Research Article

What Factors Influence the Punishment of Research Misconduct: Evidence from Hospitals in Mainland China

Gengyan Tang[1]

1. Sichuan Academy Of Social Sciences, Chengdu, China

Punishing researchers involved in research misconduct can impact their professional careers and involves issues of fairness. Previous studies (Long et al. 2023) have found that factors such as gender and academic status do not influence the severity of administrative actions taken by institutions. However, using research misconduct incidents in mainland Chinese hospitals as an example, our analysis employing tests of between-subject effects indicates that both professional title (p=0.038) and authorship order (p<0.001) significantly influence the punishment intensity. In second-order interaction effects, professional title and institutional level (p=0.025), as well as the type of research misbehavior and institutional level (p=0.006), jointly have a significant impact on punishment intensity. In third-order interaction effects, gender, professional title, and institutional level (p<0.001), as well as professional title, institutional level, and the type of research misbehavior (p=0.003), jointly have a significant impact on punishment intensity. We further discuss the specific interactions among these factors through simple effects, revealing potential "professional title bias" in addressing research misconduct in mainland Chinese hospitals. Various social factors intertwined contribute to the influence on punishment results, reflecting the complexity of addressing research misconduct as a systemic issue.

1. Introduction

Punishment is one of the means for universities, institutions, or government to address research misconduct (RM). For example, in the NIH's "*A Guide to the Handling of Research Misconduct Allegations*", it is stated that once RM is investigated and confirmed, the NIH will take measures to "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 (Dal-Ré et al. 2020). In addition to the penalties directly implemented by universities, institutions, or governments, researchers identified as having committed RM may also face public criticism and moral condemnation from society (Sugawara et al. 2017), which are forms of punishment. These combined penalties could deter researchers, thereby prompting them to take actions to ensure RI. Therefore, in some countries, punishment is an important tool for governing RM.

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 Chinese government 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). Starting from 2021, the Ministry of Science and Technology of the People's Republic of China (MOST) also regularly informs the public about the punishment results of RM incidents for supervision (MOST 2021).

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."

Inconsistent punishments for RM are not isolated cases but are widely observed, leading to dissatisfaction with the results of the sanctions. In a report on mainland Chinese biomedical researchers' views on RM between 2010 and 2015, 70.2% and 65.2% of respondents respectively believed that the penalties imposed by authorities and institutions for RM were inappropriate (Liao et al. 2018). Inconsistent punishments can lead to unfairness because the punishment may affect the career of researchers involved in RM (Galbraith 2016; Stern et al. 2014). If researchers committing the same severity of RM receive different punishments, it could lead to inequality. Although this viewpoint currently lacks sufficient evidence.

However, I would like to point out that, in addition to the impact of punishments on researchers, the factors influencing the punishments also deserve attention. Many factors in the adjudication process may influence the punishments, or the results are interrelated. Previous studies have shown that inconsistencies in the definition of RM may lead to different governance approaches (Li and Cornelis 2020). 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 status, or administrative position.

Nevertheless, due to different governance approaches to RM in different countries (Yi, Nemery, and Dierickx 2019), it is very likely that the adjudication process will be influenced by different factors, ultimately reflected in the punishment results. In mainland China, investigations and adjudications of RM are not conducted by the ORI as in the United States, but by the institutions where the researchers involved in RM are located, which conduct investigations and exercise discretion, and report the punishment results to the Ministry of Science and Technology of China. However, despite the frequent occurrence of RM in mainland China, we still lack evidence to understand and determine the factors influencing the punishments of RM in this country.

Identifying these factors is of great significance to us, as recognizing these potential influencing factors can ensure fairness in different institutions when handling such RM incidents. Additionally, providing evidence from mainland China will also help to further improve research in this field, analyzing the different challenges faced by governments and institutions in various countries in dealing with RM.

2. Research Hypothesis

2.1. Gender and Punishment of RM

Gender is not only a common factor in social life but is also found to be associated with RM (Ana-Catarina Pinho-Gomes, Carinna Hockham, and Woodward 2023). A study on retracted papers in the biomedical field showed that the majority of retractions (60.9%) involved first and last authors who were male. In research on the punishment of plagiarism, it was found that women faculty members were 36% more likely to judge cases more severely (Robinson-Zañartu et al. 2005). Gender bias is widespread in academia and can be either explicit or implicit (Llorens et al. 2021). Explicit bias involves conscious and intentional evaluations of a specific entity, with some degree of approval or

disapproval (Pratkanis, Steven James Breckler, and Greenwald 1989), and implicit bias reflects automatic judgments made by individuals without their awareness (Greenwald and Banaji 1995). Since investigations of RM often occur within the academic system, they may also be influenced by gender biases within this system. Therefore, we hypothesize that investigators or institutions might possess various forms and types of gender bias when punishing researchers guilty of RM, which in turn could influence the final punishment results. Thus, we propose Hypothesis 1:

- **H1:** Gender will influence the punishment results of RM incidents.

2.2. Professional Titles and Punishment for RM

In study on power dynamics and RM, it has been observed that younger researchers and those in lower academic status are less likely to take action and report compared to senior researchers. This reluctance is attributed to their concerns about potential harm to their career paths, and suggests a correlation between RM and academic status (Horbach et al. 2020). Furthermore, a study on the academic ecosystem in mainland China reveal that researchers with higher academic status form a "group": a socially autonomous entity that establishes its own operational rules (Weber, Colliot-Thélène, and Jean-Pierre Grossein 2000). They often take on administrative roles within schools or institutions and may participate in rule-making (Ou et al. 2021). The privileges associated with this status may enable them to mitigate the impact of RM incidents on their careers by leveraging their authority, thus minimizing the punishments. In mainland Chinese hospitals, academic status is typically represented in the form of "professional title" (职称), as these titles serve as symbolic measures of a healthcare professional's expertise and research capabilities (Zhang et al. 2017). Based on this, we propose Hypothesis 2:

- **H2:** Professional title will influence the punishment results of RM incidents.

2.3. Institutional Level and Punishment for RM

In mainland China, official institutions classify hospitals based on standards such as the quality of medical services, higher education, and research tasks, with the highest grade being tertiary A, tertiary B, secondary A, secondary B and so on (Cai et al. 2017). In existing literature, we have found no evidence suggesting a connection between these institutional level and RI or RM. In studies addressing academic integrity, scholars argue that institutional systems play a crucial role in fostering academic dishonesty (Jordan 2001). Some research indicates that lax systems (insufficient punishment or

sanctions for academic misconduct) can impact the academic integrity on campuses (Akbulut et al. 2008). We speculate that institutions with higher grades may implement more stringent RI policies to maintain their academic reputation and status, as reflected in the grading criteria, resulting in more severe consequences for RM. Consequently, we propose Hypothesis 3:

- **H3:** Institutional level will influence the punishment results of RM incidents.

2.4. Authorship Order and Punishment for RM

Generally, individuals who complete the majority of foundational work are often considered eligible to be the first author, implying that the first author may bear greater responsibility for the RI of the paper. Guidelines from the International Committee of Medical Journal Editors (ICMJE) and the Council of Science Editors (CSE) state that authors are directly responsible only for the work they contribute to, while the Committee on Publication Ethics (COPE) and the All European Academies (ALLEA) require all authors to be accountable for the integrity of the entire work (Helgesson and Eriksson 2017; Rennie 1997; Jones 2003). Although there are still divergences among different institutions or organizations regarding regulations on the responsibilities and accountabilities of authors, in specific research practices, 'individual accountability' is more prevalent than 'joint accountability'. For example, an empirical study found that: "First authors, corresponding authors, and to some extent, senior authors are more likely to be held responsible for RM than middle authors" (Hussinger and Pellens 2019).

Due to bearing greater responsibility, first authors and corresponding authors may face more severe punishments in cases of RM, while authors listed lower in the order may receive lighter punishments due to their lesser contributions and responsibilities. Thus, we propose Hypothesis 4:

- **H4:** Authorship order will influence the punishment results for RM.

2.5. Types of Research Misbehavior and Punishment for RM

In the year 2000, the U.S. federal government adopted a unified definition for RM, encompassing fabrication, falsification, or plagiarism (FFP) (KERÄNEN 2006), often considered extreme forms of RM that betray the truth (Kuroki 2018). However, Resnik (2019) found that, in addition to these types of RM, RM policies in U.S. research institutions also specify:

"'other serious deviations', 'significant or material violations of regulations', 'misuse of confidential information', 'misconduct related to misconduct', 'unethical authorship other than plagiarism', 'other deception involving data manipulation', and 'misappropriation of property/theft'".

