

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**Causal Language and Practice Recommendations in Observational Clinical Psychology  
Research**

A thesis presented in partial fulfilment of the  
requirements for the degree of

Master of Science

in

Psychology

at Massey University, Wellington,  
New Zealand.

Larry Paing

2024

## Abstract

Clinical psychology research often relies on observational designs due to constraints related to random assignments of participants. This means that observational studies are central to understanding relationships between complex variables in clinical psychology. Researchers are often motivated by causal theorising, and knowledge about causal effects is necessary for making inferences about practical implications. However, researchers often use linking language (terms and phrases that connect predictor and outcome variables) in ways that suggest causality, even when their observational designs do not support such inferences. Additionally, practice recommendations (actionable suggestions derived from study findings) frequently rely on these implied causal relationships, which raises concerns about the validity of these recommendations.

This thesis investigates the use of linking language in clinical psychology literature, specifically within observational studies. It also assesses the extent to which practice recommendations rely on the assumption of causal relationships between predictor and outcome variables. Conceptually replicating the methods of prior research by Haber et al. (2022) in public health and epidemiology, we analysed 97 articles from leading clinical psychology journals. We conducted a quantitative content analysis to assess the strength of causal implications in both linking language and practice recommendations. We also evaluated the relationship between the strength of causal implications in the linking language and those in the practice recommendations.

Our results revealed that while many linking sentences in observational clinical psychology articles implied causality, practice recommendations often carried even stronger causal implications. There was no statistically significant correlation between the causal implication strength of the linking sentences and that of the practice recommendations, indicating a disconnect between the two. Furthermore, researchers often provided practice

recommendations that implied causal effects despite including disclaimers stating they could not make causal inferences. This indicates an implicit interest in causal inference that is not explicitly articulated in their papers. Researchers should be transparent about their interests in causality and ensure that their conclusions are supported by the evidence. Practitioners must critically evaluate findings before applying their practice recommendations, while policymakers should exercise caution when interpreting findings that may exaggerate causal relationships.

## Acknowledgements

I would like to express my gratitude to my supervisors for their invaluable guidance and support throughout this thesis. To Dr Matt Williams, who has been a mentor to me since my undergraduate Honours research, I am profoundly thankful. Your expertise in meta-research, coupled with your kindness, patience, and generosity, has guided me through many challenges and inspired my own interest in meta-psychology. It has truly been a privilege to learn from your example. I am also sincerely grateful to Dr Kirsty Furness, whose clinical psychology perspective helped my research remain balanced and meaningful.

My appreciation extends to Massey University, which provided the Massey University Master's Research Scholarship that enabled me to focus more fully on this thesis. I am also thankful to the Ministry of Health for its study assistance fund and the flexibility that allowed me to undertake this thesis part-time while continuing my professional commitments.

Finally, I would like to thank my dear family and friends for their understanding and encouragement. Thank you for patiently listening to my thoughts and doubts along the way. Your emotional support and presence kept me motivated and reminded me of the larger purpose. To all who contributed, directly or indirectly to this work, thank you. Your support has made this thesis possible.

## Table of Contents

<b>Abstract.....</b>	<b>i</b>
<b>Acknowledgements .....</b>	<b>iii</b>
<b>List of Tables.....</b>	<b>viii</b>
<b>List of Figures .....</b>	<b>ix</b>
<b>1.0 Introduction.....</b>	<b>1</b>
1.1 Theories of Causality .....	1
1.2 Establishing Causal Inference .....	3
1.2.1 <i>Experimental Designs</i> .....	4
1.2.2 <i>Causal Inference in Observational Studies</i> .....	5
1.3 The Replication Crisis.....	7
1.4 Causal Language in Observational Studies.....	8
1.5 Distortion of Causal Evidence in The Media .....	11
1.6 Causal Inference and Practice Recommendations in Observational Studies .....	12
1.6.1 <i>Linking Language in Observational Studies</i> .....	14
1.7 Clinical Psychology as a Scientific Discipline .....	16
1.7.1 <i>Causal Language and Practice Recommendations in Clinical Psychology</i> .....	17
1.8 The Current Study .....	19
<b>2.0 Methods .....</b>	<b>21</b>
2.1 Phase 1: Journal Screening and Selection .....	21
2.2 Phase 2: Article Screening and Selection .....	23

2.2.1 Article Inclusion Criteria.....	23
2.2.2 Article Exclusion Criteria.....	24
2.3 Phase 3: Data Extraction Using a Review Tool .....	25
2.3.1 Predictor and Outcome Variables .....	26
2.3.2 Primary Linking Sentence.....	27
2.3.3 Primary Linking Word/Phrase.....	27
2.3.4 Modifying Word/Phrase .....	28
2.3.5 Linking Language Causal Implication .....	29
2.3.6 Practice Recommendations.....	31
2.3.7 Practice Recommendation Causal Implication .....	31
2.3.8 Additional Information.....	33
2.4 Review Procedures.....	33
2.5 Data Analysis .....	33
<b>3.0 Results .....</b>	<b>35</b>
3.1 Linking Language .....	35
3.2 Strength of Causal Implication.....	37
3.2.1 Strength of Causal Implication – Abstracts .....	37
3.2.3 Strength of Causal Implication – Concluding Sections.....	40
3.3 Practice Recommendations.....	42
3.3.1 Practice Recommendations – Abstracts .....	42
3.3.2 Practice Recommendations – Concluding Sections.....	43

3.4 Comparing Causality of Linking Language and Practice Recommendations .....	45
3.4.1 <i>Comparing Causality – Abstracts</i> .....	45
3.4.2 <i>Comparing Causality – Concluding Sections</i> .....	47
3.5 Secondary Objectives.....	49
3.5.1 <i>Intent to Draw Causal Inference</i> .....	49
3.5.2 <i>Causal Disclaimer</i> .....	49
3.5.3 <i>Discussion of Confounding</i> .....	52
<b>4.0 Discussion</b> .....	<b>52</b>
4.1 Summary of Results .....	52
4.2 Practice Recommendations.....	54
4.3 Schrödinger’s Causal Inference .....	55
4.4 Implications for Users of Clinical Psychology Research .....	57
4.4.1 <i>Practitioners</i> .....	57
4.4.2 <i>Policymaking</i> .....	58
4.4.3 <i>Media</i> .....	59
4.5 Recommendations for Better Clinical Psychology Research .....	59
4.5.1 <i>Transparency</i> .....	60
4.5.2 <i>Language Policing</i> .....	61
4.5.3 <i>Publication Process</i> .....	62
4.6 Limitations and Directions for Future Research.....	63
<b>5.0 Conclusion</b> .....	<b>65</b>

<b>References</b> .....	<b>67</b>
<b>Appendices</b> .....	<b>81</b>
Appendix A: Comparison to the Original Protocol by Haber et al. (2022) .....	81
Appendix B: Changes Made to Preregistration Document .....	89
Appendix C: Final Preregistration Document .....	91
Appendix D: Changes Made to Review Tool .....	100
Appendix E: Final Review Tool.....	102
Appendix F: List of Journals Considered During Journal Screening and Selection .....	109
Appendix G: Technology Used for Data Collection, Data Analysis, and Writing.....	114

**List of Tables**

Table 1 <i>Summary of Inclusion and Exclusion Criteria for Journals and Articles</i> .....	24
Table 2 <i>Examples of Root Word Determination</i> .....	28
Table 3 <i>Examples of Rating Scale for Linking Language Causal Implication</i> .....	30
Table 4 <i>Examples of Rating Scale for Practice Recommendations Causal Implication</i> .....	32
Table 5 <i>Distribution of Causal Implication Strength of Practice Recommendations Relative to Linking Language</i> .....	49
Table 6 <i>Comparison of Preregistered Objectives</i> .....	81
Table 7 <i>Comparison of Journal and Article Selection</i> .....	84
Table 8 <i>Comparison of Data Extraction Process</i> .....	85
Table 9 <i>Comparison of Data Analysis</i> .....	88

## List of Figures

Figure 1 <i>Clinical Psychology Articles Published by Year</i> .....	17
Figure 2 <i>Frequency of Root Words in the Concluding Sections</i> .....	37
Figure 3 <i>Strength of Causal Implication in Linking Sentences of Abstracts</i> .....	38
Figure 4 <i>Strength of Causal Implications in Linking Sentences of Concluding Sections</i> .....	41
Figure 5 <i>Strength of Causal Implication in Practice Recommendations of Abstracts</i> .....	43
Figure 6 <i>Strength of Causal Implication in Practice Recommendations of Concluding Sections</i>	45
Figure 7 <i>Comparison Between Causal Implication Strength of the Linking Language and Practice Recommendations in the Abstracts</i> .....	46
Figure 8 <i>Comparison Between Causal Implication Strength of the Linking Language and Practice Recommendations in the Concluding Sections</i> .....	48

## 1.0 Introduction

Causal inference is a basic cognitive mechanism that allows humans to discern the connections between events and predict outcomes. At a fundamental level, this process involves understanding cause-and-effect relationships. Causal inference forms the basis of decision-making across diverse domains such as law, medicine, economics, and environmental science. In the legal sphere, causal inference is essential for shaping decisions across civil and criminal cases, where determining liability or culpability often depends on understanding causal links (Moore, 2009). Likewise, in environmental science, it facilitates the identification of variables driving ecological changes, informing interventions in managing environmental challenges (Kimmel et al., 2021).

Causal inference is also significant in psychology and mental health as researchers and clinicians in the field depend on the understanding of causality to examine factors shaping mental health outcomes. Moreover, it informs the analyses of behavioural patterns, developmental trajectories, and efficacy of therapeutic interventions. Beyond scientific and professional applications, causal inference permeates everyday discussions. Contentious debates such as the impact of gun control on crime rates, the health benefits of organic foods, or the public safety implications of vaccination mandates reflect how individuals apply (and often misapply) causal reasoning. However, these discussions are complicated by cognitive biases and different interpretations, highlighting the nuanced nature of causal inference.

### 1.1 Theories of Causality

In philosophy, there are two classic views on causality. The first perspective, from thinkers like John Locke and John Mill, argues that causality can be verified through systematic investigation of empirical data (Glass et al., 2013). The systematic investigation of causality has practical applications in fields like epidemiology and public health, where rigorous methods are used to establish causal relationships. Meanwhile, the second perspective, from David Hume

(1896, as cited in Höfler, 2005), posits that although there may be cause-and-effect relationships in the natural world, these connections cannot be definitively confirmed through empirical evidence alone. Hume expressed doubts regarding the existence of causality and argued that it is not an intrinsic property of objects but rather a concept shaped by our observations of recurring patterns (Höfler, 2005).

Apart from the two classic views, another perspective on causality revolves around counterfactuals. Counterfactual theory involves imagining potential outcomes in the absence of a presumed cause (Pearl, 2009). Essentially, it involves asking "what if" situations to establish cause-and-effect relationships (Menzies & Beebee, 2024). For instance, we could rephrase the statement "Engaging in social media causes decreased happiness" as "If an individual had engaged in social media, their level of happiness would have been lower than if they had refrained from doing so". Simply put, counterfactual reasoning explains that if someone had been free of social media, their happiness levels would have remained stable provided all other factors remained unchanged. This counterfactual approach focuses on comparing two possible realities: one where the person uses social media and one where they do not, helping to clarify the causal connection.

To establish causality, modern science practices integrate the empirical approach of Locke and Mill with counterfactual reasoning (Morgan & Winship 2014). Statistical tools such as regression analysis and structural equation modelling use counterfactual reasoning to control for confounding variables<sup>1</sup> and estimate the causal effect of an independent variable on a dependent variable. In the example above, for instance, regression analysis might be applied to explore how social media use correlates with happiness, while taking into account various

---

<sup>1</sup> A confounder is a third variable that is associated with both the exposure and the outcome, affects the outcome, and is not on the causal pathway between the exposure and the outcome (Lederer et al., 2019).

factors such as age, socioeconomic status, or mental health that could influence both social media usage and happiness levels. Alternatively, we can use regression analysis to explore how happiness levels might differ if social media use were absent, assuming these other factors remained unchanged. By comparing the predicted outcomes for these two scenarios (counterfactual reasoning), regression analysis quantifies the causal effect of social media use on happiness.

## **1.2 Establishing Causal Inference**

Research has an inherent goal of establishing causal relationships by exploring how variables influence one another (Lynch, 2013). For instance, researchers may want to investigate whether introverts and extroverts demonstrate different levels of job performance because they thought that extraversion impacts how well people do at work. This example suggests a causal relationship where the predictor or independent variable, extraversion, influences the outcome or dependent variable, job performance. In social science, including psychology, causality is identified through three main criteria.

The first criterion, known as correlation or association, requires that the predictor and outcome to be linked in some way, meaning changes in the predictor correspond to changes in the outcome. Establishing this link often requires statistical methods like correlation coefficients; however, the old maxim of “correlation does not imply causation” reminds us that just because two variables are related, it does not mean that one is necessarily causing changes in the other. Therefore, we need to consider further criteria below. The second criterion is called temporal order, and it requires that the predictor must occur before the outcome, in line with the principle that causes should come before their effects.

The third criterion of non-spuriousness requires that the relationship between the predictor and outcome is not due to an external factor, which can be the true cause of both predictor and outcome. A spurious relationship between a predictor and an outcome occurs

when a third variable (confounding variable) fully explains its apparent connection, making the observed link between the predictor and the outcome misleading (Lynch, 2013). Researchers use two main methods to rule out spuriousness: one is experimental methods, and the other is through statistical control. From the example above, to establish causality, researchers would need to demonstrate that different levels of extraversion correspond to differences in job performance, that extraversion comes before job performance, and that the observed relationship is not due to a third variable.

### **1.2.1 Experimental Designs**

In a *true experiment*, also known as a randomised controlled trial (RCT), the independent variable is manipulated, and participants are randomly assigned to different conditions. This randomisation is essential as it prevents individual differences from becoming confounding variables. However, randomisation alone does not rule out the possibility of a relationship between the independent and dependent variable that may have occurred purely by random chance. This possibility can be addressed using inferential statistics, such as Null Hypothesis Significance Testing (NHST), to rule out chance findings. NHST helps determine the likelihood that the observed relationship between the predictor and the outcome (or a stronger relationship) could have occurred by chance. If the probability is very low, researchers reject the null hypothesis, which states there is no effect. By combining randomisation with inferential statistics, true experiments provide robust evidence for causal inferences, distinguishing them from observational studies and quasi-experiments, where random assignment is absent.

Another type of experimental design, known as quasi-experiments, involves manipulating a predictor without the random assignment to different conditions. Unlike true experiments, which minimise confounding variables by randomly assigning participants to groups, quasi-experiments are more vulnerable to biased estimates (Pedhazur & Schmelkin, 1991). However, quasi-experiments still offer valuable insights when paired with statistical

techniques that aim to address confounding variables. Nevertheless, true experiments are considered the gold standard for establishing causality due to their randomised design, which helps mitigate the impact of individual differences and confounding variables on the outcomes.

### **1.2.2 Causal Inference in Observational Studies**

Constraints such as ethical considerations and resource limitations often mean that randomisation or manipulation of an independent variable is unfeasible. Additionally, generalising the results of a true experiment to real-world situations can be problematic when it involves highly unrepresentative populations or is conducted under unnatural conditions (Lynch, 2013; Ohlsson & Kendler, 2020). Given these challenges, psychological researchers often turn to observational data, whether through longitudinal or cross-sectional designs, which are usually more practical.

Previous publications, like those by Cofield et al. (2010) and Adams et al. (2017), argue that causal language should not be used in observational studies. On the same note, except when marked by exceptional strengths, observational studies have been classified as low-grade evidence within certain frameworks (Guyatt et al., 2008; Prasad et al., 2013). Additionally, guidelines from the Journal of the American Medical Association recommend avoiding causal language outside of RCTs, based on the belief that only randomised trials can allow reliable causal inference (Ruich, 2017). Unfortunately, these claims are not accurate.

Drawing inferences about the causal relationship between the predictor and the outcome is heavily dependent on maintaining high internal validity. Internal validity refers to the validity of assertions regarding the effects of the independent variable(s) on the dependent variable(s), essentially addressing whether the observed phenomenon is due to the variables the researcher claims to be operating or if it can be attributed to other variables (Pedhazur & Schmelkin, 1991). True experiments or RCTs are designed to control for extraneous variables and use randomisation to ensure groups are similar except for the treatment, hence they can

achieve high internal validity which allow for causal inference (Pedhazur & Schmelkin, 1991). Observational studies, although lacking randomisation, can still suggest causal relationships through methods like statistical control and eliminating confounding variables. Thus, while RCTs serve as a gold standard for establishing causal inference, well-designed observational studies also provide valuable insights.

Researchers can use various techniques to provide credible causal claims from observational data, such as multiple regression and propensity score methods. Multiple regression is a statistical method that allows researchers to examine the relationship between one dependent variable and multiple independent variables while controlling for confounding variables.<sup>2</sup> Additionally, another set of methods called propensity score methods, involve strategies such as matching, stratification, and weighting, used to estimate the effect of a treatment or intervention by accounting for covariates (Rosenbaum & Rubin, 1983).<sup>3</sup>

The causal modelling framework linked to Judea Pearl's work using Directed Acyclic Graph (DAG) is a formal technique to visually represent causal relationships between variables (Pearl, 1993; 1995). According to Rohrer (2018), DAG structures comprise nodes representing variables connected by arrows indicating causal effects. These visual aids assist researchers in determining which additional variables, like confounders and mediators that should be controlled for in estimating causal effects. By blocking non-causal associations through methods like adding confounders as covariates, performing stratified analysis, or using propensity score matching, DAGs enable researchers to uncover causal relationships between variables and differentiate genuine causal effects from spurious correlations (Rohrer, 2018). Effectively, DAGs

---

<sup>2</sup> This approach depends on knowing and accurately measuring all the confounding variables to mitigate biased results.

<sup>3</sup> Covariates are independent variables that can influence the dependent variable and are controlled for in statistical analyses to reduce bias.

simplify the process of understanding and testing assumptions about causal processes, which leads to robust causal inference.

Another set of methods, natural experiments, offers an approach for establishing causal inferences when traditional experimental methods are not feasible. Unlike typical observational studies, natural experiments rely on events or conditions that naturally create a random or “as-if” random assignment of participants to different conditions, such as a policy change or a natural disaster (Dunning, 2012; Sieweke & Santoni, 2020). For an event to be considered a suitable natural experiment, the natural or as-if random assignment should ideally result in little to no systematic differences between individuals in the treated and untreated groups before the treatment (Grosz et al., 2024). This assignment mimics randomisation, helping to balance out confounding factors between groups and allowing researchers to attribute observed effects more confidently to the event or condition being studied, thus reducing bias in causal claims (Grosz et al., 2024). Although natural experiments have not been used widely in psychology, they have the potential to provide researchers with naturally occurring opportunities for random assignment. Effectively, natural experiments allow researchers to study causal relationships in the absence of controlled experimental conditions.

### **1.3 The Replication Crisis**

Well-replicated findings with good designs give us more confidence in causal inferences because they demonstrate that the observed effects are consistent and not due to random chance. However, many scientific fields have faced challenges in replication, often referred to as the *replication crisis*. Concerns about the reliability of scientific research were raised by John Ioannidis in his influential 2005 paper, “Why Most Published Research Findings Are False”. Ioannidis argued that biases and methodological flaws often lead to false research conclusions, impacting fields like psychology. Following this, in 2015, the *Reproducibility Project*, led by Brian Nosek, attempted to replicate 100 experimental and observational psychology studies from top

journals (Open Science Collaboration, 2015). Their findings revealed that only 36% of replication attempts were statistically significant compared to the original 97%. The stark contrast between initial results and replication outcomes raised doubts on the reliability of psychology research.

A combination of factors contributes to this crisis. Among these are publication biases, academic competition fostering questionable practices, methodological flexibility, and the pervasive practice of HARKing; hypothesising after the results are known (Grand et al., 2018). These factors incentivise researchers to pursue significant results, sometimes resorting to questionable research practices that undermine the integrity of their work (Ferguson & Heene, 2012). In response, a *credibility revolution* has emerged, advocating for more transparent, rigorous, and reproducible research practices (Vazire, 2018). Researchers have proposed metascientific<sup>4</sup> solutions, such as redefining significance levels and promoting *constraints on generality* statements in efforts to enhance research quality and restore trust in research (Benjamin et al., 2018; Simons et al., 2017). Replicability and credibility issues could reduce the value of research for evidence-based policymaking (Guay, 2018). In situations where research findings lack reliability or cannot be replicated consistently, evidence-based decisions are harder to justify. The replication crisis highlights the importance of robust research to ensure that policies are grounded in reliable scientific evidence.

#### **1.4 Causal Language in Observational Studies**

The replication crisis demonstrates the risks posed by unreliable research. While the replication crisis has focused on improving replicability of research, the appropriate use of causal language is equally important to avoid misleading or unsupported causal claims. This is

---

<sup>4</sup> Metascience is the study of the scientific process itself, focusing on improving its quality (Ioannidis, 2018).

especially important when analysing data from observational studies. Although robust statistical methods are necessary to establish causal inference using observational data, the credibility of causal inference still depends on how well its conclusions are appropriately calibrated to the evidence. Several metascientific studies have revealed issues concerning the use of causal language in observational studies.

Parra et al. (2021) evaluated the consistency of causal language in observational studies published in the British Medical Journal: BMJ in 2018. The researchers reviewed the abstracts of these papers to determine how clearly causal messages were conveyed. Their assessment involved comparing published abstracts with the full texts and originally submitted versions to identify changes introduced during peer review that led to inconsistencies. They found that 48% of the studies consistently used causal language, while 20% showed inconsistencies, often presenting causal conclusions despite initially stating an aim to evaluate associations between variables. Parra et al. (2021) called for clearer guidelines on reporting causal claims in observational research.

Several scholars have pointed out the prevalent inconsistencies of causal language used in research. *Schrödinger's causal inference*, a concept used by Tennant and Murray (2021), refers to studies where the authors simultaneously warn against making causal interpretations while implicitly making them. For example, a study might find that individuals with higher levels of stress are more likely to develop memory problems. While the authors caution that this is only a correlation and not evidence of causation (by including a disclaimer such as "correlation does not equal causation"), they might later suggest in the discussion section that reducing stress could prevent memory impairments, implying a causal relationship without explicitly stating it. This contradictory practice, where researchers both caution against and suggest causal interpretations within the same paper, creates a confusing narrative.

Moreover, in psychology, there is a general reluctance against making explicit causal inferences from non-experimental data. Grosz et al. (2020) termed this the *motte-and-bailey strategy*, where strong causal claims are initially made but then retreated from when challenged, such as during peer review, follow-up studies or press releases. For example, researchers may initially claim that self-control causes improved health, wealth, and public safety—the bailey, but when challenged, retreat to the safer position that self-control predicts these outcomes instead—the motte (Moffitt et al., 2011, cited in Grosz et al., 2020). Similarly, they may describe a study of personality and social roles as showing that investments in social roles cause changes in conscientiousness—the bailey, but later frame the findings as an association without asserting causation—the motte (Lüdtke et al., 2011, cited in Grosz et al., 2020). Some researchers defend this taboo and argue that non-experimental findings should never be used to make causal inferences (e.g., Adams et al., 2017; Cofield et al., 2010), a viewpoint that remains prevalent.

