EconomiX

# Slaying the Undead: How Long Does It Take to Kill Zombie Papers?

Marc Joëts
Valérie Mignon

2025-7 Document de Travail/ Working Paper

EconomiX - UMR 7235 Bâtiment Maurice Allais
Université Paris Nanterre 200, Avenue de la République
92001 Nanterre Cedex

Site Web : economix.fr
Contact : secreteriat@economix.fr
Twitter : @EconomixU

cnrs
Université Paris Nanterre

# Slaying the Undead: How Long Does It Take to Kill Zombie Papers?*

Marc Joëts[†] Valérie Mignon[‡]

February 10, 2025

**Abstract**

This article examines the persistent impact of zombie papers, i.e., retracted or destined-for-retraction publications that continue to influence academic discourse through ongoing citations despite their discredited status. Relying on a large sample of 25,480 retracted research articles over the 1923-2023 period, we introduce an original methodological framework combining survival analysis with the innovative Zombie Population Decay Dynamics (ZPDD) model, a theoretical approach designed to simulate the long-term persistence and decay of zombie papers under various editorial interventions. We identify key factors affecting retraction timing and zombie paper persistence. Serious misconduct, such as data fabrication, significantly delays retractions, while geographic disparities exacerbate inefficiencies, with certain regions facing prolonged processes. Journal practices, such as open-access versus subscription-based models, also influence retraction dynamics, with subscription-based journals exhibiting faster corrective actions. Developing a mathematical optimization framework derived from our ZPDD model, we determine the most effective mix of editorial policies while balancing practical feasibility and intervention intensity. The findings highlight data transparency as the most impactful intervention for reducing zombie paper persistence, followed by enhanced plagiarism detection and reproducibility measures, such as pre-registration and replication studies. Overall, a well-balanced combination of targeted editorial interventions can substantially accelerate retraction processes and limit the detrimental influence of zombie papers on academic discourse.

*JEL Classification*: C18; C80; Z0.

*Keywords*: scientific retraction, zombie papers, survival analysis, accelerated failure time model, editorial policies.

*We would like to thank Christophe Hurlin for helpful comments and suggestions.
*Corresponding author*: Valérie Mignon, EconomiX-CNRS, University of Paris Nanterre, 200 avenue de la République, 92001 Nanterre cedex, France. E-mail: valerie.mignon@parisnanterre.fr

[†]IESEG School of Management; Univ. Lille, CNRS UMR 9221 - LEM Lille Economie Management F-59000 Lille, France. E-mail: m.joets@ieseg.fr

[‡]EconomiX-CNRS, University of Paris Nanterre and CEPII, Paris, France. E-mail: valerie.mignon@parisnanterre.fr

# 1 Introduction

Scientific progress depends on the reliability and validity of published research. However, the rise in retractions due to scientific misconduct or honest errors threatens the integrity of the academia. Retractions correct the scientific record but often come too late to prevent significant damage. Flawed knowledge from retracted studies can undermine the foundations of new research (Furman et al. (2012)), distort scholarly understanding (Necker (2014)), and affect the rate and direction of scientific exploration, shifting researchers' positions in intellectual space (Azoulay et al. (2015)). Alarmingly, erroneous findings can persist in academic discourse for years, compounding their detrimental effects (Houghton (2022)).

The underlying causes of research misconduct are deeply rooted in systemic academic pressures. The "publish or perish" culture compels some researchers to prioritize productivity over integrity, leading to data fabrication, falsification, and other forms of malpractice (see Karabag & Berggren (2012), Necker (2016), Cox et al. (2018), and Le Maux et al. (2019)). Although an increasing number of top-ranked academic journals are beginning to publish replication studies, the majority still do not, which reduces the risk of exposure and lowers the cost of engaging in fraudulent research (Anderson et al. (2008), Bergh et al. (2017), Pérignon et al. (2024)). Despite the seriousness of these challenges, quantifying scientific misbehavior is difficult due to researchers' strong incentives to hide misconduct (Necker (2014)). Although retractions act as a vital self-corrective mechanism, their timing is critical:[1] prolonged delays allow false science to proliferate, influencing citations and future research direction (Fang et al. (2012), Azoulay et al. (2015)). Yet, retraction frequency reflects detection efficiency more than the true prevalence of misconduct (Hesselmann et al. (2017)), underscoring the need for improved systems.

Among the most insidious consequences of delayed retractions is the phenomenon of "zombie papers", i.e., the fact that publications that have been retracted or warrant retraction yet continue to influence academic discourse through ongoing citations, despite their discredited status.[2] These papers challenge the core function of science as a cumulative endeavor and raise questions about the efficiency of editorial and institutional safeguards. Existing literature has identified critical factors affecting retraction dynamics, including journal characteristics, geographical disparities, and misconduct types (Fang et al. (2012), Horbach & Halffman (2019)). However, gaps remain in understanding how these factors interact over time and how editorial policies could be optimized to mitigate the persistence of zombie papers.

The present study tackles these issues and addresses four key gaps in the literature. First, it provides a detailed empirical analysis of the factors influencing retraction timing, including

[1] According to the Committee on Publication Ethics (COPE) guidelines (see COPE (2022)), retractions should be initiated for data fabrication, duplication, plagiarism, and ethical violations.

[2] A striking example is the article by Mehra et al. (2020), published in May 2020 and retracted in June 2020, which had been cited 1,872 times as of February 6, 2025.

journal characteristics, geographical disparities, and the type of misconduct. Our approach integrates multiple research domains, offering a global perspective beyond the scope of prior discipline-specific studies. Second, we quantify the marginal effects of these characteristics on retraction times, contextualizing their impact on academic publishing. Third, building on these insights, we develop a theoretical model, the Zombie Population Decay Dynamics (ZPDD), to simulate the long-term effects of various editorial policy interventions on zombies. Finally, we derive the optimal mix of policy interventions to minimize the persistence of zombie papers, balancing efficiency with feasibility.

The main innovation of this paper lies in its integration of survival analysis and theoretical modeling to examine the interplay between retraction timing and the persistence of zombie papers. By emphasizing the importance of timely retractions, we highlight the critical role of institutional efficiency in maintaining the integrity of the scientific record. Additionally, we develop a mathematical optimization framework derived from our ZPDD model to determine the most effective combination of editorial policies while balancing practical feasibility and intervention intensity. In doing so, we offer actionable policy recommendations to mitigate the detrimental effects of false science. In particular, to the best of our knowledge, our study is the first to develop a theoretical model of zombie paper dynamics grounded in empirical facts, enabling the simulation of targeted editorial policy interventions and their outcomes.