But whether these types of research misbehaviors lead to different punishments still lacks evidence. In mainland China, the National Natural Science Foundation of China (NSFC)[1]released revised regulations in 2022 titled *"Procedures for Investigating and Dealing with Research Misconduct"* (National Natural Science Foundation of China 2022), which not only define types of RM but also stipulate corresponding punishments. For instance, acts of "using someone else's authorship without consent" may initially receive institutional criticism and education, while "submitting a falsified report or original records" could result in a warning[2]. We currently lack information on whether universities or institutions strictly adhere to these standards when adjudicating RM cases. Nevertheless, based on this regulation, we propose Hypothesis 5:

- **H5:** The types of research misbehavior will influence the punishment results for RM.

3. Materials and Methods

3.1. Variables

3.1.1. Independent Variables

According to the hypothesis, we have selected five independent variables:

1. Gender. Referring to some research on mainland China (Zeng et al. 2013), we divided gender into male and female categories.
2. Professional title. In mainland China, doctors, nurses, or pharmacists are evaluated for research capabilities (especially the quantity and quality of publications) during professional promotion (higher professional titles) (Xi et al. 2023). The professional title in mainland China's hospitals is divided into four levels: junior, intermediate, associate senior, and senior. Junior corresponds to resident doctors or nurses, intermediate to attending doctors or head nurses, associate senior to associate chief doctors or associate chief nurses, and senior to chief doctors and chief nurses.

3. Institutional level. As mentioned earlier, the National Health Commission of the People's Republic China (NHC) has classified hospitals in mainland China. However, during data collection, we found that almost all samples were from tertiary hospitals[3]. If we use this classification standard, the variable would lack differentiation and effectiveness. Therefore, we adopted the classification criteria of the "*China Hospital Competitiveness Report (2020-2021)*" (Cao et al. 2020). The rankings provided in this report cover a majority of hospitals in mainland China. The evaluation criteria also include indicators related to research work or RI, such as the number of papers, impact factor, and overall credit, making it more comprehensive.

4. Authorship order. We coded this variable based on the order of authors, but there are some special cases: firstly, if an author serves as both corresponding author and first author, their order is classified as "0"; secondly, in some cases, the corresponding author may be in the middle position, and we still consider the corresponding author as not occupying the authorship order. For example, if the corresponding author is in the fourth position, we still consider subsequent authors as fourth authors. Finally, if there are co-first authors, authors after them are considered second authors. Furthermore, due to the significant responsibilities undertaken by both the first author and the corresponding author for papers, and the challenge of often defining which is more important between the corresponding author and the first author (Pain 2021), we consider them of equal importance in our coding, assigning them the same code.

5. Types of research misbehavior. Due to the inconsistency in the definitions of RM across different institutions and countries (Breen 2016), it is essential to clarify the standards we adhere to. We classify and encode samples based on the categories specified in the "*Procedures for Investigating and Dealing with Research Misconduct*" released by the NSFC 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.

The coding for all variables can be found in Table 1.

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	Falsification and fabrication=1, Plagiarism=2, Violation of authorship norms=3, Ghostwriting=4, Violation of ethical standards=5, Combination of multiple research misbehaviors = 6

Table 1. Variable names, explanations, and encoding rules

3.1.2. Dependent variable

Unlike administrative actions taken by the ORI against RM, such as supervision of research or funding debarment (Banerjee, Partin, and Resnik 2022), institutions in mainland China show greater diversity in the punishments imposed on healthcare professionals involved in RM. These punishments may include financial penalties, administrative sanctions, or research sanctions (disciplinary action or warnings for researchers who are party members; administrative warnings for researchers who are not party members). These punitive measures can be implemented individually or bundled together, given the variability in this variable's categorization, making it challenging to conduct statistical analysis directly on the coded data as in previous studies. To address this issue, we propose the concept of "punishment intensity", which refers to the degree of strength or severity with which institutions and universities impose punishments on individuals for RM.

To quantify the variable of punishment intensity, we took the following steps:

1. Identify variable dimensions. Based on the collected sample's punishment results, we compiled a list of punishment results after summarization (see Table 2), consisting of five primary dimensions: 1. Administrative sanctions. This category of punishment is a unique administrative penalty system in mainland China designed to serve as a reminder and prevent recurrence. Administrative sanctions vary in severity based on the seriousness of the RM and include warnings, disciplinary actions, admonitory talks, etc. 2. Financial penalties. These measures primarily involve penalizing researchers engaged in RM financially, such as depriving them of the right to receive bonuses or performance-related pay. 3. Research penalties. This type of penalty restricts the research activities of researchers involved in RM, such as prohibiting them from applying for research funds or participating in the evaluation of research projects. 4. Professional penalties. This category restricts the opportunities for researchers committing RM to advance in their careers, such as being unable to apply for higher-level professional titles within a specified time window. 5. Teaching penalties. This penalty involves depriving researchers engaged in RM of the right to supervise graduate students.

2. Determine the valuation method. After identifying the dimensions of punishment intensity, we used the Delphi approach to assign values to different dimensions.

   "The Delphi approach is characterized by engagement and structured communication with and among experts in order to reach consensus on a topic. It is especially useful in areas where knowledge is scattered or disagreement predominates the discourse." (Lechner et al. 2023)

   For punishment intensity, we found it challenging to follow an objective standard (as there are no relevant policy documents in mainland China indicating which category of punishment intensity is stronger or weaker). Therefore, seeking consensus among experts through the Delphi approach addressed this issue.

3. Establish expert consensus. We sent valuation forms to a total of 15 experts, requesting them to assign scores for each dimension of punishment intensity. For the primary dimensions, scoring was done using percentages, and for the secondary dimensions, a 1-10 scale was used. After scoring, the final punishment intensity was calculated by multiplying the scores of secondary dimensions by the ratio (weight) of the primary dimensions. We received responses from 12 experts, representing 7 universities and 5 research institutions. Among them, 7 held Ph.D.

degrees, and 5 held master's degrees; 4 had over 8 years of experience in RI and evaluation work, 5 had over 5 years of experience, and 3 had over 3 years of experience.

4. Resolve expert discrepancies. Due to variations in scores assigned by different experts to same dimensions, we conducted two online discussion sessions on December 3 and December 18, 2023. Through negotiation and discussion, consensus was ultimately reached on the valuation of each dimension.

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 [1]	4
		Record in the RI archive or RM database [2]	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

Primary Dimension	Weight	Secondary Dimension	Score
		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

Table 2. Scores and proportions for each dimension of punishment intensity

[1] 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.

[2] 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.

3.2. Data Sources

1. Data on RM incidents are derived from the documents of investigation results disclosed by the MOST under the "Building Research Integrity (科研诚信建设)" program in 2021. The content of

these documents includes: 1. Name; 2. Title of the retracted paper involved in RM; 3. Authorship of the paper; 4. Types of research misbehavior; 5. Punishments.

2. Identity information of individuals involved in RM. As the MOST has not publicly disclosed the gender and professional title information of individuals involved in RM, we obtained relevant data through alternative channels. Initially, we retrieved information about researchers engaged in RM from institutional official websites. If the information was not available on the official website, we utilized certain apps for retrieval. In mainland China, some online medical consultation apps provide information such as the doctor's professional title, gender, name, and affiliated institution. Therefore, we conducted searches on these apps using the "affiliated unit + name" query format. For samples where gender could not be determined, we utilized the "NameSor Chinese API" (https://chinese-names.app/gender) for identification. The retrieval period was from January 3, 2023, to January 12, 2023.

The data sources for all variables are summarized in Table 3.

