Open Peer Review on Qeios

The Punishment Intensity for Research Misconduct and Its Related Factors: An Exploratory Study on Hospitals in Mainland China

Gengyan Tang[1]

1 Sichuan Academy Of Social Sciences

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

Previous studies (Long et al. 2023) have found that factors such as gender and academic positions do not influence the severity of administrative actions taken by institutions. However, this study provides inconsistent evidence; it focuses on incidents of research misconduct in hospitals across Mainland China, exploring factors related to punishment through a large cross-sectional dataset (N=815). Regression analysis revealed a significant correlation between authorship order and the punishment intensity (p<0.05). Under specific conditions, there was a significant correlation between professional title (senior) and punishment intensity (p=0.001), and an interaction between professional title and types of research misbehavior. Further analysis of simple effects showed that, in cases of fabrication and falsification, and combinations of multiple research misbehavior, researchers with senior titles received significantly lighter punishments compared to those with junior, intermediate, and associate senior titles (p<0.05). The study unveils potential accountability patterns (collective accountability and tiered punishment) that may be adopted by hospitals in Mainland China, as well as the challenges faced in ensuring fairness, emphasizing the importance of independent investigative bodies for incidents of research misconduct and advocating for fairness as a priority in governance of research misconduct.

Gengyan Tang

Independent Researcher, Chongqing, China

Email: tanggengyan@outlook.com, ORCID iD: 0000-0003-3221-3134

Keywords: Research misconduct; research integrity; punishments of research misconduct; academic status; Chinese hospitals.

Background

Punishment is one of the means by which universities, institutions, or governments address research misconduct (RM). For example, in the NIH's "*A Guide to the Handling of Research Misconduct Allegations*," it is stated that once research misconduct is investigated and confirmed, the NIH will take measures such as "removal of the responsible person from the particular project, letter of reprimand, special monitoring of future work, probation, suspension, salary reduction, or initiation of steps leading to possible rank reduction or termination of employment." (National Institutes of Health 2019)

Punishment is seen as a way to improve research integrity (RI) through deterrence (Breen 2016). In Mainland China, frequent and large-scale incidents of RM in the medical field have raised concerns about RI (Yu et al. 2020; Lancet 2015). The government of Mainland China had to announce a broad punishment regulation, including restrictions on work outside the academic field, or prohibitions on obtaining bank loans, running companies, or applying for public service jobs, to punish researchers who commit "serious" RM (Cyranoski 2018). However, as with the problems faced by other countries in the world, investigations into RM incidents in Mainland China are usually led by universities or institutions. The severity of RM incidents and the results of punishments are often determined by individual RI officers or investigation committees. As Yeo-Teh and Tang (2021) point out:

> *"Owing to the subjectivity of this process, the conclusion reached could vary between investigating officers/committees, even when adjudicating based on similar evidence. This variation would likely have an impact on the sanctions delivered." (p.279)*

Numerous factors in the adjudication process may influence the punishment results assigned for RM. Previous study indicates that discrepancies in the definition of RM can result in varying approaches to governance (Li and Cornelis 2020). Particularly, there is a strong inconsistency and controversy in the definition and understanding of research misbehavior (Resnik 2019). Research by Long et al. (2023) on the Office of Research Integrity (ORI) found that the severity of RM or the pattern of research misbehavior is related to the severity of administrative actions, but not to race, gender, academic position, or administrative position. However, due to the pervasive gender bias in the academic system (Robinson-Zañartu et al. 2005; Llorens et al. 2021) as well as factors such as academic status and professional titles (Horbach et al. 2020; Duma et al. 2019), current singular evidence is not sufficient to prove that these factors do not influence the punishment results of RM. Particularly, as different countries have varied governance approaches to RM, this may lead to different factors influencing punishment results (Shykhnenko and Sbruieva 2022). Moreover, due to the hierarchical system of hospitals in Mainland China (Cai et al. 2017), hospitals of different levels often vary in their involvement in research, which could also affect their understanding of RM and corresponding governance means, thus influencing the punishment results. Some scattered studies have found that authorship order reflects the differing degrees of responsibility that authors have for the same paper, which could also lead to varying levels of accountability (Hussinger and Pellens 2019).

However, the current body of evidence concerning factors related to punishments for RM remains limited: it often relies on

qualitative interpretations to postulate about factors associated with RM punishments, with a lack of extensive quantitative analysis. This is particularly evident in studies from Mainland China, where pertinent evidence is critically sparse. Such scarcity not only narrows our comprehension of the variance in RM punitive measures enacted by hospitals, universities, or other research institutions across diverse cultural contexts and administrative frameworks but also constrains our ability to discern the challenges that contemporary governance measures encounter. Furthermore, prior research has not accounted for certain complex scenarios, including the interaction effects among various associated factors. Consequently, to bridge the gaps identified in the literature, this study aims to (1) investigate factors correlated with RM punishments—such as gender, professional title, types of research misbehavior, authorship order, and institutional level; and (2) examine the potential interactive effect among these variables, thus contributing a more comprehensive body of evidence to the field and offering insights from a national perspective.

Methods

This is an observational, retrospective and exploratory study on the punishments of RM in hospitals across Mainland China. The purpose is to explore whether factors such as gender, institutional level, types of research misbehavior, authorship order, and professional title are associated with the punishments for RM.

Data and sample collection

Data on RM incidents are derived from the investigation result documents disclosed by the Ministry of Science and Technology of the People's Republic of China (MOST) under the "Building Research Integrity" ( ) program in 2021, spanning from January 1 to December 31. These documents include: (1) names; (2) titles of the retracted papers involved in RM; (3) authorship of the papers; (4) types of research misbehavior; (5) punishment results.

From January 3 to January 12, 2023, a search was conducted and the texts of all the RM punishment results published by the MOST were collected. As the MOST has not publicly disclosed the gender, authorship order, and professional title information of individuals involved in RM, relevant data were obtained through alternative channels.

Information about researchers engaged in RM was retrieved from institutional official websites. If the information was not available on the official website, online medical consultation apps (for example, Chunyu Doctor App, and Youlai Doctor App) were utilized for retrieval. In Mainland China, these apps provide information such as the doctor's professional title, gender, name, and affiliated institution. Therefore, searches were conducted on these apps using the "affiliated unit + name" query format. For samples where gender could not be determined, the "NameSor Chinese API" (https://chinese-names.app/gender) was utilized for identification. Moreover, for the authorship order, it was determined by searching for the paper title on PubMed. The data on institutional level comes from the "*China Hospital Competitiveness Report (2020-2021)*" (Cao et al. 2020).

As shown in Figure 1, a total of 408 cases of RM, involving 1088 researchers, were obtained. After obtaining these samples, data cleaning was conducted, and the following criteria were applied to exclude certain data: 1. Samples for

which personal information could not be obtained through publicly available ways; 2. Samples from non-hospital institutions (not within the scope of this study); 3. Samples with merged punishments (some researchers were involved in multiple RM cases, and their punishment results were merged, which could lead to outliers, thus such samples were removed). Finally, a total of 815 samples were included in the analysis.

