ir) prior to data collection (preregistration), during data collection but before peeking at data (coregistration), and after data have been analyzed for other research questions (postregistration) (Benning et al., 2019; Nosek et al., 2018). Therefore, registration continuum efforts vary in timing relative to data collection and analysis, as well as in their disclosure of study details, creating a time-disclosure continuum (Benning et al., 2019). Regarding timing, earlier registration is ideal (more data-independent). Regarding disclosure, the more study features (experimental and analytical methods) are disclosed, the greater the research transparency and the higher the likelihood of replicability by others (Benning et al., 2019). Study features that can be disclosed during registration include sample, measures, intervention, procedures, data cleaning, and statistical analyses (see Benning et al. (Benning et al., 2019) for a detailed discussion).

Coregistration prevents researcher analytical decisions from being affected by study data (Benning et al., 2019; Srivastava, 2018). For example, if a preregistered study's protocol changes during data collection, researchers can rationally coregister it (Benning et al., 2019). Postregistration is useful for confirmatory secondary analyses from large data resources (Benning et al., 2019). A challenge of postregistration is that known data idiosyncrasies bias researchers' future hypotheses, since hypotheses are informed by the data rather than theory (Benning et al., 2019). Preregistration refers to the recording of an *a priori* plan (before data collection), including claims, hypotheses, study design, data collection methods, sample size, inclusion and exclusion criteria, primary and secondary outcomes, and the statistical analysis, including specific statistical tests, all of which are publicly accessible (Benning et al., 2019; Büttner et al., 2020; Fraser et al., 2018; Kennedy, 2026; Lakens, 2019; Szollosi and Donkin, 2021; Van't Veer and Giner-Sorolla, 2016).

Registration in general, and preregistration in particular, benefits both science and scientists in several ways (Van't Veer and Giner-Sorolla, 2016). First, preregistration emphasizes developing theory and methods as the main criterion for judging research, rather than focusing only on statistically significant results (Van't Veer and Giner-Sorolla, 2016). In this way, preregistration helps to reduce publication bias (Van't Veer and Giner-Sorolla, 2016). In addition, preregistration allows others to evaluate the study transparently (Lakens, 2019) and detect selection and reporting biases (Goldacre et al., 2019) (see, for example, https://www.compare-trials.org). Second, preregistration helps distinguish confirmatory from exploratory research (Van't Veer and Giner-Sorolla, 2016). After preregistration, researchers report analysis outcomes according to a prespecified plan (Nosek et al., 2018). In this way, analyses conducted on data not part of the preregistration inform exploration, and the preregistration provides a line between exploration and confirmation (Nosek et al., 2018). With preregistration, the diagnosticity of statistical inference for confirmation is preserved, and the role of exploration in generating hypotheses for future confirmation is clarified. Third, preregistration can help to prevent some forms (but not all) of QRPs, including p-hacking, HARKing, and cherry-picking (Büttner et al., 2020; Button, 2019; Chambers et al., 2015; Fife and Rodgers, 2022; Fraser et al., 2018; Höfler et al., 2022; Yamada, 2018), as it prevents constructing hypotheses after peeking at data (Aczel, 2024), requires deeper thinking about analysis (Lakens, 2019), and provides a

commitment to analytic steps without advance knowledge of the research outcomes (Nosek et al., 2018). However, preregistration's ability to prevent QRPs depends on the level of study features disclosed; for example, registering hypotheses can prevent sHARKing but not p-hacking if statistical analyses are not clearly preregistered (Allen and Mehler, 2019). Finally, preregistration promotes communication between researchers (Van't Veer and Giner-Sorolla, 2016), makes manuscript writing easier (Benning et al., 2019), increases the chance of publishing null findings (Allen and Mehler, 2019), and, by treating reviewer-requested analyses as additional exploratory analyses, protects researchers against criticizing a study after results are known (CARKing) (Benning et al., 2019).

Regarding its advantages, preregistration is proposed as crucial to transparent science (Aczel, 2024) and serves as a method-oriented solution to the reproducibility crisis in science (Szollosi and Donkin, 2021) that should be mandated before publication (Büttner et al., 2020; Yamada, 2018). Some argue that only preregistered analyses should be accepted as confirmatory (Höfler et al., 2022). However, registration also has some limitations and drawbacks (Van't Veer and Giner-Sorolla, 2016). First, registration by itself cannot protect against poor research methodology (for example, insufficient power) and can only detect it before the study begins (Benning et al., 2019). Indeed, some authors criticize preregistration, arguing that a prespecified plan does not guarantee a good study (Richter et al., 2026) and that preregistering bad theories, methods, and analyses can harm science (Szollosi et al., 2020). Second, registration can be misused or cracked in different ways (Yamada, 2018). For example, registration may occur after the data have been obtained (Gigerenzer, 2018), which is called preregistration after results are known (PARKing), another QRP (Yamada, 2018). In this way, registered reports can be abused to market positive and suppress negative results (Höfler et al., 2022). Similarly, preregistration cannot detect infinite re-experimenting, in which researchers repeat the same experiment until they obtain desired data and then report them (Yamada, 2018). Third, preregistration increases the time and cost of research projects (Allen and Mehler, 2019). Finally, implementing preregistration in practice faces challenges, including the need to deviate from data collection and analyses and to identify assumption violations during analysis (Nosek et al., 2018; Srivastava, 2018). For example, if a pilot study showed that a specific procedure does not work, or if researchers had to modify the study protocol to recruit enough participants, they should update the registration (Benning et al., 2019). As long as they are sufficiently rationalized and transparently reported, such deviations are not problematic (Allen and Mehler, 2019). In addition, researchers can use registered flexibility to cope with contingencies that may arise over the course of the study to "transform a garden of forking paths into a map of well-trod grounds" (Benning et al., 2019; Srivastava, 2018).

Nevertheless, legitimate and transparently reported deviations from the registered plan should be distinguished from another preregistration problem: noncompliance with the prespecified plan (e.g., failing to report the prespecified outcome or silently adding new outcomes). A systematic review indicates that discrepancies between registered and published primary outcomes in clinical trials are common (~30 %) (Jones et al., 2015). In assessing randomized controlled trials (RCTs) published in high-impact journals for outcome misreporting, Goldacre et al. reported that only 9 out of 67 were perfect (Goldacre et al., 2019). Surprisingly, some scientists insist on publishing papers that are not compatible with the preplanned strategies, which remain a big challenge in the empirical sciences, as noted by Wagenmakers et al.: "The fact that researchers do not commit themselves to a plan of analysis before they see the data is the main fairy-tale factor in the empirical sciences" (Wagenmakers et al., 2012). A cornerstone of a valid clinical trial, accepted as the gold standard for testing the efficacy and safety of new drugs (Rohrig et al., 2009) and improving clinical practice (Peng et al., 2026), is compliance with quality requirements (Bhatt, 2011).

In sum, registration aims to increase research transparency and maintain the diagnostic value and credibility of confirmatory research (Benning et al., 2019) by preventing "researchers from non-transparently reducing the capacity of the test to falsify a prediction" (Lakens, 2019). Thus, registration is a mechanism for transparency and accountability that helps prevent some forms of ORPs and improves the replicability of confirmatory findings.