Variable	Data Sources
Gender	Institutional official websites, Chunyu Doctor App (https://www.chunyuyisheng.com/), and Youlai Doctor App (https://www.youlai.cn/)
Professional title	Institutional official websites, Chunyu Doctor App, and Youlai Doctor App
Institutional level	China Hospital Competitiveness Report (2020-2021)
Authorship order	Retrieved by searching PubMed for papers suspected of RM
Types of research misbehavior	MOST's reports on RM incidents (https://www.most.gov.cn/zxgz/kycxjs/kycxgzdt/)
Punishment results	MOST's reports on RM incidents and the Delphi method

Table 3. Data sources for each variable

3.3. Sample Selection

We obtained 408 cases of RM reported by the MOST in 2021, involving 1088 researchers. After obtaining these samples, we conducted data cleaning and applied the following criteria 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 our 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, so we removed such samples). The specific sample selection process is illustrated in Figure 1.
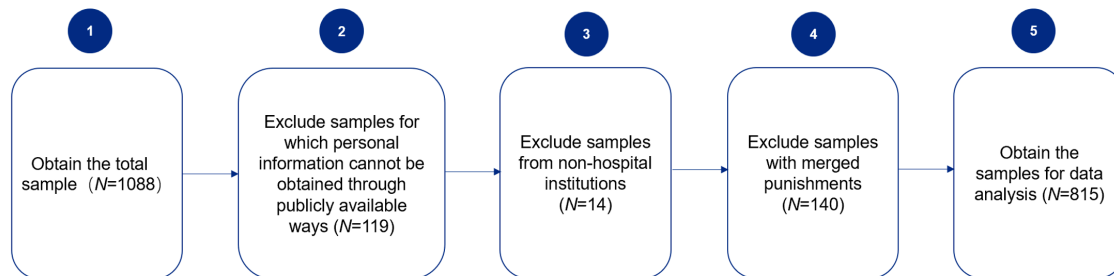
Figure 1. Sample selection process

3.4. Data Analysis

3.4.1. Coding

According to the coding table, we invited two coders to encode the data. These two coders have approximately 3 years of research experience and are familiar with research methods such as content analysis. Before coding, we provided coding guidelines and discussed coding results to reach consensus on issues where opinions differed. We used Cohen's Kappa to calculate inter-coder reliability for latent variables. The overall reliability for all latent variables exceeded 0.9.

3.4.2. Statistical Analysis

As our dependent variable is continuous and the independent variable is categorical, we employed a multivariate analysis of variance (MANOVA) to explore the impact of the independent variable on the dependent variable. Our statistical and graphical software is SPSS 27. Prior to analyzing the data, we conducted a Levene's test for homogeneity of variances using SPSS 27 to examine the equality of variances for the dependent variable. The data showed that the Levene's test statistic based on median was 1.058 with a significance level of 0.342, indicating equal variances among the data and meeting the conditions for conducting MANOVA.

4. Results

4.1. Descriptive Statistical Analysis

The results of descriptive statistics (see Table 4) indicate that, in terms of gender, the proportion of males involved in RM (64.7%) is higher than females (35.3%), supporting the findings of previous related studies (Fang, Bennett, and Casadevall 2013; Gopalakrishna et al. 2022). Researchers with the professional title of "associate senior" exhibit the highest rate of involvement in RM (42.1%), followed by researchers with intermediate titles (36.3%). In terms of institutional levels, researchers from municipal hospitals have the highest proportion of involvement in RM, with those ranked 1–250 in municipal hospitals reaching a percentage of 47.7%. Additionally, researchers from top-tier hospitals have a 14.7% involvement rate in RM incidents.

In terms of authorship order, the highest number of accountabilities falls on the first author or corresponding author, accounting for 73.6%, followed by the second author (7.9%) and the third author (6.1%). As the authorship order goes further down, the number of accountabilities decreases. Regarding the types of research misbehavior, the highest number is associated with fabrication or falsification (361), accounting for 44.3%, followed by ghostwriting (262), accounting for 32.1%. In terms of the punishment intensity (a continuous variable), the average value is 2.08, with a maximum value of 7.70 and a minimum value of 0.

Variable	Categories	N	Percentage (%)
Gender	Male	527	64.7%
	Female	288	35.3%
Professional title (PT)	Junior	27	3.3%
	Intermediate	296	36.3%
	Associate Senior	343	42.1%
	Senior	149	18.3%
Institutional level (IL)	Top-tier hospitals (T hospitals)	120	14.7%
	Provincial hospitals (P hospitals)	119	14.6%
	Municipal hospitals (MH) ranked 1-250	389	47.7%
	Municipal hospitals ranked 251-500	27	3.3%
	County hospitals (CH) ranked 1-250	36	4.4%
	County hospitals ranked 251-500	0	0%
	Not listed	124	15.2%
Authorship order (AO)	Corresponding author and first author	23	2.8%
	Corresponding author or first author	600	73.6%
	Second author	64	7.9%
	Third author	50	6.1%
	Fourth author	31	3.8%
	Fifth author	26	3.2%
	Sixth author	12	1.5%
	Seventh author	6	0.7%
	Eighth author	3	0.4%
Types of research misbehavior (TRM)	Falsification and fabrication (FF)	361	44.3%
	Plagiarism	0	0%

Variable	Categories			N	Percentage (%)
	Violation of authorship norms			1	0.1%
	Ghostwriting			262	32.1%
	Violation of ethical standards			0	0%
	Combination of multiple research misbehaviors (CMRM)			191	23.4%
	Mean	Maximum	Minimum	N	SD
Punishment results	2.08	7.70	0.00	815	1.45

Table 4. The descriptive statistical analysis results of each variable

4.2. Analysis of Between-Subjects Effects and Interactive Effects

We analyzed between-subjects effects and interactive effects among variables using a MANOVA model. Table 5 demonstrates that, in terms of between-subjects effects, professional title ($p=0.038$) and authorship order ($p<0.001$) both have statistically significant impacts on the punishment results of RM. However, gender ($p=0.252$), institutional level ($p=0.539$), and the types of research misbehavior ($p=0.634$) do not influence the punishment results of RM.

Furthermore, we conducted tests for interactive effects between pairs of variables and among three variables. Results of the bivariate interactive effects tests indicate that there is an interactive effect between professional title and institutional level ($p=0.025$), as well as between institutional level and the types of research misbehavior ($p=0.006$). However, there are no interactive effects among other pairs of variables ($p>0.05$). Results of the three-variable interactive effects show an interactive effect among gender, professional title, and institutional level ($p<0.001$). Additionally, there is an interactive effect among professional title, institutional level, and the types of research misbehavior ($p=0.003$). No interactive effects were found among the remaining three-variable combinations ($p>0.05$). To precisely elucidate the differences among variables in interactive effects, further simple effects analyses are required.

	SST	df	MSD	F-value	sig
Adjusted model	842.108	252	3.342	2.178	<0.001
Intercept	119.292	1	119.292	77.736	<0.001
Gender	2.019	1	2.019	1.316	0.252
PT	13.006	3	4.335	2.825	0.038
IL	6.252	5	1.250	0.815	0.539
AO	130.074	8	16.259	10.595	<0.001
TRM	2.632	3	0.877	0.572	0.634
PT*IL	36.152	12	3.013	1.963	0.025
PT*TRM	9.969	6	1.661	1.083	0.371
PT*AO	32.215	16	2.013	1.312	0.184
PT*Gender	2.346	3	0.782	0.510	0.676
IL*TRM	38.774	10	3.877	2.527	0.006
IL*AO	16.906	27	0.626	0.408	0.997
IL*Gender	8.654	5	1.731	1.128	0.344
TRM*AO	22.407	12	1.867	1.217	0.267
TRM*Gender	3.848	2	1.924	1.254	0.286
AO*Gender	1.813	6	0.302	0.197	0.978
Gender*PT*IL	41.080	8	5.135	3.346	<0.001
Gender*PT*TRM	16.666	5	3.333	2.172	0.056
Gender*PT*AO	14.260	5	2.852	1.858	0.100
Gender*IL*TRM	17.487	6	2.914	1.899	0.079
Gender*IL*AO	2.278	4	0.570	0.371	0.829
Gender*TRM*AO	7.246	6	1.208	0.787	0.580
PT*IL*TRM	51.379	14	3.670	2.392	0.003
PT*IL*AO	4.431	5	0.886	0.577	0.717

	SST	df	MSD	F-value	sig
PT*TRM*AO	8.201	8	1.025	0.668	0.720
IL*TRM*AO	3.761	6	0.627	0.408	0.874
Standard Error	862.434	562	1.535		
N	815				
R-square	0.494				
Adjusted R-square	0.267				

Table 5. The test results for between-subjects effects and interaction effects

Note. IL: institutional level; PT: professional title; TRM: types of research misbehavior; AO: authorship order

4.3. Analysis of Simple Effects

4.3.1. Simple Effects (Bivariate)

Due to an excessive number of tables and their length, we only report positive results; negative results are available in the Supplement Material. Table 6 reveals positive results for the simple effects of the variables. Within top-tier hospitals, samples with intermediate professional titles receive stronger punishment intensity than those with senior professional titles (*p*=0.003, MD=1.444). Under the same conditions, samples with associate senior professional titles face significantly stronger punishment intensity than those with senior professional titles (*p*=0.004, MD=1.359). Within provincial hospitals, samples with associate senior professional titles also receive significantly stronger punishment intensity than those with senior professional titles (*p*=0.035, MD=0.803).