In his 2018 commentary, Miguel Hernán contends that avoiding causal language in observational research harms scientific progress. According to Hernán, researchers and journal editors often avoid terms like “causal effect” and use ambiguous terms like “association” or “link” in reporting. Hernán argues that when the causal objective is not made explicit, readers may struggle to determine whether the chosen methods are appropriate for addressing the research question, as it is unclear whether the research questions involve causal effects. Subsequently, failing to specify causal objectives can lead to unadjusted or confounded results being misinterpreted as causal effects. Hernán concluded that clearly defining the causal goals of a study can improve the quality of research by reducing ambiguity and preventing misinterpretation of results.

Overall, the studies above indicate that translating research findings to causal understanding is not a straightforward endeavour. However, as suggested by Hernán (2018), avoiding direct discussion regarding causality by avoiding causal terminology can actually lead

to further ambiguity and misinterpretation. Decision-makers and clinicians depend on researchers to report and communicate findings reliably; thus, research communication is an important part of evidence-based practice. This is especially important considering misinterpretations can lead to errors in treatment approaches, especially among vulnerable groups (Thapa et al., 2020).

### **1.5 Distortion of Causal Evidence in The Media**

The communication chain from research findings to news articles presents many opportunities for loss of scientific nuance. Each step, progressing from results and evidence to authors' conclusions, press releases, and eventually news articles, presents opportunities for causality claims to be misrepresented. Evidence from Sumner et al. (2014) supports this point, showing that 33% of press releases for peer-reviewed journal articles with an observational design from 20 leading universities in the United Kingdom contained exaggerated causal claims. Furthermore, 81% of news stories related to these press releases also contained similar exaggerations. This exaggeration in media reporting likely arises from a desire to engage readers by employing more sensational language.

Similarly, Haber et al. (2018) found that a significant proportion of media articles (approximately 44%) exaggerated the strength of causal claims presented in the original academic research. This exaggeration is particularly pronounced in the use of strong causal language that is not supported by the methodological rigour of the studies, especially when dealing with observational designs. Additionally, in a retrospective analysis of news articles in the German medical journal *Deutsches Ärzteblatt*, 45% of news headlines exaggerated the causality claims from scientific studies, often without substantial evidence from RCTs (Buhse et al., 2018). The exaggeration in media coverage then contributes to an inflation of the overall strength of evidence.

Studies also suggest that the public often conflates causation and correlation. In a series of experiments by Adams et al. (2017), participants from the general public generally viewed news headlines as at least moderately causal, even when the headlines used classically non-causal expressions such as “is associated with” or “is related to”. Interestingly, educational background did not statistically significantly affect participants’ understanding, suggesting that people rely on common-sense ideas about causality. Similarly, Bleske-Rechek et al. (2015) showed that people often infer causation from correlation scenarios, regardless of study design, and relying on the direction of the association and intuition. For instance, their study presented a vignette depicting the relationship between video games and aggression, in which participants were more likely to infer that video game playing causes an increase in aggression than to interpret the relationship as merely correlational, even when the study design did not support a causal interpretation. Overall, the potential for exaggeration in media reporting means that journal articles should communicate internally consistent information about the causal relationships under examination, which in turn should also be firmly supported by the evidence and findings presented.

### **1.6 Causal Inference and Practice Recommendations in Observational Studies**

Academic research makes practice recommendations, and this holds true for any research methods, including observational designs. These recommendations are usually located in the *Discussion* and *Conclusion* sections of research papers, serving as a link between research findings and their potential applications. In the field of psychology, practice recommendations may vary, ranging from suggesting specific, low-level interventions for individuals facing mental health challenges to advocating for high-level changes in systems and policies based on psychological insights.

Several metascientific studies have offered important insights into the interplay between research findings, causal inference, and practice recommendations. Prasad et al. (2013)

conducted an analysis of 298 high-impact medical journals published in 2010. Their results revealed that a substantial portion of authors (56%) conducting observational studies offered recommendations for medical practice. They observed this tendency of extending findings to practice recommendations across various study designs, including those related to incidence or prevalence, testing, and correlations with non-modifiable risk factors. Prasad et al. noted that practice recommendations often depend on causal inferences, but when researchers are unclear about whether they are making causal claims, it becomes more challenging for readers to assess whether these recommendations are justified by the study's findings.

Another study by Alvarez-Vargas et al. (2020) examined the effects of using ambiguous statistical language versus explicit causal language in presenting results from observational studies. The researchers aimed to determine if the type of language used influenced how the research quality and policy recommendations were perceived. In their preregistered experiment involving 142 academics, including faculty members, post-doctoral researchers, and university students, participants reviewed hypothetical study abstracts that used either of the two linguistic approaches. These abstracts concluded with policy recommendations based on causal interpretations.

The study found that participants rated abstracts using causally ambiguous statistical language as higher in quality compared to those using explicit causal language. However, for policy recommendations, participants rated abstracts with explicit causal language as comparable to or of higher quality than those with ambiguous language. These findings highlight the impact of language choice on the perceived quality of a study and the evaluation of policy recommendations. Interestingly, it is possible that authors are aware that ambiguous language is preferred by readers. Authors may, therefore, deliberately write ambiguously to cater to this preference. This raises the question of whether studies that use explicit or clear causal

language are judged more harshly by readers compared to those that use more ambiguous language.

### **1.6.1 Linking Language in Observational Studies**

Informed by the two studies above, Haber et al. (2022) explored the use of causal language in observational health research, which includes the medical, public health, and epidemiology literature. They focused on *linking language*, defined as the language that connects the *exposure* (the independent variable of interest being investigated for a possible non-causal link to the study outcome or end-point variable) and the *outcome* (the dependent or effect variable of interest that is being investigated for a possible link to the exposure). To conduct their data analysis, the researchers reviewed 1170 articles from 18 esteemed academic journals.<sup>5</sup> Additionally, they analysed the practice recommendations<sup>6</sup> provided in these articles.

Haber et al. (2022) revealed interesting findings. While terms like "associate" were most frequently used (45.7% of cases), the level of causality conveyed in these sentences varied widely. Surprisingly, practice recommendations in these articles often carried a higher strength of causal implication than what was explicitly stated in the linking language. This means that authors may, at times, provide practice recommendations in their work, suggesting actions or interventions based on their findings. However, the level of causality implied in these recommendations often exceed what is explicitly stated in the linking language connecting the exposure and the outcome.

For example, in a study by Potter et al. (2014; as reviewed in Haber et al., 2022), the primary linking sentence in the abstract section states, "Enrollment in sixth grade in 2010 and

---

<sup>5</sup> Two reviewers initially rated these articles independently, and a third reviewer acted as an arbitrator to reconcile their ratings and offer their own judgement. These ratings involved subjective assessments, given the nuanced nature of language interpretation.

<sup>6</sup> Haber et al. (2022) have used the term *action recommendations* instead of practice recommendations in their article. These terms can be used interchangeably.

coadministration of adolescent vaccines at the first adolescent visit were significantly associated with completion of the vaccines required for Michigan's sixth graders" (p. 1526). This linking sentence does not imply a causal relationship between coadministration of vaccines (exposure) and vaccine completion (outcome). The authors further recommended that: "Education and outreach efforts, particularly regarding the importance and benefits of coadministration of all recommended vaccines in adolescents, should be directed toward health care providers, parents, and adolescents" (p. 1526). However, this practice recommendation would require a causal relationship between coadministration of vaccines (exposure) and vaccine completion (outcome) to be warranted, which does not align with the earlier linking sentence.

Mockett (2023) conducted a small-scale conceptual replication of Haber et al. (2022). Mockett's study adopted a preregistered approach and examined articles centred around psychological capital (combined psychological resources of hope, self-efficacy, resilience, and optimism). Mockett found that terms like *relate*, *influence*, *impact*, and *effect* were commonly used to describe relationships between predictors and outcomes. Even in articles that explicitly warned against drawing causal inferences, a majority of primary linking sentences still hinted at varying degrees of causality. Notably, Mockett's study discovered that in most cases, the causality implied in the practices recommendations was stronger than the causality implied in the linking sentences. Future research was advised to consider larger sample sizes and multiple reviewers to account for the subjective nature of causality ratings.

Overall, the studies above provided insight into the complex relationship between research findings, causal language, and practice recommendations in observational studies. Findings by Alvarez-Vargas et al. (2020) revealed how language choice influences the perceived quality and impact of research. Haber et al. (2022) and Mockett (2023) extended the discourse into the landscape of health and psychology, revealing a general misalignment between the causal implications of the linking language used when reporting results and that of

the subsequent practice recommendations. Further, both studies identified an implicit interest in causality among observational studies, deduced from the common discussions of causal mechanisms and the frequent adjustments for confounding variables in their samples.

### **1.7 Clinical Psychology as a Scientific Discipline**

Clinical psychology is a specialised field within psychology that involves the diagnosis and management of emotional and behavioural disorders throughout the lifespan (American Psychological Association, 2022). Clinicians in the field employ a variety of therapeutic techniques, such as cognitive-behavioural and dialectical behaviour therapy, to address issues ranging from depression and anxiety to other severe psychiatric disorders. The practice of clinical psychology includes assessment and intervention through the use of standardised tests, clinical interviews, and direct observation (Kazdin, 2008). In New Zealand, clinical psychologists have a focus on cultural competence and apply the *scientist-practitioner model*<sup>7</sup> to inform treatments, which are tailored to the local context (New Zealand Psychologists Board, 2018).

Scientific research articles are crucial in clinical psychology as they provide a key platform to share recent research findings, contributing to the knowledge base that supports clinical practice. Clinicians rely on scientific articles not only to inform their treatment choices, but also for identifying areas where more research is needed and guiding future research directions (Spring, 2007). Moreover, research findings can shape healthcare policies and mental health care standards by influencing policy decisions and clinical guidelines (Rahman 2017).

The role of scientific research in clinical psychology becomes more important as the field continues to gain prominence. Figure 1 displays the frequency of articles featuring the keywords “clinical psychology” published on Scopus (a multidisciplinary database of peer-reviewed

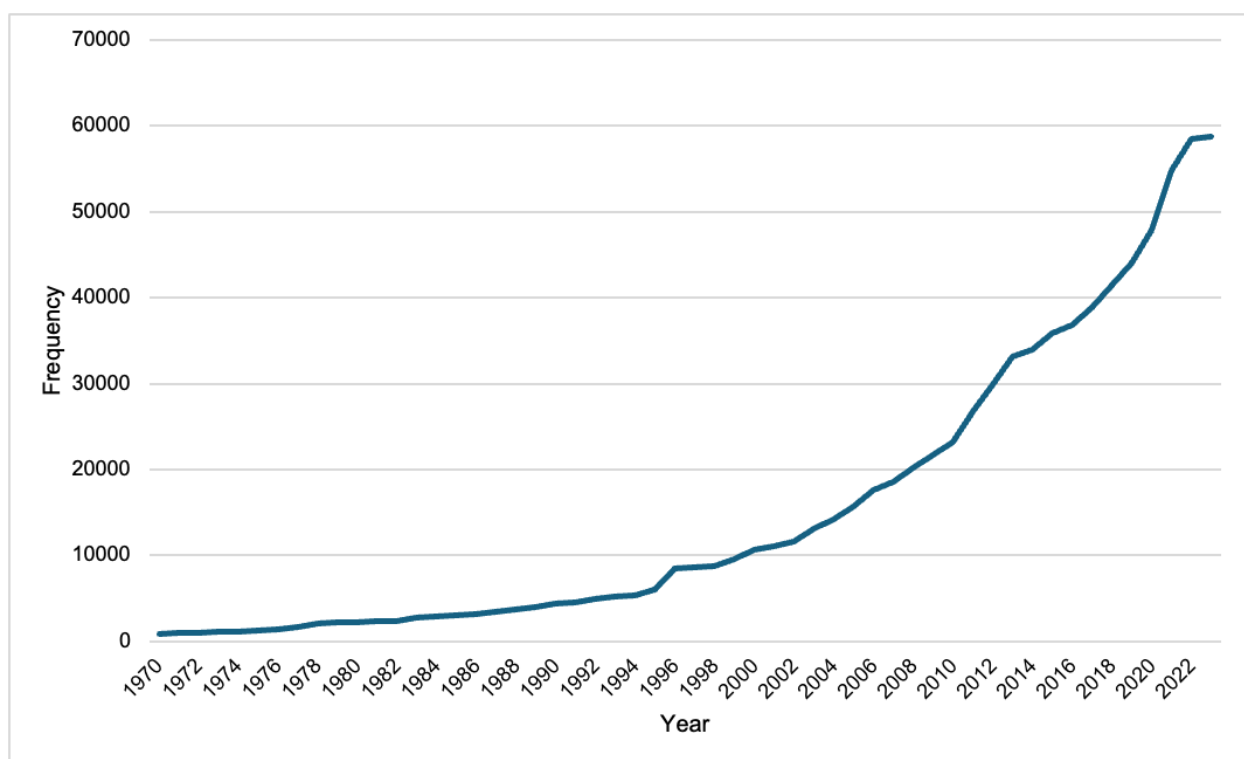
---

<sup>7</sup> The scientist-practitioner model is a framework that integrates clinical practice with scientific research for psychologists to study mental health phenomena and evaluate treatment (American Psychological Association, 2018).

literature) from 1970 to 2023. The data illustrates an increase in published works since 1970 and a notable surge starting around 1990. This pattern indicates a growth in research within clinical psychology over the last five decades and signifies an expanding body of knowledge and academic interest within this discipline.

### Figure 1

*Clinical Psychology Articles Published by Year*



*Note.* Frequency of articles featuring the keywords “clinical psychology” published on Scopus since 1970. Data obtained from Scopus search results analysis on 23 July 2024.

#### **1.7.1 Causal Language and Practice Recommendations in Clinical Psychology**

Clinical psychology addresses important and sensitive topics related to human behaviour, cognition, and emotions. Unlike some other scientific disciplines, many questions in clinical psychology are not amenable to experimental manipulation due to ethical concerns and

practical limitations. For example, researchers cannot ethically manipulate trauma or abuse to observe the effects on individuals. Therefore, researchers often rely on observational studies to explore complex relationships between variables. As discussed in the section 1.2.1 (Causal Inference in Observational Studies), observational studies can be useful for establishing causal claims, as long as they are supplemented by robust methods.

Additionally, clinical psychology research frequently goes beyond description to provide actionable recommendations. In the context of clinical psychology, causal inference becomes important because it informs aetiological models, prevention efforts, and potential interventions for psychiatric disorders (Ohlsson and Kendler, 2020). Moreover, improving outcomes for mental disorders requires research that can quickly and effectively identify causes, and a diagnostic system that can incorporate new causal knowledge (Saxe et al., 2022). Ultimately, the goal is to translate research findings into actions or solutions that clinicians and policymakers can undertake in real-world settings.

However, as highlighted by several studies, studies with observational designs often face several challenges. First, authors often used language that suggests causal relationships when reporting results, despite warning against making causal inferences (Alvarez-Vargas et al., 2020; Grosz et al., 2020; Tennant & Murray, 2021). This indicates an inherent interest in causal reasoning, for instance, authors may ask implicit causal research questions or draw conclusions that depend upon causal inference. Second, the strength of the evidence for causal effects between variables tested does not reflect the practice recommendations that follow, as the practice recommendations tend to be exaggerated (Alvarez-Vargas et al., 2020; Haber et al., 2022; Mockett, 2023). This issue could plausibly be present in clinical psychology as well, which undermines the credibility of its evidence base. To the best of our knowledge, no studies have systematically investigated the interplay between causal language and practice

recommendations in the clinical psychology literature from a metascientific or methodological standpoint.

### **1.8 The Current Study**

This study serves as a conceptual replication of the seminal research by Haber et al. (2022), who examined the language used in prominent medical, public health, and epidemiology journal articles. Haber et al. (2022) explored how language implies causality, specifically in the context of linking language and practice recommendations. Further, they assessed whether the strength of causal implication in linking language aligns with the strength of causal implication in the subsequent practice recommendations.

According to Crandall and Sherman (2016), conceptual replication is “an attempt to test the same fundamental idea or hypothesis behind the original study, but the operationalizations of the phenomenon, the independent and dependent variables, the type and design of the study, and the participant population may all differ substantially” (p. 93). In line with Crandall and Sherman’s definition, our study sought to test the same core ideas presented by Haber et al. (2022), but within the specific field of clinical psychology.

The success of decision-making in clinical psychology relies heavily on the ability to draw causal inferences from research. Furthermore, the field has increasingly focused on applying research findings in real-world settings, often derived from observational study designs. Consequently, it becomes important to assess the language used and the causality implied in such studies. This is particularly crucial given the significant growth of the clinical psychology field in recent decades.

In this study, our primary objective was to discover the prevalence of linking language within the observational studies in the clinical psychology literature (primary objective 1). We also estimated how often the linking language explicitly states or implies causality (primary objective 2). By examining the linking language used in this literature and the extent of implied

causality, we can determine whether the common practice of avoiding causal language in other fields of observational research also applies to the clinical psychology domain. Additionally, we explored practice recommendations to see how frequently they inherently require causal inference (primary objective 3). When a practice recommendation depends on causal inference, it may be invalid if the study is not properly designed to support such reasoning.

Furthermore, we were particularly interested in the mismatches between the causal implications embedded in the linking language and the resulting practice recommendations (primary objective 4). We wanted to scrutinise the inconsistencies between the *explicit* claims about causal effects in research papers and the causal inferences *assumed* by their practice recommendations. A lack of statistically significant correlation between the strength of causal implications in linking language and in practice recommendations would suggest a disconnect between the two.

In addition to the main objectives above, we were interested in exploring how often studies in clinical psychology explicitly declare their intention to make causal inferences (secondary objective 1). We also examined how frequently these studies include explicit causal disclaimers and whether they still imply causality despite these disclaimers (secondary objective 2). If causal inference is rarely acknowledged or cautioned against but still appears in the research, it implies that authors might be reluctant to explicitly state causal reasoning or may not fully understand the risks associated with making such inferences. Last, we investigated the frequency with which studies employ explicit strategies to enhance the credibility of causal inferences, such as clarifying the intent behind controlling or adjusting for specific variables (secondary objective 3). By examining these objectives, this study strives to shed light on the nuanced interplay between language, causal inference, and practice recommendations in clinical psychology literature.

## 2.0 Methods

This study constituted a systematic and small-scale examination of the causal language patterns used in the clinical psychology literature. This study was preregistered on the Open Science Framework (OSF) platform

([https://osf.io/wv23p/?view\\_only=dde5eded2af541e3a8ffccaa3f8ee41](https://osf.io/wv23p/?view_only=dde5eded2af541e3a8ffccaa3f8ee41)). Preregistration is a practice of defining research questions and analysis plans prior to data observation (Nosek et al., 2018). It is commonly associated with hypothesis-testing research, and is applicable to secondary data analysis, such as in the current study. In both contexts, preregistration offers value for transparency and reproducibility (Nosek et al., 2018).

### 2.1 Phase 1: Journal Screening and Selection

The SCImago Journal Rank (SJR; SCImago, n.d.) served as the primary reference for journal selection. SJR is a publicly available platform presenting journals and country scientific indicators derived from the Scopus database. SJR allows journal groupings by subject area, subject category, or country, and has over 30,000 titles from 5,000 international publishers and metrics from 239 countries (SCImago, n.d.). Citation data is weighted by prestige, making SJR useful for comparing journals within the same field, forming the grounds for subject category ranking (Massey University, 2024). For in-depth insights into SJR metrics, Guerrero-Bote and Moya-Anegón's (2012) article is recommended. The current study focused on the Clinical Psychology subject category (as available on SJR). The goal was to select the top 97 articles within this subject category, each assessed against specified inclusion and exclusion criteria.

The first inclusion criterion was that the journals had to be listed under the Clinical Psychology category on the SJR website. At the time of data collection, SCImago indicated that its metrics were derived from Scopus data as of April 2023. Another inclusion criterion specified that the journals had to publish original research articles, excluding those that solely provide reviews or summaries of existing research (e.g., review and commentary journals), which

resulted in the exclusion of 14 journals. This determination was made by reviewing the scope descriptions on either the SJR webpage or the individual journal's webpage, which clarified whether a journal primarily publishes original research or focuses only on reviews or summaries of existing research articles. I further verified this by examining the types of articles published in the most recent issues of these journals to ensure compliance with this inclusion criterion. Additionally, journals that primarily publish studies on topics outside of the subdiscipline of clinical psychology (e.g., political psychology and psychometrics) were excluded ( $n = 5$ ). The full list of excluded and included journals is available in Appendix F.

The primary analyses in this study involved estimating population percentages with confidence intervals (refer to section 2.5 Data Analysis). The sample size for this review was determined by considering a confidence level of 95%, a margin of error set at 10%, and a population proportion of 50%. The margin of error was set at 10% to allow for a limited tolerance of sampling error while being practical. The population proportion of 50% is a conservative assumption because for proportions either larger or smaller than 50%, the margins of error will be smaller.

Due to resource constraints, including time and availability as a sole reviewer, I had to make sure that the sample size was not only manageable but also large enough to produce meaningful results (Lakens, 2022). The specifications above led to a calculated sample size of 97, determined using the Sample Size Calculator ([Calculator.net](https://www.calculator.net)). This tool computes the minimum number of samples required to meet the desired statistical parameters. The consideration of myself as a single reviewer for a Master's thesis made the selected sample size practical. This sample size was also sufficiently extensive to cover a diverse array of journal articles.

## **2.2 Phase 2: Article Screening and Selection**

In the second phase of the study, the screening and selection of articles were conducted with the aim of selecting one article from each journal. This process began with assessing the eligibility of articles starting from the first issue of 2023. If the first article assessed did not meet the inclusion criteria, the evaluation continued with subsequent articles within the same issue. If no eligible articles were found within the first issue of 2023, the screening process moved on to the next published issue, and this iterative process continued. In cases where no articles from 2023 satisfied the inclusion criteria, the screening process reverted to the last article of the last volume (in 2022) and proceeded in reverse order from there. This continued until the predetermined sample size of 97 articles were met. Overall, one article published in 2019, and two articles published in 2022 were selected while the remaining ( $n = 94$ ) were published in 2023.

### **2.2.1 Article Inclusion Criteria**

Articles were selected based on the following criteria. First, they had to be in the English language. Second, they needed to report studies with an observational design, including cross-sectional and longitudinal designs. In particular, the studies had to not involve random assignment of participants to different conditions or interventions. Third, they had to focus on human or aggregate group observations. Fourth, the primary research question had to be quantitatively examined using primary data, excluding reviews or meta-analyses. Fifth, the research question had to pertain to the causal and/or non-causal association between one primary predictor concept and one primary outcome concept. While studies could have discussed multiple predictors and outcomes, emphasis had to be placed on one predictor and outcome pair as the primary association of interest, as reflected in the title and abstract. Last, articles had to be published, either in an issue or as an advance online publication, excluding preprints.

### **2.2.2 Article Exclusion Criteria**

Conversely, articles were excluded if they reported studies with experimental or quasi-experimental designs. We do not consider one-group pretest-posttest designs as quasi-experiments; hence, such studies were not excluded. Second, studies with secondary study designs such as reviews or meta-analyses were also excluded. Third, studies that primarily focused on mediators, moderators, or interaction effects were excluded. This exclusion was due to the complexity it would introduce to data management and review, as well as the inability to assign single relationships per paper and compare study objectives and designs with other papers. Last, studies that investigated topics outside the subdiscipline of clinical psychology were excluded. This exclusion was verified by thoroughly reading the titles and abstracts of each article. Specifically, studies that maintained a strict focus on areas unrelated to clinical psychology, such as those concentrating solely on physical illnesses or neurology, were excluded. We acknowledge that this criterion required abstract decision-making regarding the relevance of each study to the field of clinical psychology, potentially introducing a degree of subjectivity into the selection process.