7.1.2. Redefining statistical threshold

The currently used statistically significant threshold for claiming a discovery in science ($P < 0.05$) has been regarded as a leading cause of non-reproducibility in scientific studies (Benjamin et al., 2018). It has been proposed that the significance threshold be reduced to $P < 0.005$ for claiming discoveries, thereby decreasing false-positive rates (Benjamin et al., 2018; Gigerenzer, 2018). 1 in 2 p-values < 0.05, 1 in 6 p-values < 0.01, and 1 in 56 p-values < 0.001 are false positives (Sainani, 2009). More stringent standards for the significance threshold have already been adopted in genomics (5×10^{-8}; 1 in 20 million) and high-energy physics (3×10^{-7}; 1 in 3.5 million) (Benjamin et al., 2018; Horton, 2015). In addition to reducing false-positive rates (Benjamin et al., 2018; Gigerenzer, 2018), adopting lower p-value thresholds helps prevent some forms of p-hacking (Gigerenzer, 2018). Furthermore, some argue that redefining statistical significance is a simple, actionable step that would immediately improve the reproducibility of scientific research (Benjamin et al., 2018). On the other hand, some argue against redefining statistical significance (Lakens et al., 2018) because it encourages greater focus on the p-value (Gigerenzer, 2018), increases false-negative rates, and increases the required sample size (Benjamin et al., 2018). Decreasing the p-value threshold from 0.05 to 0.005, while maintaining 80 % power, would require about a 70 % increase in sample size (Benjamin et al., 2018). It has been recommended that a single alpha level is not universally applicable and that, instead of adopting a p-value, the alpha level should be justified for a given research question (Lakens et al., 2018).

7.1.3. Advancing research training

A major cause of blending is ineffective training in good research practices, particularly statistical and methodological issues (Fraser et al., 2018; Gigerenzer, 2018; Munafò et al., 2017; Nosek et al., 2018). Thus, research training and mentorship can partially address bad scientific practices (Andrade, 2021; Head et al., 2015; Horton, 2015). In this way, researchers recognize the importance of transparent exploration studies in opening new paths in science (Fife and Rodgers, 2022; Höfler et al., 2022) and may adhere to good research practices (Gigerenzer, 2018). Formal research education should distinguish between exploratory and confirmatory research, including their advantages and limitations, rather than emphasizing confirmatory research (Lange et al., 2025). Null hypothesis significance testing is the most widely used tool for statistical inference in confirmatory research (Salsburg, 2001; Turner et al., 2020), in which adherence to an evidential norm (such as P-value < α), which ensures severe testing, is used to confirm a claim with true evidence (Goeman and Solari, 2011; Höfler et al., 2022, 2023). Statistical tests were originally developed as an approach to confirmation, and today scientists are confirmation-primed through statistics education, with a nearly exclusive focus on confirmation (Höfler et al., 2023). For example, hypothesis testing assumes hypotheses are stated *a priori* (Veazie, 2006), which is not the case in exploratory research. In EDA, significance tests are used for hypothesis filtering rather than hypothesis testing (Goeman and Solari, 2011; Höfler et al., 2022). Thus, statistically significant findings in exploratory research do not carry the same weight as in a confirmatory study and should be treated with greater scrutiny; they should not be considered end results but should be followed by independent validation experiments (Goeman and Solari, 2011; Sainani, 2009).

Statistical thinking should be taught instead of statistical rituals (Ditroilo et al., 2025; Gigerenzer, 2018), as "statistical thinking enables good science" (Tong, 2019) and widespread failure to understand statistics contributes to QRPs (Colquhoun, 2014). Teaching statistical ritual leads researchers to use statistical hocus-pocus (Thistlethwaite et al., 2012) to obtain significant p-values. However, "no amount of statistical jiggery-pokery can affect the fact" (Charlton, 1997).

7.2. Changes in publication policies

7.2.1. Providing more room for exploratory research

Scientific journals should make more room for exploratory research (Lakens, 2019), recognize exploratory science, and designate it as a distinct section in published papers (Thompson et al., 2020), as *Cortex* does (i.e., Exploratory Reports) (McIntosh, 2017). Exploratory research should be accepted as a vital part of scientific practice (Aczel, 2024) and "the wellspring of discovery" (Collins et al., 2021). Journals can tag published papers as exploratory or confirmatory, and confirmatory papers as preregistered or non-registered to help transparent science (Chambers et al., 2015; Ditroilo et al., 2025; Stefan and Schönbrodt, 2023).

One strategy to prevent the blending of exploratory and confirmatory findings is adding a separate section entitled *post hoc* analyses and exploratory methods in a paper reporting confirmatory research to allow researchers to discuss findings that they did not originally expect but are of interest and write more along the line of good detective stories (Andrade, 2021; De Groot, 2014; Hollenbeck and Wright, 2017) and to bring exploratory science out of the shadows (Thompson et al., 2020). In confirmatory research, *post hoc* analyses of variables related to the study's primary outcomes are not permitted (Dirnagl, 2019; Srinivas et al., 2015). However, *Post hoc* analyses, including subgroup analyses, are recommended to generate new hypotheses (Srinivas et al., 2015). Subgroup analysis examines findings to determine whether they differ across discrete subject categories within the sample without altering the analysis (Andrade, 2023). Appropriate *post hoc* analyses can yield clinically valuable insights and associations that were not anticipated by *a priori* assumptions (Srinivas et al., 2015). Such exploration may help identify undesired effects of therapies and identify subgroups in which therapy may be more or less efficacious (Gaus et al., 2015). Another suggestion is to provide a supplementary file for discussing the exploratory nature of some findings (Büttner et al., 2020). To distinguish exploratory from confirmatory findings, authors should specify whether a result was obtained through data exploration or confirmatory analysis (Kriegeskorte et al., 2010), and reviewers/editors should ask authors to characterize exploratory and confirmatory findings (Gaus et al., 2015; Gigerenzer, 2018; Lange et al., 2025). This exploratory-confirmatory distinction helps readers better judge the findings (Lange et al., 2025). As long as authors label them transparently, they can discuss exploratory findings in confirmatory research (Van't Veer and Giner-Sorolla, 2016).

7.2.2. Changing the incentives

In the current scientific atmosphere, "no one is incentivized to be right" (Horton, 2015), which warrants reforming the reward system (Gigerenzer, 2018) to improve efficiency in scientific progress (Kimmelman and London, 2015; Nosek et al., 2012). Academic incentives should be restructured to promote truth over publishability, distinguish what is good for scientists from what is good for science, and emphasize that "published and true are not synonyms" (Nosek et al., 2012). Scientific stakeholders should reduce their emphasis on publication as the mechanism of achievement (Nosek et al., 2012) and focus on problem-solving, as we seek knowledge to solve problems (Bahadoran et al., 2025a). Funding agencies should revise their criteria for supporting research projects to provide financial incentives for exploratory quests (Höfler et al., 2023). In addition, to get a grant, applicants currently need some publications. In

some cases, this leads older scientists to exploit young ones, fostering a slave culture in which scientists publish more and more papers for the glory of their bosses and universities. As a result, they do not have time to learn the basics, including statistics (Colquhoun, 2014).