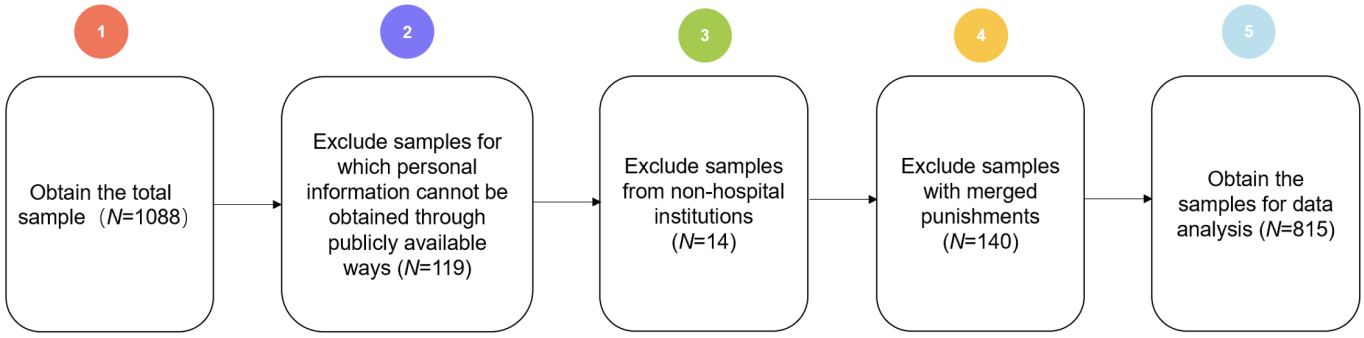


Figure 1. Sample selection process

Variable measurement

For the independent variables, which were non-numeric, collaboration with two coders was established to develop a coding scheme (see Appendix 1) and code for professional titles, gender, institutional levels, types of research misbehavior, and authorship order. The coders, each with approximately three years of research experience, are familiar with research methods such as content analysis. Before coding, coding guidelines were provided, and coding results were discussed to reach a consensus on issues where opinions differed. Cohen's Kappa was used to calculate inter-coder reliability for latent variables, with the overall reliability for all latent variables exceeding 0.9.

For the dependent variable (punishment results), a metric called "punishment intensity (PI)" was proposed to measure it. Based on the data obtained, this metric was designed to include five primary dimensions (administrative sanctions, financial penalties, research penalties, professional penalties, teaching penalties) and 31 secondary dimensions. After setting the dimensions for PI, between December 4 and December 8, 2023, the Delphi method was used to invite 12 experts who assigned scores to each secondary dimension based on the severity of the punishment (impact on the researcher's career), with the least severe being 0 points and the most severe being 10 points, and provided weights for the primary dimensions. After resolving differences of opinion through two online meetings, the final dimensions and their scores and weights were determined by collaboration with the experts, as shown in Appendix 2. Specifically, PI is a weighted index calculated with:

$$PI = \sum_{i=1}^{n} \left(W_i \times S_i \right),$$

where: W_i is the weight for the i-th primary dimension. S_i is the sum of scores of the secondary dimensions under the$i$-th primary dimension. n is the number of primary dimensions. For example, consider a case that receives organizational

criticism and admonitory talk, has its achievement bonus deducted for 10-12 months, academic awards and honors revoked for 3 years, professional promotion canceled for 5 years, and graduate supervision qualifications canceled for 3 years[1]. It is worth noting that a score of 0 for a primary dimension means that the case did not receive a punishment in that dimension. Furthermore, in some cases of RM, the MOST provides reasons for not punishing certain individuals, such as when individuals are listed as authors without being aware of it[2]. However, in other cases, some individuals are exempt from punishment without any stated justification. In these instances, these cases have been coded as 0, signifying that no punishment was received.

Data management and analysis

Data were entered and coded using Excel (Version 2403 Build 16.0.17425.20176, 2021, Microsoft Corporation, Redmond, Washington, USA, https://office.microsoft.com/excel), and statistical analysis and visualization were conducted using RStudio (Version 2023.09.1+494, Integrated Development Environment for R. RStudio, PBC, Boston, MA, https://posit.co/). In the descriptive statistics phase, all independent variables were categorical and thus described through frequency distributions and proportions. The dependent variable was continuous and described using maximum, minimum, mean, and standard deviation values. Additionally, violin plots were used to analyze the distribution of the independent variables and their within-group data.

Before initiating the regression analysis, all independent variables were converted to factor type. The Ordinary Least Squares (OLS) method was then employed during the regression analysis phase to evaluate the impact of these variables on the dependent variable. Subsequently, the Shapiro-Wilk test was utilized to assess the normality of residuals within the linear regression model. The null hypothesis of this test posits that the data follows a normal distribution. If the test statistic *W*'s significance level is less than 0.05, the null hypothesis is rejected, indicating a deviation from normal distribution (Hanusz and TarasiŃska 2014). With the Shapiro-Wilk test yielding a significance level less than 0.01, it was concluded that the residuals of the model are not normally distributed; thus, using OLS might lead to incorrect inferences and conclusions.

Subsequently, considering the non-normality of residuals and the data type, a generalized linear model (GLM) with a Tweedie distribution was selected, which does not require the residuals to be normally distributed (Pekár and Brabec 2017). The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to judge the model's goodness of fit[3], and the Variance Inflation Factor (VIF) was used to assess the model's multicollinearity issues[4]. To further investigate potential interaction effects between variables, simple slope plots were employed for a preliminary visual assessment. If the simple slope plots suggested the presence of possible interaction effects, the Wald test was applied to conduct pairwise comparisons of the marginal means of the variables. This approach aids in determining if the relationship between the variable combinations and PI differs at various levels. Additionally, a $p\text{-value} < 0.05$ was considered statistically significant.

Result

Sample characteristics

Figure 2 indicates that 64.7% of the individuals involved in RM incidents are males, a proportion higher than the 35.3% involving females, thus supporting the findings of previous studies (Fang, Bennett, Casadevall, 2013; Gopalakrishna et al., 2022). Associate senior researchers show the highest rate of RM involvement at 42.1%, followed by those holding intermediate titles at 36.3%. At the institutional level, researchers from municipal hospitals have the highest rate of involvement, accounting for 51% of cases. Regarding authorship order, the first or corresponding authors hold the greatest responsibility, comprising 73.6% of cases, with second authors at 7.9%. The incidence of accountability decreases with lower authorship positions. In terms of misbehavior types, fabrication or falsification is most common, representing 44.3%, followed by ghostwriting at 32.1%, with no reported cases of ethical violations or plagiarism. The maximum PI value is 7.70, the minimum is 0, and the average is 2.08.

Figure 2. Characteristics of included samples (N=815)

Results of Tweedie GLMs

When employing a Tweedie GLM for regression analysis, it is necessary to discuss the setting of the variance power value (where 1 < variance power < 2). Testing values within this range (results detailed in Supplementary Material Table S1) revealed that a variance power of 1.4 yielded the lowest AIC and BIC (AIC = 2586.67, BIC = 2683.44), suggesting optimal model fit; therefore, I set the variance power to 1.4. Furthermore, I selected a link power of 0, as it is the most commonly utilized link function for Tweedie GLMs (Dunn and Smyth 2018). VIF test results indicate that all independent variables have VIF values less than 10 (see Supplementary Material, Table S2), confirming the absence of multicollinearity.