**Table 1**

*Summary of Inclusion and Exclusion Criteria for Journals and Articles*

Criterion	Inclusion Criteria	Exclusion Criteria
Type		
Journals	Listed on SJR under Clinical Psychology subject. Publishing original articles.	Publishing studies outside the subdiscipline of clinical psychology (e.g., political psychology, psychometrics).

---

Articles	<p>English language.</p> <p>Observational designs (cross-sectional and longitudinal studies). Focusing on human or aggregate group observations.</p> <p>Examining primary research questions quantitatively using primary data.</p> <p>Exploring causal and/or non-causal associations between one primary predictor concept and one primary outcome concept.</p> <p>Published in an issue or as an advance online publication.</p>	<p>Providing only reviews or summaries of existing research articles.</p> <p>Experimental or quasi-experimental designs.</p> <p>Investigating topics outside the subdiscipline of clinical psychology.</p> <p>Reporting secondary study designs (e.g., reviews, meta-analyses).</p> <p>Focusing primarily on mediators, moderators, or interaction effects.</p>
----------	---	---

---

### 2.3 Phase 3: Data Extraction Using a Review Tool

A review tool developed and adapted from Haber et al. (2022) was used in this study. This review tool was designed to provide replicable and subjective assessments of journal articles. For training and piloting purposes, 10 clinical psychology articles from the years 2010 to 2013 were chosen arbitrarily. This process also helped refine the initial review tool. A copy of the final version of the review tool is available in Appendix E.

The data extraction process, conducted solely by me as the primary researcher (and reviewer), involved several steps. First, I collected an article's Digital Object Identifier (DOI) and title for identification purposes. Next, the predictor and outcome variables were recorded. In the abstract section, I extracted the primary linking sentence, which contained the key linking word or phrase, along with any modifying words or phrases. Following this, I rated the extent to which the primary linking sentence implied a causal relationship between the predictor and outcome variables. Additionally, I recorded any practice recommendations mentioned in the abstract and rated the extent to which these recommendations would require that a causal relationship had been identified.

Moving to the concluding section, I recorded the primary linking sentence and any practice recommendations. The concluding section is typically the very last section of an article and is usually titled *Conclusion*. However, in some cases, the *Discussion* section serves as the concluding section. I prioritised the first paragraph of the Conclusion section, and if no suitable sentence was found there, I examined subsequent paragraphs until I identified a primary linking sentence or a practice recommendation. If neither was present in the Conclusion section, or if there was no Conclusion section, I checked the Discussion section. I also rated the causal implication of the primary linking sentence and practice recommendations.

Additionally, I recorded any causal disclaimer statements and acknowledgments of intent to draw causal inference if they were present in the concluding section. Finally, if available, I recorded any discussion of confounders in relation to the methods, results, or interpretation of the study found anywhere in the article. Overall, the review tool contains a series of 20 questions.

### **2.3.1 Predictor and Outcome Variables**

Short descriptions of the predictor and outcome variables were directly taken from either the title or abstract, with priority given to the title when available. The predictor was defined as

the primary variable being investigated for its potential link to the outcome and was also referred to as the independent variable.<sup>8</sup> The outcome, on the other hand, was defined as the dependent variable under investigation for its relationship with the predictor. While some studies discussed multiple predictors and outcomes, our focus was on a single predictor-outcome pair, as specified in the article inclusion criteria above. This emphasis was typically reflected in the title and abstract. For example, in the article titled “The Association Between Perceived Relationship Quality and Psychological Symptoms in Refugees” by Kobayashi and Berle (2023), “perceived relationship quality” was identified as the predictor variable, and “psychological symptoms” was identified as the outcome variable.

### **2.3.2 Primary Linking Sentence**

The *primary linking sentence*, which connects the predictor and outcome variables, was identified. This sentence typically contains the predictor, outcome, a linking word, and any modifying phrases. It was collected for both the abstract and the concluding section of an article. If multiple sentences met the criteria in either section, the one making the strongest causal claim was selected. While Haber et al. (2022) chose the first causal claim in a section, this study opted to select the sentence with the strongest implication of causality. We consider that choosing the sentence with the strongest implication of causality offers greater insight because it captures the maximum potential impact that researchers attribute to the relationship between variables within their study. This approach also aligns with Mockett (2023).

### **2.3.3 Primary Linking Word/Phrase**

The primary linking word or phrase was selected based on the primary linking sentence identified above. If multiple words or phrases were present, the one implying causality most

---

<sup>8</sup> The definitions provided in this Methods section were obtained from Haber et al. (2022). These definitions were amended to suit the current research project.

strongly was chosen. A primary linking word was defined as a word describing the nature of the connection between a specified predictor and a specified outcome, typically found in a sentence containing both predictor and outcome. Before data analysis, the root of each linking word was manually identified. Table 2 demonstrates the process in determining each primary linking word, and its corresponding root word and modifiers (if available). The provided examples are sourced directly from the data collected during this study.

### 2.3.4 Modifying Word/Phrase

Modifiers, which are words or phrases that affect the nature of the primary linking word or phrase, were identified based on the primary linking sentence. These modifiers can influence various aspects such as strength, intensity, degree of uncertainty, negotiation, and direction. For example, *significant* serves as a modifier in the following sentence from Dias et al. (2023): “Previous experience was not found to have a *significant* effect on attitudes, therapeutic presence, or working alliance” (p. 439).

**Table 2**

*Examples of Root Word Determination*

Primary Linking Sentence	Primary Linking Word/Phrase	Root Word	Modifier
“There were statistically significant improvements in levels of authenticity, well-being and psychological distress over 20 sessions of therapy; a minimum of 15 sessions were required for significant change to be observed.” (Sohal & Murphy, 2023, p. 20)	improvements	improve	statistically significant

---

“Previous findings that individuals with	greater	great	N/A
--	---------	-------	-----

high trait worry endorse greater tendencies to self-report unconstructive problem-solving attitudes were replicated.” (Szabo et al., 2023, p. 1)

---

### **2.3.5 Linking Language Causal Implication**

The extent to which primary linking words or phrases (hereinafter will be referred to as *linking language*) implied causality within their respective sentences was rated using an ordinal categorical scale. This scale comprised three levels: *None*, indicating that the linking sentence implied no causal relationship; *Moderate*, indicating that the linking sentence mostly implied a causal relationship but remained ambiguous without explicit causal inference; and *Strong*, indicating that the linking sentence clearly implied causality. This three-point scale differed from the four-point scale used in Haber et al. (2022) and Mockett (2023), which included *None*, *Weak*, *Moderate*, and *Strong* responses. For this study, we decided to eliminate the *Weak response*, as it was conceptually similar to *Moderate*. This simplification helped streamline the rating and data analysis process.

Causal language in this study was defined as language suggesting one entity’s influence on another, expressed through verbs that imply movement in outcomes by predictors (e.g., increase, decrease, improve, change) and conjunctions attributing outcomes to predictors (e.g., because, due to, since). Table 3 illustrates this rating scale, providing definitions and examples for each level based on data collected during this research.

**Table 3***Examples of Rating Scale for Linking Language Causal Implication*

Causal Implication Rating Strength	Definition	Example of Primary Linking Sentence
None	The linking sentence does not imply in any way a causal relationship was identified.	“Chronic and deteriorated global sleep problems is associated with a greater risk of suicidal ideation” (Mao et al., 2023, p. 279)
Moderate	The linking sentence, while mostly implying a causal relationship, remains unclear or inconclusive in the absence of any causal inference.	“ED [eating disorder] severity decreased from pre- to posttreatment and at 1-year follow-up with a large effect size” (Levinson et al., 2023, p. 14)
Strong	The linking sentence clearly implies that causality has been identified.	Pre-school use of behavioural services was not found to affect outcomes during later childhood (Tsiplova et al., 2023, p. 2).

As Haber et al. (2022) explain, the absence of statements that have causal implications does not necessarily imply there are no effects. Causal implications should be considered separately from effect sizes. Conversely, the presence of causal implications does not necessarily imply there are effects. Strong causal implications can exist even with null effect sizes, as long as the language indicates a causal examination or relationship. As exemplified by

Tsiplova et al. (2023) in Table 3 above, even though the outcome variable showed no significant difference in relation to the predictor variable, the linking sentence still conveyed strong causal implications.

### **2.3.6 Practice Recommendations**

Practice recommendations were identified for both the abstract and concluding sections of each article, where available. Practice recommendations are descriptions of how a consumer of research might use the results and conclusions of the research. For example, authors may suggest that changing the amount of a predictor could be beneficial, assuming the predictor has a causal effect on the outcome. Alternatively, it may be suggested that a predictor need not be changed, assuming the absence of a causal effect has been established. For the purpose of this study, we excluded any practice recommendations that involved requesting additional research. If there were multiple recommendations, the one that most strongly implies causality was chosen.

### **2.3.7 Practice Recommendation Causal Implication**

The degree to which the practice recommendations imply causality was rated on an ordinal categorical scale. This scale comprises three levels: None, Moderate, and Strong. *None* indicates that the practice recommendation would be appropriately made in the absence of any causal relationship. *Moderate* suggests that the practice recommendation might be appropriate if a causal relationship had been identified, but it remains unclear or challenging to arrive at that recommendation without any causal inference. *Strong* implies that the practice recommendation could only be appropriately made if a causal relationship had been identified. Table 4 defines the rating scale, providing examples of practice recommendations rated at each level, directly drawn from the data gathered in this study.

**Table 4***Examples of Rating Scale for Practice Recommendations Causal Implication*

Causal Implication Rating Strength	Definition	Example of Practice Recommendation
None	The practice recommendation would be made appropriately in the absence of any causal relationship.	“Better understanding of cognitive and adaptive functioning in children with FASD [fetal alcohol spectrum disorders] may help in the design of tailored evidenced-based interventions”  (Martinez-Cengotitabengoa et al., 2023, p. 127)
Moderate	The practice recommendation may be made appropriately had a causal relationship been identified, but it is unclear or difficult to come to that recommendation in the absence of any causal inference.	“Monitoring dissociation and providing skills to regulate dissociative states may help to prevent symptom deterioration in traumatized individuals” (Guzman Torres et al., 2023, p. 4)
Strong	The practice recommendation could only be made appropriately had a causal relationship been identified.	“People with weaker mental capacity should avoid following relevant media information for a long time” (Zhang et al., 2023, p. 167)

### **2.3.8 Additional Information**

The final questions on the review tool assessed whether there were explicit causal disclaimer statements within the concluding section of an article, and whether there was an acknowledgment of intent to draw causal inference in that section. No strict criteria were applied to these steps; the goal was simply to determine whether a causal disclaimer statement or an acknowledgment of the intent to draw causal inference was available.

The very last question on the review tool sought an explicit discussion of confounders throughout an article, particularly in relation to the methods, results, and interpretation of the study. The goal was to identify any mention or discussion of confounders in the article. Therefore, terms such as “confounding”, “confounders”, or “third variable” related to the methods, results, and/or interpretation of the study were searched using the text-search function (Ctrl + F). This included the search term “confound\*” where \* could represent any suffix. To qualify, the terms had to be explicitly referenced in the context of the methods, results, or interpretation of the study, not in relation to other studies.

## **2.4 Review Procedures**

During the data extraction process, I, as the primary researcher independently reviewed papers and extracted data, and articles were reviewed in a random order. Additionally, a keyword scan using the *text-search* function assisted in the data extraction process. The *Results* sections were not examined as part of the data extraction process.

## **2.5 Data Analysis**

The analysis method used in this study was essentially a form of quantitative content analysis. Quantitative content analysis involves the systematic coding and quantification of content, including written material, as done in the current study (Huxley, 2020). The statistical analysis for this study primarily employed descriptive methods, with any missing data recorded as N/A and listwise exclusion applied. Proportions data were presented with a frequency count

and a 95% confidence interval. All analyses were performed within the R software (R Core Team, 2023). I also used the DescTools package for descriptive statistics in the analyses (Signorell et al., 2023). Each analysis below was conducted separately for the abstract and concluding sections.

The first primary objective was to identify the linking language. The proportions of linking language, categorised based on their root words, were calculated. Although not initially declared in the preregistration, the proportions of linking language with modifiers were calculated as an exploratory analysis.

The second primary objective aimed to generate estimates of the degree to which linking language states or implies causality. The frequency and proportions of linking language according to three categories/levels of causal implications (none, moderate, and strong) were calculated.

The third primary objective was to examine the prevalence of practice recommendations that would require causal inference to have been made. The proportions of practice recommendations' strength according to three categories/levels of causal implications (none, moderate, and strong) were calculated.

The fourth primary objective was to examine the prevalence of misalignment between linking language and practice recommendations. Three analyses were conducted to meet this objective, as described below.

1. The proportions of the causal strength of linking language that correspond with the causal strength of practice recommendations were calculated.
2. The proportion of articles in which the causal strength of the practice recommendations is stronger than the causal strength of the linking language was calculated.

3. The relationship between the causal strength of linking language and the causal strength of practice recommendations using a two-tailed Spearman's rho rank test at a 5% alpha level was calculated.

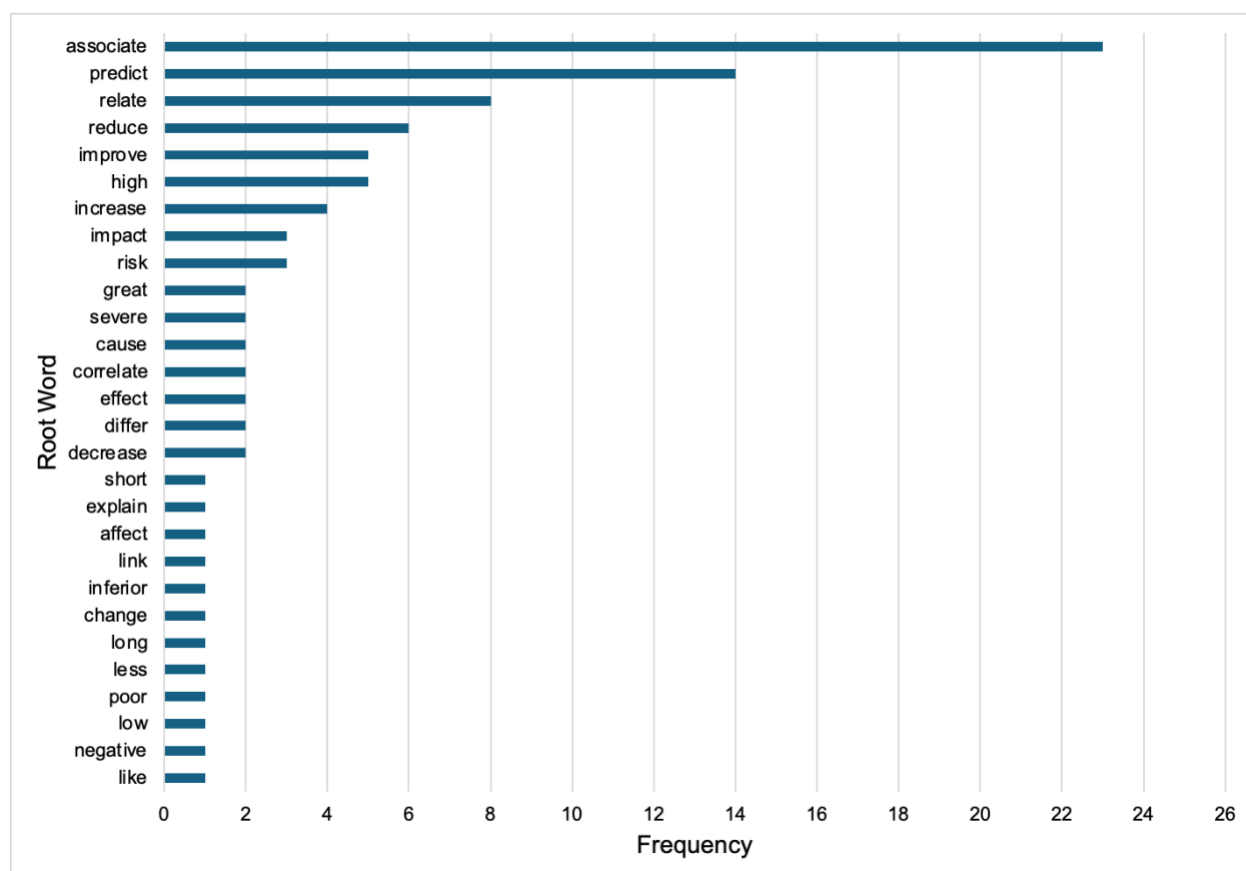
For the secondary objectives, three calculations were conducted. First, the proportions of studies that clearly stated their intention to make causal inferences (in the concluding section) were calculated. Second, the proportions of studies that offered explicit causal disclaimer statements (in the concluding section) were calculated. Last, the proportions of studies that discussed confounding factors in relation to their methods, results, and/or interpretation (anywhere throughout the article).

No ethics or low-risk notifications were pertinent to these analyses. The analyses were considered out of scope for Massey University's ethics code because they did not involve research participants, either human or animal. All materials and data are accessible on the OSF repository ([https://osf.io/pqtxd/?view\\_only=52a55a65cafc4b0285093596a791b0b6](https://osf.io/pqtxd/?view_only=52a55a65cafc4b0285093596a791b0b6)).

### **3.0 Results**

#### **3.1 Linking Language**

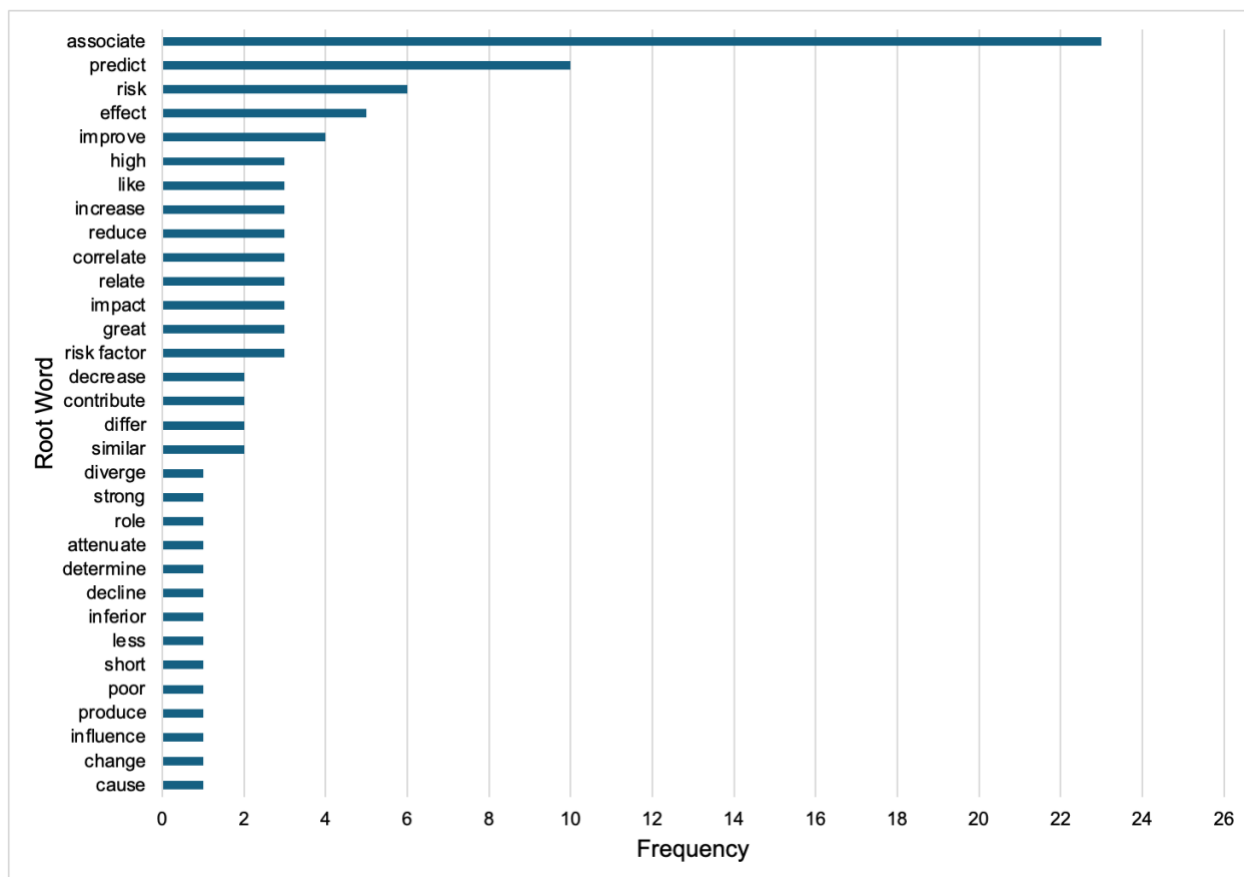
In the abstracts, the most common root word of the primary linking language (linking words or phrases) was "associate", appearing in 23.7% of cases ( $n = 23/97$ , 95% CI [15.0, 32.0]). The second most common was "predict", found in 14.4% of cases ( $n = 14/97$ , 95% CI [6.2, 23.2]), followed by "relate", found in 8.2% of cases ( $n = 8/97$ , 95% CI [0.0, 17.0]). Notably, the word "cause" was used in 2.1% of cases ( $n = 2/97$ , 95% CI [0.0, 10.8]). Figure 1 shows the frequency of root words identified in the primary linking language of abstracts.

**Figure 1***Frequency of Root Words in Abstracts*

In the concluding sections, “associate” was the most common root word of the primary linking language, appearing in 23.7% of cases ( $n = 23/97$ , 95% CI [16.5, 32.8]). The second most common was “predict”, found in 10.3% of cases ( $n = 10/97$ , 95% CI [3.1, 19.4]), followed by “risk”, found in 6.2% of cases ( $n = 6/97$ , 95% CI [0.0, 15.3]). The word “cause” was used in 1.0% of cases ( $n = 1/97$ , 95% CI [0.0, 10.1]). Figure 2 shows the frequency of root words in the primary linking language of the concluding sections.

**Figure 2**

*Frequency of Root Words in the Concluding Sections*



In the abstracts, modifiers were present in 43.3% of cases ( $n = 42/97$ , 95% CI [34.0, 53.9]). In the concluding sections, modifiers were present in 38.1% of cases ( $n = 37/97$ , 95% CI [28.9, 48.2]). “*Significant\**” was the most common modifier of the primary linking sentences in both abstracts and the concluding sections.