Relying on a large sample of 25,480 retracted research articles over the 1923-2023 period and using survival analysis and theoretical modeling, we identify key factors influencing retraction timing, including journal domains, geographical and institutional disparities, and reasons for retraction. Specifically, our main results can be summarized as follows. We find an average retraction time of approximately 1,045 days, with significant variability across disciplines and regions—systemic challenges characterizing certain regions leading to prolonged retraction times. The type of misconduct significantly influences retraction timing, with serious issues such as data falsification causing longer delays. By introducing the ZPDD model, we simulate the long-term impact of editorial policies, highlighting the potential of targeted measures such as enhanced data transparency and improved results' replication to reduce the persistence of zombie papers and the systemic propagation of erroneous knowledge. Specifically, by developing a mathematical optimization framework based on our ZPDD model, we identify the optimal combination of editorial policies that effectively balance practical feasibility and intervention intensity. Our findings reveal that data transparency is the most effective strategy for minimizing the persistence of zombie papers, followed by enhanced plagiarism detection and reproducibility initiatives, such as pre-registration and replication studies. Ultimately, a carefully calibrated mix of targeted editorial interventions can significantly expedite the retraction process and mitigate the harmful impact of zombie papers on scholarly discourse.

The remainder of this paper is organized as follows. Section 2 describes the data and methodology, including survival analysis models. Section 3 presents the empirical findings, emphasizing factors influencing retraction timing and zombie paper persistence. Section 4 develops the

ZPDD model to simulate editorial interventions and derive an optimal policy mix. Section 5 provides policy recommendations to enhance research integrity and minimize the persistence of zombie papers. Finally, Section 6 concludes the paper.

# 2 Data and empirical methodology: Hunting the undead

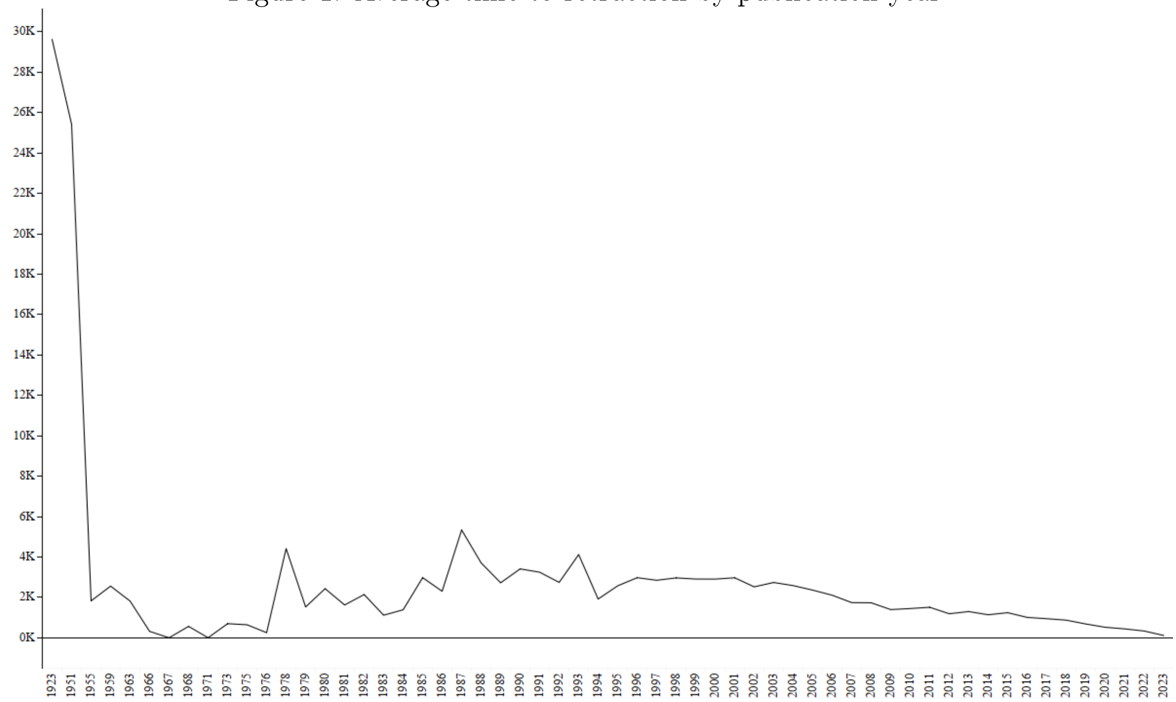
This section presents the database of retracted papers and key computed variables, highlights the characteristics of these retractions, and discusses preliminary findings. It then introduces the empirical models used in the survival analysis to examine the main factors influencing the time to retraction of zombie papers.

## 2.1 Zombies' database

To analyze zombie papers, we use the Retraction Watch Database (RWD), a comprehensive online repository of retracted scientific publications across disciplines. Covering the period from 1923 to 2023, the database includes over 48,824 records of retractions, spanning research articles, case studies, meta-analyses, books, chapters, and conference abstracts. Our analysis focuses on 25,480 retracted research articles, the most prominent form of academic communication, with each record's retraction status verified using Google Scholar. Metadata such as titles, journal names, publication and retraction dates, publishers, authors, institutions, countries, and retraction reasons were collected via the Crossref API. These data facilitate the construction of key variables, including time to retraction and explanatory factors such as journal domains, geographical zones, and retraction causes (see Table A.1 in Appendix A). The RWD's utility is further underscored by its ability to capture the dynamics of retraction dissemination through unofficial information channels, such as blogs and social media. As noted by Xu et al. (2023), these channels often achieve a broader reach than traditional, official sources, serving as intermediaries between journals and researchers. Unofficial channels, defined as platforms where content is not directly released by authorities, have several advantages: they aggregate information from multiple sources (Bushee et al. (2010)), provide enriched details about articles and authors (Drake et al. (2014)), and create new insights about scientific misconduct. By complementing traditional sources with unofficial channels, the RWD ensures a comprehensive and nuanced perspective on retraction events, making it an ideal resource for our study.

