IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

P-Hacking A Modern-Day Phenomenon: Its Impact In Various Fields, Rise Of X-Hacking, A Review

[1] Harpreet Singh, [2]MS.Bhawna Goel, [3]Sachin Kumar, [4] Harsh

[1]Student in Division of Pharmacy, SOHAS, GD Goenka University, [2]Proffesor (Assistant) at Division of B. Pharmacy SOHAS, GD Goenka University, [3] Student in Division of Pharmacy, SOHAS, GD Goenka University, [4]Student in Division of Pharmacy, SOHAS, GD Goenka University

P-Hacking is a type of data manipulation which is on rise in modern day P-hacking is when data analyst or researcher only reports statistically significant results and it is also leads to publication bias, P-hacking is common issues which started in fields of sciences and made its way to Management but since then has managed to crawl its way into other fields as of now it's overall effect till date remains low on literature remain low. Now, P-hacking can be defined as practice of data manipulation where researchers only report significant data and is achieved by changing way data is collected or reporting data which supports the research, P-hacking has led to various issues in academic research and research publications like researchers only showing data which support their studies it can lead to mistrust between publishers and researcher and p-hacking is commonly paired with Harking and data fabrication

Key Words: P-Hacking, Significant Data, P-values, Bias, Harking, Manipulation

I. INTRODUCTION

P-hacking hacking can be defined as a type of data manipulation in which data analyst or researchers only report statistically significant data [24] [5] which can lead to publication bias or simply defined as questionable practices, publication bias refers to failure to publish study data on strength of paper/study. There are 2 types of publication bias – A) Non-Publication Bias: Impacts ability to define Evidence in given area. B) Publication Bias: is bias related to reporting of data also referred to dismission bias where certain part of data is dismissed [5]. P-hacking can also be said to be done when P-values are Manipulated such as when researcher only reports low P-values while ignoring high P-values [22],Now coming back to P-hacking it is a common issue on sides Science and management and various fields . As of now P-hacking does not have significant effect on literature but a rising concern has been observed as there is struggle in replicating results of research and empirical data is being manipulated to adhere to support specific research or purposes.

As mentioned earlier P-hacking is data manipulation where researcher or data analyst present findings or data which support their cause this is achieved by changing the way of collecting data or manipulating empirical data to support research/cause [6]. The increase in P-hacking has led to rise in growing mistrust in publishing of recent research paper. This has ultimately jeopardized scientific research [5]

This is due to misleading results produces by p-hacking as they do not have any effect on real life nor they reflect in reality and other factor adding to this inability to replicate research result. Figure 1.1 shows various forms of P-hacking & Figure 1.2 shows various fields affected by P-hacking