Univariate regression results (see Appendix 3) reveal that within institutional levels, provincial hospitals mete out significantly lower PI for RM compared to top-tier hospitals (coefficient (coef.) = -0.19, $p = 0.041$, 95% confidence interval (CI) [-0.36, -0.01]). The rest did not exhibit statistical significance ($p \geq 0.05$). Regarding authorship order, positions from the second (coef. = -0.14, $p < 0.001$, 95% CI [-1.33, -0.72]) to the eighth author (coef. = -1.69,$p = 0.003$, 95% CI [-2.82, -0.63]) have significant negative coefficients compared to either the first author or the corresponding author, indicating they were subject to significantly lower PI. For variables such as gender, professional title, and types of research misbehavior, there was no significant correlation with PI ($p \geq 0.05$).

The multivariate regression results, depicted in Figure 3, show that the coefficients for authors ranked second (coef. = -1.07, $p < 0.001$, 95% CI [-1.37, -0.76]) through eighth (coef. = -1.67,$p = 0.003$, 95% CI [-2.81, -0.61]) remain significantly negative when compared to the first or corresponding author. This confirms a significant association between authorship order and PI. Notably, in the variable of professional titles, the senior group, which was not significant in the univariate regression, became significant in the multivariate regression (coef. = -0.46, $p = 0.001$, 95% CI [-0.73, -0.18]). This indicates that compared to the junior group, the senior group receives significantly lower PI. The lack of significant findings ($p \geq 0.05$) for other professional title groups suggests that the relationship between professional title and PI may only hold under certain conditions. For institutional levels, provincial hospitals (coef. = -0.004, $p = 0.959$, 95% CI [-0.17, 0.16]), significant in univariate analysis, lost significance in the multivariate model, perhaps due to explanatory power being shared with other variables. There were no significant shifts in the statistical significance of the remaining variables ($p \geq 0.05$). The diagnostic plots for the multivariate regression analysis are presented in Supplementary Material, Figure S1.

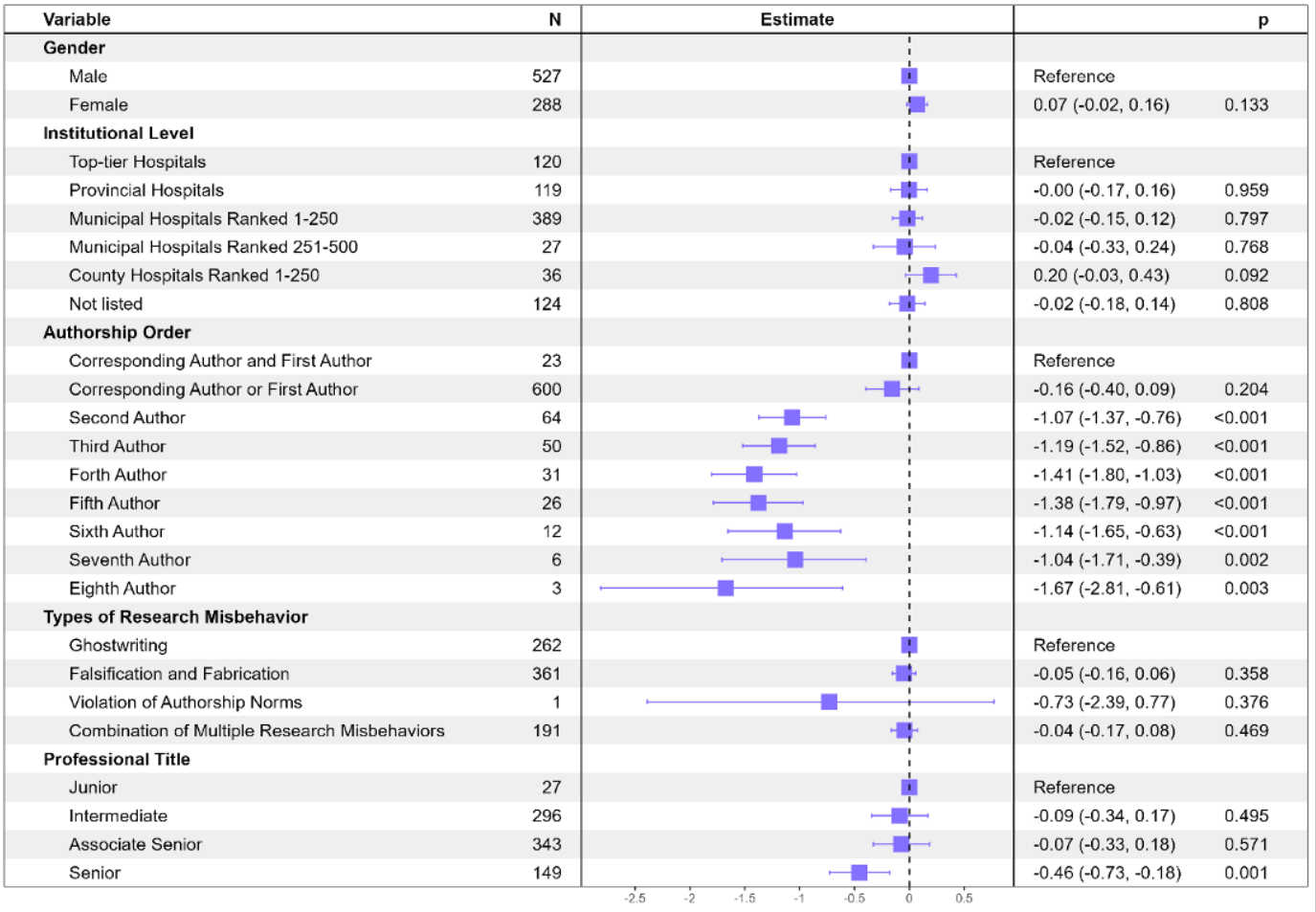


Figure 3. The results ofmultivariate regression (*N*=815)

Results of the interaction effect analysis

Due to changes in the statistical significance of certain groups within the professional title in the multivariate regression results, as compared to the univariate regression, it is posited that there may be variables capable of moderating the relationship between this variable and PI. Consequently, variables that might interact with the professional title variable were explored, revealing that only the types of research misbehavior may potentially exhibit an interaction effect with this variable.

Specifically, upon incorporating an interaction term (refer to Figure 4), the groups implicated in falsification and fabrication (coef. = 0.66, $p = 0.024$, 95% CI [0.09, 1.24]) and those engaged in multiple research misbehaviors (coef. = 0.94$p = 0.009$, 95% CI [0.23, 1.65]) exhibited a shift from non-significance to significance. Within the professional titles, the senior group's previously significant effect diminished (coef. = 0.24, $p = 0.358$, 95% CI [-0.27, 0.76]), whereas the intermediate group's effect became significant (coef. = 0.50, $p = 0.044$, 95% CI [0.02, 1.00]). Although the associate senior group's*p-value* (coef. = 0.47, $p = 0.062$, 95% CI [-0.02, 0.96]) decreased, it did not reach the threshold of significance. These findings imply that the relationship between the senior professional title and PI varies with the types of research misbehavior; such variability is pronounced to the extent that the inclusion of misbehavior types in the model obscures the distinction between the senior and the junior (reference) groups. In contrast, the intermediate group may demonstrate a

consistent relationship with PI, which becomes more apparent upon adjusting for specific misbehavior types. These findings may imply that the relationship between the senior professional title and PI varies with the types of research misbehavior; such variability is pronounced to the extent that including the interaction term in the model obscures the distinction between the senior and the junior (reference) groups. In contrast, the intermediate group may demonstrate a consistent relationship with PI, which becomes more apparent upon adjusting for specific misbehavior types. Significantly, all interactions in the model were statistically significant ($p < 0.05$), underscoring a genuine interactive effect between types of research misbehavior and professional titles, suggesting that the relationship between professional titles and PI is moderated by the specific misbehavior types.