Animal experimentation plays a vital role in scientific development (de Aguilar-Nascimento, 2005; O'Connor and Sargeant, 2014), makes major contributions to human health and welfare (Festing and Nevalainen, 2014), and remains a crucial resource for understanding the complexity of human biology (Chang and Grieder, 2024). However, many animal experiments are poorly designed, inadequately analyzed, and poorly reported (Festing and Nevalainen, 2014; Pound et al., 2004). Inappropriate animal studies may put humans at risk and waste funds (Pound et al., 2004). Many pharmaceutical interventions are first tested in animals for safety and efficacy (O'Connor and Sargeant, 2014), and valid, precise results from animal studies can inform which treatments to advance to clinical trials (Pound et al., 2004). Mogil and Macleod's proposal for *preclinical trials*, which includes a rigorous confirmatory study (Mogil and Macleod, 2017), may address the translational roadblock from preclinical to clinical studies (Dirnagl, 2016) and would help produce reproducible findings in animal studies. Journals can adopt this proposal and clearly separate exploratory and confirmatory animal studies.

In addition, data availability (access to raw data by other researchers) and open analysis (by providing access to the complete syntax used) improve transparency in research (Head et al., 2015; Höfler et al., 2023; Thompson et al., 2020; Thye and McIntosh, 2026) and support a move toward open science, which can partially address concerns about research replication (Allen and Mehler, 2019). *EXCLI Journal* (Hengstler, 2019) and *Cortex* (McIntosh, 2017) have adopted this approach. Editors and reviewers should accept or reject a paper based on scientific quality rather than significant P-values (Gigerenzer, 2018; Hengstler, 2019), reducing researchers' motivation to adopt QRPs (Büttner et al., 2020). Finally, checklists for authors, reviewers, and editors can improve scientific reports and reduce the likelihood of false positives (Nosek et al., 2012). A good proposal is to add a distinction between exploratory and confirmatory research to publication checklists (Gaus et al., 2015). The scientific community should uphold scientific integrity to help retract flawed published papers (Hengstler, 2019). In this regard, Journal editors need to provide more space for publishing letters criticizing previously published papers.

8. CONCLUSION

The ultimate aim of science, a special kind of less fallible knowledge (Bahadoran et al., 2025a), is the theoretical representation of reality, not the compilation of data (Windhorst, 1996). Research, a problem-driven effort for problem solving (Bahadoran et al., 2025a), “is the undisputed core activity defining science” (Voit, 2019). Science is most productive when ideas generated in night science exploration are followed by day science confirmation (Figure 3) (Höfler et al., 2022; Yanai and Lercher, 2021). This back-and-forth between data and hypothesis has been called “the data-hypothesis conversation”, representing the entirety of empirical science (Yanai and Lercher, 2021). In this view, the scientific method consists of an exploratory mode of science to generate ideas (generative value) and a confirmatory mode of science to test them (true value sought) (Hessels and Hooge, 2021; McIntosh, 2017). Hypothesis testing is only half the process; the other half is idea generation through exploration (Gaus et al., 2015; Yanai and Lercher, 2021). Exploratory and confirmatory research are two necessary, distinct, and valuable modes of research that mutually reinforce each other (Collins et al., 2021) but with different theoretical bases and purposes (Lange et al., 2025). Exploratory research involves flexible data examination that enables scientific discoveries (formulating hypotheses for further confirmation) but may come at the expense of accurate and replicable findings (Büttner et al., 2020). However, exploratory research is like walking a rocky path from data to truth without

knowing where one will arrive (Höfler et al., 2023), and only some exploratory results may see the light of day (Thompson et al., 2020). Confirmatory research relies on the principle of error control in hypothesis testing to make scientific claims (Ditroilo et al., 2025) that provide evidence for the practical/clinical utility of hypotheses generated in exploratory research (Kimmelman et al., 2014). Nevertheless, exploration and confirmation are equally important (Gaus et al., 2015), and neither is inherently more valuable than the other (Gaus et al., 2015; McIntosh, 2017).

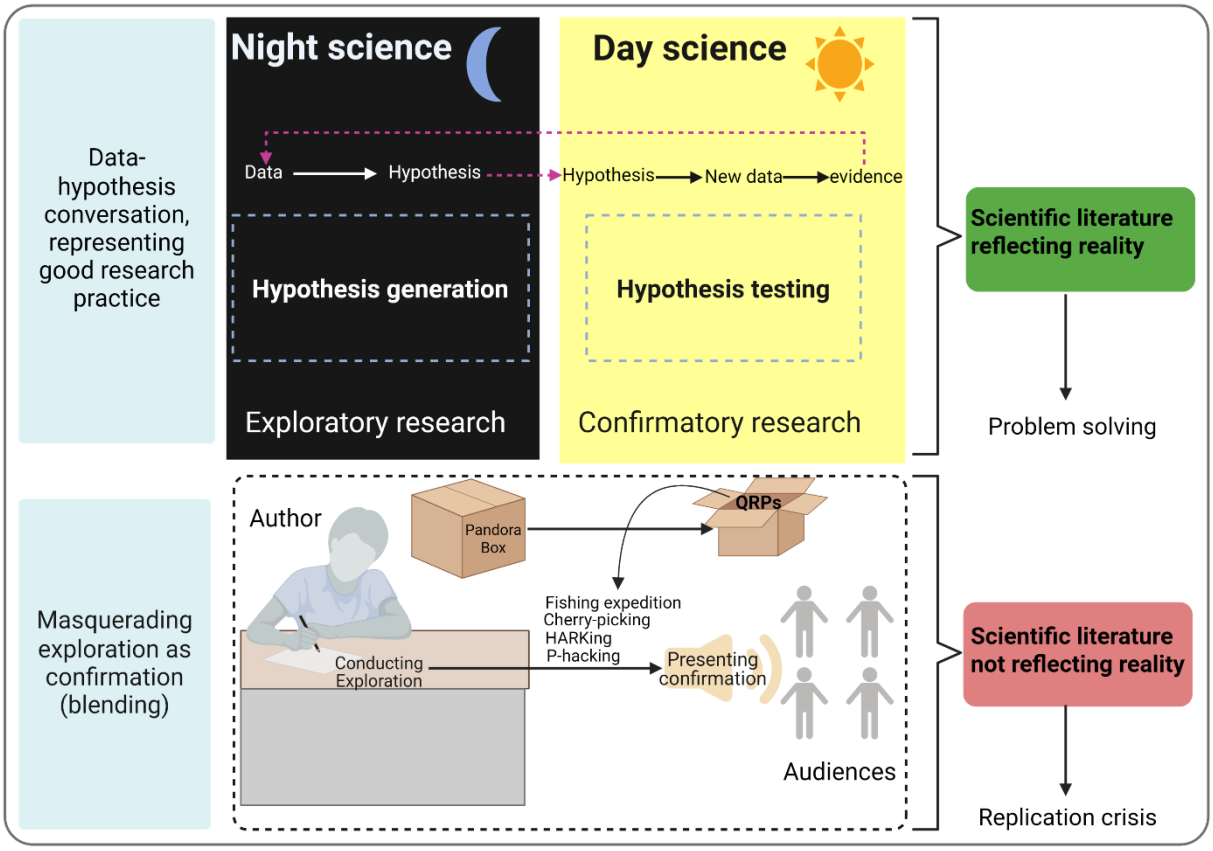

Figure 3: Data-hypothesis conversation, representing the complementary role of exploratory and confirmatory research in scientific progress. Hypotheses generated in night science (exploration) are tested in day science (confirmation) to provide evidence for practice. Misrepresenting exploration as confirmation (blending) opens Pandora's box of questionable research practices (QRPs) that bias scientific literature and lead to the replication crisis. Created in https://BioRender.com