### 3.2 Strength of Causal Implication

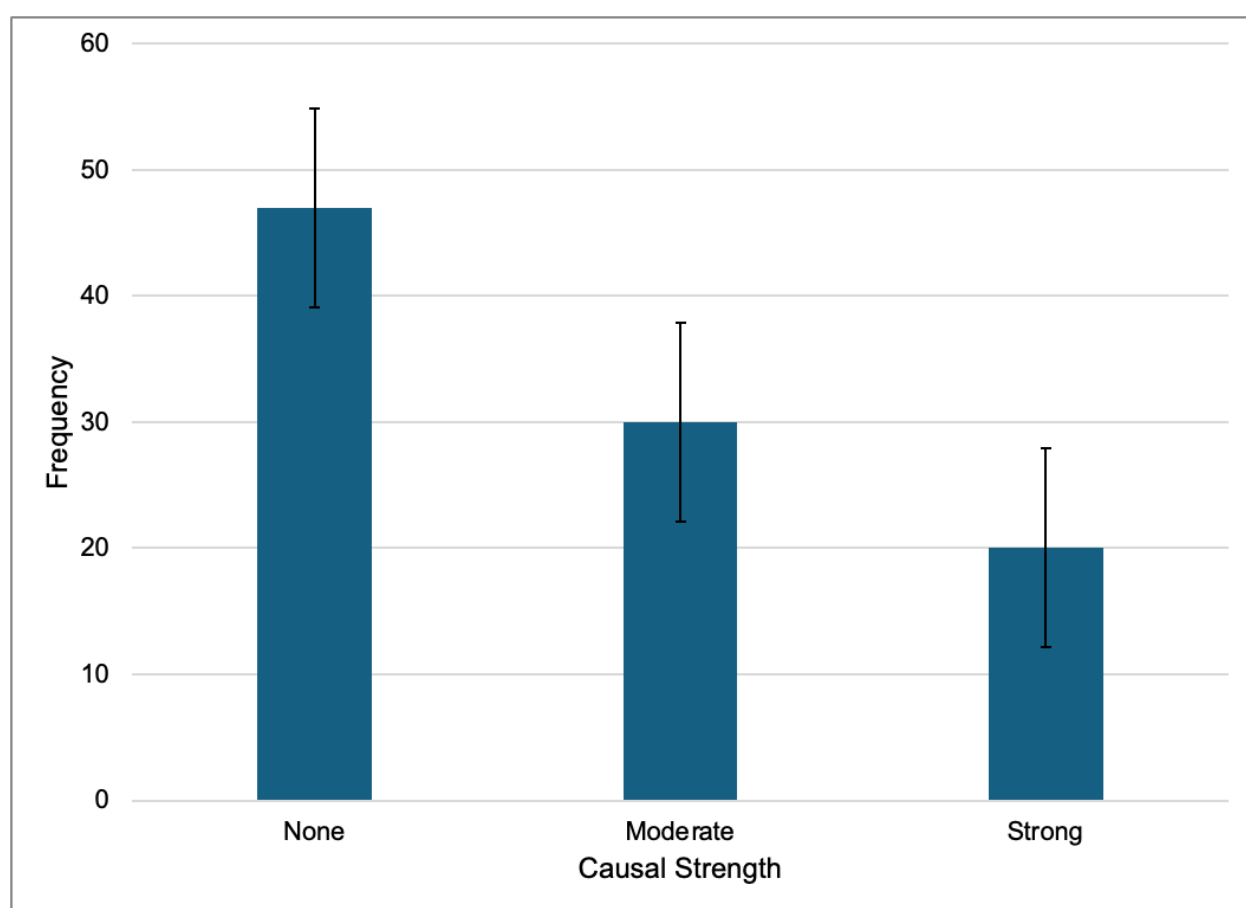
#### 3.2.1 Strength of Causal Implication – Abstracts

As provided above, the linking language was analysed to identify and group the root words. However, the causal implications of the linking language were determined by the ratings of their respective linking sentences. The strength of causal implications in the primary linking

sentence of the abstracts varied, with 48.5% ( $n = 47/97$ , 95% CI [38.1, 58.8]) of the articles rated as having no causal implication, 30.9% ( $n = 30/97$ , 95% CI [20.6, 41.3]) rated as having a moderate causal implication, and 20.6% ( $n = 20/97$ , 95% CI [10.3, 30.9]) rated as having a strong causal implication. Figure 3 shows the strength of causal implications in linking sentences of abstracts.

### Figure 3

*Strength of Causal Implication in Linking Sentences of Abstracts*



*Note.* The error bars represent the standard error of the frequency bars.

As discussed above, the most frequently identified root words in the linking sentences were “associate”, “predict”, and “relate”. For the root word “associate”, 60.9% ( $n = 14/23$ , 95%

CI [43.5%, 81.2%]) of instances had no causal implication. An example is, “Findings largely confirm earlier research, suggesting that CPTSD [complex post-traumatic stress disorder] is associated with traumatic events that start earlier in life and are perpetrated by acquaintances” (Torres et al., 2023, p. 1). “Associate” had a moderate causal implication in 39.1% ( $n = 9/23$ , 95% CI [21.7%, 59.5%]) of the articles, for example, “In a multiple linear regression model, mistrust is associated with the endorsement of a conspiracy mentality, even when accounting for other contributing factors” (Brauner et al., 2023, p. 55). The causal implication in this example is considered moderate because although “associate” suggests a link without asserting causation, the linking sentence includes an effort to account for confounding factors in the regression model, which strengthens the causal implication. No articles used “associate” with a strong causal implication.

The root word “predict” was used with no causal implication in 57.1% ( $n = 8/14$ , 95% CI [35.7%, 83.6%]) of cases. An example is, “For men only HA [harm avoidance] was a significant predictor of depression” (Ahola et al., 2023, p. 1). It had a moderate causal implication in 42.9% ( $n = 6/14$ , 95% CI [21.4%, 69.3%]) of the articles, such as in, “Greater youth engagement in SMA [screen media activity] at ages 9–10 years statistically predicted higher levels of internalizing two years later” (Zhao et al., 2023, p. 80). In this example, the causal implication is considered moderate due to the clear attempt of establishing temporal precedence with the phrase “statistically predicted” and the mention of a two-year gap. However, the language remains cautious, implying prediction rather than direct causation, and does not acknowledge confounding variables, hence the moderate rating. There were no instances of “predict” being used with a strong causal implication.

The root word “relate” had no causal implication in 87.5% ( $n = 7/8$ , 95% CI [75%, 100%]) of cases. An example of this is, “Participants generally showed heightened facial disgust towards overweight morphs, yet the strength of facial disgust was unrelated to ED [eating

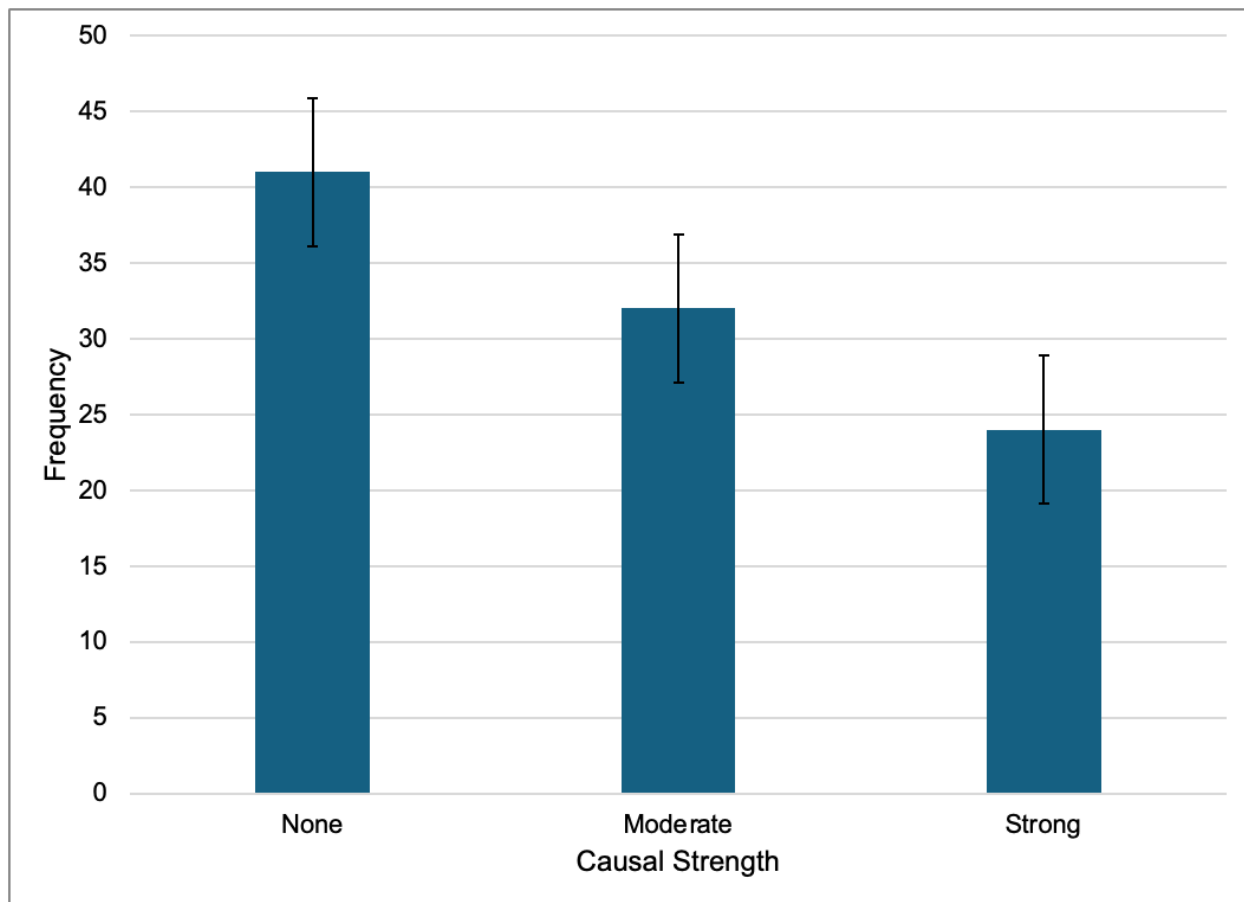
disorder] symptoms” (Masselman et al., 2023, p. 98). It had a moderate causal implication in 12.5% ( $n = 1/8$ , 95% CI [0.0%, 32.9%]) of cases, such as, “Results indicate that sexual orientation was significantly related to seeking psychological care and help from a victim services agency, and the relationship between individual and incident level characteristics varies by type of help-seeking among victims of domestic violence” (Martin et al., 2023, p. 156). There were no instances of “relate” being used with a strong causal implication.

### **3.2.3 Strength of Causal Implication – Concluding Sections**

The strength of causal implications in the primary linking sentence of concluding sections showed similar variability as the abstracts. Of the articles, 42.3% ( $n = 42/97$ , 95% CI [32.0, 53.1]) were rated as having no causal implication, 33.0% ( $n = 32/97$ , 95% CI [22.7, 43.8]) were rated as having moderate causal implication, and 24.7% ( $n = 24/97$ , 95% CI [14.4, 35.5]) were rated as having strong causal implication. Figure 4 illustrates the strength of causal implications in the linking sentences of concluding sections.

**Figure 4**

*Strength of Causal Implications in Linking Sentences of Concluding Sections*



*Note.* The error bars represent the standard error of the frequency bars.

The most frequently identified root words in these sentences were “associate”, “predict”, and “risk”. For the word “associate” in concluding sections, 82.6% ( $n = 19/23$ , 95% CI [69.6%, 95.8%]) of instances had no causal implication. An example is, “In the current study, peer support services provided in SUD [substance use disorder] systems of care were associated with lower utilization of intensive acute care” (Hutchison et al., 2023, p. 5). This term had a moderate causal implication in 17.4% ( $n = 4/23$ , 95% CI [4.3%, 30.5%]) of the articles, such as, “Overall, we observed a positive, dose-response association between smoking status and

subjective cognitive decline among adults 45 years of age and older” (Rajczyk et al., 2023, p. 220). There were no instances of “associate” being used with a strong causal implication.

The term “predict” was used with no causal implication in 20.0% ( $n = 2/10$ , 95% CI [10.0%, 48.7%]) of cases. An example is, “The current investigation replicated findings from previous work showing that brief alcohol demand can predict drinking behavior in the natural environment” (Aston and Merrill, 2023, p. 118). It had a moderate causal implication in 80.0% ( $n = 8/10$ , 95% CI [70.0%, 100.0%]) of the articles, for example, “Central to our analysis, perceived relational empathy proved to be a significant negative predictor of PTSS, such that higher levels of staff’s empathy were associated with survivors’ lower PTSS” (Bassi et al., 2023, p. 56). There were no instances of “predict” being used with a strong causal implication.

For the word “risk”, 16.7% ( $n = 1/6$ , 95% CI [0.0%, 43.0%]) of its uses had no causal implication. An example of this is, “The results suggest that individuals who fall within an EHNI [evening hyperphagia with nocturnal ingestions] subtype are at a higher risk of experiencing other, more severe eating pathology, thus potentially being at higher risk for developing obesity” (Echeverri et al., 2023, p. 10). It had a moderate causal implication in 83.3% ( $n = 5/6$ , 95% CI [66.7%, 100.0%]) of cases, such as in, “Results showed that nurses with chronic or deteriorated global sleep problems had a higher risk of developing SI [suicidal ideation]” (Mao et al., 2023, p. 286). There were no instances of “risk” being used with a strong causal implication.

### **3.3 Practice Recommendations**

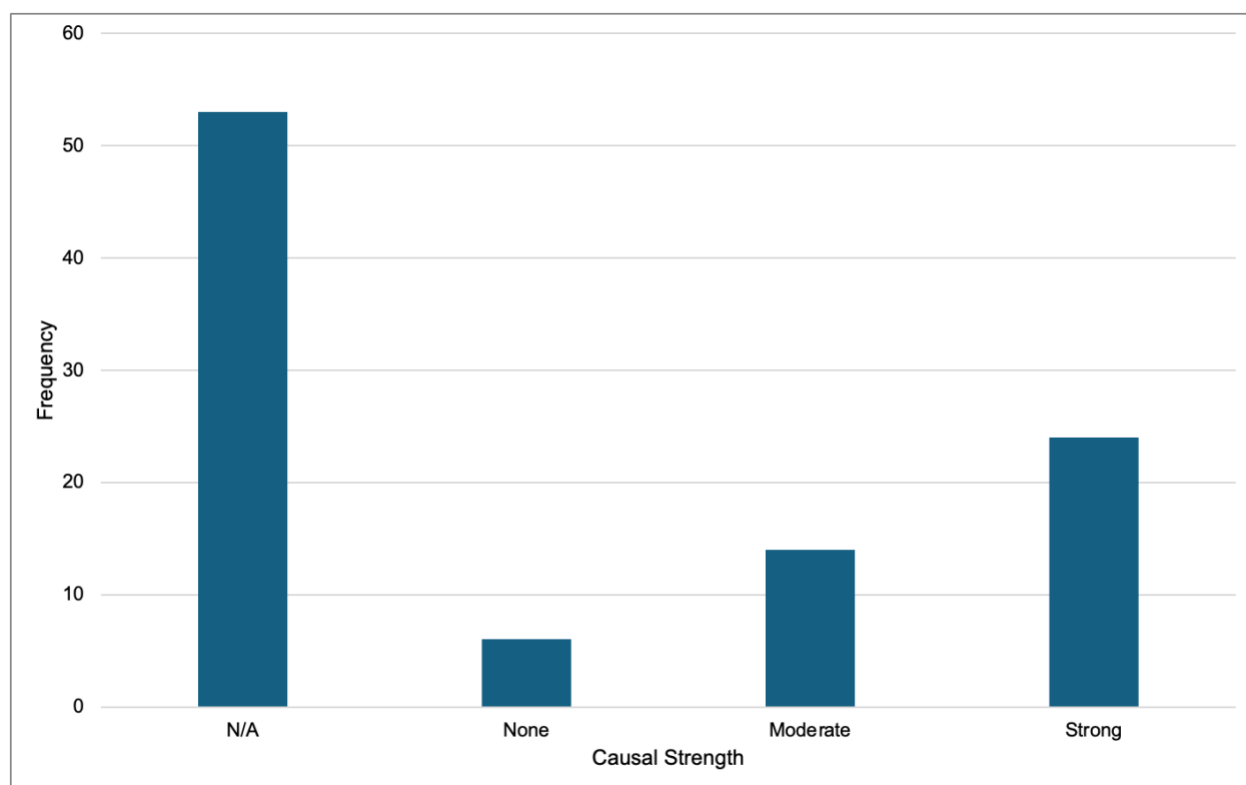
#### **3.3.1 Practice Recommendations – Abstracts**

Abstracts that included practice recommendations accounted for 45.4% ( $n = 44/97$ , 95% CI [36.1, 56.0]) of the articles. Among these, 13.6% ( $n = 6/44$ , 95% CI [0.0, 29.2]) had no causal implication. An example of this is, “These findings draw attention to the need for preventive interventions for this high-risk population” (Péter et al., 2023, p. 1). Moderate causal implications were identified in 31.8% ( $n = 14/44$ , 95% CI [18.2, 47.4]) of the recommendations, such as in, “If

clinicians can increase the interoceptive abilities of their clients at risk for MD [muscle dysmorphia], this may prevent the development of MD symptoms” (Grunewald et al., 2023, p. 1). The majority, 54.5% ( $n = 24/44$ , 95% CI [40.9, 70.1]), had strong causal implications. For example, “Addressing stress levels with the use of functional coping strategies can be beneficial to protect the general population from adverse psychological outcomes” (Kleanthous et al., 2023, p. 1). Figure 5 shows the frequency of causal implication strength of practice recommendations in the abstracts.

### Figure 5

*Strength of Causal Implication in Practice Recommendations of Abstracts*



### 3.3.2 Practice Recommendations – Concluding Sections

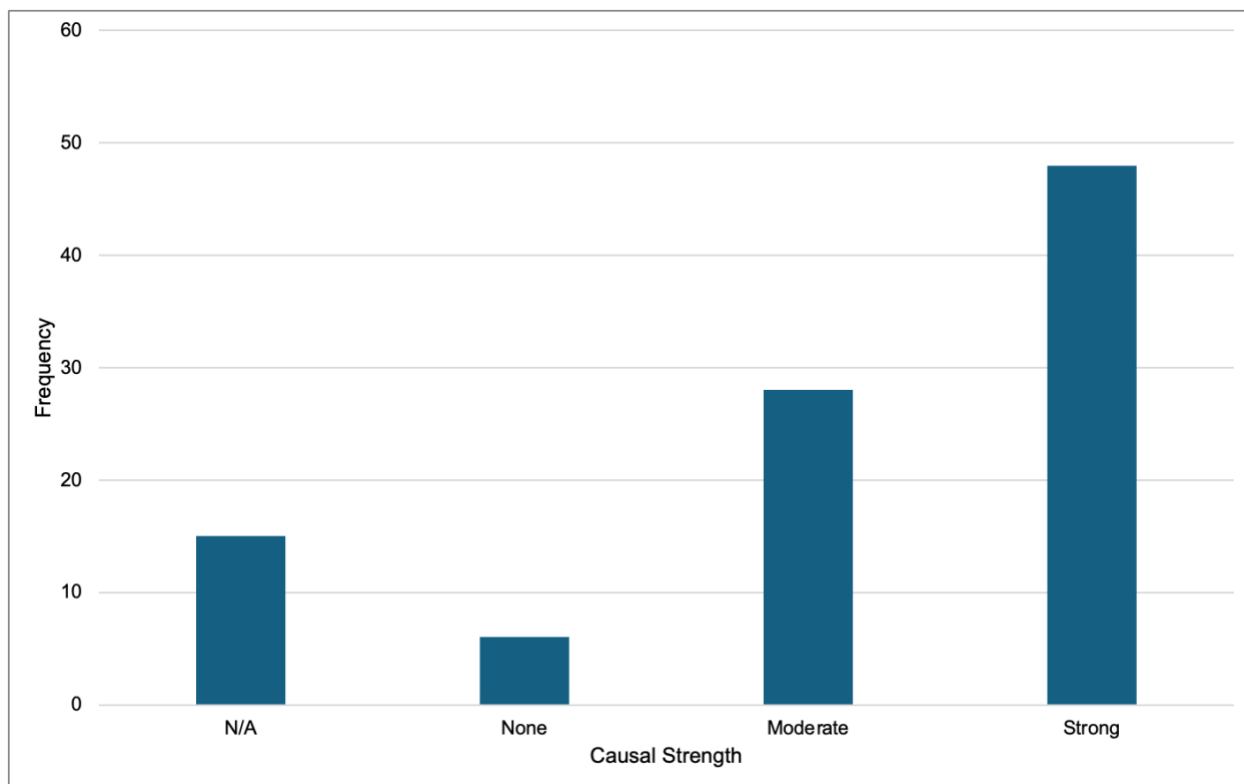
Most of the concluding sections contain practice recommendations, accounting for 84.5% ( $n = 82/97$ , 95% CI [78.4, 91.8]) of the articles. This is almost double the proportions of

practice recommendations that are present in the abstracts. Among these, 7.3% ( $n = 6/82$ , 95% CI [0.0, 18.9]) had no causal implication. An example of this is: “These results highlight the complexity of disordered eating and the importance to conduct well-rounded assessments when targeting behavioral and psychological characteristics of disordered eating” (Echeverri et al., 2023, p. 9).

Moderate causal implications were identified in 34.1% ( $n = 28/82$ , 95% CI [24.4, 45.7]) of the recommendations, such as: “Monitoring dissociation and providing skills to regulate dissociative states may help to prevent symptom deterioration in traumatized individuals” (Torres et al., 2023, p. 4). The majority, 58.5% ( $n = 48/82$ , 95% CI [48.8, 70.1]), had strong causal implications. An example is: “As such, treatment providers in similar settings can expect PTSD [post-traumatic stress disorder] treatment to be effective regardless of TST [time since trauma]” (Splaine et al., 2023, p. 91). Another example of a practice recommendation with strong causal implication is, “Therefore, identifying ways to update caregivers’ inappropriate Confucian parenting ideas and enhance the effectiveness of out-of-home support are crucial concerns for future policy” (Wan et al., 2023, p. 73). Figure 6 shows the frequency of causal implication strengths of practice recommendations in the concluding sections.