From the original database, we constructed our set of endogenous (time to retraction) and exogenous variables. To measure the time to retraction of zombie papers, we computed the difference in days between the retraction date and the publication date for each research article over the period from 1923 to 2023. As illustrated in Figure 1, the graph provides a snapshot of the time to retraction across publication years, showing a pronounced initial peak followed by a general decline and stabilization over the years. This trend suggests an improvement in retraction practices and/or possibly changes in publication standards over time. Table 1 presents descriptive statistics of retraction times, showing that the average time to retraction is approximately 1,045 days (nearly 3 years), but there is significant variability, with a stan-

Figure 1: Average time to retraction by publication year




Note: This figure displays the average time to retraction (in days) for each year of publication, highlighting temporal trends in retraction speed.

dard deviation of 1,225 days. While 25% of papers are retracted within 285 days, 75% are retracted within 1,387 days, and the median retraction time is 640 days, meaning that half of the papers are retracted within two years. However, some extreme cases take much longer, with the longest retraction occurring 81 years after publication. The high kurtosis (28.4) and positive skewness (3.3) indicate a distribution with many quick retractions but a long tail of outliers with significantly delayed retractions (see also Azoulay et al. (2015)). The time to retraction is critical in research development, as delays in retractions can have detrimental effects on the direction of scientific progress. Papers that are later retracted may continue to be cited, leading to the perpetuation of incorrect or unreliable findings, which can mislead future research efforts and waste valuable resources (Fang et al. (2012) and Azoulay et al. (2015)).

To analyze the intrinsic factors related to retraction times of zombie papers, several explanatory variables were created based on the original database. The construction of these variables is consistent with previous literature that highlights the importance of publication characteristics, author demographics, journal domains, and reasons for retraction in influencing retraction times (Fang et al. (2012), Azoulay et al. (2015), and Horbach & Halffman (2019)). For instance, the publication year has been shown to correlate with retraction timing due to in-

Table 1: Descriptive statistics of retraction times for zombie papers

| | Retraction times (days) |
|---|---|
| Count | 25,480 |
| Mean | 1,045.4 |
| Standard deviation | 1,225.9 |
| Min | 1 |
| 25th percentile | 285 |
| 50th percentile | 640 |
| 75th percentile | 1,387 |
| Max | 29,622 |
| Kurtosis | 28.4 |
| Skewness | 3.3 |

Note: This table presents basic descriptive statistics of retraction times for zombie papers.

creasing scrutiny over time and evolving publication standards (Steen (2011) and Fang et al. (2012)). This growing scrutiny over time justifies the inclusion of publication year as a key variable in our analysis of retraction times.

For journal characteristics, we created several dummy variables capturing the domain of each journal. This approach is supported by research showing that retraction rates and times can vary significantly across scientific disciplines (Budd et al. (1998)). Using resources from both Web of Science and Scimago, we matched journals with key domains, covering 24 areas, each represented by a dummy variable (see Table A.2 in Appendix A for the list of domains and occurrences). Some of the major domains include "Medicine" (17.1%), "Biochemistry, Genetics, and Molecular Biology" (19.6%), and "Computer Science, Data Science, Information Systems, and Robotics" (11.1%). Other areas, such as "Geography" (0.02%) and "Political Science" (0.2%), have significantly fewer occurrences, reflecting the diversity in journal domain representation. The inclusion of these areas allows us to capture domain-specific effects on retraction rates and patterns.[3] We thus define the following variable:

$$Area_i = \begin{cases} 1 & \text{if the zombie paper } z \text{ is published by journals in domain } i \text{ for } i = 1, ..., 24 \\ 0 & \text{otherwise.} \end{cases}$$

where $Area_i$ denotes the domain $i$ among the list of journal domains reported in Table A.2 in Appendix A. As the relationship between time to retraction and journal impact factor is modest, we decided not to consider impact factor as an exogenous variable (see Fang et al. (2012)).

For country characteristics of the authors, prior studies have demonstrated that geographic location can influence retraction outcomes, reflecting differences in research oversight, institutional practices, and integrity (Horbach & Halffman (2019)). Using Web of Science, we identified the corresponding author (or the first author) and extracted the country location

[3]As a robustness check, we also applied clustering methods to confirm our domain classification.

of their institution as referenced in the retracted paper. We then created 12 dummy variables covering different geographical locations, capturing the diversity in author locations and their potential influence on retraction patterns. The most represented regions include Asia (60.4%), North America (10.1%), and East Europe (9.2%), while regions such as Central America (0.7%) and Oceania (1.4%) have notably fewer occurrences. This geographical classification, supported by Wagner et al. (2017), enables a nuanced analysis of how institutional and geographical factors influence research quality and retraction likelihood across distinct regions. Specifically, we consider the following variable:

$$Geo_k = \begin{cases} 1 & \text{if } z \text{ is published by authors from geographical location } k \text{ for } k = 1, ..., 12 \\ 0 & \text{otherwise.} \end{cases}$$

where $Geo_k$ denotes the geographical location $k$ from the list reported in Table A.3 in Appendix A. We also captured whether journals were subscription-based (841) or open-access (24,659) using a paywalled dummy variable, as access type may influence both publication speed and retraction timing (Tennant et al. (2016)).

To account for the possibility that co-authors may come from different locations, we created a dummy variable to measure international collaboration. This reflects the growing significance of cross-border research and its impact on research quality and integrity (Cummings & Kiesler (2005), Wagner et al. (2017)). As collaboration increases, the dynamics of retraction can shift. Fanelli (2009) and Bozeman & Boardman (2014) highlight that larger, geographically diverse teams often face greater scrutiny, which can influence the likelihood of retraction. Additionally, we capture the number of authors for each zombie paper, as collaboration has been shown to affect both research transparency and the probability of retraction (Fanelli (2009)).

From publisher information, we created two dummy variables. First, we identified whether the publisher is among the top 10 largest in the world, as large publishers often follow different retraction practices. The list of the top 10 publishers is available in Table A.4 in Appendix A, with 2022 as the reference year.[4] Second, given the significant attention to predatory publishers, who tend to bypass proper peer review and quality control (Beall (2013), and Moher et al. (2017)), we merged information from Beall's List and Predatory Journals Reports to create a dummy variable indicating whether a publisher is predatory.[5]

With regard to retraction reasons, it is widely acknowledged that not all retraction causes have an equivalent impact on scientific knowledge. Fabrication (the invention of data or cases), falsification (the distortion of data and/or results), and plagiarism are generally recognized by many institutions as serious forms of scientific misconduct (LaFollette (2000), Smith (2000), and Fanelli (2009)). Plagiarism, however, differs qualitatively from fabrication and falsification. While it does not distort scientific knowledge itself, it still has significant

[4]The information is sourced from https://www.peeref.com.

[5]Sources: https://beallslist.net and https://predatoryjournals.org/the-list. The full list of predatory publishers included in our database is available upon request.

consequences for the careers of those involved and ultimately impacts the integrity of the scientific enterprise (Steneck (2006)). To systematically capture these distinctions, we created ten dummy variables, grouping related terms as recommended by Retraction Watch's taxonomy.[6] These categories include: "data issues" (11,556 occurrences), "results issues" (8,411), "authorship issues" (1,882), "peer review/editorial issues" (7,030), "duplication/plagiarism issues" (6,913), "referencing issues" (2,787), "withdrawal" (1,125), "miscommunication" (1,983), "non-reportable" (2,572), and "various ethical violations" (4,251) (refer to Tables A.1 and A.5 in Appendix A for further details). Additionally, we created two dummy variables to indicate whether a post-publication investigation was conducted by the publisher/journal (10,013 instances) or by an external organization, such as a company, institution, or third party (6,197 instances).