To further ascertain the interaction effect, a simple slopes graph was constructed (see Appendix 4). The graph illustrates that, in cases of falsification and fabrication, the trend in PI decreases from the junior to the senior group. Conversely, under ghostwriting activities, the trend in PI increases, suggesting an interactive association between types of research misbehavior and professional titles. Additionally, the results of pairwise comparisons (simple effects) revealed complex interactions between the two variables (see Figure 5). When controlling for falsification and fabrication, significant differences in PI were observed between the associate senior and senior (coef. = 0.41, $p < 0.001$, 95% CI [0.22, 0.60]), intermediate and senior (coef. = 0.33, $p = 0.008$, 95% CI [0.13, 0.53]), and junior and senior (coef. = 0.61,$p = 0.009$, 95% CI [0.23, 0.98]). Under the condition of ghostwriting, no significant differences were noted between the four professional title groups ($p \geq 0.05$). However, significant differences in PI emerged under the scenario involving a combination of multiple research misbehaviors, specifically between associate senior and senior (coef. = 0.51, $p < 0.001$, 95% CI [0.24, 0.77]), intermediate and senior (coef. = 0.55, $p < 0.001$, 95% CI [0.28, 0.82]), and junior and senior (coef. = 1.01,$p = 0.003$, 95% CI [0.44, 1.59]). This may indicate that the association between professional titles and PI could be strengthened or weakened depending on the specific types of research misbehavior, such as falsification and fabrication, as well as combinations of multiple misbehaviors.

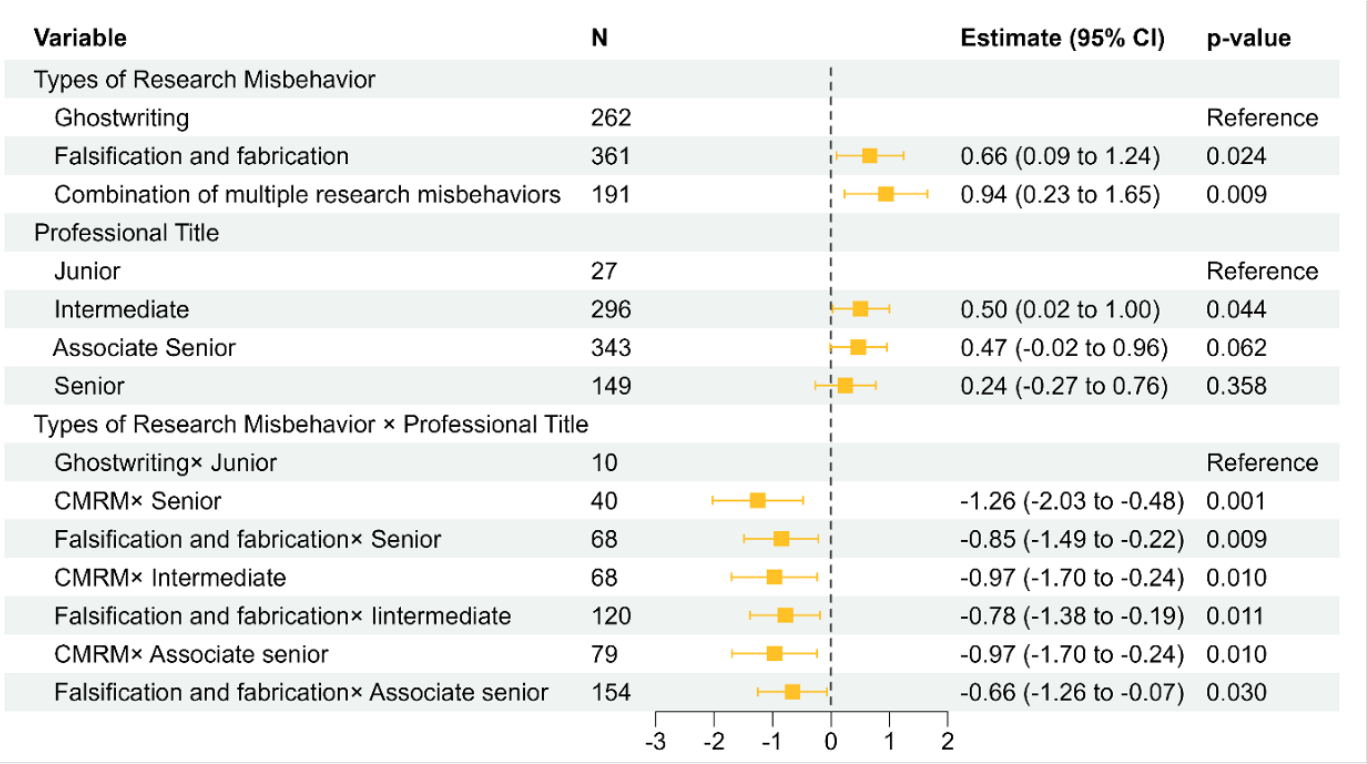


Figure 4. The results of the regression analysis with interaction terms (*N*=815)

Note: CMRM: Combinations of multiple misbehaviors. To enhance clarity, the figure does not report other variables such as authorship order, gender, and institutional level. Additionally, the “violation of authorship norms” category within types of research misbehavior was omitted from the analysis due to an insufficient number of samples. This exclusion also occurred in some combinations of interaction terms.