**Figure 6**

*Strength of Causal Implication in Practice Recommendations of Concluding Sections*



### 3.4 Comparing Causality of Linking Language and Practice Recommendations

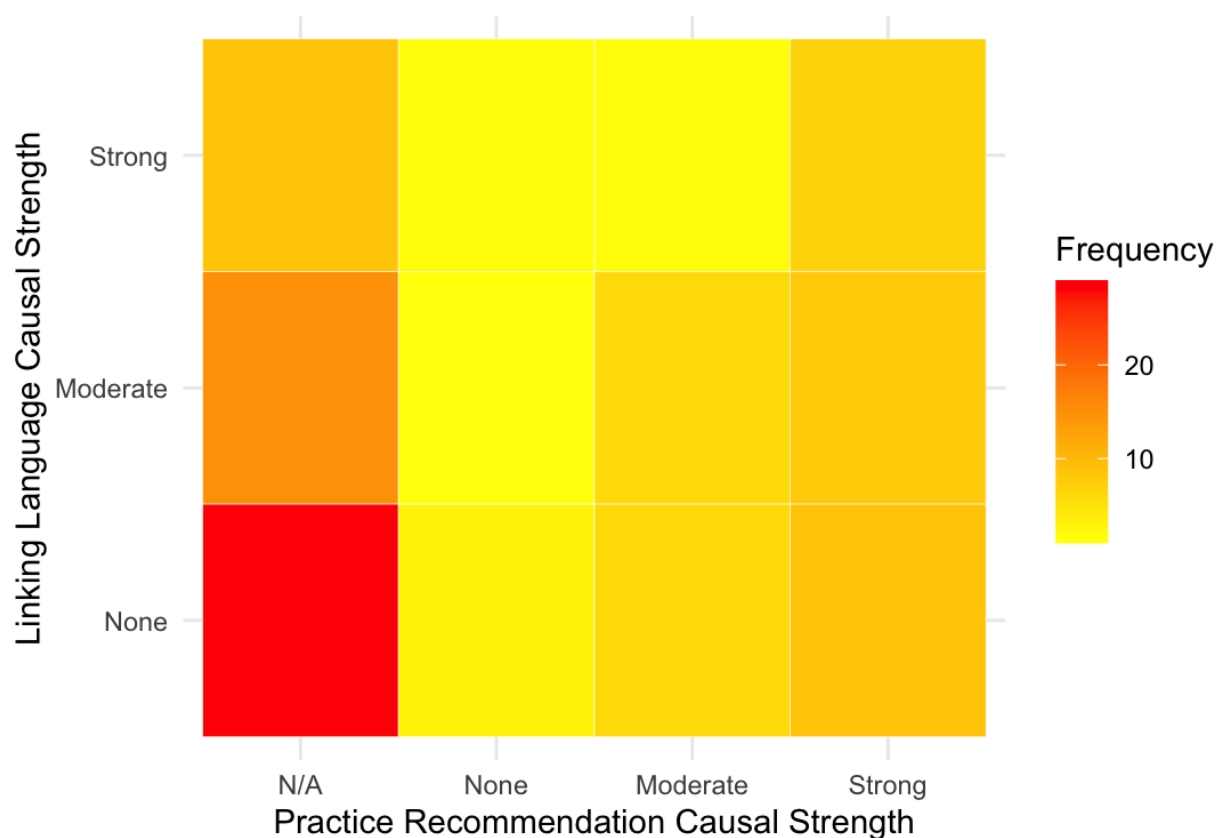
#### 3.4.1 Comparing Causality – Abstracts

A two-tailed Spearman's rank correlation ( $\rho$ ) at a 5% alpha level was conducted to estimate the strength of the causal implications in the linking language and practice recommendations. This analysis was performed separately for the abstracts and the concluding sections of the articles. In the abstracts, no statistically significant relationship was found ( $r_s = .09$ ,  $p = .57$ , 95% CI [-.21, .38]). The strength of the causal implications in the linking language that corresponded with the strength of the causal implications in practice recommendations was found in 16.5% ( $n = 16/97$ , 95% CI [7.2, 27.3]) of the articles. Additionally, the strength of the causal implications in the practice recommendations was greater than that of the linking

language in 23.7% ( $n = 23/97$ , 95% CI [14.4, 34.6]) of the articles. A specific example is from Bokoch et al. (2023). The linking sentence in the abstract states: “Motivation to change health and eating behaviours significantly increased, but confidence in managing emotional eating did not” (p. 117). This carried no causal implication. However, this was followed by a practice recommendation in the same section: “Art therapists should consider using psychosocial art-based community workshops to generate motivation for change and prevent obesity” (p. 117), which carried a strong causal implication. Figure 7 shows the difference in the distributions of the strength of causal implications in the linking language and practice recommendations of the abstracts.

### Figure 7

*Comparison Between Causal Implication Strength of the Linking Language and Practice Recommendations in the Abstracts*



### 3.4.2 Comparing Causality – Concluding Sections

Similarly, in the concluding sections, no statistically significant relationship was found ( $r_s = .16$ ,  $p = .16$ , 95% CI [-.06, .36]) between the strength of causal implications of the linking language and those in the practice recommendations. The strength of causal implications of the linking language that corresponded with the strength of causal implications of the practice recommendations was observed in 29.9% ( $n = 29/97$ , 95% CI [20.6, 40.7]) of the articles. Additionally, the strength of causal implications in the practice recommendations was greater than that in the linking language in 46.4% ( $n = 45/97$ , 95% CI [36.1, 56.7]) of the articles.

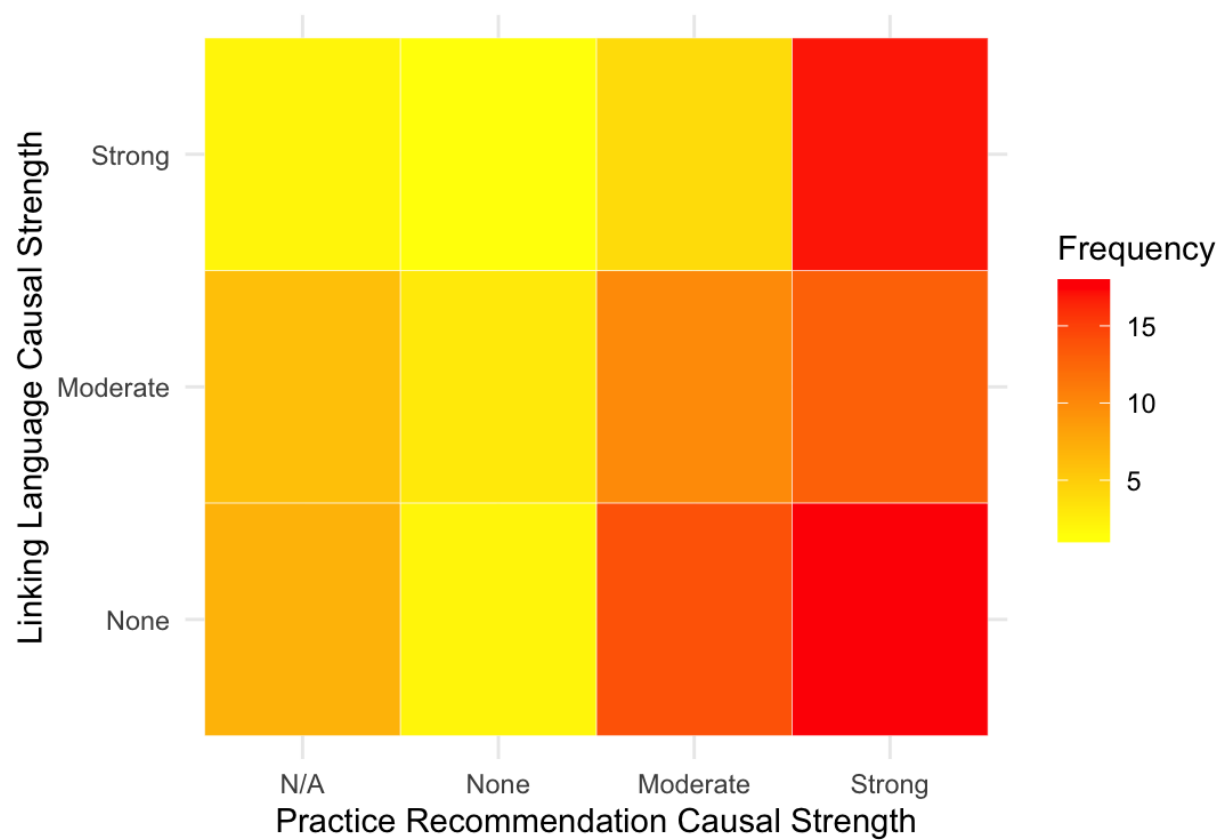
A specific example is Tekgunduz et al. (2023), who examined the correlation between engagement in specific sexual positions and outcomes related to sexual satisfaction and dysfunction. The linking sentence in their “Conclusion and recommendations” section states: “Those who engaged in female-on-top positions had higher mean scores in desire, arousal, lubrication, orgasm, satisfaction subscales, overall index, and lower mean score in the pain subscale” (p. 62). This carried no causal implication. However, the practice recommendation that followed, located in the same section, carried a strong causal implication: “Therefore, in the light of these results, it can be recommended; for midwives, nurses, counselors, and other healthcare professionals providing sexuality-related services to provide training on sexual dysfunction and sex positions, to give information to couples about the positions they prefer, to conduct similar studies on sex positions and dysfunction in larger groups, and to determine their rates in Turkey” (p. 62). This practice recommendation carried a strong causal implication because it assumes that interventions based on knowledge of specific sexual positions can directly lead to measurable improvements in patient well-being and satisfaction.

Figure 8 shows the difference in the distributions of the strength of causal implications in the linking language and practice recommendations of the concluding sections.

**Figure 8**

*Comparison Between Causal Implication Strength of the Linking Language and Practice*

*Recommendations in the Concluding Sections*



**Table 5**

*Distribution of Causal Implication Strength of Practice Recommendations Relative to Linking Language*

Strength of practice recommendations	Percentage of articles	
	Abstracts	Concluding sections
No practice recommendation	54.6%	15.5%
Less than linking language	5.2%	8.2%
Corresponded with linking language	16.5%	29.9%
Stronger than linking language	23.7%	46.4%

### **3.5 Secondary Objectives**

#### **3.5.1 Intent to Draw Causal Inference**

The expression of the intent to draw causal inferences was examined in the concluding sections of the articles, and it was exceedingly rare. Only 1.0% ( $n = 1/97$ , 95% CI [0.0, 2.8]) of the articles explicitly stated such an intent. A singular example of this can be found in the study by Gurm et al. (2023, p. 234), where the authors stated, "In this study, we have demonstrated the application of PSM [propensity score matching] for making valid inferences about treatment effectiveness in a naturalistic setting".

#### **3.5.2 Causal Disclaimer**

In the dataset, 30.9% ( $n = 30/97$ , 95% CI [22.7, 40.8]) of the articles included a causal disclaimer in their concluding sections. These disclaimers were intended to caution readers against drawing causal inferences from the study's findings. A clear example includes: "As we did not conduct an RCT, we have to be cautious with drawing causal inferences" (Schmeck et al., 2023, p. 157). This disclaimer is very straightforward in its presentation of the cautionary statement regarding causal inferences.

However, other disclaimers are less clear. For example, Raines et al. (2023, p. 13) stated that: “First, given the pilot nature, the current study lacked a control or comparison group. Although reductions in overall anxiety, depression, impairment, and safety aid usage were found, we cannot conclusively rule out that veterans merely improved due to the passage of time or because of nonspecific aspects of the treatment/study procedure”. The example above is an instance where authors implied a causal disclaimer instead of explicitly stating it. Specifically, the authors noted limitations in their study’s design, which is the absence of a control or comparison group, affecting causal interpretations. Despite not directly stating a causal disclaimer, the implication is clear in their discussion of the limitations.

Of the articles that did include a causal disclaimer statement, a significant portion still implied some degree of causality through the linking language. Specifically, 46.7% ( $n = 14/30$ ) of these articles implied causality through linking language in the abstracts, with 33.3% (95% CI [16.7, 51.6]) rated as having a moderate causal implication, and 13.3% rated as having a strong causal implication (95% CI [0, 31.6]). For example, Schmeck et al. (2023) included the following causal disclaimer in their “Discussion” section: “As we did not conduct an RCT, we have to be cautious with drawing causal inferences” (p. 157). However, their primary linking sentence, located in the abstract, has a strong causal implication: “Both DBT-A [dialectical behavior therapy for adolescents] and AIT [adolescent identity treatment] significantly improved adolescents’ psychosocial functioning and personality functioning” (p. 148).

Additionally, 63.3% ( $n = 19/30$ ) of the articles implied some degree of causality through the linking language in their concluding sections. Among them, 40.0% were rated as having a moderate causal implication (95% CI [23.3, 60.0]), and 23.3% were rated as having a strong causal implication (95% CI [6.7, 43.3]). For example, Xie et al. (2023) included the following causal disclaimer in their “Discussion” section: “First, due to the cross-sectional study design, we were not able to determine a causal relationship between functional disability and marijuana

use disorder” (p. 26). However, their primary linking sentence states: “The prevalence of marijuana use disorder increased significantly among PWD [people with disabilities], and this upward trend may continue in the coming years, as more states legalize the use of medical and recreational marijuana” (p. 34). This linking sentence, located in their “Conclusions” section, carried a moderate causal implication. It suggests a correlation between marijuana use disorder and its increased prevalence among people with disabilities, with a temporal element indicating that this rise may continue as marijuana legalisation expands, which contributes to the moderate causal implication.

A significant proportion of articles with causal disclaimer statements also made practice recommendations that would only be appropriate if a causal relationship were identified. In the abstracts, 94.4% ( $n = 17/18$ ) of such articles included practice recommendations. Of these practice recommendations, 33.3% (95% CI [16.7, 58.6]) were rated as having a moderate causal implication, and 61.1% (95% CI [44.4, 86.3]) were rated as a strong causal implication. For example, Levinson et al. (2023) made the following practice recommendation in their abstract: “These data suggest that network-informed personalized treatment has high acceptability and feasibility and can decrease ED [eating disorder] and related pathology, possibly serving as a feasible alternative to existing treatments” (p. 14). This practice recommendation carried a strong causal implication. However, the authors went on to include the following causal disclaimer: “First, this was not a randomized controlled trial, which means we do not know if network-informed personalized treatment will outperform other types of ED treatments” (p. 25)”. This causal disclaimer is another example of a less clear disclaimer, which hints at the impossibility of drawing causal conclusions from their findings by pointing out a limitation in study design.

Similarly, 89.7% ( $n = 26/29$ ) of articles with causal disclaimer statements made similar practice recommendations in the concluding sections. Among them, 31.0% (95% CI [17.2,

51.4]) were rated as having a moderate causal implication, while 58.6% (95% CI [44.8, 79.0]) were rated as having a strong causal implication. For example, Marino et al. (2023) made the following practice recommendation in their “Conclusions” section: “Women and men with problematic Internet pornography use can benefit from intervention programs aimed at reducing DT [desire thinking]” (p. 7). This practice recommendation has a strong causal implication. However, in their “Discussion” section, the authors included the following disclaimer: “The causality of the associations cannot be inferred from cross-sectional studies” (p. 5).

### **3.5.3 Discussion of Confounding**

In our dataset analysis, 30.9% ( $n = 30/97$ , 95% CI [22.7, 40.8]) of the articles included a discussion of confounding factors found anywhere in the text. The majority of the articles did not address confounding factors: 69.1% ( $n = 67/97$ , 95% CI [60.8, 78.9]). We searched for “confound\*” where \* can be any suffix, anywhere in the articles. Examples include “confounding”, “confounders”, or “third variable”, made in reference to the methods, results, or interpretation of the study, and not in reference to other studies. However, we did not examine whether these terms were used in a way that would support causal inference; we only identified the existence of such a strategy.

## **4.0 Discussion**

### **4.1 Summary of Results**

The current study is a systematic evaluation and small-scale examination of the causal language patterns used in the clinical psychology literature. As a result, we have identified several key findings. The most common linking word identified in the abstracts was “associate”, with other frequently used words including “predict” and “relate”. In the concluding sections, “associate” was again the most common, followed by “predict” and “risk”. Notably, “associate” was also the most common linking word identified by Haber et al. (2022), who examined causal language in medicine and epidemiology. Interestingly, while these terms are traditionally

considered non-causal, authors frequently used them in ways that implied causality to varying degrees.

This study found a greater variety of linking words across abstracts and concluding sections compared to Mockett (2023). This difference could reflect genuine variability in the use of linking words within clinical psychology; however, it is more likely due to differences in sample sizes, with the current study analysing 97 articles compared to Mockett's 50 articles. Mockett noted that linking words with traditionally non-causal meanings did not often have causal implications in the psychological capital literature. For instance, although "relate" was the most frequently used linking word in their study, it rarely implied causality. This finding is consistent with the present study, which also found that traditionally non-causal linking words such as "associate", "predict", and "relate" often did not suggest causality. In contrast, Haber et al. (2022) observed that traditionally non-causal linking words usually carried causal implications, with "associate" being rated as having at least some causal strength 78.6% of the time.

Explicit statements of intent to draw causal inferences were exceedingly rare in this study. This suggests a reluctance amongst clinical psychology authors to openly acknowledge causal goals, perhaps due to the observational nature of their studies or concerns about methodological scrutiny. In contrast, such statements were more prevalent in the psychological literature examined by Mockett (2023). However, it is worth considering that some of the papers reviewed by Mockett might not have included such statements as a thoughtful acknowledgment of an intent to investigate causal relationships. Rather, the authors may not have been fully aware of the limitations inherent in observational research. Hence, the higher prevalence of explicit statements of intent to draw causal inferences in Mockett does not necessarily indicate a more mature engagement with causal inference.

Overall, our findings broadly replicate those reported in the medical (Haber et al., 2022) and psychological capital literature (Mockett, 2023). We identified an implicit interest in

causality, demonstrated by the widespread use of language that implied causality and frequent discussions of confounding factors. Additionally, we identified a misalignment between the causal implications of the linking language used when reporting results and the subsequent practice recommendations.

#### **4.2 Practice Recommendations**

Most practice recommendations identified in the current study have strong causal implications. This pattern was observed in both the abstracts and the concluding sections, where more than half of the practice recommendations had strong causal implications. Despite the frequent strong causal implications in practice recommendations, our analysis revealed no significant relationship between the causal implication strength of the practice recommendations and the causal implication strength of the linking sentences.

The lack of evidence for a correlation here highlights the disconnect between the language researchers use to describe the relationships between predictor and outcome and the language they use when offering practice recommendations. Such a disconnect suggests that researchers may exercise caution when describing the relationships between predictor and outcome variables, due to the taboo against making causal inference in observational studies (Grosz et al., 2020). However, they may still feel compelled to provide actionable guidance that assumes a strong causal link.

Researchers often face pressure from funding bodies, journal publishers, and the broader scientific community to demonstrate the practical significance of their work (Ioannidis, 2016). Consequently, they may feel obligated to suggest applications of their work to real-world problems. Researchers may also feel inclined to overstate the applicability of findings even when the causal evidence is not robust. As a result, practice recommendations offered could be misguided, as the recommendations may implicitly exaggerate the evidence for causality

presented in their studies. Exaggerated practice recommendations can then lead to ineffective or unsuitable interventions that harm vulnerable or at-risk groups (Thapa et al., 2020).

### **4.3 Schrödinger's Causal Inference**

We found that while many articles included causal disclaimers (30.9%), most of the articles with causal disclaimers still implied some degree of causality through their linking language. Moreover, although many studies included causal disclaimers, an implicit interest in causality was evident through the frequent discussions of confounding in the articles. Most studies also offered practice recommendations that would only be appropriate if a causal relationship were identified. As postulated by Tennant and Murray (2021), this phenomenon can be referred to as Schrödinger's causal inference, where authors simultaneously imply and deny causality within the same articles, leading to confusion about the strength of the evidence presented.

Notably, a large proportion of concluding sections in this study had causal disclaimers. For the concluding sections, Schrödinger's causal inference complicates interpretation because these sections often summarise and emphasise the most important findings, potentially leading readers to overestimate the strength of the evidence. When causal disclaimers are present alongside strong causal implications in practice recommendations, it creates a confusing narrative for readers, as the disclaimers suggest caution while the practice recommendations imply robust causal relationships.

Furthermore, our study identified causal disclaimers that clearly argue against making causal inferences from observational data. A specific example is Raines et al. (2023, p. 13), who stated: "First, given the pilot nature, the current study lacked a control or comparison group. Although reductions in overall anxiety, depression, impairment, and safety aid usage were found, we cannot conclusively rule out that veterans merely improved due to the passage of time or because of nonspecific aspects of the treatment/study procedure (i.e. therapist contact)".

The excerpt above exemplifies the “taboo against causal inference in nonexperimental psychology”, as put forward by Grosz et al. (2020). This taboo refers to the reluctance among psychology researchers to draw causal conclusions from observational data, reflecting a general scepticism about the validity of causal claims derived from observational data. Despite this causal disclaimer by Raines et al., the practice recommendation located in their discussion section carried a strong causal implication: “Further, we now have support for the utility of safety aid reduction treatments across a wider array of diagnoses, namely OCD [obsessive-compulsive disorder], PTSD, and MDD [major depressive disorder]” (p.13). This practice recommendation carried a strong causal implication because it suggests that safety aid reduction treatments can directly improve mental health outcomes for individuals with OCD, PTSD, and MDD.

Motte-and-bailey strategies (Grosz et al., 2020) are also relevant here, which involve authors making bold causal claims but retreating to a safer position when challenged. In the context of our findings, authors would use strong causal language to emphasise their practice recommendations; however, they would also use weaker linking language to shield themselves from criticism, within the same paper. This is evident in the concluding sections, as the strength of causal implications in the practice recommendations was greater than that of its linking language in the majority of the articles. We argue that this strategy allows the authors to make strong statements that can attract attention and suggest significant contributions to the field, while also providing a fallback position that protects them from scrutiny over their causal claims. The gap between the linking language and the stronger causal implications in practice recommendations creates risks for research consumers, including clinicians, policymakers, and other decision-makers, who often rely on journal articles to guide critical decisions (Haber et al., 2018).

#### **4.4 Implications for Users of Clinical Psychology Research**

The findings of this study have important implications for users of clinical psychology research, including practitioners, policymakers, and researchers. The identified misalignment between the linking language and practice recommendations, as well as the prevalent use of causal implications despite disclaimers, highlight the need for carefully interpreting and evaluating research findings in clinical psychology.

##### **4.4.1 Practitioners**

Practitioners should approach research with caution, especially when interpreting clinical implications. It is essential to understand that research decisions, including study design, methods, and analyses, impact the validity of causal inferences. For instance, while longitudinal studies can provide stronger evidence for causal relationships compared to cross-sectional studies, they require careful control for confounders that may vary over time (Rohrer & Murayama, 2023). Practitioners should seek studies that offer a clear rationale for their assumptions (Rohrer, 2018) and be cautious of those that present strong practice recommendations without robust causal evidence.

When using a journal article as a basis for clinical decision-making, practitioners should, when possible, read beyond practice recommendations presented in the abstract or concluding sections. In this study, the majority of practice recommendations had strong causal implications. Our findings suggest that authors often offer practice recommendations that are inconsistent with their claims about causal inference, and consequently unlikely to be warranted by their empirical findings. Thus, it would be prudent to review the methods and results underlying specific practice recommendations, even briefly, rather than assuming that practice recommendations are justified by evidence simply because they appear within peer-reviewed journal articles.

Practitioners may also look for explicit statements of causal intent and examine the methods used to support these claims. However, our analysis revealed that clinical psychology articles rarely acknowledge an explicit intent to draw causal inferences. Hence, it is helpful for practitioners to examine the statistical methods and assumptions underlying them, such as the handling of unmeasured confounders, to determine whether the causal conclusions (if offered) are justified (Parra et al., 2021). However, we acknowledge that practitioners may have limited time and resources to conduct such thorough evaluations. Collaborating with colleagues who have expertise in research methods can help bridge knowledge gaps by providing support in understanding study designs and data analysis.

#### **4.4.2 Policymaking**

This study identified a disconnect between the linking language and the stronger causal implications in practice recommendations. This disconnect is a significant concern for policymakers who rely on research to inform decisions. As observed in the current study, practice recommendations may implicitly exaggerate the evidence for causality. Consequently, misinterpretations of these practice recommendations can lead to an overstatement of evidence strength, as they often assume causal effects that the study design does not provide sufficient evidence to justify.

To reduce the risk of allocating resources to interventions without robust causal support, policymakers should carefully evaluate practice recommendations to ensure they are well-calibrated to the evidence presented. This includes identifying studies with robust methods and assessing whether the study design justifies the assumed causal effects claimed by the authors. This process may also involve consulting independent experts in research methods and the relevant field of study to ensure informed decision-making. Aligning interventions with appropriately interpreted evidence can reduce the likelihood of wasted resources or unintended consequences (Buhse et al., 2018; Glass et al., 2014; Rossouw et al., 2002).