The database spans publications from 1923 to 2023, including 5,755 journals, 1,281 publishers, 82,226 authors, and 164 countries, providing a comprehensive view of retracted articles over time. The full list of variables is detailed in Table A.5 in Appendix A.

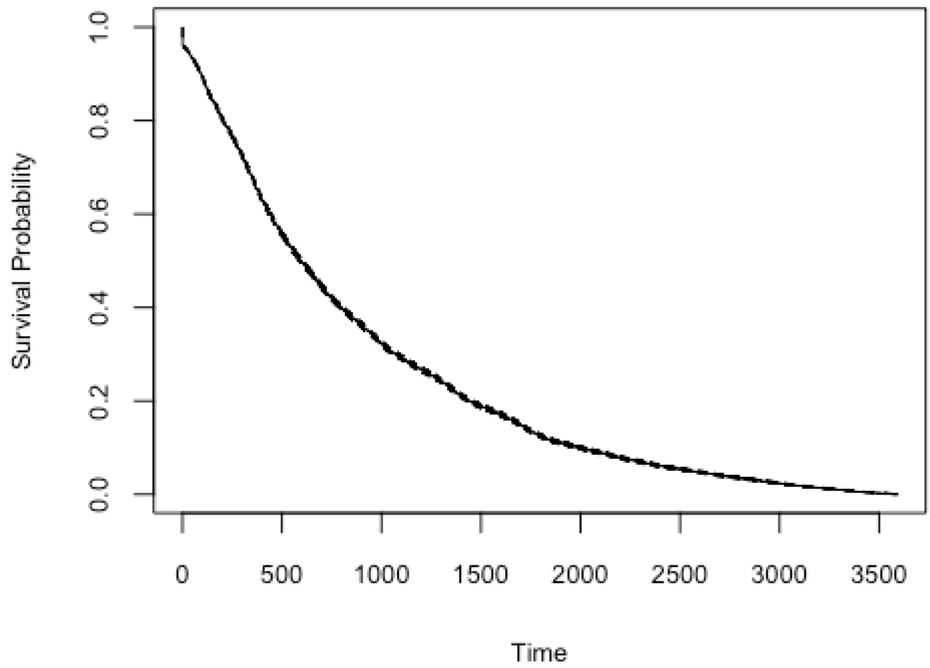
## 2.2 Survival times and empirical methodology

To model the survival time of zombie papers, we employed a range of approaches, including non-parametric, semi-parametric, and parametric models, to determine the most effective method for analyzing time to retraction. As a preliminary step, for descriptive purposes and to approximate the distribution of retraction times, we applied the Kaplan-Meier (KM) non-parametric method (see Kaplan & Meier (1958)). This method does not assume any specific model or distribution for survival times and does not account for covariates. Although limited in its application to our context, the KM approach provides valuable initial insights that inform the development of more sophisticated models. Figure 2 illustrates the estimated survival function using the KM method. The curve shows a steady decline in survival probability over time, suggesting that the likelihood of a paper remaining unretracted decreases as time passes. The steep drop in the early stages indicates that a significant number of retractions occur within the first few years following publication. After this initial decline, the curve flattens but continues to decrease, reflecting a slower retraction rate over time. This preliminary analysis highlights that most zombie papers are retracted relatively early in their lifespan, but a subset persists for much longer periods before being retracted.

Since our primary goal is to identify the factors influencing the time to retraction, we go beyond the KM approach and explore two commonly used classes of models: proportional hazards models and non-proportional alternatives. We first implement the semi-parametric Cox proportional hazards model (Cox (1972)), which expresses the hazard rate as a function of covariates:

[6] The complete list of retraction reasons can be found at https://retractionwatch.com/retraction-watch-database-user-guide/retraction-watch-database-user-guide-appendix-b-reasons/. More details on our categorization process are available upon request.

Figure 2: Survival time of zombie papers: Non-parametric Kaplan-Meier method




Note: This figure presents the survival time of zombie papers using the Kaplan-Meier non-parametric method.

$$h(t|X) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p) \tag{1}$$

where $h(t|X)$ represents the hazard function of time to retraction, conditional on the covariates $X$, which are discussed in Section 2.1 and detailed in Appendix A. The term $h_0(t)$ is the baseline hazard function, and $\beta_j$ $(j = 1, ..., p)$ denotes the vector of coefficients corresponding to the covariates $X$.

The Cox model assumes proportional hazards, meaning that the hazard ratios between groups remain constant over time. However, as shown in Table B.1 in Appendix B, this assumption does not hold for our dataset of zombie papers. To address this issue while maintaining the Cox model framework, we employed the weighted Cox proportional hazards model, which is specifically designed to handle non-proportional hazards. This model accounts for time-varying effects by applying weights to the covariates, offering greater flexibility in estimating hazard ratios without the need for interaction terms. It produces unbiased average hazard ratios even when the proportional hazards assumption is violated. Compared to models with complex interaction terms, the weighted Cox model remains computationally efficient and provides a robust alternative for managing non-proportional covariates (see Schemper et al. (2009); Dunkler et al. (2018)). Given the large number of covariates in our analysis, it offers a more feasible and interpretable solution.

Another well-established alternative, which departs from the proportional hazards framework, is the Accelerated Failure Time (AFT) parametric model (see Wei (1992)). The AFT model directly estimates survival time, rather than hazard rates, within a non-proportional framework. The AFT model is defined as:

$$log(t) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \sigma\epsilon \quad (2)$$

In this equation, $\log(t)$ represents the log-transformed survival time, $\beta_0$ is the intercept, $\beta_j$ ($j = 1, ..., p$) are the covariate coefficients, $\sigma$ is the scale parameter, and $\epsilon$ is an error term which follows a specific distribution (Weibull, Log-Normal, or Gaussian, in our analysis).

To account for potential non-linear relationships between covariates and survival times, we also implemented a random forest survival model, a machine learning technique capable of capturing complex interactions. Additionally, to address unobserved heterogeneity or random effects, we estimated Cox frailty models (see Therneau et al. (1990), and Wienke (2010)), which incorporate random effects into the Cox model to better account for clustering or unmeasured covariates.