However, in the current academic atmosphere, confirmatory research is more likely to be published (Höfler et al., 2022; McIntosh, 2017) and is often considered more valuable than exploratory research (Hessels and Hooge, 2021; McIntosh, 2017; Thompson et al., 2020). This idealization of the confirmatory mode of scientific inquiry (McIntosh, 2017) leads to a publish-or-perish culture (Büttner et al., 2020), under which some authors force their exploratory work into a confirmatory framework to maximize their publication counts. To increase the attractiveness and publishability of findings (Nosek et al., 2018), authors have to open the mysterious Pandora's Box, "whereupon ten thousand evils fly out of it" (Durant, 2011), including QRPs (e.g., p-hacking, sHARKing, cherry-picking, and fishing expedition) (Nosek et al., 2018; Thompson et al., 2020). Indeed, some researchers "torture the data until they confess" (Wagenmakers et al., 2012), "sculpt data to fit their preferred theory of the world" (Horton, 2015), or "retrofit hypotheses to fit their data" (Horton, 2015), and "data will eventually confess

if interrogated sufficiently" (Cook and Ranstam, 2017). The final product of these QRPs is one of science's worst failings (Horton, 2015): the replication crisis (Horton, 2015).

Can we do anything? Yes, "Hope alone remains" in Pandora's Box (Durant, 2011). Changes in research practice (preregistration, redefining statistical thresholds, and advancing research training) and in publication policies (providing more room for exploratory research and changing the wrong incentives in science practice) may partially solve the problem of blending and its associated QRPs (Andrade, 2021; Benjamin et al., 2018; Button, 2019; Gigerenzer, 2018; Höfler et al., 2022; Horton, 2015; Nosek et al., 2018). These changes should be directed so that researchers are "rewarded more for the marathon than for sprint" (Mogil and Macleod, 2017). The best way to achieve good research practice may be to promote research training, as Socrates says that "no one errs knowingly" (Segvic, 2006).

Acknowledgment

I would like to thank one of the anonymous reviewers of this manuscript (Reviewer A) for his/her insightful and constructive comments and for providing precise definitions for p-value, type I error, and statistical power.

Conflict of interest

The author declares no conflict of interest.

Funding

This manuscript received no funding.

AI declaration

Artificial intelligence (such as Large Language Models, chatbots, or image creators) has not been used in the production of this work. Grammarly App was used for English editing of the manuscript, but its generative AI was not used.

REFERENCES

Aczel B. Let the data talk: embrace exploratory research. Nature. 2024;635:788. https://doi.org/10.1038/d41586-024-03826-z

Agnoli F, Wicherts JM, Veldkamp CL, Albiero P, Cubelli R. Questionable research practices among Italian research psychologists. PloS one. 2017;12: e0172792. https://doi.org/10.1371/journal.pone.0172792

Allen C, Mehler DM. Open science challenges, benefits and tips in early career and beyond. PLoS Biology. 2019;17:e3000246. https://doi.org/10.1371/journal.pbio.3000246

Altman DG, Moher D, Schulz KF. Harms of outcome switching in reports of randomised trials: CONSORT perspective. Bmj. 2017;356:j396. https://doi.org/10.1136/bmj.j396

Anderson DR, Burnham KP, Gould WR, Cherry S. Concerns about Finding Effects That Are Actually Spurious. Wildlife Society Bulletin 2001;29(1):311–6. https://doi.org/10.2307/3784014

Andrade C. HARKing, Cherry-Picking, P-Hacking, Fishing Expeditions, and Data Dredging and Mining as Questionable Research Practices. J Clin Psychiatry. 2021;82:20f13804. https://doi.org/10.4088/JCP.20f13804

Andrade C. Types of analysis: planned (prespecified) vs post hoc, primary vs secondary, hypothesis-driven vs exploratory, subgroup and sensitivity, and others. Indian Journal of Psychological Medicine. 2023;45: 640–1. https://doi.org/10.1177/02537176231216842

Andrew DP, Pedersen PM, McEvoy CD. Research methods and design in sport management. Champaign: Human Kinetics; 2011.

Ayala FJ. On the scientific method, its practice and pitfalls. History and Philosophy of the Life Sciences. 1994;16:205–40.

Ayala FJ. Darwin and the scientific method. Proc Natl Acad Sci U S A. 2009;106(Suppl 1):10033–9. https://doi.org/10.1073/pnas.0901404106

Bahadoran Z, Mirmiran P, Ghasemi A. Biomedical research: The research problem matters. Addiction & Health. 2025a;17:1542. https://doi.org/10.34172/ahj.1542

Bahadoran Z, Mirmiran P, Kashfi K, Ghasemi A. Biomedical Research: Formulating a well-built and worth-answering research question. Addiction & Health. 2025b;17:1564.
https://doi.org/10.34172/ahj.1564

Banks GC, Rogelberg SG, Woznyj HM, Landis RS, Rupp DE. Evidence on questionable research practices: The good, the bad, and the ugly. Journal of Business and Psychology. 2016;31:323–38.
https://doi.org/10.1007/s10869-016-9456-7

Bender R, Lange S. Adjusting for multiple testing--when and how? J Clin Epidemiol. 2001;54:343–9.
https://doi.org/10.1016/s0895-4356(00)00314-0

Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, et al. Redefine statistical significance. Nat Hum Behav. 2018;2:6–10.
https://doi.org/10.1038/s41562-017-0189-z

Benning SD, Bachrach RL, Smith EA, Freeman AJ, Wright AGC. The registration continuum in clinical science: A guide toward transparent practices. J Abnorm Psychol. 2019;128:528–40.
https://doi.org/10.1037/abn0000451

Bhatt A. Quality of clinical trials: A moving target. Perspectives in clinical research. 2011;2:124.
https://doi.org/10.4103/2229-3485.86880

Booth A. Clear and present questions: formulating questions for evidence based practice. Library hi tech. 2006:24:355–368.
https://doi.org/10.1108/07378830610692127

Brian Haynes R. Forming research questions. J Clin Epidemiol. 2006;59:881–6.
https://doi.org/10.1016/j.jclinepi.2006.06.006

Büttner F, Toomey E, McClean S, Roe M, Delahunt E. Are questionable research practices facilitating new discoveries in sport and exercise medicine? The proportion of supported hypotheses is implausibly high. British Journal of Sports Medicine. 2020;54:1365–71.
https://doi.org/10.1136/bjsports-2019-101863