Figure 1.1 Various Forms of P-hacking

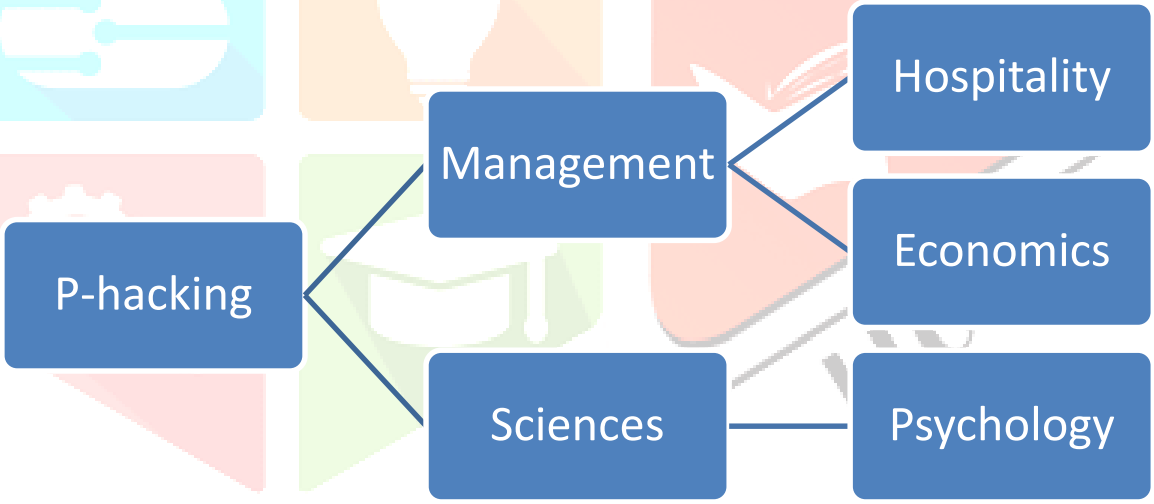


Figure 1.2 Fields Affected by P-hacking

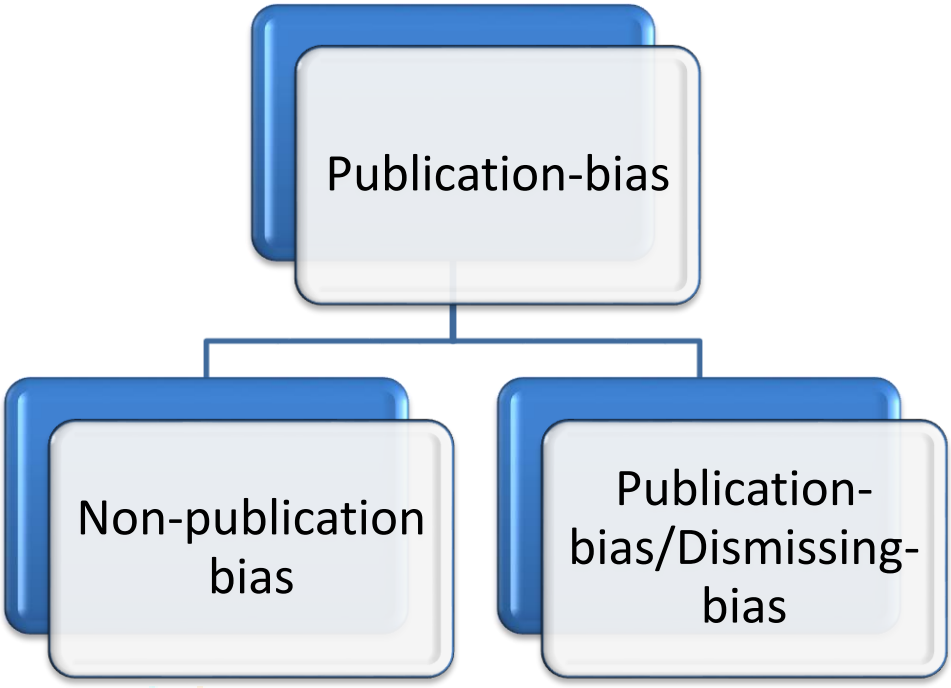


figure 1.3 Types of Publication-Bias [5]

1.1.1 P-hacking In Management:

P-hacking in management is done mostly due to academic pressure as in management paper publishing of great importance, now it was observed that 2 authors display many significant variables / coefficients (i.e., .04< p ≤.05) Example (Banks et al., 2017; Bettis 2012; Charness, Masclet &Villeval 2013; Schwab & Starbuck 2017) [6]. While p-hacking has made its way into management it has not affected restaurant nor field of hospitality yet [25]

Baum, Joel A.C., Bromiley, Philip also suggested that people who obtained degrees at middle schools are more prone to engage in acts of p-hacking while females or those who acquired degrees via elite or prestige institutions are less like likely to engage in P-hacking or Q.R. P'S [5]

1.1.2 P-hacking in Business:

One of the key factors for presence of p-hacking in business research papers is its reliance on statistical method and hypothesise testing one of major issues in here is that scholars and researchers heavily rely on the historical and empirical data which gives them opportunity to try out Villeda various variables or model specifications until there data seems significant. Here it can be caught as if P-value has exceeded or preceded the threshold (i.e. P>0.05) it can then be considered that data is not real and is thus insignificant [22]

1.1.3 P-hacking In Psychology:

P-hacking is said to be done when a researcher engages in Q.R.P or reports statistically significant data only it was observed by Crede, Marcus Sotola, Lukas K. That only 37% of papers published in psychology were legit and had real life value [7] To do this they used