#### **4.4.3 Media**

Journalists need to be aware that even rigorous, peer-reviewed studies may include practice recommendations that are not fully supported by evidence, as observed in the current study. These recommendations require careful scrutiny before being reported. Exaggeration in health-related science news is a well-documented issue, with researchers like Sumner et al. (2014) emphasising the need for journalists to cross-check their sources. However, journalists face a challenging role, particularly when press releases and study abstracts are misleading. Hence, critical evaluation skills are essential, and there is an opportunity for journalists to improve these skills through workshops. For example, the German Network for Evidence-based Medicine regularly offers workshops to support journalists in enhancing their ability to critically assess research (Buhse et al., 2018).

In addition to these recommendations, we contend that journalists should critically question academic researchers about how well their evidence supports their claims. Where possible, journalists should incorporate expert opinions that provide alternative interpretations to guide readers to understand the nuances and range of possible explanations for the findings. This includes asking whether researchers have sufficient evidence to justify practice recommendations that rely on causal assumptions. Applying a high level of scrutiny and accountability to researchers means that we expect their claims to be well-supported, which in turn can reduce the risk of exaggeration in reporting (Sumner et al., 2014). Additionally, journalists should include explanations of study limitations, such as potential confounders, sample size, and study design, when reporting on research.

#### **4.5 Recommendations for Better Clinical Psychology Research**

Improving the quality of clinical psychology research requires addressing the issues identified in this study. Precise language can support clear causal thinking, which is necessary for effective clinical practice (Saxe et al., 2022).

### **4.5.1 Transparency**

The findings of this study emphasise the importance of transparency in communicating research findings in journal articles. Specifically, we observed that many articles used linking language that had no causal implication in their abstracts and concluding sections. Despite this, over half of the practice recommendations in both sections carried strong causal implication. This inconsistency indicates a lack of transparency among authors, as they offer practice recommendations that require strong causal assumptions about predictor-outcome relationships while being careful in describing such relationships throughout an article. Furthermore, this issue is made worse by unspoken norms and formal internal guidelines that often discourage the use of “causal language” in observational studies (Hernán, 2018). We argue that instead, where applicable, researchers should clearly state their causal aims and be transparent about the level of evidence they provide. These practices improve clarity for both readers and writers, which in turn strengthens the connection between research methods and theory (Alvarez-Vargas et al., 2020; Hernán, 2018; Parra et al., 2021).

Moreover, we identified an implicit interest in causality, demonstrated by the widespread discussions of confounding factors. When dealing with quantitative data, researchers should clearly define their estimand, which is the specific quantity the research aims to estimate (Lundberg et al., 2021). In the context of psychological research, this means articulating the precise causal effect of interest, such as the effect of a specific behaviour on an outcome, and considering the relevant conditions and assumptions necessary for estimating this effect accurately (Rohrer and Murayama, 2023). A well-defined estimand, such as the average causal effect or population mean, links statistical analysis to theoretical questions and ensures clarity in what the study targets. When there is a lack of precision in describing this causal effect, the study’s findings become difficult to challenge or falsify, making it hard to assess the validity of the results (Smaldino, 2017).

Transparency also means being explicit and precise about assumptions, which allows for a clearer understanding and evaluation of the research (Lederer et al., 2019; Rohrer, 2018). For example, researchers can explicitly state assumptions about confounding variables and use appropriate statistical controls to make the causal inference process more explicit. However, justifying the selection of control variables is challenging without a clear understanding of the causal effects being studied. Many researchers select control variables based on statistical associations rather than clear causal reasoning, despite the importance of careful selection of control variables to achieve accurate causal effect estimation (Wysocki et al., 2022).

Statistical control is effective only when control variables are chosen based on a justified causal framework, as controlling inappropriate variables can lead to biased estimates (Wysocki et al., 2022). However, given the complexity of causal relationships, no single study can address all underlying assumptions or definitively eliminate all possible alternative causal explanations. Nevertheless, authors should be clear about how confident they are in the credibility of their causal evidence, particularly in titles and abstracts (Thapa et al., 2020).

#### ***4.5.2 Language Policing***

Previous researchers have recommended strict guidelines for the use of causal language in observational research. For example, Adams et al. (2017) made specific recommendations: (1) authors should use direct causal expressions when reporting findings from rigorous experimental designs; (2) the word “can” should be added before direct causal expressions when there is uncertainty about experimental findings, such as when using small sample sizes or testing a new drug only on healthy participants; and (3) moderate causal expressions should be used when discussing observational findings. Similarly, Thapa et al. (2020) emphasised that causal language should clearly indicate how one variable affects another, including the direction of causality, while non-causal language should describe

relationships without implying causation. They noted that choosing appropriate language is important for accurately reporting study results and their causal relationships.

However, this thesis suggests that focusing solely on policing language overlooks the deeper issue, which is the need for improved causal inference methods in observational research. Observational researchers interested in causality should use suitable methods and research designs to strengthen their claims. For example, techniques like propensity score matching, instrumental variable analysis, and structural equation modelling can help control for confounding variables, enabling stronger causal claims. Additionally, as discussed in section 1.2.2 (Causal Inference in Observational Studies), DAGs are useful for visualising relationships between variables and identifying confounders for more accurate interpretations (Pearl, 2009; Rohrer, 2018).

#### **4.5.3 Publication Process**

Journal article reviewers should pay closer attention to implied causality in research articles, particularly when practice recommendations are based on findings that lack robust causal evidence, as observed in this study. Grosz et al. (2022) noted that many psychologists conducting observational research often fail to justify their selection of control variables explicitly, and very few use formalised frameworks that support causal reasoning. Reviewers should check whether practice recommendations are based on assumptions about causal effects and whether such assumptions are consistent with the rest of the given article. This involves examining the methodology, statistical controls, and the assumptions underlying the causal inferences. Even a simple step, such as checking for the presence of both causal inferences and causal disclaimers within an article, can help verify that the findings are accurately represented.

Journal editors and publishers also have an important role in research dissemination by ensuring its quality and providing support to publish strong work. Schiavone et al. (2023)

introduced *Seaboat* (<https://www.seaboat.io>), an online tool that helps reviewers systematically evaluate potential threats to the validity of empirical research, particularly in the social sciences. *Seaboat* is based on the “Four Validities” framework (Shadish et al., 2002) and assists peer reviewers in identifying common threats to construct, internal, external, and statistical validity. The tool is freely available online, providing a structured approach to reviewing and enabling users to generate detailed reports on validity issues, which can be shared during peer review or used post-publication.

To enhance tools like *Seaboat*, additional checks informed by the current study could be incorporated, such as verifying the presence of explicit causal inferences, evaluating the causal implications of practice recommendations, and assessing the consistency between causal inferences and practice recommendations within an article. Editors and publishers could also consider developing a separate tool specifically for these purposes to help ensure that published work accurately reflects the evidence. Such measures would support a publication process that prioritises transparency and robust causal reasoning.

#### **4.6 Limitations and Directions for Future Research**

This study has several limitations and offers directions for future meta-psychological research on this topic. One limitation is that the ratings were subjective and conducted by a single reviewer, which may introduce bias. The lack of multiple independent reviewers also prevented the calculation of interrater reliability. Future research should involve multiple reviewers to enhance reliability and enable the calculation of interrater reliability statistics.

Another limitation is that this study did not evaluate the specific study designs or methods used in the reviewed articles, meaning that the quality and rigour of the included studies were not assessed. While we could identify and describe how researchers have used causal language, and pick up inconsistencies in their statements, we could not actually determine whether the causal claims were justified based on the data they collected. Future

research should evaluate the study designs and methods to better understand the quality of the evidence behind the causal language used.

This study was designed to be highly replicable, with clear and detailed methods, including the use of a review tool and explicit criteria for article selection and data extraction. All materials and data are available on the [OSF repository](#) to ensure transparency. Although this study's methods could be easily replicated by others, the findings of this study may not be generalisable to other research domains, as the sample came from the social sciences. The focus on clinical psychology means that the results may not apply to other fields that have different research practices and standards, such as sociology or public health. Expanding the analysis to a broader range of journals, including those from related fields, could provide a more comprehensive picture of causal language use in research. Future research may also replicate this study in other closely-related domains such as health psychology and social psychology, to assess whether our findings apply more broadly and understand how causal language and practice recommendations may vary across various psychology disciplines.

Next, this study mainly reviewed specific sections of the articles (abstract, discussion, and conclusion) rather than the full text, which limits our understanding of how causal language is used throughout entire articles. Future research could expand the review to include the full texts for a more comprehensive analysis. For example, the introduction and literature review sections might reveal how authors frame their causal hypotheses and the theoretical underpinnings that guide their research questions. Additionally, while this study featured a larger sample size compared to Mockett's (2023) research and expectations at the Master's level, the sample size may still be considered small when compared to large-scale systematic reviews or meta-analyses in the field. A larger sample size would increase statistical power and precision, resulting in more accurate confidence intervals and potentially more reliable and generalisable findings (Morey & Lakens, 2016).

Lastly, future studies could investigate how causal language and practice recommendations impact readers' interpretations and decision-making. For instance, experimental designs could be used to test how different phrasings influence clinicians' perceptions and the actions they intend to take in treatments. Clinicians have different levels of training or skills in research methods, which could influence how they interpret and apply research findings, potentially leading to differences in clinical practice. Similarly, future studies could examine how causal language influences powerful actors like government policymakers. Since their decisions can have wide-reaching consequences, it's crucial to know how different ways of presenting information may affect their choices. For example, by investigating how causal language is used in government evidence briefs, we could improve the way research findings are communicated to policymakers, helping them make informed decisions that are based on accurate interpretations of the evidence.

## **5.0 Conclusion**

Clinical psychology is a rapidly growing field that relies on robust evidence to advance, and much of the knowledge in this field directly impacts vulnerable populations. However, researchers in clinical psychology often cannot employ experimental designs, which are useful for establishing causal relationships between variables. This study examined the use of linking language in clinical psychology literature, specifically within observational studies, and assessed the extent to which practice recommendations rely on assumed causal relationships between predictor and outcome variables.

Analysis of 97 journal articles revealed that while many linking sentences implied causality, practice recommendations often carried even stronger causal implications. Notably, there was no statistically significant correlation between the strength of causal implications in linking language and those of practice recommendations. These findings suggest that there is a disconnect between the linking language used to describe predictor-outcome relationships and

the stronger causal assumptions often implied in practice recommendations that follow. Our findings also indicate that researchers have an implicit interest in causality, which manifests in the high prevalence of linking language that implies causality and the frequent discussions of confounding factors. However, researchers are generally not transparent about their causal aims, and causal assumptions are rarely made explicit.

The inconsistencies identified within clinical psychology research articles emphasise the need for researchers to be clear about their causal intentions and the evidence supporting their practice recommendations. Additionally, practitioners should critically evaluate research findings before adopting practice recommendations, and policymakers should be cautious when interpreting research findings that may overstate causal relationships. Addressing the concerns raised in this study is important so that practice recommendations in clinical psychology are based on reliable evidence, ensuring effective and safe interventions for vulnerable populations and contributing to better clinical outcomes and progress in the field.

## References

- Adams, R. C., Challenger, A., Bratton, L., Boivin, J., Bott, L., Powell, G., Williams, A., Chambers, C. D., & Sumner, P. (2019). Claims of causality in health news: A randomised trial. *BMC Medicine*, *17*(1), 1-11. <https://doi.org/10.1186/s12916-019-1324-7>
- Adams, R. C., Sumner, P., Vivian-Griffiths, S., Barrington, A., Williams, A., Boivin, J., Chambers, C. D., & Bott, L. (2017). How readers understand causal and correlational expressions used in news headlines. *Journal of Experimental Psychology: Applied*, *23*(1), 1–14. <https://doi.org/10.1037/xap0000100>
- Ahola, A., Rautio, N., Timonen, M., Nordström, T., Jääskeläinen, E., & Miettunen, J. (2023). Premorbid temperament as predictor of onset of depression: 23-year follow-up. *Comprehensive Psychiatry*, *121*, 1-7. <https://doi.org/10.1016/j.comppsy.2022.152359>
- Alvarez-Vargas, D., Braithwaite, D. W., Lortie-Forgues, H., Moore, M. M., Castro, M., Wan, S., Martin, E. A., & Bailey, D. H. (2020). *Hedges, mottes, and baileys: Causally ambiguous statistical language can increase perceived study quality and policy relevance* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/nkf96>
- American Psychological Association. (2018). *Scientist-practitioner model*. <https://dictionary.apa.org/scientist-practitioner-model>
- American Psychological Association. (2022). *Clinical psychology*. <https://www.apa.org/ed/graduate/specialize/clinical>
- Aston, E. R., & Merrill, J. E. (2023). Alcohol demand assessed daily as a predictor of same day drinking. *Psychology of Addictive Behaviors*, *37*(1), 114–120. <https://doi.org/10.1037/adb0000890>
- Bassi, M., Carissoli, C., Tonelli, F., Trombetta, L., Magenta, M., Delle Fave, A., & Cogliati, C. (2023). Stress and mental health of COVID-19 survivors and their families after hospital

- discharge: Relationship with perceived healthcare staff empathy. *Psychology, Health & Medicine*, 28(1), 48–59. <https://doi.org/10.1080/13548506.2021.2019811>
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., ... Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6–10. <https://doi.org/10.1038/s41562-017-0189-z>
- Bleske-Rechek, A., Morrison, K. M., & Heidtke, L. D. (2015). Causal Inference from Descriptions of Experimental and Non-Experimental Research: Public Understanding of Correlation-Versus-Causation. *The Journal of General Psychology*, 142(1), 48–70. <https://doi.org/10.1080/00221309.2014.977216>
- Bokoch, R., Hass-Cohen, N., Fowler, G., & Liu, L. (2023). Obesity prevention: A pilot study on community workshops delivered by art therapists. *International Journal of Art Therapy*, 28(3), 117–124. <https://doi.org/10.1080/17454832.2022.2123010>
- Brauner, F., Fonagy, P., Campbell, C., Griem, J., Storck, T., & Nolte, T. (2023). “Trust me, do not trust anyone”: How epistemic mistrust and credulity are associated with conspiracy mentality. *Research in Psychotherapy: Psychopathology, Process and Outcome*, 26(3), 55-67. <https://doi.org/10.4081/ripppo.2023.705>
- Buhse, S., Rahn, A. C., Bock, M., & Mühlhauser, I. (2018). Causal interpretation of correlational studies – Analysis of medical news on the website of the official journal for German physicians. *PLOS ONE*, 13(5), 1-12. <https://doi.org/10.1371/journal.pone.0196833>
- Cofield, S. S., Corona, R. V., & Allison, D. B. (2010). Use of Causal Language in Observational Studies of Obesity and Nutrition. *Obesity Facts*, 3(6), 353–356. <https://doi.org/10.1159/000322940>

- Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology, 66*, 93–99. <https://doi.org/10.1016/j.jesp.2015.10.002>
- Dias, M. J., Rathenau, S., Sousa, D., & Santos, J. M. (2023). Therapeutic presence, working alliance, and clients' attitudes toward online therapy. *Journal of Psychotherapy Integration, 33*(4), 439–456. <https://doi.org/10.1037/int0000314>
- Dunning, T. (2012). *Natural experiments in the social sciences: A design-based approach*. Cambridge University Press.
- Echeverri, B., Kozak, A. T., Gildner, D. J., & Pickett, S. M. (2023). Night eating syndrome subtypes: Differences in binge eating and food addiction symptoms. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity, 28*(1), 1-12. <https://doi.org/10.1007/s40519-023-01534-7>
- Ferguson, C. J., & Heene, M. (2012). A Vast Graveyard of Undead Theories: Publication Bias and Psychological Science's Aversion to the Null. *Perspectives on Psychological Science, 7*(6), 555–561. <https://doi.org/10.1177/1745691612459059>
- Glass, T. A., Goodman, S. N., Hernán, M. A., & Samet, J. M. (2013). Causal Inference in Public Health. *Annual Review of Public Health, 34*(1), 61–75. <https://doi.org/10.1146/annurev-publhealth-031811-124606>
- Grand, J. A., Rogelberg, S. G., Allen, T. D., Landis, R. S., Reynolds, D. H., Scott, J. C., Tonidandel, S., & Truxillo, D. M. (2018). A Systems-Based Approach to Fostering Robust Science in Industrial-Organizational Psychology. *Industrial and Organizational Psychology, 11*(1), 4–42. <https://doi.org/10.1017/iop.2017.55>
- Grosz, M. P., Ayaita, A., Arslan, R. C., Buecker, S., Ebert, T., Hünermund, P., Müller, S. R., Rieger, S., Zapko-Willmes, A., & Rohrer, J. M. (2024). Natural experiments: Missed

- opportunities for causal inference in psychology. *Advances in Methods and Practices in Psychological Science*, 7(1), 1–15. <https://doi.org/10.1177/25152459231218610>
- Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science*, 15(5), 1243–1255. <https://doi.org/10.1177/1745691620921521>
- Grunewald, W., Fogelberg, S., Ferguson, W., Hines, S., Fortenberry, B., & Smith, A. R. (2023). Longitudinal relationships between specific domains of interoception and muscle dysmorphia symptoms. *Eating Behaviors*, 48, 2-7. <https://doi.org/10.1016/j.eatbeh.2022.101686>
- Guay, J. (2018). Evidence-based policymaking: Is there room for science in politics? *Apolitical*. <https://apolitical.co/en/solution-articles/en/evidence-based-policymaking-is-there-room-for-science-in-politics>
- Guerrero-Bote, V. P., & Moya-Anegón, F. (2012). A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *Journal of Informetrics*, 6(4), 674–688. <https://doi.org/10.1016/j.joi.2012.07.001>
- Gurm, K., Wampold, B. E., Piatt, C., Jagodzinski, R., Caperton, D. D., & Babins-Wagner, R. (2023). Effectiveness of telemental health during the COVID-19 pandemic: A propensity score noninferiority analysis of outcomes. *Psychotherapy*, 60(2), 231–236. <https://doi.org/10.1037/pst0000472>
- Guyatt, G. H., Oxman, A. D., Kunz, R., Vist, G. E., Falck-Ytter, Y., & Schünemann, H. J. (2008). What is “quality of evidence” and why is it important to clinicians? *BMJ*, 336(7651), 995–998. <https://doi.org/10.1136/bmj.39490.551019.BE>
- Guzman Torres, E., Krause-Utz, A., & Sack, M. (2023). Predictors of complex PTSD: The role of trauma characteristics, dissociation, and comorbid psychopathology. *Borderline*

*Personality Disorder and Emotion Dysregulation*, 10(1), 1–6.

<https://doi.org/10.1186/s40479-022-00208-7>

Haber, N. A., Smith, E. R., Moscoe, E., Andrews, K., Audy, R., Bell, W., Brennan, A. T., Breskin, A., Kane, J. C., Karra, M., McClure, E. S., Suarez, E. A., & on behalf of the CLAIMS research team. (2018). Causal language and strength of inference in academic and media articles shared in social media (CLAIMS): A systematic review. *PLOS ONE*, 13(5), 1-21. <https://doi.org/10.1371/journal.pone.0196346>

Haber, N. A., Wieten, S. E., Rohrer, J. M., Arah, O. A., Tennant, P. W. G., Stuart, E. A., Murray, E. J., Pilleron, S., Lam, S. T., Riederer, E., Howcutt, S. J., Simmons, A. E., Leyrat, C., Schoenegger, P., Booman, A., Dufour, M.-S. K., O'Donoghue, A. L., Baglini, R., Do, S., ... Fox, M. P. (2022). Causal and Associational Language in Observational Health Research: A Systematic Evaluation. *American Journal of Epidemiology*, 191(12), 2084–2097. <https://doi.org/10.1093/aje/kwac137>

Hernán, M. (2018). The C-Word: The More We Discuss It, the Less Dirty It Sounds. *American Journal of Public Health*, 108(5), 625–626. <https://doi.org/10.2105/AJPH.2018.304392>

Höfler, M. (2005). Causal inference based on counterfactuals. *BMC Medical Research Methodology*, 5(1), 1-12. <https://doi.org/10.1186/1471-2288-5-28>

Hutchison, S. L., MacDonald-Wilson, K. L., Karpov, I., Herschell, A. D., & Carney, T. (2023). Peer support to reduce readmission in Medicaid-enrolled adults with substance use disorder. *Journal of Substance Abuse Treatment*, 144, 1-6. <https://doi.org/10.1016/j.jsat.2022.108901>

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8), 696-701. <https://doi.org/10.1371/journal.pmed.0020124>

Ioannidis, J. P. A. (2016). Why Most Clinical Research Is Not Useful. *PLOS Medicine*, 13(6), 1-10. <https://doi.org/10.1371/journal.pmed.1002049>

- Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist*, 63(3), 146–159. <https://doi.org/10.1037/0003-066X.63.3.146>
- Kimmel, K., Dee, L. E., Avolio, M. L., & Ferraro, P. J. (2021). Causal assumptions and causal inference in ecological experiments. *Trends in Ecology & Evolution*, 36(12), 1141–1152. <https://doi.org/10.1016/j.tree.2021.08.008>
- Kleanthous, E., Evangelou, E., Georgiadou, A., Galanis, P., Andreadaki, E., Tzavara, C., Kaitelidou, D., & Kallergis, G. (2023). Stress and coping strategies in the general population of Greece and Cyprus in response to the COVID-19 pandemic: A cross-sectional study. *European Journal of Trauma & Dissociation*, 7(1), 1-8. <https://doi.org/10.1016/j.ejtd.2022.100306>
- Kobayashi, H., & Berle, D. (2023). The Association Between Perceived Relationship Quality and Psychological Symptoms in Refugees. *Journal of Social and Clinical Psychology*, 42(2), 103–124. <https://doi.org/10.1521/jscp.2023.42.2.103>
- Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology*, 8(1), 1-28. <https://doi.org/10.1525/collabra.33267>
- Lederer, D. J., Bell, S. C., Branson, R. D., Chalmers, J. D., Marshall, R., Maslove, D. M., Ost, D. E., Punjabi, N. M., Schatz, M., Smyth, A. R., Stewart, P. W., Suissa, S., Adjei, A. A., Akdis, C. A., Azoulay, É., Bakker, J., Ballas, Z. K., Bardin, P. G., Barreiro, E., ... Vincent, J.-L. (2019). Control of Confounding and Reporting of Results in Causal Inference Studies. Guidance for Authors from Editors of Respiratory, Sleep, and Critical Care Journals. *Annals of the American Thoracic Society*, 16(1), 22–28. <https://doi.org/10.1513/AnnalsATS.201808-564PS>
- Levinson, C. A., Williams, B. M., Christian, C., Hunt, R. A., Keshishian, A. C., Brosof, L. C., Vanzhula, I. A., Davis, G. G., Brown, M. L., Bridges-Curry, Z., Sandoval-Araujo, L. E., &