When the target of punishment holds a intermediate professional title, county hospitals ranked 1-250 exhibit significantly stronger punishment intensity than municipal hospitals ranked 1-250 (*p*=0.039, MD=0.945). Similarly, when the target of punishment holds an associate senior professional title, county hospitals ranked 1-250 impose significantly stronger punishment intensity than municipal hospitals ranked 1-250 (*p*=0.021, MD=1.043), and also more substantial punishment intensity than hospitals not listed in the ranking (*p*=0.032, MD=1.023). In cases where the target of punishment holds a senior professional title, municipal hospitals ranked 1-250 (*p*=0.049, MD=0.833) and county

hospitals ranked 1–250 (p=0.019, MD=1.387) both impose punishments significantly stronger than top-tier hospitals. Additionally, hospitals not listed in the ranking impose greater punishment intensity than top-tier hospitals (p=0.006, MD=1.464) and provincial hospitals (p=0.038, MD=0.990)

For research misbehavior involving fabrication and falsification, provincial hospitals (p=0.047, MD=0.674) and municipal hospitals ranked 1–250 (p=0.028, MD=0.645) impose significantly stronger punishment intensity than hospitals not listed in the rankings. When the research misbehavior involves ghostwriting, county hospitals impose significantly stronger punishment intensity than top-tier hospitals (p<0.001, MD=2.182), provincial hospitals (p<0.001, MD=1.945), municipal hospitals ranked 1–250 (p<0.001, MD=1.893), municipal hospitals ranked 251–500 (p=0.002, MD=1.643), and hospitals not listed in the rankings (p=0.003, MD=1.450). For combination of multiple research misbehaviors, top-tier hospitals impose significantly stronger punishment intensity than provincial hospitals (p=0.022, MD=1.033) and municipal hospitals ranked 1–250 (p=0.027, MD=0.943). Additionally, county hospitals ranked 1–250 impose significantly stronger punishment intensity than provincial hospitals (p=0.046, MD=1.511). Hospitals not listed in the rankings impose significantly stronger punishment intensity than provincial hospitals (p=0.013, MD=1.074) and municipal hospitals ranked 1–250 (p=0.016, MD=0.984).

In top-tier hospitals, research misbehavior involving combinations receives significantly stronger punishment intensity than ghostwriting (p=0.018, MD=1.357). In county hospitals ranked 1–250, research misbehavior involving ghostwriting receives significantly stronger punishments than fabrication and falsification (p=0.003, MD=1.560). In hospitals not listed in the ranking, research misbehavior involving combinations receives significantly stronger punishment intensity than fabrication and falsification (p=0.002, MD=1.303).

Paired Comparison (Examining differences in punishment intensity of PT at different levels when IL is at a certain level)						
IL	(I) PT	(J) PT	MD	SE	sig	95% CI
T hospitals	Intermediate	Senior	1.444	0.481	0.003	0.500~2.388
	Associate Senior	Senior	1.359	0.470	0.004	0.435~2.282
P hospitals	Associate Senior	Senior	0.803	0.379	0.035	0.059~1.548
Paired Comparison (Examining differences in punishment intensity of IL at different levels when PT is at a certain level)						
PT	(I) IL	(J) IL	MD	SE	sig	95% CI
Intermediate	CH ranked 1–250	MH ranked 1–250	0.945	0.457	0.039	0.047~1.843
Associate Senior	CH ranked 1–250	MH ranked 1–250	1.043	0.452	0.021	0.155~1.932
	CH ranked 1–250	Not listed	1.023	0.477	0.032	0.086~1.959
Senior	MH ranked 1–250	T hospitals	0.833	0.423	0.049	0.03~1.664
	CH ranked 1–250	T hospitals	1.387	0.591	0.019	0.226~2.547
	Not listed	T hospitals	1.463	0.526	0.006	0.429~2.497
	Not listed	P hospitals	0.990	0.477	0.038	0.054~1.927
Paired Comparison (Examining differences in punishment intensity of IL at different levels when the TRM is at a certain type)						
TRM	(I) IL	(J) IL	MD	SE	sig	95% CI
FF	P hospitals	Not listed	0.674	0.339	0.047	0.008~1.339
	MH ranked 1–250	Not listed	0.645	0.292	0.028	0.072~1.218
Ghostwriting	CH ranked 1–250	T hospitals	2.182	0.595	<0.001	1.013~3.352
	CH ranked 1–250	P hospitals	1.945	0.483	<0.001	0.996~2.894
	CH ranked 1–250	MH ranked 1–250	1.893	0.443	<0.001	1.022~2.764
	CH ranked 1–250	MH ranked 251–500	1.643	0.530	0.002	0.602~2.683
	CH ranked 1–250	Not listed	1.450	0.488	0.003	0.491~2.410

	T hospitals	P hospitals	1.033	0.449	0.022	0.150~1.915
	T hospitals	MH ranked 1-250	0.943	0.425	0.027	0.107~1.778
CMRM	CH ranked 1-250	P hospitals	1.511	0.756	0.046	0.026~2.995
	Not listed	P hospitals	1.074	0.433	0.013	0.223~1.925
	Not listed	MH ranked 1-250	0.984	0.408	0.016	0.183~1.786

Paired Comparison (Examining differences in punishment intensity of the TRM at different types when IL is at a certain level)						
IL	(I) TRM	(J) TRM	MD	SE	sig	95% CI
T hospitals	CMRM	Ghostwriting	1.357	0.573	0.018	0.231~2.482
CH ranked 1-250	Ghostwriting	FF	1.560	0.529	0.003	0.522~2.599
Not listed	CMRM	FF	1.303	0.418	0.002	0.482~2.124

Table 6. The test results for the simple effects of the bivariate analysis

Note. IL: institutional level; PT: professional title; TRM: types of research misbehavior; AO: authorship order; FF: falsification and fabrication; CMRM: combination of multiple research misbehaviors; T hospitals: top-tier hospitals; P hospitals: provincial hospitals; MH ranked 1-250: municipal hospitals ranked 1-250; MH ranked 251-250: municipal hospitals ranked 251-500; CH ranked 1-250: county hospitals ranked 1-250.

4.3.2. Simple Effects (Trivariate)

4.3.2.1. Simple Effects of Gender, Professional Titles, and Institutional Level

Table 7 presents the results of simple effects for three variables: gender, professional title, and institutional level. The results indicate that, within top-tier hospitals, male researchers with associate senior professional titles experience significantly stronger punishment intensity than their female counterparts (p=0.015, MD=1.471). Conversely, in hospitals not listed in the ranking, female researchers with senior professional titles receive significantly stronger punishment intensity than male researchers (p=0.014, MD=2.025).

Within top-tier hospitals, male researchers with intermediate professional titles (p=0.004, MD=1.695) and associate senior professional titles (p<0.001, MD=1.979) experience significantly stronger punishment intensity than those with senior professional titles. In provincial hospitals, female researchers with associate senior professional titles receive significantly stronger punishment intensity than those with senior professional titles (p=0.027, MD=1.271).

When male researchers with intermediate professional titles are the target of punishment, top-tier hospitals impose significantly stronger punishment intensity than provincial hospitals (p=0.026, MD=1.131) and hospitals not listed in the ranking (p=0.019, MD=1.187). For male researchers with associate senior professional titles as the target, top-tier hospitals implement significantly stronger punishment intensity than provincial hospitals (p=0.012, MD=1.197), municipal hospitals ranked 1-250 (p<0.001, MD=1.363), and hospitals not listed in the ranking (p<0.001, MD=1.621). Under the same conditions, county hospitals ranked 1-250 impose significantly stronger punishment intensity than municipal hospitals (p=0.035, MD=1.146) and hospitals not listed in the ranking (p=0.018, MD=1.404). When the target is female researchers with senior professional titles, hospitals not listed in the ranking implement significantly stronger punishment intensity than top-tier hospitals (p=0.004, MD=2.908), provincial hospitals (p=0.007, MD=2.283), and municipal hospitals ranked 1-250 (p=0.029, MD=1.717).