Z-curve which is analytical technique which is like p-curve but is used in heterogenous condition it is used to predict replication rate and suggest that STJ publication will show low replication [**7]'**

1.1.4 P-hacking In Social Sciences:

At present field of social sciences is seeing a surge in activities such as replication and pre-registrations which makes it difficult and prevents researchers from adjusting the data thus prevent them from generating statistically insignificant data or fabricating data and forces them to publish real/significant data which has real life value and prevent p-hacking [19]

1.1.5 Future Variant, X-hacking/ Sciences:

XAI stands for explainable ai it is used in machine learning to build trust in model predictions but in case of x-ai which is variant of p-hacking P shap values of x-ai

Are manipulated to rig the result or give a pre-specified result [23]

It has always been difficult to study tissue and organ biology especially in mammal cells but the rapid progression and advancements in the field of stem cell culture have made it possible to derive in vitro 3D tissues called organoids. It also demonstrated vast potential that organoids have and that they can be used

LITERATURE REVIEW:

Table 1. Advancement in p-hacking. Showcases evolution of term P-hacking how it came about and how definition of p-hacking evolved overtime and slow growth of p-hacking and its consequences and how it led to rise in publication bias, its slow transmission to fields of science and ai, various form of p-hacking its effect on fields and their respective publications aswell the Rise of x-hacking.

Author	Year	Advancement	
Baum, Joel A.C., Bromiley, Philip	2019	One if the first papers which suggested that p-hacking was an endemic in science which eventually spread out to Management.	
Oleg Rytchkov, Xun Zhong	2019	Studied relation between information aggregation and p-hacking and how they can used to manipulate stock market predictions	
Oleg Rytchkov, Xun Zhong	2019	Emphasises on how due to lack of data availability and reliance on historical and empirical data led to increase in p-hacking in field of business.	
Jasper Brinker ink	2021	Defined P-hacking as Q.R.P And submission of incomplete data with increase in use in case of publishing literature	
Robert J. Macoun	2022	Defined p-hacking as an act in which researchers enter into concerning / questionable practices	
Angelika M. Stefan, Felix D. Schönbrodt	2023	Discusses jeopardy p-hacking has led to name of research, discusses publication bias	
Han-fen Hu, Dennis F. Galletta, Gregory D. Moody	2023	Discuses p-hacking as repeated tinkering of particular and emphasises on “Harking”	
Abel Brodeur Scott Carrell David Figlio Lester Lusher	2023	Studies publication-bias which is a type of p-hacking. Involves dismissing data which does not support research or only submitting data which supports the research, Also, discusses how unintentionally researcher can prefer a data	
Ian D. Gow	2023	Emphasis on empirical accounting data and how to reduce p-hacking and how to prevent it and effects of p-hacking on empirical accounting data	
Angelika Marlene Stefan, Felix D. Schönbrodt	2023	Discusses and analyses compound strategies looking to render non-significant hypothesis as significant.	
Alisha Gupta	2023	Used long curated database to I identify p-values Discussed P-hackings history and methods use to identify it and prevent it	

Marko Sarstedt, Susanne J. Adler	2023	Introduced a new value known as "$P_{ointless\ value}$" which shall ensure integrity of data affects of p-hacking in social sciences	
Robin Chark, Michael Man Him	2023	Talks about how p-hacking has made way into management but has not impacted the field of hospitality, tried to study p-hacking in field of hospitality via restaurant	
Lihan Chen, Rachele Benjamin, Yingchi Guo, Addison Lai, Steven J. Heine	2023	Analysed various data and concluded that there are various biases which arise due to p-hacking and can be understood via P-curve	
Susanne Jana Adler, Lukas Röseler , Martina Katharina Schöniger	2023	Emphasise on increasing demand of researchers to replicate studied to assure data integrity and prevent p-hacking and to maintain trust of publication and suggest use of meta-analyses	
Jiaqi Guo, Peng Li	2023	Talks about boot-strap method developed to evaluate both sampling and empirical design variations.	
Jelena Cerar	2023	Talks about use of p-hacking in I.B (international business) and suggest use of methods like Nulls Hypothesis to prevent P-hacking and talks about increase in use of Harking which is a Type of P-hacking.	
Rahul Sharma, Sergey Redyuk, Sumantrak Mukherjee, Andrea Sipka, Sebastian Vollmer, David Selby	2023	Talk about variant P-hacking Known as "X-hacking" p-hacking is referred to as x-hacking when used in X-ai and used to manipulate Shap values	
D YOKUM, J BOWERS	2023	Discusses use of PAP and talk about how it can be used to counter P-hacking, Harking and other Q.R. P's.	