Button KS. Double-dipping revisited. Nat Neurosci. 2019;22:688–90.
https://doi.org/10.1038/s41593-019-0398-z

Chambers CD, Dienes Z, McIntosh RD, Rotshtein P, Willmes K. Registered reports: realigning incentives in scientific publishing. Cortex. 2015;66:A1–2.
https://doi.org/10.1016/j.cortex.2015.03.022

Chang MCJ, Grieder FB. The continued importance of animals in biomedical research. Lab Anim (NY). 2024; 53:295–7.
https://doi.org/10.1038/s41684-024-01458-4

Charlton BG. Restoring the balance: evidence-based medicine put in its place. Journal of evaluation in clinical practice. 1997;3:87–98.
https://doi.org/10.1046/j.1365-2753.1997.00097.x

Chen JJ, Robeson PK, Schell MJ. The false discovery rate: a key concept in large-scale genetic studies. Cancer Control. 2010;17:58–62.
https://doi.org/10.1177/107327481001700108

Chuard PJ, Vrtílek M, Head ML, Jennions MD. Evidence that nonsignificant results are sometimes preferred: Reverse P-hacking or selective reporting? PLoS Biology. 2019;17:e3000127.
https://doi.org/10.1371/journal.pbio.3000127

Coccia M, Benati I. Comparative models of inquiry. In: Farazmand A, editor. Global encyclopedia of public administration, public policy, and governance. Switzerland: Springer; 2022. p. 2112–8.

Collins HK, Whillans AV, John LK. Joy and rigor in behavioral science. Organizational Behavior and Human Decision Processes. 2021;164:179–91.
https://doi.org/10.1016/j.obhdp.2021.03.002

Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. Royal Society open science. 2014;1:140216.
https://doi.org/10.1098/rsos.140216

Cook JA, Ranstam J. Spurious findings. BJS. 2017; 104:97.
https://doi.org/10.1002/bjs.10239

Cooper DR, Schindler P. Business research methods. Columbus: Mcgraw-hill; 2014.

Curran-Everett D. CORP: Minimizing the chances of false positives and false negatives. Journal of Applied Physiology. 2017;122:91–5.
https://doi.org/10.1152/japplphysiol.00937.2016

Daly LE, Bourke GJ. Interpretation and uses of medical statistics. 5th ed. Oxford: Blackwell Sciences Ltd; 2000.

de Aguilar-Nascimento JE. Fundamental steps in experimental design for animal studies. Acta cirurgica brasileira. 2005;20:2–8.
https://doi.org/10.1590/s0102-86502005000100002

De Groot AD. The meaning of "significance" for different types of research [translated and annotated by eric-jan wagenmakers, denny borsboom, josine verhagen, rogier kievit, marjan bakker, angelique cramer, dora matzke, don mellenbergh, and han lj van der maas]. Acta psychologica. 2014;148:188–94.
https://doi.org/10.1016/j.actpsy.2014.02.001

Dirnagl U. Thomas Willis lecture: is translational stroke research broken, and if so, how can we fix it? Stroke. 2016;47:2148–53. https://doi.org/10.1161/STROKEAHA.116.013244

Dirnagl U. Resolving the tension between exploration and confirmation in preclinical biomedical research. Good Research Practice in Non-Clinical Pharmacology and Biomedicine. Cham: Springer International Publishing; 2019. p. 71–9.

Dirnagl U. Preregistration of exploratory research: Learning from the golden age of discovery. PLoS Biology. 2020;18:e3000690. https://doi.org/10.1371/journal.pbio.3000690

Ditroilo M, Mesquida C, Abt G, Lakens D. Exploratory research in sport and exercise science: Perceptions, challenges, and recommendations. Journal of Sports Sciences. 2025;43:1108–20. https://doi.org/10.1080/02640414.2025.2486871

Djulbegovic B, Guyatt GH, Ashcroft RE. Epistemologic inquiries in evidence-based medicine. Cancer control. 2009;16:158–68. https://doi.org/10.1177/107327480901600208

Due A. Are 'phase IV' trials exploratory or confirmatory experiments? Stud Hist Philos Sci. 2022;95:126–33. https://doi.org/10.1016/j.shpsa.2022.07.006

Durant W. The life of Greece: the story of civilization, Volume II. New York: Simon and Schuster; 2011.

Durham TA, Koch GG, LaVange LM. Introductory Statistics in Medical Research. In: Principles of Research Methodology: A Guide for Clinical Investigators. Cham: Springer; 2012. p. 207–31. https://doi.org/10.1007/978-1-4614-3360-6_11

Edwards A. Research and the Goal of Improving Patient Care. Forsch Komplementarmed. 1998;5 Suppl S1:116–20. https://doi.org/10.1159/000057330

Eldredge J. Evidence-Based Librarianship: levels of evidence. Hypothesis. 2002;16:10–4.

Elliott KC. Varieties of exploratory experimentation in nanotoxicology. Hist Philos Life Sci. 2007;29:313–36.

Emmert-Streib F. Severe testing with high-dimensional omics data for enhancing biomedical scientific discovery. npj Systems Biology and Applications. 2022;8:40. https://doi.org/10.1038/s41540-022-00251-8

Epstein L, King G. The rules of inference. The University of Chicago Law Review. 2002;69:1–133.

Festing MF, Nevalainen T. The design and statistical analysis of animal experiments: introduction to this issue. ILAR journal. 2014;55:379–82. https://doi.org/10.1093/ilar/ilu046

Fife D. The eight steps of data analysis: A graphical framework to promote sound statistical analysis. Perspectives on Psychological Science. 2020;15:1054–75. https://doi.org/oi.org/10.1177/17456916209173

Fife DA, Rodgers JL. Understanding the exploratory/confirmatory data analysis continuum: Moving beyond the “replication crisis”. Am Psychol. 2022;77:453–66. https://doi.org/10.1037/amp0000886

Frambach JM, van der Vleuten CP, Durning SJ. AM last page. Quality criteria in qualitative and quantitative research. Academic Medicine. 2013;88:552. https://doi.org/10.1097/ACM.0b013e31828abf7f

Fraser H, Parker T, Nakagawa S, Barnett A, Fidler F. Questionable research practices in ecology and evolution. PLoS One. 2018;13:e0200303. https://doi.org/10.1371/journal.pone.0200303

Fudge DS. Fifty years of J. R. Platt's strong inference. J Exp Biol. 2014;217:1202–4. https://doi.org/10.1242/jeb.104976

Furchgott RF, Zawadzki JV. The obligatory role of endothelial cells in the relaxation of arterial smooth muscle by acetylcholine. Nature. 1980;288:373–6. https://doi.org/10.1038/288373a0

Gaus W. Which level of evidence does the US National Toxicology Program provide? Statistical considerations using the Technical Report 578 on Ginkgo biloba as an example. Toxicology Letters. 2014;229:402–4. https://doi.org/10.1016/j.toxlet.2014.06.017

Gaus W, Mayer B, Muche R. Interpretation of statistical significance-exploratory versus confirmative testing in clinical trials, epidemiological studies, meta-analyses and toxicological screening (using Ginkgo biloba as an example). Clin Exp Pharmacol. 2015;5:1000182. https://doi.org/10.4172/2161-1459.1000182

Ghasemi A, Kashfi K. Nitric oxide: A Brief History of Discovery and Timeline of its Research. In: Ghasemi A, Kashfi K, Bahadoran Z, editors. The Role of Nitric Oxide in Type 2 Diabetes. Singapore: Bentham Science Publisher; 2022. p. 27–38.