Pairwise comparisons (detecting differences in punishment intensity between different Genders at certain levels or types of IL and PT)							
PT	IL	(I) Gender	(J) Gender	MD	SE	sig	95% CI
Associate Senior	T hospitals	Male	Female	1.471	0.604	0.015	0.285~2.656
Senior	Not listed	Female	Male	2.025	0.821	0.014	0.411~3.638

Pairwise comparisons (detecting differences in punishment intensity between different PT at certain levels or types of IL and Gender)							
Gender	IL	(I) PT	(J) PT	MD	SE	sig	95% CI
Male	T hospitals	Intermediate	Senior	1.695	0.594	0.004	0.528~2.861
Male	T hospitals	Associate Senior	Senior	1.979	0.534	<0.001	0.931~3.027
Female	P hospitals	Associate Senior	Senior	1.271	0.575	0.027	0.141~2.400

Pairwise comparisons (detecting differences in punishment intensity between different ILs at certain types of PT and Gender)							
Gender	PT	(I) IL	(J) IL	MD	SE	sig	95% CI
Male	Intermediate	T hospitals	P hospitals	1.131	0.508	0.026	0.133~2.130
		T hospitals	Not listed	1.187	0.506	0.019	0.192~2.182
	Associate Senior	T hospitals	P hospitals	1.197	0.472	0.012	0.270~2.124
		T hospitals	MH ranked 1–250	1.363	0.409	<0.001	0.599~2.166
		T hospitals	Not listed	1.621	0.469	<0.001	0.700~2.542
		CH ranked 1–250	MH ranked 1–250	1.146	0.543	0.035	0.079~2.213
		CH ranked 1–250	Not listed	1.404	0.589	0.018	0.246~2.562
Female	Senior	Not listed	T hospitals	2.908	1.011	0.004	0.922~4.895
		Not listed	P hospitals	2.283	0.839	0.007	0.636~3.931

		Not listed	MH ranked 1–250	1.717	0.787	0.029	0.172~3.262

Note. IL: institutional level; PT: professional title; TRM: types of research misbehavior; AO: authorship order; FF: falsification and fabrication; CMRM: combination of multiple research misbehaviors; T hospitals: top-tier hospitals; P hospitals: provincial hospitals; MH ranked 1–250: municipal hospitals ranked 1–250; MH ranked 251–250: municipal hospitals ranked 251–500; CH ranked 1–250: county hospitals ranked 1–250.

4.3.2.2. Simple Effects of Professional Title, Institutional Level, and Types of Research Misbehavior

Table 8 presents the results of the simple effects of three variables: professional title, institutional level, and types of research misbehavior. The results indicate that, within top-tier hospitals, researchers with intermediate professional titles (p=0.007, MD=2.253) and associate senior professional titles (p=0.029, MD=2.037) receive significantly stronger punishment intensity when involved in combinations of multiple research misbehavior compared to researchers with senior professional titles. In provincial hospitals, when researchers with associate senior professional titles are found to commit fabrication and falsification, their punishment intensity is significantly stronger than that of researchers with intermediate professional titles (p=0.029, MD=1.366) and senior professional titles (p=0.002, MD=1.877).

In municipal hospitals ranked 1–250, researchers with junior professional titles receive significantly stronger punishment intensity for engaging in combinations of multiple research misbehavior compared to those with intermediate professional titles (p<0.001, MD=4.493), associate senior professional titles (p<0.001, MD=4.616), and senior professional titles (p<0.001, MD=3.854). In county hospitals ranked 1–250, researchers with intermediate professional titles engaging in fabrication and falsification face significantly stronger punishment intensity than those with senior professional titles (p=0.022, MD=1.969). Additionally, researchers with senior professional titles who engage in ghostwriting receive significantly stronger punishment intensity than those with intermediate professional titles (p=0.028, MD=2.087). When researchers with associate senior professional titles

are involved in combinations of multiple research misbehavior, they face significantly stronger punishment intensity than researchers with senior professional titles (p=0.016, MD=3.650).

In hospitals not listed in the ranking, researchers with senior professional titles engaging in fabrication and falsification receive stronger punishment intensity compared to those with associate senior professional titles (p=0.022, MD=1.601). However, when researchers with associate senior professional titles are involved in combinations of multiple research misbehavior, they receive significantly stronger punishment intensity than those with senior professional titles (p=0.018, MD=1.983).

For researchers with junior professional titles, engaging in combinations of multiple research misbehavior in county hospitals results in significantly stronger punishment intensity compared to provincial hospitals (p=0.011, MD=3.850) and hospitals not listed in the ranking (p=0.002, MD=4.750). Researchers with intermediate professional titles, when involved in ghostwriting in county hospitals, receive significantly stronger punishment intensity than in provincial hospitals (p=0.001, MD=1.884), municipal hospitals ranked 1–250 (p<0.001, MD=1.735), municipal hospitals ranked 251–500 (p=0.014, MD=1.648), and hospitals not listed in the ranking (p=0.010, MD=1.555). In top-tier hospitals, they face significantly stronger punishment intensity than in municipal hospitals ranked 1–250 (p=0.012, MD=1.695).

For researchers with associate senior professional titles, engaging in fabrication and falsification in top-tier hospitals results in significantly stronger punishment intensity than in hospitals not listed in the ranking (p=0.034, MD=1.066); in provincial hospitals, the punishment intensity is stronger than in top-tier hospitals (p=0.035, MD=1.158), municipal hospitals ranked 1–250 (p=0.019, MD=1.245), and hospitals not listed in the ranking (p<0.001, MD=2.224). When involved in various combinations of multiple research misbehavior, the punishment intensity in top-tier hospitals is significantly stronger than in municipal hospitals ranked 1–250 (p=0.034, MD=1.603). Under similar conditions, county hospitals ranked 1–250 exhibit significantly stronger punishment intensity than provincial hospitals (p=0.004, MD=2.784) and municipal hospitals ranked 1–250 (p<0.001, MD=3.066). Similarly, in hospitals not listed in the ranking, the punishment intensity is stronger than in provincial hospitals (p=0.002, MD=2.218) and municipal hospitals ranked 1–250 (p<0.001, MD=2.499).

For researchers with senior professional titles, engaging in fabrication and falsification in hospitals not listed in the ranking results in stronger punishment intensity compared to top-tier hospitals (p=0.044, MD=1.580) and county hospitals ranked 1–250 (p=0.024, MD=2.137). When they hire

someone to ghostwrite their papers, the punishment intensity in county hospitals is significantly stronger than in top-tier hospitals ($p<0.001$, MD=3.842), provincial hospitals ($p<0.001$, MD=3.270), municipal hospitals ranked 1–250 ($p<0.001$, MD=3.112), and municipal hospitals ranked 251–500 ($p=0.011$, MD=2.900). Under the same conditions, the punishment in hospitals not listed in the ranking is significantly stronger than in top-tier hospitals ($p=0.027$, MD=2.104).

For researchers with junior professional titles, engaging in fabrication and falsification in municipal hospitals ranked 1–250 results in significantly stronger punishment intensity than engaging in ghostwriting ($p=0.008$, MD=1.662). When involved in combinations of multiple research misbehavior, the punishment intensity is significantly stronger than engaging in fabrication and falsification ($p=0.001$, MD=3.192) and ghostwriting ($p<0.001$, MD=4.854). For researchers with intermediate professional titles, engaging in fabrication and falsification in provincial hospitals results in stronger punishment intensity compared to engaging in ghostwriting ($p=0.049$, MD=1.170) and combinations of multiple research misbehavior ($p=0.004$, MD=1.660). Under the same professional title category, engaging in combinations of multiple research misbehavior in municipal hospitals ranked 1–250 results in significantly stronger punishment intensity than engaging in fabrication and falsification ($p=0.017$, MD=2.663). Similarly, engaging in combinations of multiple research misbehavior in hospitals not listed in the ranking results in significantly stronger punishment intensity than engaging in fabrication and falsification ($p<0.001$, MD=2.781) and ghostwriting ($p=0.010$, MD=1.779). For researchers with senior professional titles, hiring someone to ghostwrite their papers in county hospitals results in significantly stronger punishment intensity than engaging in fabrication and falsification ($p<0.001$, MD=4.083) and combinations of multiple research misbehavior ($p=0.007$, MD=3.850).