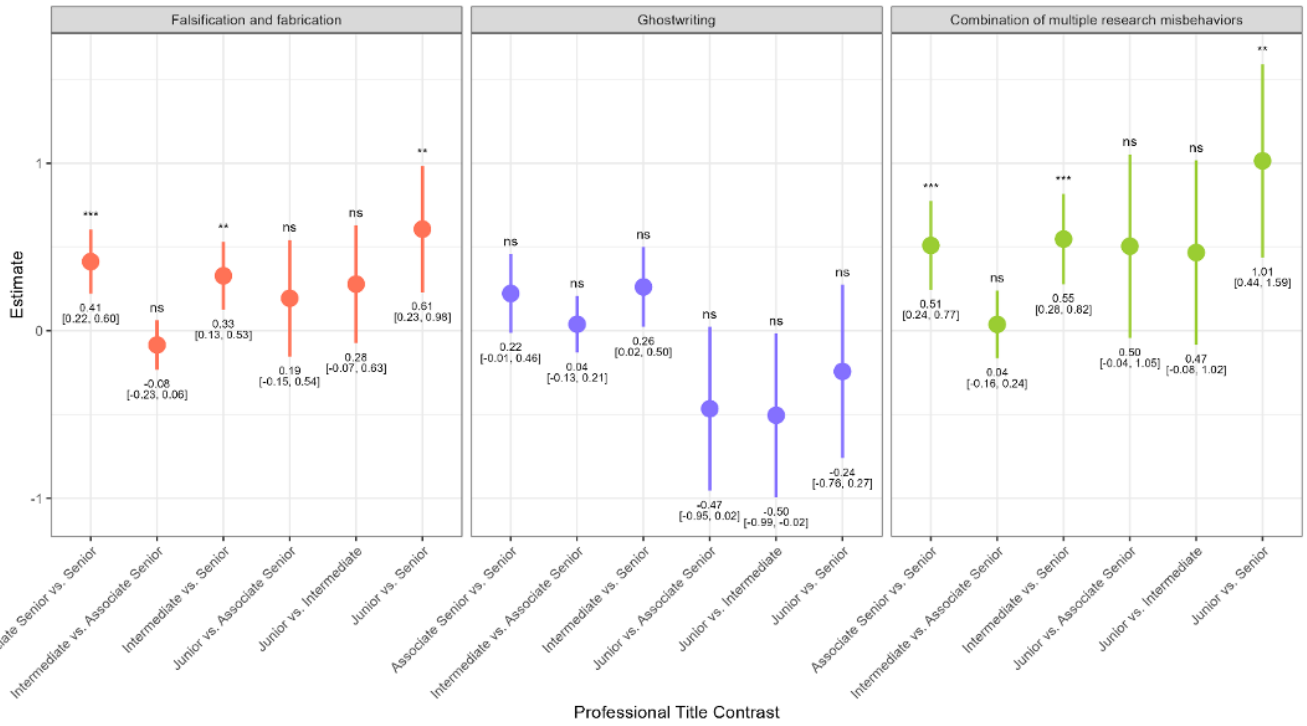


Figure 5. The results of the analysis of simple effects (*N*=815)

Note: ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$; ns: $p \geq 0.05$.

Discussion

In previous studies on RM, scholars have mostly focused on how researchers perceive RM (Li and Cornelis 2020b; Paruzel-Czachura, Baran, and Spendel 2020) and how to prevent it (Pratt et al. 2019). Over the past decade, with the increase in RM incidents—evidenced by the growing number of retractions—there has been an intensifying punitive atmosphere within the academic community. There have been increasing calls for more sanctions and even discussions about classifying them as criminal offenses (Dal-Ré et al. 2020). However, sanctions for RM often take place in a "black box," and our understanding of the punishment process remains very limited. As Hesselmann and Reinhart (2020) noted:

> *"We conclude that sanctions for scientific misconduct operate within a cycle of invisibility: The challenges in making visible the processes that precede the sanctions, such as their rare and incidental character, as well as confidentiality concerns, result in a low visibility for the investigation procedures." (p.416)*

This exploratory study offers insights into the factors related to punishment through a large cross-sectional data set, further considering the challenges institutions face in dealing with incidents of RM. Contrary to previous literature, the study found a significant negative relationship between authorship order and PI. This suggests that hospitals in Mainland China may adopt a collective punishment model (where paper authors share responsibility) and also implement a tiered accountability, with the first and corresponding authors receiving the most severe punishment, while other authors

significantly less so. The study also revealed nuanced relationships between professional title and PI under certain conditions. While no significant differences were found in the PI received by intermediate and associate senior researchers compared to those with junior titles involved in RM, senior title holders received significantly lighter penalties. A possible explanation is that unlike the rule-based governance systems in the West, Mainland China operates on a relationship-based governance system (Li, Park, and Bao 2019). Within such systems, governance within organizations tends to be hierarchical and paternalistic (Lau and Young 2013). Individuals with more social capital may leverage relationships and networks to mitigate PI (Mooijman and Graham 2018), though more evidence is needed.

Interestingly, the study also discovered an interaction effect between types of research misbehavior and professional title, especially under conditions of fabrication and falsification and combination of misconduct behaviors, where the interaction held statistical significance. Further analysis of simple effects revealed that when involved in fabrication and falsification, researchers with senior titles received less severe punishments compared to other groups. Similarly, when involved in combination of misconduct behaviors, senior title holders also received lighter punishments. This differentiation more specifically reveals how the relationship between professional title and PI might be strengthened or weakened under certain research misbehaviors, suggesting that institutions might need to consider the appropriateness of punitive measures under specific conditions. This is an issue not sufficiently emphasized in the existing literature (Were, Kaguiri, and Kiplagat 2020), which seems to overlook that RM is a “wicked problem,” often focusing on how external factors (such as incentive systems or institutional culture) promote the occurrence of RM (Bouter 2015; Olesen, Amin, and Mahadi 2017). As Eaton (2023) points out:

> *“When we talk about the study and practice of scholarly integrity, we are talking about wicked problems. The term ‘wicked problem’ refers to that class of social system problems which are ill-formulated, where the information is confusing, where there are many clients and decision makers with conflicting values, and where the ramifications in the whole system are thoroughly confusing.” (p.4)*

Therefore, if we view the issue of RM as a continuum, then in addition to the factors that cause it and the measures or systems to prevent it, the governance process, particularly the process of implementing punishment, also deserves attention. This study goes beyond the purview of existing literature, revealing the challenges inherent in the governance process of RM itself. From a practical standpoint, it suggests that institutions might need to consider the involvement of independent investigative bodies (Resnik, Hosseini, and Rasmussen 2024) to avoid interference from certain factors during the investigation and governance of RM. However, Mainland China lacks independent bodies to investigate incidents of RM. Investigations of such incidents almost entirely rely on administrative staff employed by universities, hospitals, or research institutions themselves (Yi, Nemery, and Dierickx 2019), who have long worked within these entities and cannot ensure the absence of conflicts of interest or connections with the investigated parties. Some studies have even identified the politicization of RM incidents, using governance of RM as a sufficient and effective reason to oust competitors (Tsai 2018). Therefore, this research underscores the importance of independence in adjudicating incidents of RM, as this may weaken the association between professional titles and the PI.

Furthermore, this study's exploration touches upon a pressing unresolved question: how to ensure fairness in the investigation of RM incidents. Previous literature has analyzed the impact of retractions on academic careers, finding that prominent scholars suffer more severe consequences after retractions involving fraud or misconduct than their less distinguished counterparts (Azoulay, Bonatti, and Krieger 2017). Conversely, other studies have found that authors with a history of retraction, particularly those with less research experience, are more likely to abandon their careers after a retraction (Memon, Makovi, and AlShebli 2023). These studies powerfully unveil the structural unfairness lurking behind the punishments for RM incidents. This research supports these findings. If previous literature focused on the unfairness in career progression for researchers of different academic positions post-punishment (with retractions seen as a form of punishment), then this study to some extent reveals the potential unfairness within the punishment process itself: researchers with different professional titles face significantly varying intensities of punishment for the same types of research misbehavior. Although stronger evidence is required to further corroborate this point, ensuring fairness in punishment should be considered a priority. This is not only a matter concerning individuals' career trajectories but also pertains to the values we uphold regarding RI.

The limitations of this study include: (1) It is a cross-sectional retrospective study, and although the sample size is large, it cannot infer causative factors that might affect the PI from a causal perspective. Instead, it can only emphasize correlational relationships; (2) The study's sample exclusively comes from hospitals in Mainland China, which may not allow for generalization to universities or research institutions. There may be institutional differences in the governance of incidents of RM, thus limiting the universality of the conclusions; (3) Given that the independent variables are categorical, the sample size of some groups varies significantly, which may have amplified the range of error and limited the statistical detection power for these groups.

Conclusion

In summary, this study provides the first evidence from a large cross-sectional dataset to uncover factors associated with the handling of RM incidents in hospitals across Mainland China. The findings illustrate the collective accountability pattern adopted by Chinese hospitals and the tiered punishment intensity based on authorship order that is applied during adjudication. Furthermore, the study identifies an interaction between professional title and types of research misbehavior, discovering significant differences in the punishments received by senior title holders for specific types of research misbehavior compared to other researchers. The study emphasizes the importance of third-party institutions' involvement in the investigation of scientific misconduct cases, as well as the need to prioritize fairness in the execution of punishments.