- Ralph-Nearman, C. (2023). Personalizing eating disorder treatment using idiographic models: An open series trial. *Journal of Consulting and Clinical Psychology, 91*(1), 14-28. <https://doi.org/10.1037/ccp0000785>
- Lundberg, I., Johnson, R., & Stewart, B. M. (2021). What is your estimand? Defining the target quantity connects statistical evidence to theory. *American Sociological Review, 86*(3), 532–565. <https://doi.org/10.1177/00031224211004187>
- Lynch, S. M. (2013). Introduction to multiple regression. In *Using statistics in social research* (pp. 143-173). Springer.
- Mao, F., Wan, J., Sun, Y., Yang, B., Wang, Y., & Cao, F. (2024). Association between transition patterns of sleep problems and suicidal ideation in Chinese female nurses: A prospective study. *Journal of Clinical Psychology, 80*(2), 279–290. <https://doi.org/10.1002/jclp.23612>
- Marino, C., Melodia, F., Pivetta, E., Mansueto, G., Palmieri, S., Caselli, G., Canale, N., Vieno, A., Conti, F., & Spada, M. M. (2023). Desire thinking and craving as predictors of problematic Internet pornography use in women and men. *Addictive Behaviors, 136*, 1-10. <https://doi.org/10.1016/j.addbeh.2022.107469>
- Martin, T. E., Gover, A. R., & Langton, L. (2023). The Relationship Between Formal Help-Seeking and Sexual Orientation Among Victims of Domestic Violence. *Journal of Interpersonal Violence, 38*(1–2), 156–182. <https://doi.org/10.1177/08862605221080150>
- Martinez-Cengotitabengoa, M., Jackson, C. L., Pugh, E. J., Long-Martinez, D., Sanchez-Martinez, A., Sanchez-Martinez, M., Bermudez-Ampudia, C., & Martinez-Cengotitabengoa, M.-T. (2023). Global intellectual ability and adaptive functioning in children with FASD with and without sentinel facial features. *The American Journal of Drug and Alcohol Abuse, 49*(1), 123–128. <https://doi.org/10.1080/00952990.2022.2164503>

- Masselman, I., De Jong, P. J., & Glashouwer, K. A. (2023). Subjective disgust and facial electromyography responses towards unedited and morphed overweight self-pictures in women with varying levels of eating disorder symptomatology. *European Eating Disorders Review*, 31(1), 98–109. <https://doi.org/10.1002/erv.2940>
- Massey University. (2024). *Journal ranking and impact*. Massey University Library. <https://www.massey.ac.nz/study/library/researcher-support/publish-and-share-your-research/journal-ranking-and-impact>
- Menzies, P., & Beebee, H. (2024). Counterfactual theories of causation. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Summer 2024 Edition). <https://plato.stanford.edu/archives/sum2024/entries/causation-counterfactual>
- Mockett, E. J. (2023). *Examining linking language and causal implications in observational psychological capital research* [Master's thesis, Massey University]. Massey Research Online. <http://hdl.handle.net/10179/19062>
- Moore, M. S. (2009). *Causation and responsibility: An essay in law, morals, and metaphysics*. Oxford University Press.
- Morgan, S. L., & Winship, C. (2014). *Counterfactuals and causal inference: Methods and principles for social research* (2nd ed.). Cambridge University Press.
- New Zealand Psychologists Board. (2018). *New Zealand Psychologists Board (2018) NZPB Core Competencies.pdf*. [https://psychologistsboard.org.nz/wp-content/uploads/2023/08/Core\\_Competerencies.pdf](https://psychologistsboard.org.nz/wp-content/uploads/2023/08/Core_Competerencies.pdf)
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>

- Ohlsson, H., & Kendler, K. S. (2020). Applying Causal Inference Methods in Psychiatric Epidemiology: A Review. *JAMA Psychiatry*, 77(6), 1-16.  
<https://doi.org/10.1001/jamapsychiatry.2019.3758>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). <https://doi.org/10.1126/science.aac4716>
- Parra, O., Bertizzolo, L., Schroter, S., Dechartres, A., & Goetghebeur, E. (2021). Consistency of causal claims in observational studies: A review of papers published in a general medical journal. *BMJ Open*, 11(5), 1-7. <https://doi.org/10.1136/bmjopen-2020-043339>
- Pearl, J. (1993). [Bayesian Analysis in Expert Systems]: Comment: Graphical Models, Causality and Intervention. *Statistical Science*, 8(3). <https://doi.org/10.1214/ss/1177010894>
- Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4), 669–688.  
<https://doi.org/10.1093/biomet/82.4.669>
- Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.
- Pedhazur, E., & Schmelkin, L. (1991). *Measurement, design, and analysis: An integrated approach*. Lawrence Erlbaum Associates.
- Potter, R. C., DeVita, S. F., Vranesich, P. A., & Boulton, M. L. (2014). Adolescent Immunization Coverage and Implementation of New School Requirements in Michigan, 2010. *American Journal of Public Health*, 104(8), 1526–1533.  
<https://doi.org/10.2105/AJPH.2014.301910>
- Prasad, V., Jorgenson, J., Ioannidis, J. P. A., & Cifu, A. (2013). Observational studies often make clinical practice recommendations: An empirical evaluation of authors' attitudes. *Journal of Clinical Epidemiology*, 66(4), 361-366.  
<https://doi.org/10.1016/j.jclinepi.2012.11.005>

- Rahman, H. (2017). Influence of Research on Health Policy and Clinical Practice. *The Journal of Obstetrics and Gynecology of India*, 67(5), 319–323. <https://doi.org/10.1007/s13224-017-1013-4>
- Raines, A. M., Chambliss, J. L., Norr, A. M., Sanders, N., Smith, S., Walton, J. L., True, G., Franklin, C. L., & Schmidt, N. B. (2023). Acceptability, feasibility, and utility of a safety aid reduction treatment in underserved veterans: A pilot investigation. *Cognitive Behaviour Therapy*, 52(1), 1–17. <https://doi.org/10.1080/16506073.2022.2130819>
- Rajczyk, J. I., Ferketich, A., & Wing, J. J. (2023). Relation Between Smoking Status and Subjective Cognitive Decline in Middle Age and Older Adults: A Cross-Sectional Analysis of 2019 Behavioral Risk Factor Surveillance System Data. *Journal of Alzheimer's Disease*, 91(1), 215–223. <https://doi.org/10.3233/JAD-220501>
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27-42. <https://doi.org/10.1177/2515245917745629>
- Rohrer, J. M., & Murayama, K. (2023). These Are Not the Effects You Are Looking for: Causality and the Within-/Between-Persons Distinction in Longitudinal Data Analysis. *Advances in Methods and Practices in Psychological Science*, 6(1), 1-14. <https://doi.org/10.1177/25152459221140842>
- Rosenbaum, P. R., & Rubin, D. B. (1983). *The central role of the propensity score in observational studies for causal effects*. <https://doi.org/10.1093/biomet/70.1.41>
- Rossouw, J. E., Anderson, G. L., Prentice, R. L., LaCroix, A. Z., Kooperberg, C., Stefanick, M. L., Jackson, R. D., Beresford, S. A. A., Howard, B. V., Johnson, K. C., Kotchen, J. M., Ockene, J., & Writing Group for the Women's Health Initiative Investigators. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women:

- Principal results from the Women's Health Initiative randomized controlled trial. *JAMA*, 288(3), 321–333. <https://doi.org/10.1001/jama.288.3.321>
- Ruich, P. (2017). The use of cause-and-effect language in the JAMA Network journals. *AMA Style Insider*. <https://amastyleinsider.com/2017/09/19/use-cause-effect-language-jama-network-journals>
- Saxe, G. N., Bickman, L., Ma, S., & Aliferis, C. (2022). Mental health progress requires causal diagnostic nosology and scalable causal discovery. *Frontiers in Psychiatry*, 13, 1-24. <https://doi.org/10.3389/fpsy.2022.898789>
- Schiavone, S. R., Quinn, K. A., & Vazire, S. (2023, March 8). *A consensus-based tool for evaluating threats to the validity of empirical research* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/fc8v3>
- Schmeck, K., Weise, S., Schlüter-Müller, S., Birkhölzer, M., Fürer, L., Koenig, J., Krause, M., Lerch, S., Schenk, N., Valdes, N., Zimmermann, R., & Kaess, M. (2023). Effectiveness of adolescent identity treatment (AIT) versus DBT-a for the treatment of adolescent borderline personality disorder. *Personality Disorders: Theory, Research, and Treatment*, 14(2), 148–160. <https://doi.org/10.1037/per0000572>
- SCImago. (n.d.). *SCImago Journal & Country Rank*. SCImago. <http://www.scimagojr.com>
- Sieweke, J., & Santoni, S. (2020). Natural experiments in leadership research: An introduction, review, and guidelines. *The Leadership Quarterly*, 31(1), 1-25. <https://doi.org/10.1016/j.leaqua.2019.101338>
- Signorell, A. (2023). *DescTools: Tools for descriptive statistics* (R package version 0.99.49). <https://CRAN.R-project.org/package=DescTools>
- Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, 12(6), 1123–1128. <https://doi.org/10.1016/j.leaqua.2019.101338>

- Sohal, A., & Murphy, D. (2023). A longitudinal analysis of person-centred therapy with suicidal clients. *Counselling and Psychotherapy Research*, 23(1), 20–30.  
<https://doi.org/10.1002/capr.12588>
- Splaine, C., Smith, D. L., & Held, P. (2023). The role of time since trauma on treatment outcomes of veterans in two intensive posttraumatic stress disorder treatment programs. *Journal of Traumatic Stress*, 36(1), 83–93. <https://doi.org/10.1002/jts.22881>
- Spring, B. (2007). Evidence-based practice in clinical psychology: What it is, why it matters; what you need to know. *Journal of Clinical Psychology*, 63(7), 611–631.  
<https://doi.org/10.1002/jclp.20373>
- Sumner, P., Vivian-Griffiths, S., Boivin, J., Williams, A., Venetis, C. A., Davies, A., Ogden, J., Whelan, L., Hughes, B., Dalton, B., Boy, F., & Chambers, C. D. (2014). The association between exaggeration in health related science news and academic press releases: Retrospective observational study. *BMJ*, 349(dec09 7), 1-8.  
<https://doi.org/10.1136/bmj.g7015>
- Szabo, J., Koerner, N., & Carney, C. E. (2023). Social problem-solving attitudes and performance as a function of differences in trait and state worry. *Journal of Behavior Therapy and Experimental Psychiatry*, 78, 1-10.  
<https://doi.org/10.1016/j.jbtep.2022.101792>
- Tekgunduz, S., Ozturk Altinayak, S., Ejder Apay, S., Ozdemir, F., & Mohammad Alipour, M. (2023). The Relationship of Females with Sexual Dysfunctions to Different Sex Positions. *Journal of Sex & Marital Therapy*, 49(1), 56–64.  
<https://doi.org/10.1080/0092623X.2022.2067090>
- Tennant, P. W. G., & Murray, E. J. (2021). The Quest for Timely Insights into COVID-19 Should not Come at the Cost of Scientific Rigor. *Epidemiology*, 32(1), e2.  
<https://doi.org/10.1097/EDE.0000000000001258>

- Thapa, D. K., Visentin, D. C., Hunt, G. E., Watson, R., & Cleary, M. (2020). Being honest with causal language in writing for publication. *Journal of Advanced Nursing*, 76(6), 1285–1288. <https://doi.org/10.1111/jan.14311>
- Tsiplova, K., Ungar, W. J., Szatmari, P., Cost, K., Pullenayegum, E., Duku, E., Volden, J., Smith, I. M., Waddell, C., Zwaigenbaum, L., Bennett, T. A., Elsabbagh, M., Georgiades, S., & Zaidman-Zait, A. (2023). Measuring the association between behavioural services and outcomes in young children with autism spectrum disorder. *Research in Developmental Disabilities*, 132, 1-11. <https://doi.org/10.1016/j.ridd.2022.104392>
- Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*, 13(4), 411–417. <https://doi.org/10.1177/1745691617751884>
- Wan, G., Pei, T., He, X., & Zhang, B. (2023). The Association Between Child Maltreatment and Family Structures: Evidence from Children in Rural China. *Journal of Family Violence*, 38(1), 63–75. <https://doi.org/10.1007/s10896-022-00358-w>
- Wysocki, A. C., Lawson, K. M., & Rhemtulla, M. (2022). Statistical control requires causal justification. *Advances in Methods and Practices in Psychological Science*, 5(2), 1–19. <https://doi.org/10.1177/25152459221095823>
- Xie, Z., Stallings-Smith, S., Stetten, N., Hamadi, H. Y., & Marlow, N. M. (2023). Marijuana use disorder among adults with functional disabilities—A US population-based cross-sectional study. *The American Journal on Addictions*, 33(1), 26–35. <https://doi.org/10.1111/ajad.13485>
- Zhang, S., He, J., Yang, Q., Du, Y., Xiao, W., Gao, J., & Li, H. (2023). Anxiety during the COVID-19 pandemic in prisoners who had high risks to suffer from mood disorders: A longitudinal study before and during the COVID-19. *Stress and Health*, 39(1), 162–168. <https://doi.org/10.1002/smi.3176>

Zhao, Y., Paulus, M. P., & Potenza, M. N. (2023). Brain structural co-development is associated with internalizing symptoms two years later in the ABCD cohort. *Journal of Behavioral Addictions*, 12(1), 80–93. <https://doi.org/10.1556/2006.2023.00006>

## Appendices

### Appendix A: Comparison to the Original Protocol by Haber et al. (2022)

#### Objectives

The current study replicates and extends several objectives from Haber et al. (2022). Haber et al.'s objectives were shared on the Open Science Framework under the [Creative Commons Attribution 4.0 International \(CC BY 4.0\) license](https://creativecommons.org/licenses/by/4.0/). The original protocol is available at <https://osf.io/jtdaz>.

#### Table 6

##### *Comparison of Preregistered Objectives*

	Haber et al. (2022)	Current study
	Primary objectives	
1.1	Identify the associational linking words and phrases used to describe relationships between exposures and outcomes examined in the high impact published health literature.	Identify the linking language (linking words and phrases) used to describe the relationships between predictors and outcomes in the clinical psychology literature.
1.2	Generate estimates of the degree to which linking words and phrases state or imply causality.	Generate estimates of the degree to which linking language states or implies causality.
1.3	Examine the prevalence of recommendations and action claims,	Examine the prevalence of practice recommendations, and other

as well as other implications, that would require causal inference to have been made.

implications that would require causal inference to have been made.

- |     |   |  |
|-----|---|--|
| 1.4 | Examine the prevalence of disconnections between the language used to describe relationships and implications subsequently drawn. | Examine the prevalence of misalignments between causal implications of linking language and subsequent practice recommendations. |
|-----|---|--|

---

Secondary objectives

---

- |     |   |   |
|-----|---|---|
| 2.1 | Develop a list of what linking words could be considered to imply causality, based on a guided framework.                             | Examine how often studies clearly state their intention to make causal inferences.  |
| 2.2 | Document how the type and strength of language changes between the title, abstract, discussion, and action claim of journal articles. | Examine how often studies offer explicit causal disclaimer statements, and check if they still imply causality despite these disclaimers. |
| 2.3 | Identify differences in the types and strength of language used across different types of studies, clinical areas, and journals.      | Examine how often studies discuss confounding factors in relation to their methods, results, and/or interpretation.                       |

---

Tertiary objectives

---

- |     |                                 |     |
|-----|---------------------------------|-----|
| 3.1 | Examine trends in language over | N/A |
|-----|---------------------------------|-----|

---

	time.	
3.2	Examine whether studies contain causal disclaimer statements.	N/A
3.3	Examine language differences across journals.	N/A
3.4	Examine language differences across substantive medical topic areas.	N/A
3.5	Examine whether and how observational studies control/adjust for variables, and how that relates to language used.	N/A
3.6	Examine and characterize modifying phrases used to describe relationships.	N/A

---

*Note.* Primary objectives 1.1–1.4 replicate primary objectives 1.1–1.4 from Haber et al. (2022). Secondary objective 2.1 replicates a secondary objective from Mockett (2023). Secondary objective 2.2 replicates tertiary objective 3.2 from Haber et al. (2022). Secondary objective 2.3 replicates tertiary objective 3.5 from Haber et al. (2022).

### **Journal and Article Selection**

Compared to Haber et al. (2022), the approach to selecting journals and articles in the current study was narrower in scope due to the time and resource constraints associated with a Master's thesis research.

**Table 7***Comparison of Journal and Article Selection*

<b>Aspect</b>	<b>Haber et al. (2022)</b>	<b>Current study</b>
Domain	Health, medicine, and epidemiology.	Clinical psychology
Journal selection	Selected the top 18 high-profile journals based on impact factor, h-index, and SCImago Journal Rank (SJR).	Selected journals under the Clinical Psychology category on SJR.
Data source	PubMed (a free search engine accessing primarily the MEDLINE database).	SJR and individual journal websites.
Selection criteria for journals	Journals from general health, medicine, or epidemiology categories, excluding narrow specialty journals.	Journals publishing original research articles, excluding journals exclusively publishing review/commentary articles and topics outside clinical psychology.
Article inclusion criteria	<ul style="list-style-type: none"> <li>● Quantitative studies examining the relationship between a main exposure and outcome</li> <li>● Focus on human or aggregate group observations</li> <li>● Excludes studies with multiple exposure/outcome sets, reviews, or meta-analyses</li> </ul>	<ul style="list-style-type: none"> <li>● English language</li> <li>● Observational design (cross-sectional, longitudinal)</li> <li>● Focus on human or aggregate group observations</li> <li>● One primary predictor and outcome pair</li> </ul>

Article exclusion criteria	<ul style="list-style-type: none"> <li>• Reviews or meta-analyses</li> <li>• Studies examining more than one exposure/outcome pair</li> </ul>	<ul style="list-style-type: none"> <li>• Experimental or quasi-experimental designs</li> <li>• Secondary study designs (reviews/meta-analyses)</li> <li>• Studies focusing on mediators, moderators, or interaction effects</li> <li>• Studies outside of the clinical psychology subdiscipline</li> </ul>
Article selection process	Screened a fixed number of articles (65 non-RCTs and six RCTs) from each journal for the 2010-2019 time frame.	Iterative process starting from the first issue of 2023 and moving backward if needed to find one eligible article per journal.

**Table 8***Comparison of Data Extraction Process*

Aspect	Haber et al. (2022)	Current study
Extent of review	Reviewed abstracts of all articles ( $n = 1,275$ ). Only one-third underwent full-text review.	All articles ( $n = 97$ ) were reviewed in full, including both abstract and concluding sections.
Reviewer process	<ul style="list-style-type: none"> <li>• Used multiple reviewers (two independent and one arbitrating) with systematic training and guidelines.</li> </ul>	<ul style="list-style-type: none"> <li>• Primary researcher as the single reviewer due to resource constraints associated with a Master's thesis research.</li> </ul>

	<ul style="list-style-type: none"> <li>● Reviewers included doctoral-level students, postdoctoral fellows, faculty, and others from diverse fields.</li> <li>● Training included a one-hour session and discussion on Slack to clarify guidelines and foster standardisation.</li> </ul>	<ul style="list-style-type: none"> <li>● Review tool piloted with 10 articles from 2010 to 2013 for training and to refine the review tool.</li> </ul>
Selection of linking sentences	Selected the first occurrence of a sentence containing the linking word between the exposure and outcome.	Selected the sentence that maximally implied causation to provide greater insight.
Questions on the review tool	<p>The review tool extracts:</p> <ul style="list-style-type: none"> <li>● pop-out sections</li> <li>● causal theory explanations</li> <li>● formal causal models</li> <li>● confounders in methods, results, or discussion</li> <li>● linking language across sections</li> <li>● strength of causal language</li> <li>● action recommendations</li> <li>● mediation and moderation analysis.</li> <li>● causal implication disclaimers</li> </ul>	<p>The review tool extracts:</p> <ul style="list-style-type: none"> <li>● primary linking sentence and modifying phrases</li> <li>● causal implications of linking sentences</li> <li>● practice recommendations</li> <li>● causal disclaimer statements</li> <li>● explicit intent to draw causal inference</li> <li>● discussion of confounders throughout the article</li> </ul>

Mediation analysis	No specific questions or considerations for mediation in the review process.	No specific mediation questions, but studies focusing on mediators, moderators, or interaction effects were excluded due to complexity.
Causal theory explanations	Included questions on the presence of causal theory explanations in the introduction and methods sections.	No direct questions on causal theory explanations but considered any statements indicating an intent to draw causal inference.
Pop-out sections	Reviewed pop-out sections separately if available, examining for linking sentences and action recommendations.	Did not specifically consider pop-out sections; focused on abstract and concluding sections only.
Causal implication ratings for linking language	Used a four-point scale to rate the causal implication of linking sentences: <i>None, Weak, Moderate, Strong</i> .	Used a simplified three-point scale: <i>None, Moderate, and Strong</i> . Eliminated the <i>Weak</i> category for clarity and efficiency.
Causal implication ratings for practice recommendations	Rated practice recommendations for their causal implications on a four-point scale ( <i>None, Weak, Moderate, Strong</i> ) and selecting those that most strongly implied causality.	Rated practice recommendations for their causal implications on a three-point scale ( <i>None, Moderate, Strong</i> ) and selecting those that most strongly implied causality.

---

**Table 9***Comparison of Data Analysis*

<b>Aspect</b>	<b>Haber et al. (2022)</b>	<b>Current Study</b>
Comparison of root words and sentences	Made a comparison between decontextualised root word ratings and contextual ratings of linking sentences.	N/A
Data analysis techniques	Descriptive statistics, Spearman's correlation coefficients (using the <i>pspearman</i> package), ordinal logistic regression (using the <i>MASS</i> package), and block bootstrapping for estimating statistical uncertainty were performed using R software.	Descriptive statistics using R software and the <i>DescTools</i> package; two-tailed Spearman's rho rank test to examine the relationship between linking language and practice recommendations.

## Appendix B: Changes Made to Preregistration Document

This section provides a list of changes made between the original and revised versions of the preregistration document. The changes were made after the study was performed and during the writing of the thesis manuscript, for the purpose of improving clarity. Both versions are available on the Open Science Framework repository

([https://osf.io/pqtxd/?view\\_only=52a55a65cafc4b0285093596a791b0b6](https://osf.io/pqtxd/?view_only=52a55a65cafc4b0285093596a791b0b6)).

- **Section B. Objectives.** The title “Objectives/Research Questions” was shortened to “Objectives” for accuracy.
- **Section D. Methods.** In Phase 3: Data extraction, “Discussion and conclusion sections” was changed to “Concluding section (discussion or conclusion section)” for clarity and consistency.
- **Section E. Analysis.** The secondary objectives were expanded into three detailed points:
  - i. Examine how often studies clearly state their intention to make causal inferences.
  - ii. Examine how often studies offer explicit causal disclaimer statements, and check if they still imply causality despite these disclaimers.
  - iii. Examine how often studies discuss confounding factors in relation to their methods, results, and/or interpretation.
- **Minor language changes.** We adjusted from plural to singular forms in several instances (for example, “practice recommendations” to “practice recommendation”). Additionally, some sentences were restructured slightly for clarity.
- **Deviation from preregistration.** I have not employed this preregistered review procedure in the current study: *The primary researcher will denote uncertain questions during the review process by marking the answers with 42. At some point during or after*

*the review concludes the primary researcher will address these uncertainties together with the supervisor, with a focus on achieving the most accurate data extraction.*

## **Appendix C: Final Preregistration Document**

### **Topic: Causal Language and Practice Recommendations in Clinical Psychology Research**

#### **A. Introduction/Background**

The use of causal language in psychology research is fundamental. This is true in clinical psychology research, especially in understanding the causes of mental distress and the effectiveness of interventions. Clinical psychology research aims to provide actionable practice recommendations for practitioners and policymakers, going beyond merely describing the phenomena. Hence, the use of causal language is essential for translating research findings into effective treatments, which can have a significant impact on real-world mental health outcomes. Due to ethical constraints and practical limitations in experimental manipulation, clinical psychology relies heavily on observational studies to explore the complex relationships between variables. However, the dissemination of findings from observational studies has been a subject of concern due to the potential for misrepresentation and distortion.