Pairwise comparisons (detecting differences in punishment intensity between different PTs at certain ILs and TRM)							
IL	TRM	(I) PT	(J) PT	MD	SE	sig	95% CI
T hospitals	CMRM	Intermediate	Senior	2.253	0,834	0.007	0.614~3.892
		Associate Senior	Senior	2.037	0.929	0.029	0.213~3.862
P hospitals	FF	Associate Senior	Intermediate	1.366	0.625	0.029	0.138~2.594
		Associate Senior	Senior	1.877	0.601	0.002	0.698~3.057
MH ranked 1–250	CMRM	Junior	Intermediate	4.493	0.955	<0.001	2.616~6.369
		Junior	Associate Senior	4.616	0.927	<0.001	2.796~6.436
		Junior	Senior	3.854	0.956	<0.001	1.975~5.732
CH ranked 1–250	FF	Intermediate	Senior	1.969	0.855	0.022	0.290~3.648
	Ghostwriting	Senior	Associate Senior	2.087	0.946	0.028	0.229~3.946
	CMRM	Associate Senior	Senior	3.650	1.517	0.016	0.670~6.630
Not listed	FF	Senior	Associate Senior	1.601	0.698	0.022	0.231~2.971
	CMRM	Associate Senior	Senior	1.983	0.834	0.018	0.344~3.622
Pairwise comparisons (detecting differences in punishment intensity between different ILs at certain PTs and TRM)							
PT	TRM	(I) IL	(J) IL	MD	SE	sig	95% CI
Junior	CMRM	MH ranked 1–250	P hospitals	3.850	1.517	0.011	0.870~6.830
		MH ranked 1–250	Not listed	4.750	1.517	0.002	1.770~7.730
Intermediate	Ghostwriting	CH ranked 1–250	P hospitals	1.884	0.574	0.001	0.757~3.011
		CH ranked 1–250	MH ranked 1–250	1.735	0.511	<0.001	0.732~2.738

		CH ranked 1-250	MH ranked 251-500	1.648	0.669	0.014	0.334~2.962
		CH ranked 1-250	Not listed	1.555	0.605	0.010	0.366~2.744
	CMRM	T hospitals	MH ranked 1-250	1.695	0.677	0.012	0.367~3.024
Associate Senior	FF	T hospitals	Not listed	1.066	0.502	0.034	0.080~2.052
		P hospitals	T hospitals	1.158	0.549	0.035	0.080~2.236
		P hospitals	MH ranked 1-250	1.245	0.528	0.019	0.208~2.282
		P hospitals	Not listed	2.224	0.522	<0.001	1.199~3.249
		MH ranked 1-250	Not listed	0.979	0.479	0.041	0.039~1.919
	CMRM	T hospitals	MH ranked 1-250	1.603	0.756	0.034	0.119~3.087
		CH ranked 1-250	P hospitals	2.784	0.972	0.004	0.876~4.693
		CH ranked 1-250	MH ranked 1-250	3.066	0.927	<0.001	1.246~4.886
		Not listed	P hospitals	2.218	0.700	0.002	0.884~3.592
		Not listed	MH ranked 1-250	2.499	0.635	<0.001	1.251~3.747
Senior	FF	Not listed	T hospitals	1.580	0.783	0.044	0.042~3.118
		Not listed	CH ranked 1-250	2.137	0.942	0.024	0.287~3.986
	Ghostwriting	CH ranked 1-250	T hospitals	3.842	0.968	<0.001	1.940~5.744
		CH ranked 1-250	P hospitals	3.270	0.888	<0.001	1.527~5.013
		CH ranked 1-250	MH ranked 1-250	3.112	0.814	<0.001	1.513~4.712
		CH ranked 1-250	MH ranked 251-500	2.900	1.131	0.011	0.679~5.121

		Not listed	T hospitals	2.104	0.952	0.027	0.235~3.974

Pairwise comparisons (detecting differences in punishment intensity between different TRM at certain PTs and ILs)							
PT	IL	(I) TRM	(J) TRM	MD	SE	sig	95% CI
Junior	MH ranked 1–250	FF	Ghostwriting	1.662	0.626	0.008	0.433~2.891
		CMRM	FF	3.192	0.973	0.001	1.282~5.102
		CMRM	Ghostwriting	4.854	0.990	<0.001	2.909~6.799
Associate Senior	P hospitals	FF	Ghostwriting	1.170	0.594	0.049	0.004~2.336
		FF	CMRM	1.660	0.581	0.004	0.519~2.801
	CH ranked 1–250	CMRM	FF	2.663	1.117	0.017	0.469~4.856
	Not listed	CMRM	FF	2.781	0.651	<0.001	1.502~4.060
		CMRM	Ghostwriting	1.779	0.691	0.010	0.421~3.137
Senior	CH ranked 1–250	Ghostwriting	FF	4.083	1.011	<0.001	2.097~6.070
		Ghostwriting	CMRM	3.850	1.430	0.007	1.040~6.660

Table 8. The test results for the simple effects of TRM, PT, and IL

Note. IL: institutional level; PT: professional title; TRM: types of research misbehavior; AO: authorship order; FF: falsification and fabrication; CMRM: combination of multiple research misbehaviors; T hospitals: top-tier hospitals; P hospitals: provincial hospitals; MH ranked 1–250: municipal hospitals ranked 1–250; MH ranked 251–250: municipal hospitals ranked 251–500; CH ranked 1–250: county hospitals ranked 1–250.

4.5. Robustness Test

To assess the robustness of results, we altered the weight of administrative penalties (independent variable) in the dimensions of penalty intensity from 15% to 10% and increased the weight of research penalties from 25% to 30%. Through MANOVA (results in Appendix 1), we found that the significance of variables did not undergo major changes, indicating a certain level of robustness in our results.

5. Conclusion and Discussion

5.1. Main findings

Using MANOVA, we analyzed the impact of gender, institutional level, professional title, author order, and types of research misbehavior on punishment intensity in cases of RM in mainland Chinese hospitals. The results of between-subject effects indicated that gender, institutional level, and types of research misbehavior did not have statistically significant effects on punishment intensity, indicating that hypotheses 1, 3, and 5 were not supported. Professional title and author order had statistically significant effects on punishment intensity, supporting hypotheses 2 and 4. Professional title and authorship order have been confirmed in previous studies to influence researchers' perceptions of RM or accountability for RM incidents (Haven et al. 2019). However, there is no evidence yet to suggest that these two factors affect the punishment results. Therefore, our study provides the first evidence for this hypothesis and offers a large-scale sample of data from hospitals in mainland China. Furthermore, we found interactions between variables in the analysis of between-subject effects, suggesting potential biases if conclusions were drawn solely from main effects. Further analysis of simple effects between variables was conducted to understand the specific nature of these interactions.

Despite gender, institutional level, and types of research misbehavior not individually affecting punishment intensity in the between-subject effects tests, their interactions had a joint impact on punishment intensity. In a second-order interaction, the interaction between institutional level and professional title affected punishment intensity. For example, in provincial hospitals, researchers with intermediate and associate professional titles faced significantly stronger punishment intensity for RM than those with senior professional titles. This revealed a potential "professional title bias" in handling RM cases in specific hierarchical hospitals, where researchers with senior professional titles might receive milder punishment compared to those with lower titles. This bias has been previously identified in the introductions of speakers at oncology surgery conferences (Stewart et al. 2020). However, in the field of RI, there is still a lack of evidence regarding whether this factor interacts with other factors. Our study addresses this gap in knowledge. Interestingly, when professional titles were in a certain type, different hospitals showed distinct punishment orientations. For instance, when researchers with junior professional titles engaged in RM, county hospitals ranked 1-250 imposed significantly harsher punishment than municipal hospitals ranked 1-250. This finding suggests

variations in punishment for researchers with the same professional title in different levels of hospitals.

In another interaction, we found that institutional level and types of research misbehavior interacted to influence punishment intensity. On one hand, for the same type of research misbehavior, hospitals of different levels applied inconsistent punishment intensities. For instance, for ghostwriting, punishments in county hospitals ranked 1-250 were significantly higher than in all other levels of hospitals. Even within this category of hospitals, the punishment intensity for ghostwriting was significantly stronger than for fabrication and falsification of research misbehavior. This reflects a potential issue: despite official guidelines from institutions like NSFC governing the handling of RM cases, hospitals might have significant interpretive and operational autonomy. They may adopt different standards in dealing with RM incidents, resulting in significant variations in the punishment intensity for the same type of research misbehavior across different hospitals. Previous literature has suggested that inconsistencies in how institutions define types of research misbehavior may affect the investigative process of such cases (Catano and Turk 2007). Our study further confirms that, under the combined influence of multiple factors, varying attitudes towards the definition of types of research misbehavior can indeed lead to significant differences in punishment results.