Appendix

Appendix 1. Variable names, explanations, and encoding rules

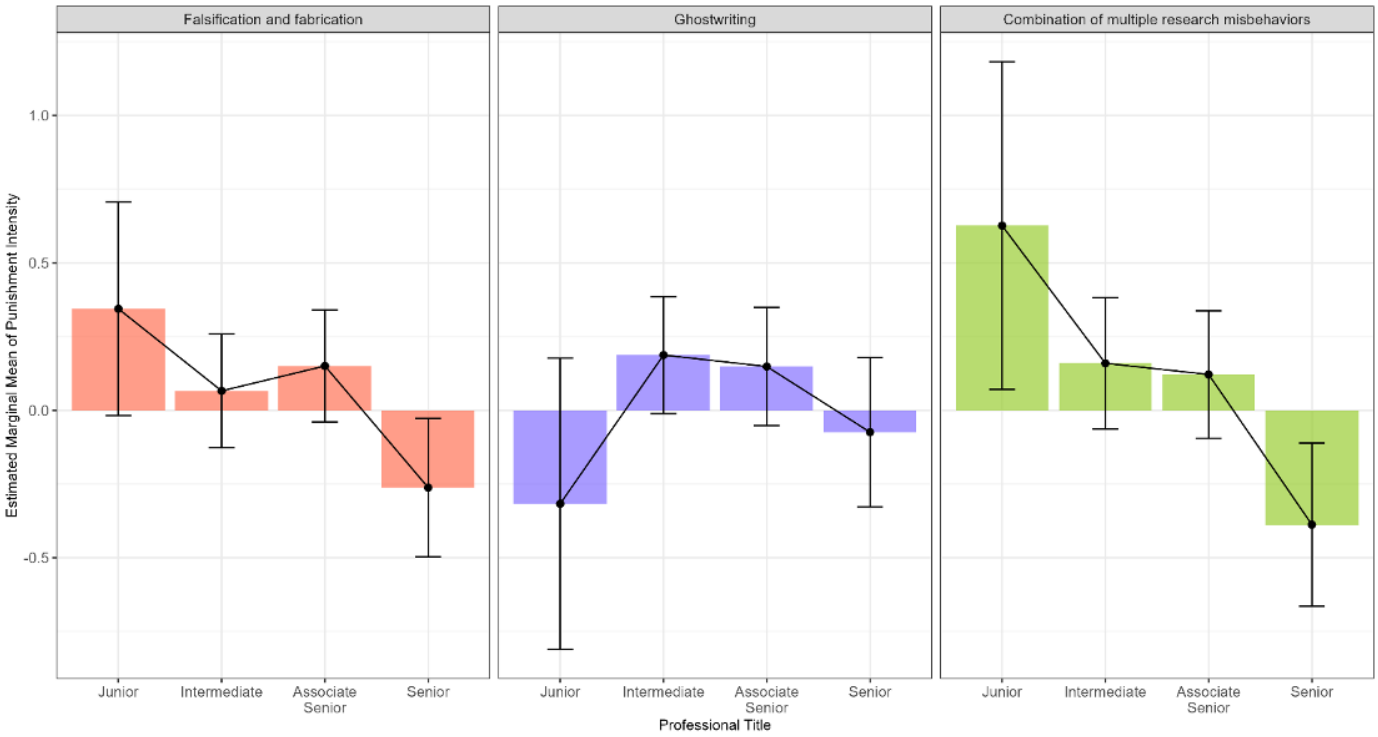
Variables	Meanings	Encoding
Gender	Gender of the sample	Male=0, Female=1
Professional title	The level of professional title for the sample	Junior=1, Intermediate=2, Associate Senior=3, Senior=4
Institutional level	Ranking of the institution affiliated with the sample	Top-tier hospitals = 1, Provincial hospitals = 2, Municipal hospitals ranked 1-250 = 3; Municipal hospitals ranked 251-500 = 4; County hospitals ranked 1-250 = 5; County hospitals ranked 251-500 = 6; Not listed = 7.
Authorship order	Position of the sample in the paper's authorship	Corresponding author and first author=0, First author=1, Corresponding author=1, Second author=2, Third author=3, and so on
Types of research misbehavior	The types of RM engaged in by the sample[5]	Falsification and fabrication=1, Plagiarism=2, Violation of authorship norms=3, Ghostwriting=4, Violation of ethical standards=5, Combination of multiple research misbehaviors = 6

Appendix 2. Scores and proportions for each dimension of PI

Primary Dimension	Weight	Secondary Dimension	Score
Administrative sanctions	15%	Admonitory talk	1
		Criticism	2
		Warning (Party members receive internal party warning, non-party members receive administrative warning)	3
		Record a demerit [6]	4
		Record in the RI archive or RM database[7]	5
Financial penalties	15%	Deduct achievement bonus for 1-3 months	1
		Deduct achievement bonus for 4-6 months	2
		Deduct achievement bonus for 7-9 months	3
		Deduct achievement bonus for 10-12 months	4
		Deduct achievement bonus for more than 1 year	6
Research penalties	25%	Revoke academic awards, scholarly honors, and project application qualifications for a period of 1 year	1
		Revoke academic awards, scholarly honors, and project application qualifications for a period of 2 years	2
		Revoke academic awards, scholarly honors, and project application qualifications for a period of 3 years	3
		Revoke academic awards, scholarly honors, and project application qualifications for a period of 4 years	4
		Revoke academic awards, scholarly honors, and project application qualifications for a period of 5 years	5
		Revoke academic awards, scholarly honors, and project application qualifications for a period of 6-10 years	7
		Permanently revoke academic awards, scholarly honors, and project application qualifications	10
Professional penalties	25%	Cancel professional promotion for 1 year	1
		Cancel professional promotion for 2 years	2
		Cancel professional promotion for 3 years	3
		Cancel professional promotion for 4 years	4
		Cancel professional promotion for 5 years	5
		Cancel professional promotion for 6-10 years	7
		Revoke professional title	10
Teaching penalties	20%	Cancel supervision of graduate qualifications for 1 year	1
		Cancel supervision of graduate qualifications for 2 years	2
		Cancel supervision of graduate qualifications for 3 years	3
		Cancel supervision of graduate qualifications for 4 years	4
		Cancel supervision of graduate qualifications for 5 years	5
		Cancel supervision of graduate qualifications for 6-10 years	7
		Revoke supervision of graduate qualifications	10