After excluding outliers to ensure the robustness of our analysis, resulting in a final sample of 24,332 observations, we set "Area23: Geography", "Withdrawn", and "Central America" dummies as reference categories.[7] Multicollinearity was checked (see Figure A.1 in Appendix A), and we compared the following models: standard Cox, Cox-Frailty, weighted Cox, AFT-Weibull, AFT-Log Normal, AFT-Gaussian, and random forest. A detailed comparison and sensitivity analysis (see Table B.2 in Appendix B) evaluated the models using key metrics: (i) Bayesian Information Criterion (BIC), (ii) Concordance Index (C-index, Harrell et al. (1982)), (iii) Brier Score (Brier (1950)), (iv) Integrated Absolute Error (IAE, Graf et al. (1999)), and (v) Integrated Squared Error (ISE, Graf et al. (1999)). The results consistently show that the AFT-Weibull model outperforms the others in accurately capturing the survival time of zombie papers, establishing it as the core model of this study.

## 3 Empirical results: The clock ticks on the zombies

This section begins by presenting the empirical results of the AFT-Weibull model across all retractions, disaggregated by total, early-stage (under 2 years), mid-stage (2 to 5 years), and late-stage (over 5 years) retractions. These findings provide an overview of how different covariates influence the retraction timing across various stages. Following this, we examine the marginal effects, offering a practical interpretation of the AFT-Weibull model coefficients. The marginal effects shed light on the relative impact of each covariate on the retraction timing, allowing for a clearer contextualization of how specific factors accelerate or delay retraction at different stages.

[7] These dummies were chosen as baseline categories due to their smaller representation in the sample. All results should be interpreted as differences relative to these baselines.

### 3.1 Survival times of zombies

We model the survival time of zombie papers using the AFT model, where the error term $\epsilon$ in Equation (2) follows a Weibull distribution, and the scale parameter is defined as $\sigma = 1/\lambda$. $\lambda$ plays a critical role in interpreting survival times: if $\lambda < 1$, survival times are compressed, indicating faster retractions, while $\lambda > 1$ suggests stretched survival times, or slower retractions. Although our primary focus is on the total sample, previous studies have highlighted distinct retraction patterns across different time periods due to variations in editorial practices and external pressures (e.g., Steen (2011), Fang et al. (2012), Azoulay et al. (2015)). Accordingly, we examine early, mid, and late-stage retractions, defined as less than 2 years, between 2 and 5 years, and more than 5 years, respectively. The results of the AFT-Weibull model are presented in Table 2.[8] To facilitate interpretation, all coefficients (excluding the intercept) are presented in exponential form and should be understood as percentage changes, calculated as $(\exp^{\beta_j} - 1) \times 100$. A coefficient less than 1 indicates a percentage decrease in time to retraction, while a coefficient greater than 1 refers to an increase.

In our total model (first column), the estimated scale parameter $\lambda$ is 0.913, which is less than 1. This suggests that retractions of zombie papers tend to occur more rapidly over time, reflecting an increasing hazard rate. This acceleration may be attributed to enhanced scrutiny enabled by digital archives and systematic reviews of historical publications. Additionally, the proactive efforts of journals and organizations such as Retraction Watch have improved the detection and publicity of misconduct or errors, increasing the likelihood of retractions even for older publications. The baseline (intercept) coefficient of 152 days implies that, on average, a zombie paper in the "Geography" journal domain, authored by a corresponding author from "Central America" and retracted for "Withdraw," is retracted after 152 days. This result is statistically significant at the 1% level and serves as the reference point for interpreting the effects of other covariates.

Regarding publication year, the coefficient of 0.930 ($< 1$) indicates that more recently published papers are retracted faster, corresponding to a 7% reduction in time to retraction per additional year. This finding is highly significant ($p < 0.001$) and aligns with previous research (Fang et al. (2012), Azoulay et al. (2015)) attributing this trend to heightened scrutiny, the rise of post-publication peer review, and evolving journal standards. These results extend across all domains of science, reflecting increased vigilance within the academic community in addressing problematic publications.

The journal characteristics reveal interesting patterns. Papers published in paywalled journals have a coefficient of 0.708, indicating they are retracted 29.2% faster than those in open-access journals. This finding aligns with studies (Tennant et al. (2016)) suggesting that subscription-based journals often impose stricter editorial standards and have better mechanisms to detect errors early. However, no statistically significant differences are observed across journal do-

[8] Our findings remain robust, as demonstrated through the bootstrap AFT-Weibull estimates presented in Table B.3 in Appendix B.

mains compared to the baseline, suggesting that retraction times are more influenced by editorial policies than by disciplinary differences (Horbach & Halffman (2017)).

Significant regional effects are evident. Papers from Asia (0.985), South America (0.708), North Europe (0.762), and Central Europe (0.737) are retracted faster than the baseline region (Central America), with time reductions up to 29.2%. These differences likely reflect variations in institutional oversight, research integrity practices, and journal quality, consistent with prior studies (Fang et al. (2012), Horbach & Halffman (2017)). In contrast, papers from Eastern Europe (1.506) take 50.6% longer to retract, possibly due to weaker institutional frameworks or less stringent journal policies. The coefficient for the log of the number of authors (1.086) indicates that papers with more authors take approximately 8.6% longer to retract. This delay likely arises from the complexities of coordinating investigations and resolving disputes among larger author teams, often involving multiple institutions, which can slow the retraction process.

The coefficient for predatory publishers (0.762) indicates that papers in such journals are retracted approximately 23.8% faster than those in reputable venues. While this may seem counterintuitive, predatory journals often publish papers with evident deficiencies that are more easily detectable by external parties, such as watchdog organizations or institutions, leading to quicker retractions when misconduct is identified. Additionally, the lack of formal editorial procedures in predatory journals may result in a less structured and faster retraction process (Beall (2013), and Moher et al. (2017)).

Retraction reasons exhibit distinct patterns in terms of their impact on retraction timelines. Papers retracted for “Data Issues” (1.509) and “Plagiarism” (1.125) take significantly longer to retract than those withdrawn for unspecified reasons (“Withdraw”), increasing retraction times by 50.9% and 12.5%, respectively. This delay reflects the complexity of investigating these cases, as they often involve verifying raw data, replicating analyses, or confirming instances of misconduct. “Results Issues” (1.038) and “Miscommunication” (1.033) result in a moderate (3.8% and 3.3%, respectively) increase in retraction time, possibly due to the need for detailed reviews of analytical errors or incorrect conclusions for the former, and because such cases are often procedural and easier to address administratively for the latter. Procedural issues like “Peer Review Problems” (0.969) and “Referencing Errors” (0.834) are resolved more quickly, with reductions in retraction times of 3.1% and 16.6%, respectively, reflecting their less investigative nature. Papers retracted with no specific reason provided (“None”, 0.644) are resolved 35.6% faster than the baseline, likely due to the absence of complex or contentious disputes. These patterns suggest that the severity and complexity of the retraction reason play a significant role in determining the time to resolution.