Ghasemi A, Mirmiran P, Kashfi K, Bahadoran Z. Scientific Publishing in Biomedicine: A Brief History of Scientific Journals. International Journal of Endocrinology and Metabolism. 2023;21:e131812. https://doi.org/10.5812/ijem-131812

Ghasemi A, Hosseinpanah F, Kashfi K, Bahadoran Z. Research hypothesis: a brief history, central role in scientific inquiry, and characteristics. Addiction & Health. 2025;17:1–13.
https://doi.org/10.34172/ahj.1623

Ghasemi P, Ghasemi A, Kashfi K. The purpose of research: Clarifying aims, goals, objectives, and targets. Addiction & Health. 2026;18:1–5.
https://doi.org/10.34172/ahj.1748

Gigerenzer G. Statistical rituals: The replication delusion and how we got there. Advances in Methods and Practices in Psychological Science. 2018;1:198–218.
https://doi.org/10.1177/25152459187713

Giray L. Is any research ever truly bias-free? The myth of objectivity in science. Naunyn-Schmiedeberg's Archives of Pharmacology. 2026;399:14647–57.
https://doi.org/10.1007/s00210-026-05362-1

Goeman JJ, Solari A. Multiple testing for exploratory research. Statistical Science. 2011;26:584–97.
https://doi.org/10.1214/11-STS356

Goldacre B, Drysdale H, Dale A, Milosevic I, Slade E, Hartley P, et al. COMPare: a prospective cohort study correcting and monitoring 58 misreported trials in real time. Trials. 2019;20:118.
https://doi.org/10.1186/s13063-019-3173-2

Goodman SN, Royall R. Evidence and scientific research. American journal of public health. 1988;78: 1568–74.
https://doi.org/10.2105/ajph.78.12.1568

Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. European journal of epidemiology. 2016;31: 337–50.
https://doi.org/10.1007/s10654-016-0149-3

Grimes DA, Schulz KF. Descriptive studies: what they can and cannot do. Lancet. 2002;359:145–9.
https://doi.org/10.1016/s0140-6736(02)07373-7

Hayes BK, Heit E, Swendsen H. Inductive reasoning. Wiley interdisciplinary reviews Cognitive science. 2010;1:278–92.
https://doi.org/10.1002/wcs.44

Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-hacking in science. PLoS Biology. 2015;13:e1002106.
https://doi.org/10.1371/journal.pbio.1002106

Hengstler J. Current editorial challenges: Criteria for selecting articles for publication. EXCLI journal. 2019; 18:1101.
https://doi.org/10.17179/excli2019-2076

Hessels RS, Hooge IT. Dogmatic modes of science. Perception. 2021;50:913–6.
https://doi.org/10.1177/03010066211047826

Höfler M, Scherbaum S, Kanske P, McDonald B, Miller R. Means to valuable exploration: I. The blending of confirmation and exploration and how to resolve it. Meta-Psychology. 2022;6:MP.2021.2837.
https://doi.org/10.15626/MP.2021.2837

Hollenbeck JR, Wright PM. Harking, sharking, and tharking: Making the case for post hoc analysis of scientific data. 2017;43:5–18.
https://doi.org/10.1177/0149206316679487

Horton R. Offline: What is medicine's 5 sigma. Lancet. 2015;385:1380.
https://doi.org/10.1016/S0140-6736(15)60696-1

Hulley SB, Newman TB, Cummings SR. Getting started: The anatomy and physiology of clinical research. In: Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB, editors. Designing clinical research. 4 ed. China: Lippincott Williams & Wilkins; 2013. p. 2–13.

Ioannidis JP. Why most published research findings are false. PLoS Med. 2005;2:e124.
https://doi.org/10.1371/journal.pmed.0020124

Ioannidis JPA, Klavans R, Boyack KW. Thousands of scientists publish a paper every five days. Nature. 2018;561:167–9.
https://doi.org/10.1038/d41586-018-06185-8

Jacobucci R. A critique of using the labels confirmatory and exploratory in modern psychological research. Frontiers in Psychology. 2022;13:1020770.
https://doi.org/10.3389/fpsyg.2022.1020770

Jaeger RG, Halliday TR. On confirmatory versus exploratory research. Herpetologica. 1998;54:S64–S6.

John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. Psychol Sci. 2012;23:524–32.
https://doi.org/10.1177/0956797611430953

Jones CW, Keil LG, Holland WC, Caughey MC, Platts-Mills TF. Comparison of registered and published outcomes in randomized controlled trials: a systematic review. BMC Med. 2015;13:282.
https://doi.org/10.1186/s12916-015-0520-3

Kaufman MB. Pharmaceutical Approval Update. P & T : a peer-reviewed journal for formulary management. 2018;43:22–60.

Kennedy JE. Planning falsifiable confirmatory research. Psychol Methods. 2026;31:627–42.
https://doi.org/10.1037/met0000639

Kennedy-Shaffer L. Before p< 0.05 to beyond p< 0.05: using history to contextualize p-values and significance testing. The American Statistician. 2019;73:82–90.
https://doi.org/10.1080/00031305.2018.1537891

Kimmelman J, London AJ. The structure of clinical translation: efficiency, information, and ethics. Hastings Cent Rep. 2015;45:27–39.
https://doi.org/10.1002/hast.433

Kimmelman J, Mogil JS, Dirnagl U. Distinguishing between exploratory and confirmatory preclinical research will improve translation. PLoS Biology. 2014; 12:e1001863.
https://doi.org/10.1371/journal.pbio.1001863

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci. 2009;12:535–40.
https://doi.org/10.1038/nn.2303

Kriegeskorte N, Lindquist MA, Nichols TE, Poldrack RA, Vul E. Everything you never wanted to know about circular analysis, but were afraid to ask. J Cereb Blood Flow Metab. 2010;30:1551–7.
https://doi.org/10.1038/jcbfm.2010.86

Lakens D. The value of preregistration for psychological science: A conceptual analysis. Japanese Psychological Review. 2019;62:221–30.
https://doi.org/10.31234/osf.io/jbh4w

Lakens D, Adolfi FG, Albers CJ, Anvari F, Apps MAJ, Argamon SE, et al. Justify your alpha. Nature Human Behaviour. 2018;2:168–71.
https://doi.org/10.1038/s41562-018-0311-x

Lange FJD, Wilcke JC, Hoffmann S, Herrmann M, Boulesteix AL. On "Confirmatory" Methodological Research in Statistics and Related Fields. Stat Med. 2025;44:e70303.
https://doi.org/10.1002/sim.70303

Lawrence PA. The mismeasurement of science. Curr Biol. 2007;17:R583–5.
https://doi.org/10.1016/j.cub.2007.06.014

Leavy P. Research design: Quantitative, qualitative, mixed methods, arts-based, and community-based participatory research approaches. New York: Guilford Publications; 2023.