Variable	N	Estimate	p
Gender			
Male	527	Reference	
Female	288	0.03 (-0.07, 0.13)	0.559
Institutional Level			
Top-tier Hospitals	120	Reference	
Provincial Hospitals	119	-0.19 (-0.36, -0.01)	0.041
Municipal Hospitals Ranked 1-250	389	-0.07 (-0.21, 0.07)	0.310
Municipal Hospitals Ranked 251-500	27	-0.19 (-0.49, 0.11)	0.215
County Hospitals Ranked 1-250	36	0.09 (-0.16, 0.34)	0.474
Not listed	124	-0.13 (-0.30, 0.04)	0.140
Authorship Order			
Corresponding Author and First Author	23	Reference	
Corresponding Author or First Author	600	-0.14 (-0.37, 0.11)	0.269
Second Author	64	-1.03 (-1.33, -0.72)	<0.001
Third Author	50	-1.14 (-1.47, -0.82)	<0.001
Forth Author	31	-1.37 (-1.75, -0.98)	<0.001
Fifth Author	26	-1.35 (-1.76, -0.94)	<0.001
Sixth Author	12	-1.11 (-1.63, -0.60)	<0.001
Seventh Author	6	-0.91 (-1.57, -0.27)	0.006
Eighth Author	3	-1.69 (-2.82, -0.63)	0.003
Types of Research Misbehavior			
Ghostwriting	262	Reference	
Falsification and Fabrication	361	-0.02 (-0.13, 0.09)	0.774
Violation of Authorship Norms	1	-0.91 (-2.75, 0.73)	0.317
Combination of Multiple Research Misbehaviors	191	-0.04 (-0.17, 0.09)	0.560
Professional Title			
Junior	27	Reference	
Intermediate	296	0.08 (-0.20, 0.36)	0.590
Associate Senior	343	0.18 (-0.09, 0.46)	0.193
Senior	149	-0.23 (-0.52, 0.07)	0.135

Appendix 3. The results of the univariate retrospective analysis



Appendix 4. Simple slopes graphs for professional titles and types of research misbehavior

Statements and Declarations

Acknowledgments

I would like to express my gratitude for the constructive suggestions provided by the anonymous reviewers. Their invaluable advice has greatly improved the quality of this study.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

The author of this study did not receive any funding neither from industry nor from research institutions.

Data availability statement

The data that support the findings of this study are available at:

The programming code involved in the data analysis can be found in the supplementary materials.

Ethical considerations

The raw data, which involves human subjects and is publicly available, has been handled in accordance with the principles of minimal harm and privacy protection in research. Specifically, to avoid potential additional harm, names and institutional details of the cases have been replaced with numerical identifiers in the raw data. Furthermore, the order of sample collection has been randomized to ensure that personal information cannot be identified.

Footnotes

[1] The PI for this case would be calculated as follows: (1+2) × 0.15 + 4 × 0.15 + 3 × 0.25 + 5 × 0.25 + 3 × 0.20 = 3.65.

[2] In this case, since these instances do not involve RM, they were not initially collected as samples.

[3] Generally, the smaller the values of AIC and BIC, the better the fit of the model. (Burnham and Anderson 2002, 302–5)

[4] A VIF value greater than 10 is considered to indicate severe multicollinearity. (Marquaridt 1970)

[5] I classify and encode samples based on the categories specified in the "*Procedures for Investigating and Dealing with Research Misconduct*" released by the Natural Science Foundation of China regarding types of research misbehavior. This classification does not merge FPP into a single category; instead, it treats plagiarism as one category and considers

fabrication and falsification as another. Additionally, it includes research misbehaviors such as ghostwriting, violating authorship norms, and violating ethical standards.

[6] A “demerit recording” is one of the administrative sanctions in Mainland China. It is more severe than a warning because it not only serves as a warning but also implies that the research personnel at fault must undergo certain punitive measures.

[7] In Mainland China, the personal archives of citizens document information such as their educational background, disciplinary actions in education, and honors received. When citizens seek employment, employers access this file to obtain relevant information. The integrity file for research is similar to this; it accompanies researchers throughout their lives, leaving a permanent “stamp” on their career.

References

- Azoulay, Pierre, Alessandro Bonatti, and Joshua L. Krieger. 2017. “The Career Effects of Scandal: Evidence from Scientific Retractions.” *Research Policy* 46 (9): 1552–69. https://doi.org/10.1016/j.respol.2017.07.003
- Bouter, Lex M. 2015. “Commentary: Perverse Incentives or Rotten Apples?” *Accountability in Research* 22 (3): 148–61. https://doi.org/10.1080/08989621.2014.950253
- Breen, K. J. 2016. “Research Misconduct: Time for a Re-Think?” *Internal Medicine Journal* 46 (6): 728–33. https://doi.org/10.1111/imj.13075
- Burnham, Kenneth P., and David R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd ed. New York: Springer.
- Cai, Miao, Echu Liu, Hongbing Tao, Zhengmin Qian, Xiaojun Lin, and Zhaohui Cheng. 2017. “Does Level of Hospital Matter? A Study of Mortality of Acute Myocardial Infarction Patients in Shanxi, China.” *American Journal of Medical Quality* 33 (2): 185–92. https://doi.org/10.1177/1062860617708608
- Cao, Ronggui, Nanshan Zhong, Yiqiang Zhuang, Xinbo Liao, Xinglin Wang, and Quangguang Xu. 2020. “China Hospital Competitiveness Rankings 2017-2019.” In *Annual Report on China’s Hospital Competitiveness*, 1–172. Beijing: Social Sciences Academic Press (China).
- Cyranoski, David. 2018. “China Introduces ‘Social’ Punishments for Scientific Misconduct.” *Nature* 564 (7736): 312–12. https://doi.org/10.1038/d41586-018-07740-z
- Dal-Ré, Rafael, Lex M Bouter, Pim Cuijpers, Christian Gluud, and Søren Holm. 2020. “Should Research Misconduct Be Criminalized?” *Research Ethics* 16 (1-2): 174701611989840. https://doi.org/10.1177/1747016119898400
- Duma, Narjust, Urshila Durani, Cynthia B. Woods, Lionel A. Kankeu Fonkoua, Joselle M. Cook, Christopher Wee, Harry E. Fuentes, et al. 2019. “Evaluating Unconscious Bias: Speaker Introductions at an International Oncology Conference.” *Journal of Clinical Oncology* 37 (36): 3538–45. https://doi.org/10.1200/jco.19.01608
- Eaton, Sarah Elaine. 2023. “Academic and Research Integrity as Transdisciplinary Fields of Scholarship and Professional Practice.” In *Handbook of Academic Integrity*, edited by Sarah Elaine Eaton, 1–18. Singapore: Springer. https://doi.org/10.1007/978-981-287-079-7_165-1