Building on these findings, we conducted third-order interaction analysis and found that gender, along with professional title and institutional level, jointly influenced punishment intensity. Specifically, the impact of gender on punishment intensity needed to be considered under specific conditions. For instance, in top-tier hospitals, when research misbehavior was confirmed, male researchers with associate senior professional titles faced significantly different punishment intensities compared to their female counterparts with the same professional titles. However, besides gender differences, variations also existed within each gender group. For example, in provincial hospitals, female researchers with associate senior professional titles faced significantly harsher punishments than their same-gender counterparts with senior professional titles if research misbehavior was confirmed. Interestingly, no such differences were observed among male researchers under the same conditions. On the other hand, when gender and professional title were in a certain type, the institutional level also influenced punishment intensity. For instance, male researchers with junior professional titles received significantly harsher punishments in top-tier hospitals than in provincial hospitals. This finding further reveals inconsistent actions taken by hospitals of different levels in handling RM cases and the resulting diverse punishment results.

Finally, we identified interactions between professional title, institutional level, and types of research misbehavior. This interaction effect reflected, under certain conditions (professional title), the attitudes of hospitals at the same level toward different types of research misbehavior. For example, in municipal hospitals ranked 1-250, researchers with junior professional titles engaging in fabrication and falsification faced significantly harsher punishments than those engaged in ghostwriting. Similarly, in county hospitals ranked 1-250, researchers with senior professional titles faced significantly harsher punishments for ghostwriting than for fabrication and falsification. Furthermore, under the same professional title conditions, hospitals of different levels adopted different punishment strategies for the same type of research misbehavior. For instance, when researchers with associate senior professional titles engaged in fabrication and falsification, provincial hospitals imposed significantly stronger penalties than top-tier hospitals. This suggests that hospitals of different levels have different attitudes toward the same type of research misbehavior (although under specific conditions), confirming the significant impact of the joint action of professional title and institutional level on the punishment results.

In conclusion, our second-order interaction effects revealed title biases and disparities in the perceived severity of different types of research misbehavior, leading to inconsistent punishment results. Third-order interaction effects further revealed the complex interactions between gender, professional title, and institutional level and their joint influence on punishment results, demonstrating intricate interactions among professional title, institutional level, and types of research misbehavior.

5.2. Theoretical Contributions and Practical Significance

In the preceding text, we mentioned that many studies have found that factors influencing the actions taken by institutions or universities in response to RM cases are often singular, such as the type of research misbehavior or gender. However, they overlook a crucial aspect: if we perceive the handling of RM incidents by institutions or universities as a more complex system (involving different stakeholders and interconnections), the governance approach may involve the combined influence of multiple factors rather than the independent effect of a single factor. As Sarah Elaine Eaton (2023) pointed 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."

Based on this, we believe that when addressing the governance of RM cases, we should consider it as a complex problem and further explore the role of multiple factors. We undoubtedly advance this perspective, providing updated theoretical insights: factors such as gender, type of research misbehavior, institutional level, professional title, etc., may not act independently or have no impact on the punishment of RM cases in hospitals. Instead, they intertwine and interact in a complex manner to jointly influence punitive outcomes, leading to significant inconsistencies in punishment methods adopted by hospitals under specific conditions.

We also offer insights into the practical implications of RM governance policies, emphasizing the importance of conducting investigations and handling RM cases by unified or independent institutions, rather than institutions conducting self-examinations or rendering judgments independently. During the writing of this article, the Ministry of Education of the People's Republic of China called for universities to conduct self-investigations into whether retractions in the large-scale Hindawi retraction incident (involving over 7,000 articles by mainland Chinese authors) were related to RM (LV 2024). We cannot confirm whether such self-investigations are also subject to the combined influence of various social factors (such as gender, academic title, institution), leading to inconsistent punishments for RM cases of similar severity. We urge institutions such as universities and hospitals to exercise caution in implementing punishments for RM cases, assess the consistency of punishment results, and consider more adaptive alternative measures to ensure fairness.

5.3. Research Limitations and Future Directions

Due to the limitation of our sample selection to RM events in mainland Chinese hospitals, we cannot confirm whether our conclusions can be generalized to a broader range of institutions, such as universities. We hope to include more diverse samples in future studies to determine if our research findings are more robust. Additionally, we aim to further explore the causes of inconsistent punitive outcomes. In other words, besides policies, we want to investigate whether other factors contribute to institutions adopting inconsistent governance approaches. Through enhancing these studies, we aspire to bring forth more compelling evidence and insights to the field of RI.

Appendix

	SST	df	MSD	F-value	sig
Adjusted model	902.892	252	3.583	2.200	<0.001
Intercept	114.402	1	114.402	70.239	<0.001
Gender	1.843	1	1.843	1.132	0.288
PT	14.142	3	4.714	2.894	0.035
IL	6.127	5	1.225	0.752	0.585
AO	140.011	8	17.501	10.745	<0.001
TRM	2.814	3	0.938	0.576	0.631
PT*IL	35.902	12	2.992	1.837	0.040
PT*TRM	9.255	6	1.542	0.947	0.461
PT*AO	33.940	16	2.121	1.302	0.190
PT*Gender	1.876	3	0.625	0.384	0.765
IL*TRM	38.774	10	3.877	2.527	0.006
IL*AO	19.291	27	0.714	0.439	0.994
IL*Gender	8.079	5	1.616	0.992	0.422
TRM*AO	24.392	12	2.033	1.248	0.246
TRM*Gender	3.968	2	1.984	1.218	0.297
AO*Gender	2.474	6	0.412	0.253	0.958
Gender*PT*IL	40.666	8	5.083	3.121	0.002
Gender*PT*TRM	17.938	5	3.588	2.203	0.053
Gender*PT*AO	14.442	5	2.888	1.773	0.116
Gender*IL*TRM	17.854	6	2.976	1.827	0.092
Gender*IL*AO	1.978	4	0.494	0.304	0.876
Gender*TRM*AO	7.563	6	1.261	0.774	0.591

	SST	df	MSD	F-value	sig
PT*IL*TRM	54.294	14	3.878	2.381	0.003
PT*IL*AO	4.372	5	0.874	0.537	0.748
PT*TRM*AO	9.617	8	1.202	0.738	0.658
IL*TRM*AO	4.168	6	0.695	0.427	0.861
Standard Error	915.360	562	1.629		
N	815				
R-square	0.497				
Adjusted R-square	0.271				

Appendix 1. The results of robustness test

Note. IL: institutional level; PT: professional title; TRM: types of research misbehavior; AO: authorship order.

Statements and Declarations

Data availability statement

The data that support the findings of this study are available at: https://doi.org/10.5281/zenodo.10544715.

Ethical Considerations

As we have made our raw data publicly available, and the raw data involves human subjects, we have adhered to the principles of minimal harm and privacy protection in research. Specifically, all samples involving research misconduct have undergone institutional review or punishments. Disclosing their names and institutional information could potentially cause harm, so we have replaced their names and institutional details with numerical identifiers in the raw data. Additionally, we have randomized the order of sample collection to ensure that their personal information cannot be identified.

Footnotes

[1] NSFC is an organization similar to the National Science Foundation (NSF) in the United States and currently represents the highest level, with the most substantial funding and the strongest authority among the national project categories in Mainland China's natural fund programs.

[2] In Mainland China's administrative punishment system, criticism and education is a less severe form of punishment compared to a warning.

[3] The reason for this might be that in this classification standard, only tertiary hospitals are mandated to carry out higher education and research tasks. Consequently, personnel in such hospitals are more likely to be involved in research activities, leading to a higher occurrence of RM cases.

[4] 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.

[5] 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

- Akbulut, Yavuz, Serkan Şendağ, Gürkay Birinci, Kerem Kılıçer, Mehmet C Şahin, and Hatice F Odabaşı. 2008. "Exploring the Types and Reasons of Internet-Triggered Academic Dishonesty among Turkish Undergraduate Students: Development of Internet-Triggered Academic Dishonesty Scale (ITADS)." *Computers & Education* 51 (1): 463–73. https://doi.org/10.1016/j.compedu.2007.06.003.