Investigations conducted by journals or publishers (1.159) significantly increase retraction times, as these processes often involve detailed fact-finding and institutional collaboration (Steen et al. (2013)). These findings underscore the complexity of the retraction process when formal investigations are initiated.

Overall, the AFT-Weibull model confirms established trends in the literature, such as faster retractions for administrative or technical errors and slower retractions for misconduct cases. However, our study extends the analysis by systematically examining the impact of geographical regions, journal types, and retraction reasons on retraction times. These findings underscore the role of editorial policies and institutional practices in shaping the retraction process and provide a foundation for exploring potential editorial interventions in subsequent sections.

In addition to the total model, we conducted a sensitivity analysis by segmenting the data into three retraction time windows: less than 2 years (second column), 2 to 5 years (third column), and more than 5 years (fourth column). This segmentation allows us to observe how retraction dynamics evolve over time and identify time-specific patterns. The analysis reveals important temporal differences and similarities when compared to the total column, offering deeper insights into retraction behavior over time. The scale parameter ($\lambda$) highlights the periods of 2 to 5 years and more than 5 years as having the highest speed of retraction, with values of 0.230 and 0.168, respectively, compared to the total model (0.913). This indicates that these periods experience accelerated retractions, possibly due to improved detection mechanisms and growing pressure on journals to retract problematic papers as they become more scrutinized over time.

While the publication year coefficient remains consistent across all time periods indicating faster retractions for recently published papers, results for journal domains reveal significant deviations in the segmented periods compared to the total column. For example, “Computer Science; Data Science; Information Systems; and Robotics” exhibits significantly slower retractions in the 2 to 5 years window, with a 63.9% longer retraction time compared to “Geography.” Similarly, “Dentistry” shows a 19.7% longer retraction time in the more than 5 years period. These results contrast with the total column, where journal domains were largely insignificant, suggesting that disciplinary differences in addressing older or mid-aged publications become more pronounced over time. The geographical location appears less influential overall in the segmented models, with fewer significant regional effects compared to the total column. However, papers from Eastern Europe consistently take longer to retract, with significant delays in both the 2 years (+47.8%) and 2 to 5 years (+11.1%) periods. This persistent trend underscores systemic challenges in retraction practices within the region, particularly for more recent publications. The reasons for retraction exhibit relative consistency across time periods. Serious scientific misconduct, such as “Data Issues”, “Results Issues”, and “Plagiarism”, continues to result in significantly longer retraction times, reflecting the complexity of investigations. For example, “Data Issues” increase retraction times by 10.3% within the 2 years window and 4.1% in the more than 5 years period, while “Results Issues” shows significant delays of 5.8% in 2 years. These findings align with the total column, confirming that the severity of the issue plays a critical role in determining retraction timelines.

Overall, while some patterns—such as the effect of publication year and reasons for retraction—remain stable over time, significant differences emerge in journal domains and geographical regions, particularly for older publications. This temporal analysis sheds light on the evolving dynamics of retraction processes, emphasizing the need to account for time-specific factors when assessing retraction practices.

### 3.2 Marginal effects of zombie papers

To contextualize our estimated coefficients from the AFT-Weibull model, we compute the marginal effects of selected covariates on the time to retraction of zombie papers. This allows us to quantify how changes in each covariate affect retraction times in academic publishing, highlighting the magnitude of these impacts. Figures 3 and 4 display the marginal effects of retraction issues over time across different geographical regions and journal domains, respectively, for various retraction stages: total sample, early-stage (2 years), mid-stage (2 to 5 years), and late stage (over 5 years). For simplicity, we focus on serious forms of scientific misconduct, including data fabrication, results falsification, and plagiarism, as emphasized by LaFollette (2000), Smith (2000), and Fanelli (2009). Additional results for other variables are available upon request.

Figure 3 illustrates how data, results, and plagiarism issues affect predicted retraction times across significant regions, including "South America", "Eastern Europe", and the "Middle East", over various time frames. Across these regions, a consistent pattern emerges between 1980 and 2020, with retraction times generally decreasing over the decades, likely reflecting gradual improvements in editorial standards and retraction processes. However, the presence of retraction issues (blue lines) leads to consistently higher predicted retraction times compared to the absence of issues (red lines), with the magnitude of these effects varying by region and issue type. In regions such as "Eastern Europe" and the "Middle East", retraction times remain substantially higher when data fabrication and plagiarism issues are present. This suggests that in regions where retraction processes are historically slower, these types of misconduct contribute to prolonged retraction delays. For example, in Eastern Europe, the gap between retractions with and without issues persists over time, indicating that improvements in retraction speed may not fully mitigate the effects of serious misconduct. Conversely, "South America" exhibits a pronounced downward trend in retraction times, suggesting notable advancements in editorial standards and response mechanisms. However, the presence of issues still leads to a delay, particularly in the early and mid-stages, underscoring the importance of early detection mechanisms in regions where editorial robustness is evolving.

Taking the example of "Economics, Finance, Econometrics", Figure 4 presents the marginal effects of retraction issues within specific journal domains over the mid-range retraction period (2 to 5 years).[9] In domains where empirical rigor and data integrity are critical, such as in

[9] Results for other journal domains are available upon request to the authors.