- Fang, Ferric C., Joan W. Bennett, and Arturo Casadevall. 2013. “Males Are Overrepresented among Life Science Researchers Committing Scientific Misconduct.” Edited by Françoise Dromer. *MBio* 4 (1). https://doi.org/10.1128/mbio.00640-12
- Gopalakrishna, Gowri, Gerben ter Riet, Gerko Vink, Ineke Stoop, Jelte M. Wicherts, and Lex M. Bouter. 2022. “Prevalence of Questionable Research Practices, Research Misconduct and Their Potential Explanatory Factors: A Survey among Academic Researchers in the Netherlands.” Edited by Sergi Fàbregues. *PLOS ONE* 17 (2): e0263023. https://doi.org/10.1371/journal.pone.0263023
- Hanusz, Zofia, and Joanna TarasiŃska. 2014. “Simulation Study on Improved Shapiro–Wilk Tests for Normality.” *Communications in Statistics - Simulation and Computation* 43 (9): 2093–2105. https://doi.org/10.1080/03610918.2013.844835
- Hesselmann, Felicitas, and Martin Reinhart. 2020. “Cycles of Invisibility: The Limits of Transparency in Dealing with Scientific Misconduct.” *Social Studies of Science* 51 (3): 414–38. https://doi.org/10.1177/0306312720975201
- Horbach, Serge P. J. M., Eric Breit, Willem Halffman, and Svenn-Erik Mamelund. 2020. “On the Willingness to Report and the Consequences of Reporting Research Misconduct: The Role of Power Relations.” *Science and Engineering Ethics* 26 (3). https://doi.org/10.1007/s11948-020-00202-8
- Hussinger, Katrin, and Maikel Pellens. 2019. “Scientific Misconduct and Accountability in Teams.” Edited by Giuseppe Sartori. *PLOS ONE* 14 (5): e0215962. https://doi.org/10.1371/journal.pone.0215962
- Lancet, The. 2015. “China’s Medical Research Integrity Questioned.” *The Lancet* 385 (9976): 1365. https://doi.org/10.1016/S0140-6736(15)60700-0
- Lau, Kun Luen Alex, and Angus Young. 2013. “Why China Shall Not Completely Transit from a Relation Based to a Rule Based Governance Regime: A Chinese Perspective.” *Corporate Governance: An International Review* 21 (6): 577–85. https://doi.org/10.1111/corg.12037
- Li, Dan, and Gustaaf Cornelis. 2020a. “Defining and Handling Research Misconduct: A Comparison between Chinese and European Institutional Policies.” *Journal of Empirical Research on Human Research Ethics* 15 (4): 302–19. https://doi.org/10.1177/1556264620927628
- ———. 2020b. “Differing Perceptions Concerning Research Misconduct between China and Flanders: A Qualitative Study.” *Accountability in Research* 28 (2): 63–94. https://doi.org/10.1080/08989621.2020.1802586
- Li, Shaomin, Seung Ho Park, and Rosey Shuji Bao. 2019. “The Transition from Relation-Based to Rule-Based Governance in East Asia.” *International Journal of Emerging Markets* 14 (1): 171–86. https://doi.org/10.1108/ijoem-10-2016-0287
- Llorens, Anaïs, Athina Tzovara, Ludovic Bellier, Ilina Bhaya-Grossman, Aurélie Bidet-Caulet, William K. Chang, Zachariah R. Cross, et al. 2021. “Gender Bias in Academia: A Lifetime Problem That Needs Solutions.” *Neuron* 109 (13): 2047–74. https://doi.org/10.1016/j.neuron.2021.06.002
- Long, Brandon R, Savannah Laux, Benjamin Lemon, Alexa Guarente, Mark S Davis, Arturo Casadevall, Ferric C Fang, Min Shi, and David B Resnik. 2023. “Factors Related to the Severity of Research Misconduct Administrative Actions: An Analysis of Office of Research Integrity Case Summaries from 1993 to 2023.” *Accountability in Research*, November, 1–22. https://doi.org/10.1080/08989621.2023.2287046

- Marquaridt, Donald W. 1970. “Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation.” *Technometrics* 12 (3): 591–612. https://doi.org/10.1080/00401706.1970.10488699
- Memon, Shahan Ali, Kinga Makovi, and Bedoor AlShebli. 2023. “Characterizing the Effect of Retractions on Scientific Careers.” *arXiv*, January. https://doi.org/10.48550/arxiv.2306.06710
- Mooijman, Marlon, and Jesse Graham. 2018. “Unjust Punishment in Organizations.”*Research in Organizational Behavior* 38: 95–106. https://doi.org/10.1016/j.riob.2018.10.001
- National Institutes of Health. 2019. “A Guide to the Handling of Research Misconduct Allegations.” December 2019. https://oir.nih.gov/system/files/media/file/2021-08/guide-handling_research_misconduct_allegations.pdf
- Olesen, Angelina P., Latifah Amin, and Zurina Mahadi. 2017. “Malaysian Researchers Talk about the Influence of Culture on Research Misconduct in Higher Learning Institutions.” *Accountability in Research* 24 (8): 469–82. https://doi.org/10.1080/08989621.2017.1399358
- Paruzel-Czachura, Mariola, Lidia Baran, and Zbigniew Spendel. 2020. “Publish or Be Ethical? Publishing Pressure and Scientific Misconduct in Research.” *Research Ethics* 17 (3): 375–97. https://doi.org/10.1177/1747016120980562
- Pekár, Stano, and Marek Brabec. 2017. “Generalized Estimating Equations: A Pragmatic and Flexible Approach to the Marginal GLM Modelling of Correlated Data in the Behavioural Sciences.” Edited by R. Bshary. *Ethology* 124 (2): 86–93. https://doi.org/10.1111/eth.12713
- Pratt, Travis C., Michael D. Reisig, Kristy Holtfreter, and Katelyn A. Golladay. 2019. “Scholars’ Preferred Solutions for Research Misconduct: Results from a Survey of Faculty Members at America’s Top 100 Research Universities.” *Ethics & Behavior* 29 (7): 510–30. https://doi.org/10.1080/10508422.2019.1612748
- Resnik, David B, Mohammad Hosseini, and Lisa Rasmussen. 2024. “Responding to Research Misconduct Allegations Brought against Top University Officials.” *Accountability in Research*, February, 1–6. https://doi.org/10.1080/08989621.2024.2321179
- Resnik, David B. 2019. “Is It Time to Revise the Definition of Research Misconduct?”*Accountability in Research* 26 (2): 123–37. https://doi.org/10.1080/08989621.2019.1570156
- Robinson-Zañartu, Carol, Elizabeth D. Peña, Valerie Cook-Morales, Anna M. Peña, Rosalyn Afshani, and Lynda Nguyen. 2005. “Academic Crime and Punishment: Faculty Members’ Perceptions of and Responses to Plagiarism.” *School Psychology Quarterly* 20 (3): 318–37. https://doi.org/10.1521/scpq.2005.20.3.318
- Shykhnenko, Kateryna, and Alina Sbruieva. 2022. “The Code of Conduct for Research Integrity, Governance, and Ethics in Education in the USA, Europe, and Ukraine: Comparative Analysis.” *European Journal of Educational Research* 11 (4): 2195–2207. https://doi.org/10.12973/eu-jer.11.4.2195
- Tsai, Daniel Fu-Chang. 2018. “Reflections on the Research Misconduct Cases in East Asia.”*East Asian Science, Technology and Society* 12 (2): 181–84. https://doi.org/10.1215/18752160-6577762
- Were, Edwin, Eunice Kaguiri, and Jepchirchir Kiplagat. 2020. “Perceptions of Occurrence of Research Misconduct and Related Factors among Kenyan Investigators Engaged in HIV Research.” *Accountability in Research* 27 (6): 372–89. https://doi.org/10.1080/08989621.2020.1759425
- Yeo-Teh, Nicole Shu Ling, and Bor Luen Tang. 2021. “A Research Misconduct Severity Matrix That Could Serve to Harmonize Adjudication of Findings.” *Accountability in Research* 29 (5): 279–93.

https://doi.org/10.1080/08989621.2021.1917398

- Yi, Nannan, Benoit Nemery, and Kris Dierickx. 2019. “How Do Chinese Universities Address Research Integrity and Misconduct? A Review of University Documents.” *Developing World Bioethics* 19 (2): 64–75. https://doi.org/10.1111/dewb.12231
- Yu, Lingling, Miao Miao, Wei Liu, Bi Zhang, and Pengjun Zhang. 2020. “Scientific Misconduct and Associated Factors: A Survey of Researchers in Three Chinese Tertiary Hospitals.” *Accountability in Research* 28 (2): 95–114. https://doi.org/10.1080/08989621.2020.1809386