Lewith GT, Chan J. An exploratory qualitative study to investigate how patients evaluate complementary and conventional medicine. Complement Ther Med. 2002; 10:69–77.
https://doi.org/10.1054/ctim.2002.0524

Lipton P. Testing hypotheses: prediction and prejudice. Science. 2005;307:219–21.
https://doi.org/10.1126/science.1103024

Ludbrook J. Multiple comparison procedures updated. Clinical and Experimental Pharmacology and Physiology. 1998;25:1032–7.
https://doi.org/10.1111/j.1440-1681.1998.tb02179.x

Luijken K, Dekkers OM, Rosendaal FR, Groenwold RHH. Exploratory analyses in aetiologic research and considerations for assessment of credibility: mini-review of literature. BMJ. 2022;377:e070113.
https://doi.org/10.1136/bmj-2021-070113

Macnaughton RJ. Evidence and clinical judgement. Journal of Evaluation in Clinical Practice. 1998;4:89–92.

Maedche A, Gregor S, Morana S, Feine J. Conceptualization of the problem space in design science research. In: Extending the Boundaries of Design Science Theory and Practice: 14th International Conference on Design Science Research in Information Systems and Technology, DESRIST 2019, Worcester, MA, USA, June 4–6, 2019, Proceedings 14. Cham: Springer; 2019. p. 18–31.
https://doi.org/10.1007/978-3-030-19504-5_2

Mari L. Epistemology of measurement. Measurement. 2003;34:17–30.
https://doi.org/10.1016/s0263-2241(03)00016-2

Marshall T. Scientific knowledge in medicine: a new clinical epistemology? J Eval Clin Pract. 1997;3:133–8.
https://doi.org/10.1046/j.1365-2753.1997.00075.x

Martini C. What "Evidence" in Evidence-Based Medicine? Topoi. 2020;40:299–305.
https://doi.org/10.1007/s11245-020-09703-4

Mayo NE, Asano M, Barbic SP. When is a research question not a research question? Journal of rehabilitation medicine. 2013;45:513–8.
https://doi.org/10.2340/16501977-1150

McGaghie WC, Bordage G, Crandall S, Pangaro L. Research design. Academic Medicine. 2001;76:929–30.
https://doi.org/10.1097/00001888-200109000-00024

McGregor SL. Philosophical underpinnings of the transdisciplinary research methodology. Transdisciplinary Journal of Engineering & Science. 2018;9:182–198.
https://doi.org/10.22545/2018/00109

McIntosh RD. Exploratory reports: A new article type for Cortex. Cortex. 2017;96:A1–A4. https://doi.org/10.1016/j.cortex.2017.07.014

McMullin E. The inference that makes science. Zygon®. 2013;48:143–91. https://doi.org/10.1111/j.1467-9744.2012.01319.x

Menyhart O, Weltz B, Győrffy B. MultipleTesting.com: A tool for life science researchers for multiple hypothesis testing correction. PloS one. 2021;16:e0245824. https://doi.org/10.1371/journal.pone.0245824

Mogil JS, Macleod MR. No publication without confirmation. Nature. 2017;542:409–11. https://doi.org/10.1038/542409a

Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, et al. A manifesto for reproducible science. Nature Human Behaviour. 2017;1:0021. https://doi.org/10.1038/s41562-016-0021

Newman I, Ridenour CS, Newman C, DeMarco GMP. A typology of research purposes and its relationship to mixed methods. In: Tashakkori A, Teddlie C, editors. Handbook of mixed methods in social and behavioral research. Thousand Oaks: Sage Publications; 2003. p. 167–88.

Nosek BA, Spies JR, Motyl M. Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. Perspect Psychol Sci. 2012;7:615–31. https://doi.org/10.1177/1745691612459058

Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. Proc Natl Acad Sci U S A. 2018;115:2600–6. https://doi.org/10.1073/pnas.1708274114

O'brien P. The Appropriateness of Analysis of Variance and Multiple Comparison Procedure. Biometrics. 1983;39:787–94. https://doi.org/10.2307/2531110

O'Connor AM, Sargeant JM. Critical appraisal of studies using laboratory animal models. ILAR journal. 2014;55:405–17. https://doi.org/10.1093/ilar/ilu038

O'Boyle Jr EH, Banks GC, Gonzalez-Mulé E. The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. Journal of Management. 2017;43:376–99. https://doi.org/10.1177/0149206314527133

Onwuegbuzie AJ, Leech NL. Linking research questions to mixed methods data analysis procedures. The qualitative report. 2006;11:474–98. https://doi.org/10.46743/2160-3715/2006.1663

Pangaro L, McGaghie WC. Relevance. Academic Medicine. 2001;76:927–9. https://doi.org/10.1097/00001888-200109000-00023

Peng Y, Lyu P, Peng X. Improving research transparency: An interpretation of the updated consolidated standards of reporting trials 2025 guideline from the perspective of clinical trials in oncology. Cancer Pathog Ther. 2026;4:75–80. https://doi.org/10.1016/j.cpt.2025.07.002

Perneger TV. What's wrong with Bonferroni adjustments. Bmj. 1998;316:1236–8. https://doi.org/10.1136/bmj.316.7139.1236

Polit DF, Beck CT. Nursing research: Principles and methods. Philadelphia: Lippincott Williams & Wilkins; 2004.

Pound P, Ebrahim S, Sandercock P, Bracken MB, Roberts I. Where is the evidence that animal research benefits humans? BMJ. 2004;328:514–7. https://doi.org/10.1136/bmj.328.7438.514

Rajić V. Statistical Hypothesis Testing: A Comprehensive Review of Theory, Methods, and Applications. Mathematics. 2026;14:300. https://doi.org/10.3390/math14020300

Rasmussen P, Bailey DM, Stewart A. Transparency as the foundation of discovery. Experimental Physiology. 2026;111:3431. https://doi.org/10.1113/EP093801

Raymond J, Darsaut TE. Understanding statistical populations and inferences. Neurochirurgie. 2025;71:101608. https://doi.org/10.1016/j.neuchi.2024.101608

Richter EA, Kiens B, Mackey AL. Prespecified analysis plans are not suitable for exploratory experimental physiology research. Experimental Physiology. 2026;111:3427. https://doi.org/10.1113/EP093575

Ringsted C, Hodges B, Scherpbier A. 'The research compass': an introduction to research in medical education: AMEE Guide no. 56. Medical teacher. 2011;33:695–709. https://doi.org/10.3109/0142159x.2011.595436

Roberts P, Priest H. Reliability and validity in research. Nurs Stand. 2006;20:41–5. https://doi.org/10.7748/ns2006.07.20.44.41.c6560

Rohrig B, du Prel JB, Wachtlin D, Blettner M. Types of study in medical research: part 3 of a series on evaluation of scientific publications. Dtsch Arztebl Int. 2009;106:262–8. https://doi.org/10.3238/arztebl.2009.0262

Rothchild I. Induction, deduction, and the scientific method. Society for the Study of Reproduction. 2006: 1–11.