Table 2: Survival time of zombie papers: AFT-Weibull model

| | Total | 2-years | 2 to 5-years | >5-years |
|---|---|---|---|---|
| Baseline (intercept) | 152.0*** | 27.75*** | 25.83*** | 36.90*** |
| **Time period** | | | | |
| Year publication | 0.930*** | 0.989*** | 0.990*** | 0.985*** |
| **Journal characteristics** | | | | |
| Computer Science; Data Science, Information Systems; Robotics | 1.481 | 0.828 | 1.639** | 1.084 |
| Pharmacology, Toxicology and Pharmaceutics | 1.149 | 0.722 | 1.476* | 1.150** |
| Medicine | 1.119 | 0.669 | 1.495* | 1.157** |
| Engineering | 1.347 | 0.759 | 1.449 | 1.129** |
| Biochemistry, Genetics and Molecular Biology | 1.311 | 0.730 | 1.465* | 1.163*** |
| Chemistry engineering; Chemistry; Material Sciences | 1.259 | 0.666 | 1.422 | 1.173*** |
| Physics | 0.954 | 0.659 | 1.440 | 1.150** |
| Neuroscience | 1.209 | 0.676 | 1.493* | 1.188*** |
| Earth and Planetary Sciences; Environmental Science; Ecology; Energy | 0.853 | 0.556 | 1.505* | 1.081 |
| Arts and Humanities; History | 1.328 | 0.811 | 1.538* | 1.099 |
| Immunology; Microbiology; Virology | 1.062 | 0.646 | 1.418 | 1.164** |
| Business; Management; Accounting | 1.324 | 0.708 | 1.577** | 1.079 |
| Sociology; Anthropology; Ethnology | 1.258 | 0.693 | 1.557* | 1.071 |
| Multidisciplinary | 1.409 | 0.736 | 1.477* | 1.174*** |
| Psychology | 1.200 | 0.700 | 1.513* | 1.204*** |
| Mathematics | 1.126 | 0.699 | 1.605** | 1.059 |
| Agricultural and Biological Sciences | 0.995 | 0.635 | 1.435 | 1.175*** |
| Economics; Finance; Econometrics | 1.222 | 0.824 | 1.582** | 1.112 |
| Dentistry | 1.343 | 0.767 | 1.414 | 1.197*** |
| Veterinary | 1.095 | 0.723 | 1.579* | 1.155 |
| Education | 1.158 | 0.784 | 1.473* | 1.047 |
| Law; Administrative Science; Military Science | 1.348 | 0.825 | 1.536* | 1.055 |
| Political Science | 1.509 | 0.807 | 1.462 | 1.040 |
| Paywalled | 0.708*** | 1.088** | 0.938*** | 0.962*** |
| **Authors information** | | | | |
| Asia | 0.985** | 1.113* | 1.003 | 1.043 |
| Oceania | 0.800 | 0.985 | 0.944 | 1.108** |
| Middle East | 1.136*** | 1.154** | 1.025 | 1.121*** |
| Africa | 0.792 | 1.089 | 0.985 | 1.006 |
| North America | 0.820 | 1.025 | 0.976 | 1.056* |
| South America | 0.708*** | 0.843** | 1.042 | 1.024 |
| North Europe | 0.762*** | 1.083 | 1.048 | 1.013 |
| Central Europe | 0.737** | 0.946 | 0.956 | 1.008 |
| South Europe | 0.919 | 0.969 | 1.012 | 1.048 |
| West Europe | 0.827 | 0.970 | 0.995 | 1.044 |
| East Europe | 1.506*** | 1.478*** | 1.111*** | 0.988 |
| International collaboration | 0.998 | 1.006 | 1.007 | 1.011 |
| log(number authors +1) | 1.086*** | 1.055*** | 1.023*** | 1.012 |
| **Publishers characteristics** | | | | |
| Top 10 publishers | 1.045*** | 1.019 | 0.991* | 1.035*** |
| Predatory | 0.762*** | 0.963* | 0.949*** | 0.959*** |
| **Reasons of retraction** | | | | |
| Data | 1.509*** | 1.103*** | 1.046*** | 1.041*** |
| Results | 1.038*** | 1.058*** | 1.018*** | 0.985* |
| Authorship | 1.023 | 1.042 | 0.995 | 1.027* |
| Peer | 0.968* | 0.969 | 0.983* | 1.038*** |
| Plagiarism | 1.125*** | 0.918*** | 1.019** | 1.039*** |
| References | 0.834*** | 1.139*** | 0.904*** | 0.949** |
| Miscommunication | 1.033* | 1.023 | 1.037*** | 0.982 |
| Ethics | 1.002 | 1.042** | 1.000 | 0.993 |
| None | 0.644*** | 0.545*** | 1.006 | 0.988 |
| **Investigation** | | | | |
| Journal/Publishers | 1.159*** | 1.182*** | 0.993 | 1.006 |
| Company/Institutions/Third Party | 0.999 | 0.977 | 0.999 | 0.986 |
| $\lambda$ (scale) | 0.913 | 0.837 | 0.230 | 0.168 |
| Bayesian Information Criteria | 371891 | 186520 | 107418 | 44771 |
| Sample size | 24,332 | 13,903 | 7,493 | 2,936 |

Note: This table presents the estimated coefficients in exponential form for the covariates affecting the time to retraction of zombie papers across the total sample, as well as early, mid, and late stages of retraction. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Economics and Finance, data fabrication and results falsification (blue lines) lead to significantly longer retraction times compared to cases without these issues (red lines), as shown by the wider confidence intervals. This effect is especially pronounced in fields heavily reliant on quantitative data, suggesting that these domains face unique challenges when addressing retraction-worthy misconduct. The gap between retraction times with and without issues in Economics and Finance indicates persistent delays when dealing with severe problems, emphasizing the need for more stringent checks and faster responses in these fields.

In summary, these marginal effects plots provide a nuanced view of how retraction issues and regional or domain characteristics influence retraction times over the years. The steady decline in retraction times between 1980 and 2020 across most regions and domains suggests that editorial practices and retraction mechanisms have improved globally. However, the impact of serious issues like data fabrication and plagiarism remains significant, particularly in regions and domains with specific vulnerabilities. This analysis supports the development of a theoretical model by identifying key areas where policy interventions could be most effective in reducing retraction times. In particular, fields and regions where issues such as data fabrication and plagiarism are prevalent would benefit from targeted policies focused on early detection and rapid response mechanisms. By integrating these empirical insights, the theoretical model can simulate editorial interventions to reduce the persistence of zombie papers, addressing both immediate retraction needs and longer-term improvements in publication standards. These empirical findings lay the groundwork for our theoretical model, detailed in Section 4, which explores the dynamics of zombie paper persistence and decay under various editorial interventions.