Abel Brodeur Nikolai Cook, Carina Neisser	2024	discusses data sharing does not decrease p-hacking	
Marcos Lopez de Prado, Vincent Zoonekynd	2024	Emphasises P-hacking is main cause of False positives in factor investigation.	
Crede, Marcus Sotola, Lukas K.	2024	Discusses about how only 37% papers published in psychology were replicable in result, conducted over 244 samples.	
Federico Echenique, Kevin He	2024	Motivated by recent malpractice in behavioural sciences developed method knows as "dissemination noise" as a way to screen -hacked findings that arise from data mining and fabrication	
Abel Brodeur Nikolai M. Cook Jonathan S. Hartley Anthony Heyes	2024	Suggest pre-registration itself does not reduce p-hacking but pre-registration paired with P.A.P can reduce P-hacking	
Maya B Mathur	2024	Suggest use of simple method like meta-analysis it can prevent publication bias and see direction of research.	
Adam McCloskey, Pascal Michaillat	2024	Talks about method they have developed for testing hypothesis and how to cross examine p-values	
Kristof Meding & Thilo Hagendorff	2024	Talks about developing M.L to ensure integrity and develop algorithm to check data for publishing articles	

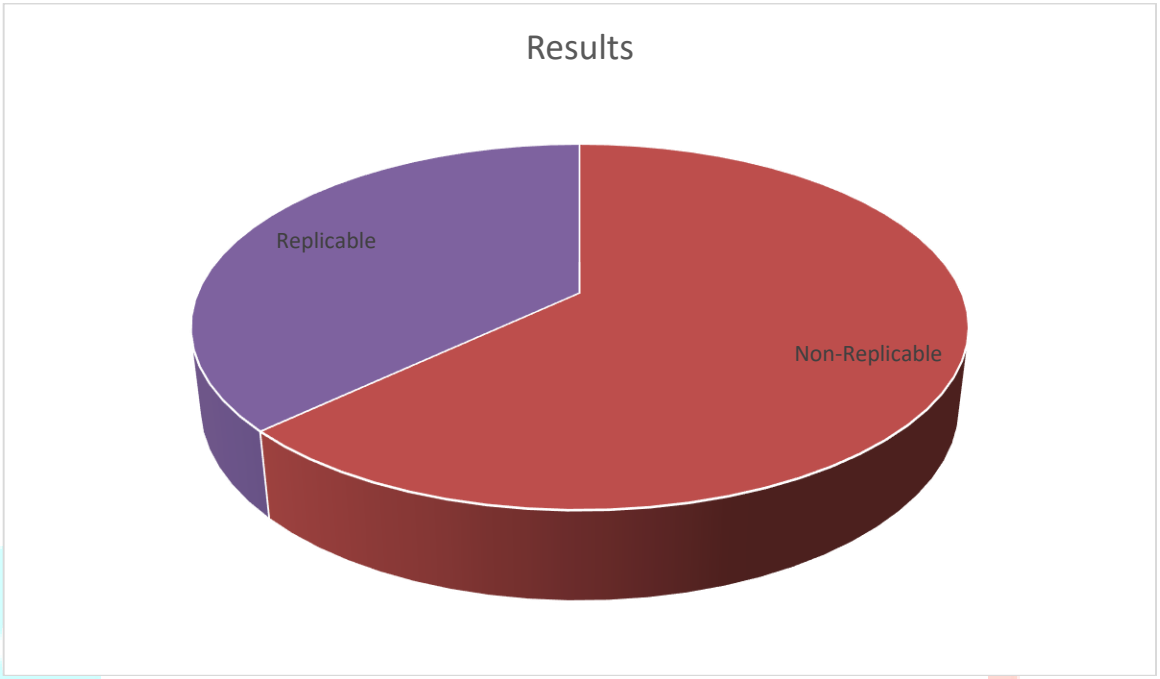
Stephan B. Bruns, Teshome K. Deressa, T. D. Stanley, Chris Doucouliagos, John P. A. Ioannidis	2024	Used sample of p-values from over 192 meta-databases and determined 57.7% p-values were selectively reported	
William M. Cready	2024	EVALUATED IMPLICATION OF RESEARCHER HYPOTHESIS WHERE RESEARCHER USED FABRICATED DATA TO SUPPORT THEIR RESEARCH, USING ERRONEOUS REJECTIONS OF TRUE NULL HYPOTHESES	
Jonny Karunia Fajar	2024	Discusses impact of publication bias on meta-data and suggest to use graphical data to verify integrity of data using methods like T.E. S.	
Moses Mayondi Richard Mulenga	2024	emphasizes on using Bayesian methods to control p-hacking in fields of management and economics and they discovered 97% papers wrote one or two lines about this issue. Also promote publishers to verify meta data and check integrity of research .	