Rubin M. When does HARKing hurt? Identifying when different types of undisclosed post hoc hypothesizing harm scientific progress. Review of General Psychology. 2017;21:308–20. https://doi.org/10.1037/gpr0000128

Sainani KL. The problem of multiple testing. PM R. 2009;1:1098–103. https://doi.org/10.1016/j.pmrj.2009.10.004

Salsburg D. The lady tasting tea: How statistics revolutionized science in the twentieth century. New York: W. H. Freeman and Company; 2001.

Salsburg D. Hypothesis Testing, with an Historical Perspective. Wiley StatsRef: Statistics Reference Online. 2014. https://doi.org/10.1002/9781118445112.stat05865

Schickore J. "Exploratory experimentation" as a probe into the relation between historiography and philosophy of science. Studies in History and Philosophy of Science Part A. 2016;55:20–6. https://doi.org/10.1016/j.shpsa.2015.08.007

Schwab S, Held L. Different worlds Confirmatory versus exploratory research. Significance. 2020;17:8–9. https://doi.org/10.1111/1740-9713.01369

Segvic H. No one errs willingly: The meaning of Socratic intellectualism. In: Ahbel-Rappe S, Kamtekar R, editors. A Companion to Socrates. Cornwall: Blackwell Publishing; 2006. p. 171–85.

Smaldino PE, McElreath R. The natural selection of bad science. R Soc Open Sci. 2016;3:160384. https://doi.org/10.1098/rsos.160384

Srinivas TR, Ho B, Kang J, Kaplan B. Post hoc analyses: after the facts. Transplantation. 2015;99:17–20. https://doi.org/10.1097/tp.0000000000000581

Srivastava S. Sound inference in complicated research: A multi-strategy approach. PsyArXiv. 2018. https://doi.org/10.31234/osf.io/bwr48

Stebbins RA. Exploratory research in the social sciences. Thousand Oaks: Sage Unversity Paper Series on Qualitative Research Methods; 2001.

Stefan AM, Schönbrodt FD. Big little lies: a compendium and simulation of p-hacking strategies. R Soc Open Sci. 2023;10:220346. https://doi.org/10.1098/rsos.220346

Steinhorn RH. Therapeutic approaches using nitric oxide in infants and children. Free radical biology & medicine. 2011;51:1027–34. https://doi.org/10.1016/j.freeradbiomed.2011.01.006

Sterne JA, Smith GD. Sifting the evidence—what's wrong with significance tests? Physical therapy. 2001; 81:1464–9. https://doi.org/10.1093/ptj/81.8.1464

Swaraj A. Exploratory research: Purpose and process. Parisheelan Journal. 2019;15.2:665–70.

Szollosi A, Donkin C. Arrested Theory Development: The Misguided Distinction Between Exploratory and Confirmatory Research. Perspect Psychol Sci. 2021; 16:717–24. https://doi.org/10.1177/1745691620966796

Szollosi A, Kellen D, Navarro DJ, Shiffrin R, van Rooij I, Van Zandt T, et al. Is Preregistration Worthwhile? Trends Cogn Sci. 2020;24:94–5. https://doi.org/10.1016/j.tics.2019.11.009

Thistlethwaite J, Davies H, Dornan T, Greenhalgh T, Hammick M, Scalese R. What is evidence? Reflections on the AMEE symposium, Vienna, August 2011. Medical teacher. 2012;34:454–7. https://doi.org/10.3109/0142159X.2012.671561

Thomas D, Siemiatycki J, Dewar R, Robins J, Goldberg M, Armstrong B. The problem of multiple inference in studies designed to generate hypotheses. American Journal of Epidemiology. 1985;122:1080–95. https://doi.org/10.1093/oxfordjournals.aje.a114189

Thompson WH, Wright J, Bissett PG. Open exploration. Elife. 2020;9:e52157. https://doi.org/10.7554/eLife.52157

Thye M, McIntosh RD. Scientific transparency at Cortex: Guidelines for authors. 2026;195:A1–A5. https://doi.org/10.1016/j.cortex.2026.01.006

Tong C. Statistical inference enables bad science; statistical thinking enables good science. The American Statistician. 2019;73:246–61. https://doi.org/10.1080/00031305.2018.1518264

Tully MP. Research: articulating questions, generating hypotheses, and choosing study designs. The Canadian journal of hospital pharmacy. 2014;67:31–4. https://doi.org/10.4212/cjhp.v67i1.1320

Turner DP, Deng H, Houle TT. Statistical Hypothesis Testing: Overview and Application. Headache. 2020; 60:302–8. https://doi.org/10.1111/head.13706

Uygun Tunç D, Tunç MN, Lakens D. The epistemic and pragmatic function of dichotomous claims based on statistical hypothesis tests. Theory & Psychology. 2023;33:403–23. https://doi.org/.org/10.1177/09593543231160112

van Diepen M, Ramspek CL, Jager KJ, Zoccali C, Dekker FW. Prediction versus aetiology: common pitfalls and how to avoid them. Nephrol Dial Transplant. 2017; 32:ii1–ii5. https://doi.org/10.1093/ndt/gfw459

Van't Veer AE, Giner-Sorolla R. Pre-registration in social psychology—A discussion and suggested template. Journal of experimental social psychology. 2016; 67:2–12. https://doi.org/10.1016/j.jesp.2016.03.004

Veazie PJ. When to combine hypotheses and adjust for multiple tests. Health services research. 2006;41:804–18. https://doi.org/10.1111/j.1475-6773.2006.00512.x

Voit EO. Perspective: Dimensions of the scientific method. PLoS computational biology. 2019;15: e1007279. https://doi.org/10.1371/journal.pcbi.1007279

Wagenmakers EJ, Wetzels R, Borsboom D, van der Maas HL, Kievit RA. An agenda for purely confirmatory research. Perspect Psychol Sci. 2012;7:632–8. https://doi.org/10.1177/1745691612463078

Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond "p< 0.05". 2019;73:1–19. https://doi.org/10.1080/00031305.2019.1583913

Waters CK. The nature and context of exploratory experimentation: An introduction to three case studies of exploratory research. History and philosophy of the life sciences. 2007;29:275–84.

Waters CK. Presidential address, PSA 2016: An epistemology of scientific practice. Philosophy of Science. 2019;86:585–611. https://doi.org/10.1086/704973

Windhorst U. Regulatory principles in physiology. In: Comprehensive Human Physiology: From Cellular Mechanisms to Integration. New York: Springer; 1996. p. 21–42.

Worrall J. Evidence: philosophy of science meets medicine. J Eval Clin Pract. 2010;16:356–62. https://doi.org/10.1111/j.1365-2753.2010.01400.x

Yamada Y. How to crack pre-registration: Toward transparent and open science. Frontiers in psychology. 2018;9:1831. https://doi.org/10.3389/fpsyg.2018.01831

Yanai I, Lercher M. Night science. Genome Biol. 2019;20:179. https://doi.org/10.1186/s13059-019-1800-6

Yanai I, Lercher M. The data-hypothesis conversation. Genome Biol. 2021;22:58. https://doi.org/10.1186/s13059-021-02277-3