The present study is a conceptual replication and extension of Haber et al.'s (2022) study—which investigated the use of (causal) linking language in high-profile health literature. In the current study, we aim to analyse the language used in clinical psychology literature to identify linking language, assess its causal implications, and examine the alignment between causal language and practice recommendations. Secondary objectives involve examining how often studies explicitly state their intention to make causal inferences, causal disclaimer statements, and discussing confounding factors. This study aims to contribute significantly to understanding the interplay between language, causality, and practice recommendations in clinical psychology literature.

## **B. Objectives**

### ***Primary objectives***

1. Identify the linking words and phrases (hereinafter referred to as linking language) used to describe the relationships between predictors and outcomes in the clinical psychology literature.
2. Generate estimates of the degree to which linking language states or implies causality.
3. Examine the prevalence of practice recommendations, and other implications that would require causal inference to have been made.
4. Examine the prevalence of misalignments between causal implications of linking language and subsequent practice recommendations.

### ***Secondary objectives***

5. Examine how often studies clearly state their intention to make causal inferences.
6. Examine how often studies offer explicit causal disclaimer statements, and check if they still imply causality despite these disclaimers.
7. Examine how often studies discuss confounding factors in relation to their methods, results, and/or interpretation.

## **C. Definitions**

The definitions below are drawn from Haber et al. (2022) and Mockett (2023). These definitions are provided for brief reference.

1. Predictor: The primary or antecedent variable (in a regression sense) being investigated for a possible (non-)causal link to the study outcome, or resulting or end-point variable. It is also known as the independent variable.

2. Outcome: The dependent or effect variable of interest that is being investigated for a possible link to the predictor variable. It is typically assumed or known to be preceded by the predictor. It is sometimes called the study endpoint variable, dependent variable, response variable, consequence, result, and other recognised terms.
3. Linking word/phrase: A word or phrase that describes the nature of the connection between a specified predictor and a specified outcome, generally used in a sentence containing both predictor and outcome. For our purposes, the phrase may contain one to three words, with one of the words possibly being a preposition to link the predictor and outcome.
4. Linking word/phrase modifier: A word or phrase that modifies the relationship between the predictor and outcome. This includes adding signals of direction, strength, or doubt to the relationship.
5. Causal linking word/phrase: Causal language implies that one entity influences another. This can be expressed through various means, including verbs that imply movement (or lack thereof) in the outcome was impelled by the predictor of interest (e.g., increase, decrease, improve, changed) as well as conjunctions that imply attribution of the outcome to the predictor (e.g. because, due to, since).
6. Practice recommendation: A practice recommendation is a description of how a consumer of research might utilise the results and conclusions of the research. A recommendation may often imply a causal interpretation of a finding. For example, an author may suggest that it could be beneficial to change the amount of a predictor, which rests on the assumption that the predictor has a causal effect on the outcome. As a variation, it may also be suggested that a predictor need not be changed, which rests on the assumption that the absence of a causal effect has been established. For the

purpose of this study, we will exclude any practice recommendation that involves requesting additional research.

#### **D. Methods**

This study is a small-scale analysis of the language used in the clinical psychology literature.

The target sample includes studies published in clinical psychology journals.

##### ***Journal inclusion criteria***

1. Journals listed on the SCImago Scientific Journal Rankings (SJR) under Clinical Psychology subject. The metrics are based on Scopus data as of April 2023.
2. Journals that publish original articles.

##### ***Journal exclusion criteria***

1. Journals that provide only reviews and/or summaries of existing research articles (i.e. review journals) will be excluded.
2. Journals that primarily publish studies on topics outside of the clinical psychology subdiscipline.

##### ***Article inclusion criteria***

1. English language.
2. Observational (i.e., non-experimental) design, including longitudinal studies.
3. Human or aggregate group of humans unit level observations.
4. The primary research question must be examined quantitatively using primary data (i.e. not a review or meta-analysis).
5. The primary research question must examine the causal and/or non-causal association between one primary predictor concept and one primary outcome concept.
  - 5.1. One primary predictor or outcome can encompass multiple measures of the same or similar concept.

- 5.2. While the articles may discuss multiple predictors and outcomes, they should particularly emphasise one predictor and outcome pair as their primary association of interest. This emphasis should be reflected in the title and abstract.
6. The article must be published, whether in an issue or as an advance online publication (i.e., not a preprint).

***Article exclusion criteria***

1. Experimental or quasi-experimental studies.
2. Secondary study design (e.g. review and meta-analysis).
3. Studies that are primarily concerned with mediators, moderators, or interaction effects.  
This is because (i) such studies would not allow for the assignment of single relationships per paper, (ii) study objectives and designs could not easily be compared with other papers, and (iii) would unnecessarily complicate data management and review.
4. Studies that investigated topics outside of the subdiscipline of clinical psychology.

***Sample size determination and rationale***

1. The primary analyses in this study will involve estimating population percentages with confidence intervals. To determine the sample size for this review, the parameters considered were a confidence level of 95%, a margin of error set at 10%, and a population proportion of 50%. Using the Sample Size Calculator with these specifications resulted in a calculated sample size of 97.
2. The chosen sample size is practical for a Master's thesis with a single reviewer and sufficiently extensive to cover a diverse array of journal articles, as detailed below.

***Phase 1: Journal screening and selection***

1. Select the top 97 articles according to SCImago Scientific Journal Rankings (SJR) - Clinical Psychology subject.
2. Assess each journal against the inclusion and exclusion criteria outlined above.
3. Determine the number of excluded journals and substitute them with additional journals from the rankings.

### ***Phase 2: Article screening and selection***

1. The objective is to select one article from each journal.
2. Start with the first issue of 2023 and assess the eligibility of the first article.
3. If the first article is not eligible, move on to the second article of the first issue, and so forth. If none are eligible, proceed to the second issue, and continue this process.
4. In cases where no articles from 2023 meet the eligibility criteria, backtrack to the last article of the last volume (in 2022) and proceed in reverse order from that point.

### ***Phase 3: Data extraction***

1. Review tool.
  - 1.1. A review tool adapted from Haber et al. (2021) will be used. A draft of the review tool is attached to this preregistration and is subject to changes.
2. Training and administration/piloting.
  - 2.1. 10 clinical psychology articles from the years 2010 to 2013 would be chosen arbitrarily during the refinement of the review tool.
3. Data to be extracted by the primary researcher.
  - 3.1. Article DOI and title.
  - 3.2. Predictor and outcome variables.
  - 3.3. Abstract section: Primary linking sentence (containing primary linking word/phrase and modifying word/phrase). If available, practice recommendations.

- 3.4. Concluding section (discussion or conclusion section): Primary linking sentence (containing primary linking word/phrase and modifying word/phrase). If available, practice recommendations, causal disclaimer statement, and acknowledgement of intent to draw causal inference.
- 3.5. Additional information: If available, discussion of confounders in relation to the methods, results, and/or interpretation of the study, located anywhere in the article.
4. Review procedures.
  - 4.1. The primary researcher will independently review papers and extract data.
  - 4.2. The primary researcher will denote uncertain questions during the review process by marking the answers with 42. At some point during or after the review concludes the primary researcher will address these uncertainties together with the supervisor, with a focus on achieving the most accurate data extraction.
  - 4.3. Articles will be reviewed in random order.
  - 4.4. A keyword scan using the text-search function will assist in the data extraction process.
  - 4.5. The results sections will not be examined.

## **E. Analysis**

### ***Statistical analysis***

1. The analyses are largely descriptive in nature.
2. In cases where no relevant data are available, this will be recorded as N/A.
3. Listwise exclusion for missing data will be applied during the analysis.
4. All proportions presented will include a 95% confidence interval.
5. No ethics or low-risk notifications are applicable to the analyses.

6. All materials and data will be made available on the Open Science Framework repository.

***Main outcomes measures for each objective***

***Primary objectives***

1. Primary objective 1: Identify the linking language.
  - 1.1. Determine the frequency and proportions of linking language, categorised based on their root words.
  - 1.2. These will be presented separately for the abstracts and the concluding sections.
2. Primary objective 2: Generate estimates of the degree to which linking language states or implies causality.
  - 2.1. Determine the proportions of linking language according to three categories/levels of causal implications (none; moderate; and strong).
  - 2.2. These will be presented separately for the abstracts and the concluding sections.
3. Primary objective 3: Examine the prevalence of practice recommendations that would require causal inference to have been made.
  - 3.1. Determine the proportions of practice recommendations' strength according to three categories/levels of causal implications (none; moderate; and strong).
  - 3.2. These will be presented separately for the abstracts and the concluding sections.
4. Primary objective 4: Examine the prevalence of misalignment between linking language and practice recommendations:
  - 4.1. Determine the proportions of the causal strength of linking language that corresponds with the causal strength of practice recommendations.
  - 4.2. Estimate the relationship between the causal strength of linking language and the causal strength of practice recommendations using a two-tailed Spearman's rho rank test at a 5% alpha level.

- 4.3. Determine the proportion of articles in which the causal strength of the practice recommendations is stronger than the causal strength of the linking language.
- 4.4. These will be presented separately for the abstracts and the concluding sections.

***Secondary objectives***

1. Determine the proportion of articles (in the concluding sections) that explicitly acknowledge an intent to draw causal inference.
2. Determine the proportion of articles (in the concluding sections) that explicitly mention causality as a cautionary/disclaimer statement.
3. Determine the proportion of articles (anywhere in the article) that discuss confounding factors.

## Appendix D: Changes Made to Review Tool

This section provides a list of changes made between the original (Review Tool 5.0) and revised (Review Tool 7.0) versions of the document. Review Tool 5.0 was used for training and piloting purposes; however, Review Tool 7.0 was used for the actual data extraction. Both versions were preregistered and are available on the Open Science Framework repository. We note that several iterations occurred before and after Review Tool 5.0, before arriving at the final version (7.0)

- The original review tool had 23 questions, but the final version was reduced to 20 after removing or merging similar questions to streamline the review process.
- The “Discussion/Conclusion” section was renamed to “Concluding sections (discussion or conclusion)” for clarity.
- Review Tool 7.0 combined questions 17 and 18 from the Review Tool 5.0 into a single question (question 17) about causal disclaimer statements to reduce redundancy.
- Review Tool 7.0 combined questions 19 and 20 from Review Tool 5.0 into a single question (Question 19) about explicit intent to draw causal inference to avoid redundancy.
- Review Tool 5.0 questions 19 and 20 were both about identifying statements that explicitly acknowledge an intent to draw causal inference. In the revised version, these were merged into one question (Question 19 in Review Tool 7.0) to avoid duplication.
- Review Tool 7.0 question 19 includes a new response option (“Yes, in the results section”) to provide clearer identification of where confounding factors are discussed in the study.
- Several instructions were added or revised for clarity. For example, in question 11, “If none are found in this section, proceed to the second *last* section of the article” was

rephrased to “If none are found in the conclusion section, proceed to the *discussion* section of the article”.

## Appendix E: Final Review Tool

# Review Tool

Topic: Causal Language and Practice Recommendations in Clinical Psychology Research

### 1. Article DOI

Your answer<sup>9</sup>

### 2. Article Title

Your answer

### 3. Predictor/Independent Variable(s)

Copy a few-word description directly extracted from the title and/or abstract. Extract it from the title if available. Otherwise, check the introduction section, discussion section, or results section of the abstract, in that order. There may be multiple predictors/independent variables.

Your answer

### 4. Outcome/Dependent Variable(s)

Copy a few-word description directly extracted from the title and/or abstract. Extract it from the title if available. Otherwise, check the introduction section, discussion section, or results section of the abstract, in that order. There may be multiple outcome/dependent

---

<sup>9</sup> The phrase “Your answer” constitutes a free text field on the Google Forms platform.

variables.

Your answer

### **Abstract: Linking Language**

#### 5. Abstract: Primary Linking Sentence

Identify the key sentence or phrase that links the predictor and outcome variables, focusing on a sentence containing the independent variable, outcome, linking word, and any modifying phrases. Copy and paste this sentence from the abstract. If multiple sentences meet these criteria, select the one with the most robust causal language.

Your answer

#### 6. Abstract: Primary Linking Word/Phrase

Based on the sentence copied above, please select the primary linking word or phrase and copy it here. The word or phrase should be a maximum of three words and can include a preposition. It should not contain any modifying phrases. If there are multiple words or phrases, choose the one that most strongly implies causality.

Your answer

#### 7. Abstract: Modifying Word/Phrase

Based on the sentence copied above, copy any modifiers here. These modifiers include any words or phrases that affect the nature (e.g., strength, intensity, room for doubt, negotiation, direction, etc.) of the primary linking word/phrase. If there are multiple modifiers, separate them with a semicolon. Enter "N/A" if not available.

Your answer

### 8. Abstract: Linking Language Causal Implication

How strongly does the language in this sentence imply that the authors identified a causal relationship between the primary predictor/independent variable of interest and the primary outcome/dependent variable of interest?

- None: The linking sentence does not imply in any way a causal relationship was identified.
- Moderate: The linking sentence, while mostly implying a causal relationship, remains unclear or inconclusive in the absence of any causal inference.
- Strong: The linking sentence clearly implies that causality has been identified.

### **Abstract: Practice Recommendation**

#### 9. Abstract: Practice Recommendation

Copy a claim about how a consumer of this research might apply its results and conclusions. If there are multiple recommendations, choose the one that most strongly implies causality. Exclude any recommendations that involve requesting additional research. Enter "N/A" if not available.

Your answer

#### 10. Abstract: Practice Recommendation Causal Implication

Choose the claim that most strongly implies or requires that the evidence from this study was of a causal nature. How strongly is this implication made?

- N/A: No practice recommendation(s) provided in this abstract.
- None: The practice recommendation would be made appropriately in the absence of any causal relationship.

- O Moderate: The practice recommendation may be made appropriately had a causal relationship been identified, but it is unclear or difficult to come to that recommendation in the absence of any causal inference.
- O Strong: The practice recommendation could only be made appropriately had a causal relationship been identified.

## **Concluding Sections (Discussion or Conclusion): Linking Language**

### 11. Concluding Sections: Primary Linking Sentence

Identify the primary sentence or phrase containing linking words that connect the independent variable to the outcome variable. Copy and paste the identified sentence from the very last section of the article, regardless of the heading title.

Give preference to the first paragraph of the section. If multiple sentences equally meet these criteria, select the one with the strongest causal language. If none are found in the first paragraph, proceed to the second paragraph, and so on until you find the linking sentence.

If none are found in the Conclusion section, proceed to the Discussion section of the article. Pay particular attention to sentences that include the exposure, outcome, linking words, and any modifying phrases.

Your answer

### 12. Concluding Sections: Primary Linking Word/Phrase

Based on the sentence copied above, select the primary linking word or phrase and copy it here. The word or phrase should be a maximum of three words, often including a preposition. It should not contain any modifying phrases. If there are multiple words or phrases, choose the one that most strongly implies causality.

Your answer

### 13. Concluding Sections: Modifying Word/Phrase

Based on the sentence copied above, copy any modifiers here. These modifiers include any words or phrases that affect the nature (e.g., strength, intensity, room for doubt, negotiation, direction, etc.) of the primary linking word/phrase. If there are multiple modifiers, separate them with a semicolon. Enter "N/A" if not available.

Your answer

### 14. Concluding Sections: Linking Language Causal Implication

How strongly does the language in this sentence imply that the authors identified a causal relationship between the primary predictor/independent variable of interest and the primary outcome/dependent variable of interest?

- None: The linking sentence does not imply in any way a causal relationship was identified.
- Moderate: The linking sentence might imply a causal relationship was identified, but it is unclear or difficult to come to that conclusion in the absence of any causal inference.
- Strong: The linking sentence clearly implies that causality has been identified.

## **Concluding Sections: Practice Recommendation**

### 15. Concluding Sections: Practice Recommendation

If available, copy any major claims made about how a consumer of this research might utilise its results and conclusions from the very last section of the article, regardless of the heading title. If there are multiple claims, choose the one that most strongly implies causality. If none are found in the Conclusion section, proceed to the Discussion section of the article.

Note: Actions calling for more research do not apply here. Enter "N/A" if not available.

Your answer

### 16. Concluding Sections: Practice Recommendation Causal Implication

Regarding the selection of a claim that most strongly implies or requires that the evidence from this study was of a causal nature, how strongly is this implication made?

- N/A: No practice recommendation(s) provided in the discussion/conclusion.
- None: The practice recommendation would be made appropriately in the absence of any causal relationship.
- Moderate: The practice recommendation may be made appropriately had a causal relationship been identified, but it is unclear or difficult to come to that recommendation in the absence of any causal inference.
- Strong: The practice recommendation could only be made appropriately had a causal relationship been identified.

### 17. Concluding Sections: Causal Disclaimer Statements

Are there any statements in the discussion/conclusion which explicitly mention causality as a cautionary or disclaimer statement? Examples may include phrases such as "correlation does not equal causation", or "the observational nature of this study means that causal inferences cannot be drawn". Copy and paste the identified statement(s) above if available. Enter "N/A" if not available.

Your answer

### 18. Concluding Sections: Intent to Draw Causal Inference

If there are any statements in the discussion/conclusion which explicitly acknowledge an intent to draw causal inference, paste them here. Enter "N/A" if not available.

Your answer

### 19. Anywhere In the Article: Confounding Factors

Is "confounding", "confounders", or "third variable" discussed or mentioned in relation to the methods, results, and/or interpretation of this study? Includes "confound\*" where \* can be any suffix. Explicit word(s) must be included. This must be in reference to the methods, results, and/or interpretation of this study, and not in reference to other studies. Check all that apply.

- No, this does not appear in the study manuscript.
- Yes, in the introduction.
- Yes, in the methods section.
- Yes, in the results section.
- Yes, in the discussion/conclusion section.

### 20. Additional Comments

Enter "N/A" if not applicable.

Your answer

## **Appendix F: List of Journals Considered During Journal Screening and Selection**

### **Journals excluded due to belonging to another subdiscipline**

- Political Psychology (other subdiscipline)
- Social Psychological and Personality Science
- Psychological Assessment
- Assessment
- Journal of Personality Assessment

### **Journals excluded due to being review journals**

- Aggression and Violent Behavior
- Alcohol research: current reviews
- Annual Review of Clinical Psychology
- Clinical Child and Family Psychology Review
- Clinical Psychology Review
- Clinical Psychology: Science and Practice
- Current Addiction Reports
- Current Treatment Options in Psychiatry
- Health Psychology Review
- International Journal of Psychoanalysis
- Journal of the American Psychoanalytic Association
- Perspectives on Behavior Science
- Psychoanalytic Quarterly
- Sleep Medicine Clinics

### **Journals included**

- Psychotherapy and Psychosomatics
- Journal of Anxiety Disorders

- Depression and Anxiety
- International Journal of Clinical and Health Psychology
- Journals of Gerontology - Series B Psychological Sciences and Social Sciences
- Journal of Consulting and Clinical Psychology
- Clinical Psychological Science
- Psychological Trauma: Theory, Research, Practice, and Policy
- Journal of Abnormal Psychology
- Journal of Affective Disorders
- Journal of Behavioral Addictions
- Comprehensive Psychiatry
- Cognitive Behaviour Therapy
- Journal of Clinical Child and Adolescent Psychology
- Behaviour Research and Therapy
- Schizophrenia
- Journal of Traumatic Stress
- Journal of Substance Abuse Treatment
- Clinical Neuropsychologist
- Family Process
- Borderline Personality Disorder and Emotion Dysregulation
- Journal of Counseling Psychology
- Behavior Therapy
- Anxiety, Stress and Coping
- Addictive Behaviors
- Psychology and Psychotherapy: Theory, Research and Practice
- International Psychogeriatrics

- Psychotherapy Research
- Clinical Psychology and Psychotherapy
- Personality Disorders: Theory, Research, and Treatment
- Stress and Health
- Cognitive Therapy and Research
- Addiction science & clinical practice
- British Journal of Clinical Psychology
- Chronic Stress
- Journal of Alzheimer's Disease
- Depression Research and Treatment
- Journal of Psychosomatic Research
- Journal of Applied Research in Memory and Cognition
- Addictive Behaviors Reports
- Journal of Trauma and Dissociation
- American Journal of Drug and Alcohol Abuse
- Journal of Attention Disorders
- Psychology of Addictive Behaviors
- Journal of Interpersonal Violence
- European Eating Disorders Review
- Behavioral Disorders
- Journal of Clinical Psychology
- Psychopathology
- American Journal on Addictions
- Rehabilitation Psychology
- Death Studies

- Psychological Services
- Neuroscience of Consciousness
- Psychology, Health and Medicine
- Cognitive and Behavioral Practice
- Journal of Psychotherapy Integration
- Psychotherapy
- Stigma and Health
- Behavior Modification
- Counselling Psychology Quarterly
- Journal of Psychosomatic Obstetrics and Gynaecology
- Journal of Sex and Marital Therapy
- Behavioural and Cognitive Psychotherapy
- Research in Autism Spectrum Disorders
- Journal of the International Neuropsychological Society
- Journal of Personality Disorders
- Clinical Gerontologist
- ADHD Attention Deficit and Hyperactivity Disorders
- Journal of Behavior Therapy and Experimental Psychiatry
- Eating Disorders
- Health Psychology Open
- Research in Developmental Disabilities
- Suicide and Life-Threatening Behavior
- European Journal of Trauma and Dissociation
- Clinical Psychology in Europe
- Counselling and Psychotherapy Research

- Archives of Clinical Neuropsychology
- Journal of Social and Clinical Psychology
- Journal of Psychopathology and Behavioral Assessment
- Journal of Contemporary Psychotherapy
- Research in Psychotherapy: Psychopathology, Process and Outcome
- Eating and Weight Disorders
- Palliative and Supportive Care
- Eating Behaviors
- American Journal of Alzheimer's Disease and other Dementias
- Neuropsychiatrie
- Journal of Family Violence
- Cognitive Behaviour Therapist
- Journal of Addictive Diseases
- Indian Journal of Psychological Medicine
- International Journal of Art Therapy: Inscape
- Journal of Aggression, Maltreatment and Trauma
- Alzheimer Disease and Associated Disorders
- Journal of Experimental Psychopathology
- Journal of the Academy of Consultation-Liaison Psychiatry
- Psychology of Consciousness: Theory Research, and Practice

**Appendix G: Technology Used for Data Collection, Data Analysis, and Writing**

- Google Forms: Platform for review tool.
- Grammarly: Writing assistant for spelling, grammar, and punctuation.
- Microsoft Excel and Google Sheets: Data management and generating graphs.
- Microsoft Word and Google Docs: Writing manuscript and editing with supervisor feedback.
- Open Science Framework: Platform for preregistration and repository for study materials and data.
- R Program: Statistical analysis and data visualisation, including the DescTools Package (Signorell et al., 2023) for descriptive statistics.
- Sample Size Calculator (Calculator.net): Sample size calculation.
- Zotero: Reference management software.