P-HACKING IS DEFINED IN FOLLOWING WAYS SUCH AS QUESTIONABLE RESEARCH PRACTICE, DATA MANIPULATION, P-VALUE MANIPULATION [6] ALSO REFERRED TO QRP (QUESTIONABLE RESEARCH PRACTICES) WHICH CAN VARY FROM DEGREE OF IDEAL TO MALPRACTICE CAN RANGE FROM SIMPLE MANIPULATION TO FULL ON FABRICATION OF DATA [12], CAN BE EXPRESSED AS DATA MANIPULATION OR DATA ANALYSES TO REACH THE CONVENTIONAL THRESHOLDS FOR STATISTICAL SIGNIFICANCE

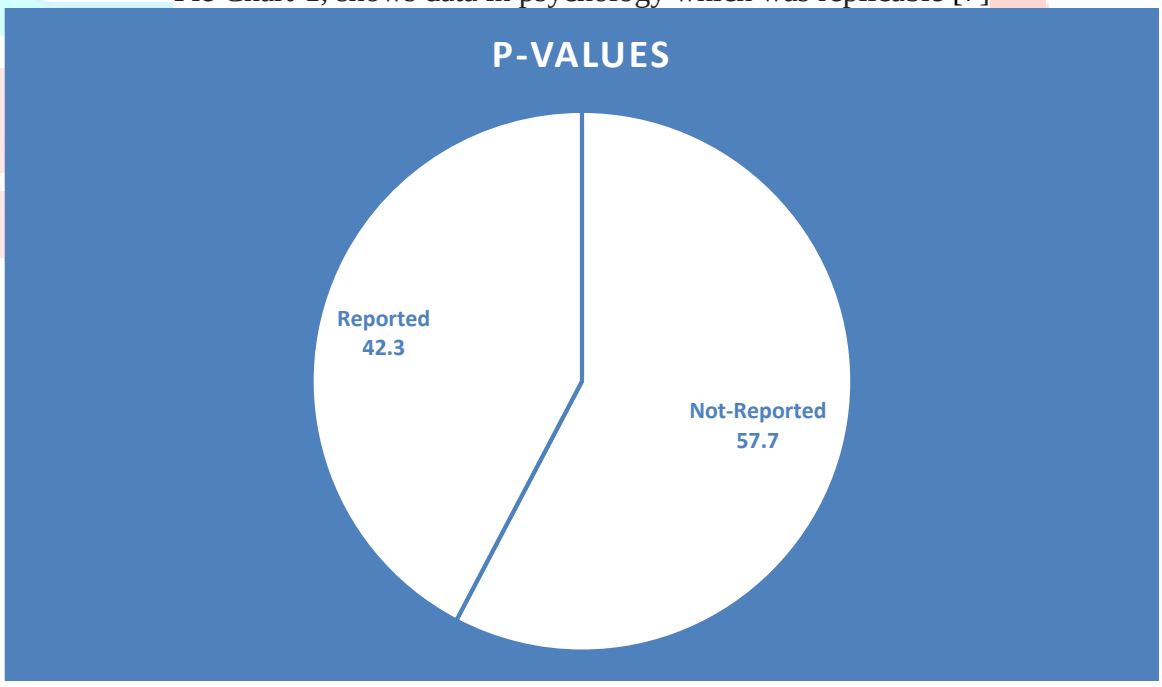
Example ($p < 0.04$). P-hacking is used to render non-significant hypothesis as significant hypothesis also in modern day increasing demand of literature publication has led to increase in demand of p-hacking. practice of p-hacking involves providing data which is significant to research or dismissing data which does not favor the research it can range from dismissing data to manipulating data or using empirical data which may not exist in reality.