- Ana-Catarina Pinho-Gomes, Carinna Hockham, and Mark Woodward. 2023. "Women's Representation as Authors of Retracted Papers in the Biomedical Sciences." *PLOS ONE* 18 (5): e0284403–3. https://doi.org/10.1371/journal.pone.0284403.

- Banerjee, Taraswi, Kathy Partin, and David B. Resnik. 2022. "Authorship Issues When Articles Are Retracted due to Research Misconduct and Then Resubmitted." *Science and Engineering Ethics* 28 (4). https://doi.org/10.1007/s11948-022-00386-1.

- 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.

- 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).

- Catano, Victor M, and James Turk. 2007. "Fraud and Misconduct in Scientific Research: A Definition and Procedures for Investigation." *Medicine and Law* 26 (3): 465–76.

- 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.

- Sarah Elaine Eaton. 2023. "Academic and Research Integrity as Transdisciplinary Fields of Scholarship and Professional Practice." *Handbook of Academic Integrity*, January, 1–18. 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.

- Galbraith, Kyle L. 2016. "Life after Research Misconduct." *Journal of Empirical Research on Human Research Ethics* 12 (1): 26–32. https://doi.org/10.1177/1556264616682568.

- 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.

- Greenwald, Anthony G., and Mahzarin R. Banaji. 1995. "Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes." *Psychological Review* 102 (1): 4–27. https://doi.org/10.1037/0033-295X.102.1.4.

- Haven, Tamarinde L., Joeri K. Tijdink, Brian C. Martinson, and Lex M. Bouter. 2019. "Perceptions of Research Integrity Climate Differ between Academic Ranks and Disciplinary Fields: Results from a Survey among Academic Researchers in Amsterdam." Edited by Daniele Fanelli. *PLOS ONE* 14 (1): e0210599. https://doi.org/10.1371/journal.pone.0210599.

- Helgesson, Gert, and Stefan Eriksson. 2017. "Responsibility for Scientific Misconduct in Collaborative Papers." *Medicine, Health Care and Philosophy* 21 (3): 423–30. https://doi.org/10.1007/s11019-017-9817-7.

- 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.

- Jones, Anne Hudson. 2003. "Can Authorship Policies Help Prevent Scientific Misconduct? What Role for Scientific Societies?" *Science and Engineering Ethics* 9 (2): 243–56. https://doi.org/10.1007/s11948-003-0011-3.

- Jordan, Augustus E. 2001. "College Student Cheating: The Role of Motivation, Perceived Norms, Attitudes, and Knowledge of Institutional Policy." *Ethics & Behavior* 11 (3): 233–47. https://doi.org/10.1207/s15327019eb1103_3.

- KERÄNEN, LISA. 2006. "Assessing the Seriousness of Research Misconduct: Considerations for Sanction Assignment." *Accountability in Research* 13 (2): 179–205. https://doi.org/10.1080/08989620500440261.

- Kuroki, Toshio. 2018. "New Classification of Research Misconduct from the Viewpoint of Truth, Trust, and Risk." *Accountability in Research* 25 (7-8): 404–8. https://doi.org/10.1080/08989621.2018.1548283.

- Lancet, The. 2015. "China's Medical Research Integrity Questioned." *The Lancet* 385 (9976): 1365. https://doi.org/10.1016/S0140-6736(15)60700-0.

- Lechner, Iris, Lidwine B Mokkink, Geert Ridder, R. van Woudenberg, Lex M Bouter, and Joeri K Tijdink. 2023. "The Core Epistemic Responsibilities of Universities: Results from a Delphi Study." *Accountability in Research*, September, 1–21. https://doi.org/10.1080/08989621.2023.2255826.

- Li, Dan, and Gustaaf Cornelis. 2020. "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.

- Liao, Qing-Jiao, Yuan-Yuan Zhang, Yu-Chen Fan, Ming-Hua Zheng, Yu Bai, Guy D. Eslick, Xing-Xiang He, Shi-Bing Zhang, Harry Hua-Xiang Xia, and Hua He. 2018. "Perceptions of Chinese Biomedical Researchers towards Academic Misconduct: A Comparison between 2015 and 2010." *Science and Engineering Ethics* 24: 629–45. https://doi.org/10.1007/s11948-017-9913-3.

- 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.

- LV, Xinwen. 2024. "Ministry of Education of the People's Republic of China: Conducting Self-Inspection of Retracted Papers in Universities." *The Paper*, January 4, 2024. https://www.thepaper.cn/newsDetail_forward_25897368.

- MOST. 2021. "The Joint Conference on Research Integrity Has Established a Mechanism for Reporting Research Integrity Cases and Publicly Disclosing the Investigation Results." MOST. August 23, 2021. https://www.most.gov.cn/zxgz/kycxjs/kycxgzdt/202108/t20210823_176553.html.

- 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.

- National Natural Science Foundation of China. 2022. "Procedures for Investigating and Dealing with Research Misconduct." 2022. https://www.nsfc.gov.cn/publish/portal0/tab475/info88329.htm.

- Ou, Guiyan, Mingliang Yue, Jiang Wu, and Tingcan Ma. 2021. "An Analysis of the Evolution of Research Collaboration according to the Academic Careers of Distinguished Researchers in Chemistry." *Journal of the China Society for Scientific and Technical Information* 40 (7): 756–67. https://doi.org/10.3772/j.issn.1000-0135.2021.07.008.

- Pain, Elisabeth. 2021. "How to Navigate Authorship of Scientific Manuscripts." *Science*, May. https://doi.org/10.1126/science.caredit.abj3459.

- Pratkanis, Anthony R, Steven James Breckler, and Anthony G Greenwald. 1989. *Attitude Structure and Function.* Hillsdale, N.J.: L. Erlbaum Associates.

- Rennie, D. 1997. "When Authorship Fails. A Proposal to Make Contributors Accountable." *JAMA: The Journal of the American Medical Association* 278 (7): 579–85. https://doi.org/10.1001/jama.278.7.579.

- 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.

- Stern, Andrew M, Arturo Casadevall, R Grant Steen, and Ferric C Fang. 2014. "Financial Costs and Personal Consequences of Research Misconduct Resulting in Retracted Publications." *ELife* 3 (August). https://doi.org/10.7554/elife.02956.

- Stewart, Camille L, James P De, Narjust Duma, Oliver Ralph, Kevin Choong, Lorena Gonzalez, Natalie C McClintock, Ethan Sandoval, Laleh G Melstrom, and Susanne G Warner. 2020. "Unconscious Bias in Speaker Introductions at a Surgical Oncology Meeting: Hierarchy Reigns over Gender." *Annals of Surgical Oncology* 27 (10): 3754–61. https://doi.org/10.1245/s10434-020-08906-8.

- Sugawara, Yuya, Tetsuya Tanimoto, Shoko Miyagawa, Masayasu Murakami, Atsushi Tsuya, Atsushi Tanaka, Masahiro Kami, and Hiroto Narimatsu. 2017. "Scientific Misconduct and Social Media:

Role of Twitter in the Stimulus Triggered Acquisition of Pluripotency Cells Scandal." *Journal of Medical Internet Research* 19 (2): e57. https://doi.org/10.2196/jmir.6706.

- Weber, Max, Colliot-ThélèneCatherine, and Jean-Pierre Grossein. 2000. *Confucianisme et Taoïsme.* Paris: Gallimard.

- Xi, Xinxue, Jing Mao, Tieying Zeng, and Dandan Ju. 2023. "Validation of Publication Pressure Questionnaire-Revised into Chinese Nurses from a University Hospital." *Nursing Open* 10 (12): 7713–24. https://doi.org/10.1002/nop2.2012.

- 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.

- Zeng, Junxia, Xiaopeng Pang, Linxiu Zhang, Alexis Medina, and Scott Rozelle. 2013. "Gender Inequality in Education in China: A Meta-Regression Analysis." *Contemporary Economic Policy* 32 (2): 474–91. https://doi.org/10.1111/coep.12006.

- Zhang, Fengfan, Zhenni Luo, Ting Chen, Rui Min, and Pengqian Fang. 2017. "Factors Affecting Turnover Intentions among Public Hospital Doctors in a Middle-Level City in Central China." *Australian Health Review* 41 (2): 214. https://doi.org/10.1071/ah15238.

Supplementary data: available at https://doi.org/10.32388/8WGIT6

Declarations