It discusses various definitions of p-hacking it can be defined as Q.R.P it can also be defined manipulation of p values or selectively reporting the data, it started in field of sciences but has managed to make its way into field of Management, Hospitality, Psychology and has impacted the data integrity and value of empirical data and how industries have managed to degrade it reinforce their research and data empirical data is data we observe ourselves and evaluate apart from data it can involves acts of harking , Fabrication, Selective reporting , and dismissing the data . P-hacking managed to wreak havoc in publication as current academic publications have observed decrease in trust and lack reliability and data presented cannot be replicated into real life thus making paper useless as data present in it is hypothetical with no real-life implication thus defeating the purpose of academic research also it has been there is new form of P-hacking on the rise known

as X-hacking which involves manipulation of posh values in X-ai (X-ai is explainable ai)[23]. Not only that but we are seeing increase in number of academic papers being published which means p-hacking epidemic is on the rise it may not have ground breaking impact right now but it may present itself in future below are 2 pie chart showing percentage of papers in psychology whose data was not reproducible in real life and pie chart 2 showcases p-values which were reported and not reported. This trend shows that papers being published in these fields have no real-life value as they are hypothetical and can't be produced in real thus making then Null's Hypothesis.



Pie Chart-1, shows data in psychology which was replicable [7]



Pie Chart-2, shows % of p-values which were reported and not-reported [26]

METHODOLOGY:

For this Literature review on topic p-hacking in Management authors used various online databases and repositories to conduct comprehensive research are

- NCBI
- Research Gate
- Oxford Academics
- Google Scholars

In total Authors found 142 articles online out of which 100 were in English and 80 were relevant and 28 Papers were analyzed and reviewed.

Future Discussion:

P-hacking has managed to wreak havoc in various fields including –

- Science
- Management
- Hospitality
- Psychology
- Economics

Apart from that p-hacking may not have an impact on state of publication right now doesn't

a) Mean it won't have in future there has been increase in lack of trust on modern day academic publication apart from that we are Seeing rise of new form of P-hacking known as x-hacking which involves manipulating data of posh values in X-ai. Publications have themselves taken the note and have started to check their publication by trying to recreate data/research or by trying to integrity of data. Various methods and algorithms are being developed to check/catch p-hacking, which are but not limited to –

1. Nulls Hypothesis
2. P.A.P

CONCLUSION:

In conclusion it can be said can say that p-hacking are malpractices or data manipulation where a researcher or data analyst does not represent or publish their whole data but publish a specific part or portion of data which support their research or cause and this leads to mistrust in papers published now as data in them in them cannot be replicated nor be represented in real world as it didn't exist in reality or was hypothetical as of now the impact is not major on publication but it leads to several issues in management as data meant to be managed is hypothetical or not existent it can't be managed in some cases people manipulate empirical data which leads to various issues in management like loss of actual data and disruption in empirical status which disrupts management from top to bottom defeating the whole purpose of academic research as it it revolves around finding real data which has affect in real life and has application in real life P-hacking can caught / reduced by checking data integrity or simply rerunning the experiment or check margin of error or checking whether data provided is within or has surpassed suggested threshold i.e.…(P>0.05). There are various algorithms & methods which are recommended to detect p-hacking: -

- Bayesian methods to control p-hacking in fields of management and economics
- Using T.E.S to verify integrity of Meta Data
- Replication: process of simply replicating the research and cross checking with reported data
- Pre analysis
- Meta Analysis
- Bonferroni adjustment
- Back-testing of predictors

Glossary:

1. Null's Hypothesis: defined as statistical concept that states that there is no relationship between variables and the research
2. Non-Publication Bias: Impacts ability to define Evidence in given area
3. Publication Bias: is bias related to reporting of data also referred to dismission bias where certain part of data is dismissed
4. Empirical Data: Data/Information acquired via Observation
5. Harking: defined as presenting Priori Hypothesis

6. Replication: process of simply replicating the research and cross checking with reported data

Reference:

1. Abel Brodeur, Scott Carrell, David Figlio, Lester Lusher (2023). "UNPACKING P-HACKING AND PUBLICATION BIAS"
2. Abel Brodeur, Nikolai Cook, Carina Neisser, Jonathan Hartley, Anthony Heyes (2024). "Do Pre-registration and Pre-analysis Plans Reduce p-Hacking and Publication Bias?"
3. Adam McCloskey, Pascal Michaillat (2024). "Critical Values Robust to P-hacking"
4. Alisha Gupta (2023). "Tempest in a teacup: An analysis of p-Hacking in organizational research"
5. Angelika M. Stefan, Felix D. Schönbrodt (2023). "Big little lies: a compendium and simulation of p-hacking strategies"
6. Baum, Joel A.C., Bromiley, Philip (2019). "P-hacking in Top-tier Management Journals"
7. Crede, Marcus, Sotola, Lukas K. (2024). "All is well that replicates well: The replicability of reported moderation and interaction effects in leading organizational sciences journals"
8. D Yokum, J Bowers (2023). "Value of Pre-Analysis as it assures data integrity and promotes good project practices"
9. Federico Echenique, Kevin He (2024). "Screening p-Hackers: Dissemination Noise as Bait"
10. Han-fen Hu, Dennis F. Galletta, Gregory D. Moody (2023). "Harking and P-Hacking: A Call for More Transparent Reporting of Studies in the Information Systems Field"
11. Ian D. Gow (2023). "The Elephant in the Room: p-hacking and Accounting Research"
12. Jasper Brinker ink (2021). "When Shooting for the Stars Becomes Aiming for Asterisks: P-Hacking in Family Business Research"
13. Jelena Cerar (2023). "What's the P(roblem)? The State of Hypothesis-Testing Research in International Business"
14. Jiaqi Guo, Peng Li (2023). "Empirical Decisions and Replicating Anomalies"
15. Jonny Karunia Fajar (2024). "Approaches for identifying and managing publication bias in meta-analysis"
16. Kristof Meding, Thilo Hagendorff (2024). "Fairness Hacking: The Malicious Practice of Shrouding Unfairness in Algorithms"
17. Lihan Chen, Rachele Benjamin, Yingchi Guo, Addison Lai, Steven J. Heine (2023). "Managing the terror of publication bias: A comprehensive p-curve analysis of the Terror Management Theory literature"
18. Marcos Lopez de Prado, Vincent Zoonekynd (2024). "Why has Factor Investing Failed? The Role of Specification Errors"
19. Marko Sarstedt, Susanne J. Adler (2023). "An advanced method to streamline p-hacking"
20. Maya B Mathur (2024). "Assessing robustness to worst case publication bias using a simple subset meta-analysis"
21. Moses Mayondi, Richard Mulenga (2024). "The Scientology of Hypothesis Testing in Empirical Research: Emphasizing Economic Significance"
22. Oleg Rytchkov, Xun Zhong (2019). "Information Aggregation and P-hacking"
23. Rahul Sharma, Sergey Redyuk, Sumantrak Mukherjee, Andrea Sipka, Sebastian Vollmer, David Selby (2023). "X Hacking: The Threat of Misguided AutoML"
24. Robert J. Macoun (2022). "P-hacking: A Strategic Analysis"
25. Robin Chark, Michael Man Him (2023). "Is menu design effective? A p-curving analysis"
26. Stephan B. Bruns, Teshome K. Deressa, T. D. Stanley, Chris Doucouliagos, John P. A. Ioannidis (2024). "Estimating the extent of selective reporting: An application to economics"
27. Susanne Jana Adler, Lukas Röseler, Martina Katharina Schöniger (2023). "A toolbox to evaluate the trustworthiness of published findings"

28. William M. Cready (2024). "Incentives, Assessment, and the Reliability of Statistical Significance Examinations of Evidence"