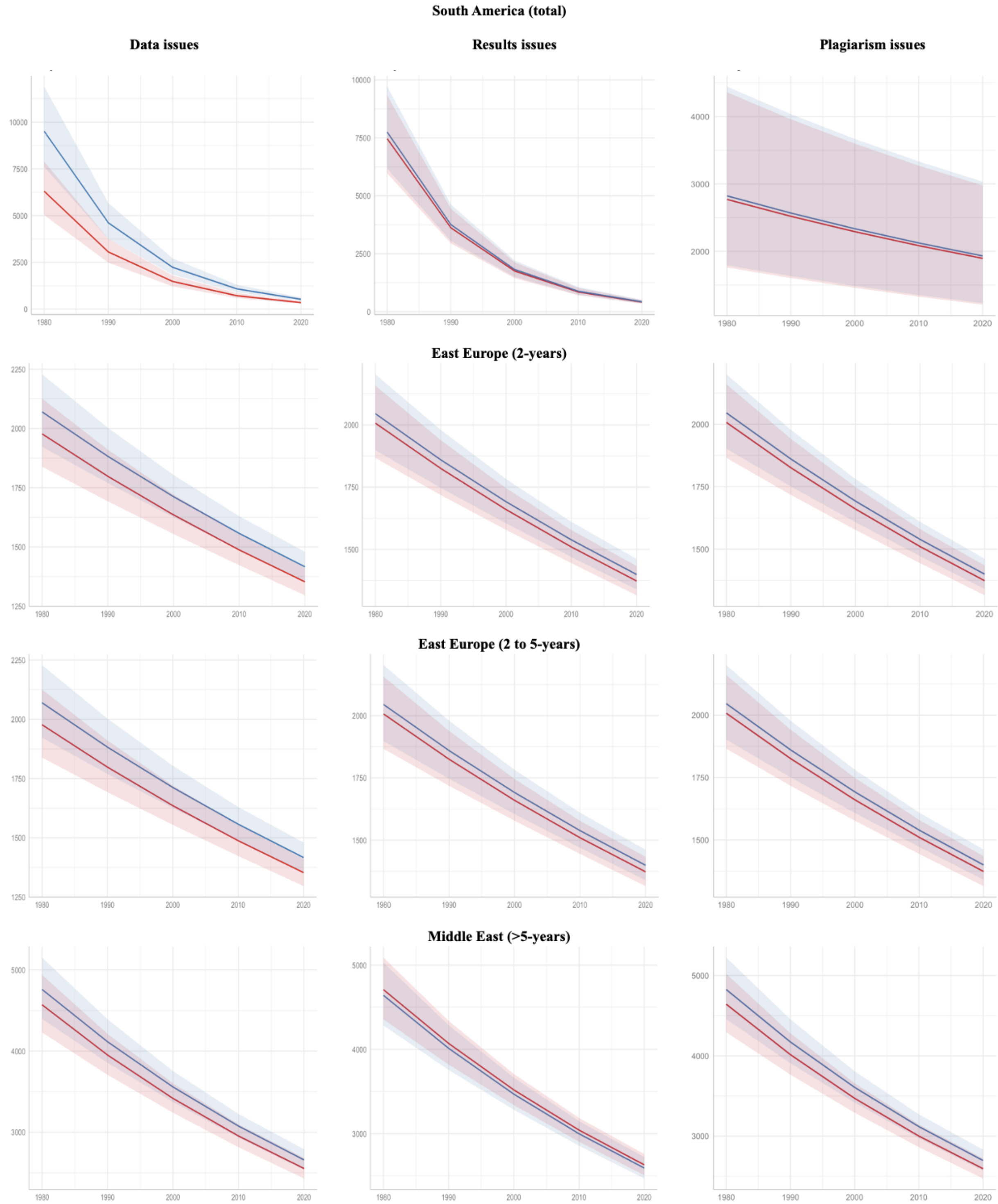
# 4 Theoretical model and simulation: Turning back the clock on the zombies

This section introduces a theoretical population framework to model the dynamics of the entire zombie paper population. Additionally, it presents a series of simulated experiments to assess the impact of editorial interventions on the persistence of zombie papers.

## 4.1 Theoretical model: Zombies Population Decay Dynamics (ZPDD)

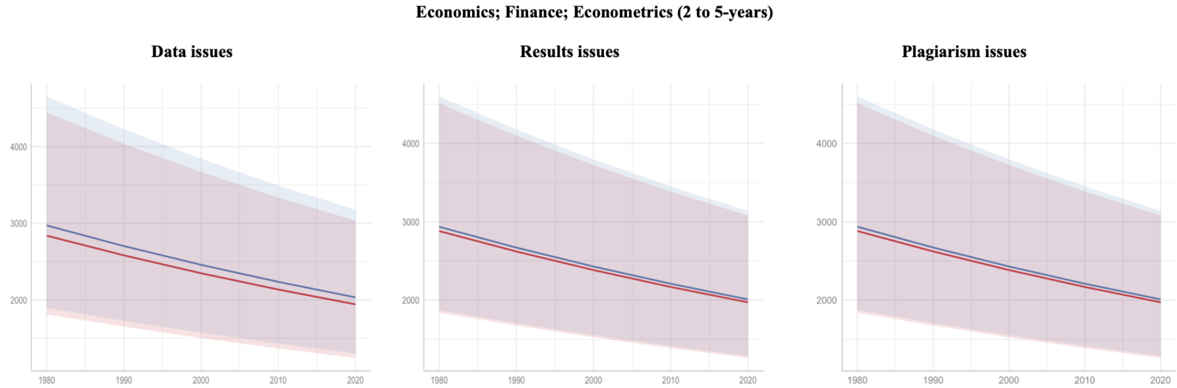
To simulate the long-term effects of editorial policy interventions on the persistence and decay of zombie papers, we adopt a theoretical population dynamics framework, traditionally used in ecological modeling, and apply it to the context of science and research integrity. This approach, which we term Zombie Population Decay Dynamics (ZPDD), conceptualizes the body of zombie papers as a population subject to external pressures, specifically, editorial policies. In doing so, we can analyze how these interventions influence the broader scientific ecosystem, capturing not only individual retraction events (akin to the marginal effects discussed in the previous section) but also the system-wide evolution of scientific knowledge. Population dynamics models are widely employed in both ecological and social sciences to examine how

Figure 3: Marginal effects of retraction issues by geographical region




Note: This figure presents the marginal effects of various retraction issues—such as data, results, and plagiarism—over time across significant geographical regions. Results are shown for different time scales (total, 2 years, 2 to 5 years, and over 5 years). The absence of the issue (red) is compared with its presence (blue), with 95% confidence bands included.

Figure 4: Marginal effects of retraction issues by journal domain




Note: This figure shows the marginal effects of various retraction issues—such as data, results, and plagiarism—over time for the "Economics, Finance and Econometrics" domain. Results are presented for the mid-range retraction period (2 to 5 years). The absence of each issue (red) is compared with its presence (blue), with 95% confidence bands provided.

populations respond to varying external pressures and interventions (Tilman et al. (2014), Bruch & Atwell (2015)), and some studies indicate that such models can also provide valuable insights into the dynamics of scientific knowledge and research retractions (Bettencourt et al. (2008), Vitanov & Ausloos (2012)). The derivation and illustration of the model is provided in Appendix C.1.

The KM survival analysis (see Figure 2) supports our choice of a negative logistic model for decay as illustrated by Figure C.1 in Appendix C. The survival curve indicates a sharp initial drop in the survival probability of zombie papers, as many retractions occur within the first few years following publication. This rapid initial decay is followed by a long, gradual decline, suggesting that some papers persist for significantly extended periods before being retracted. This pattern reflects a two-phase decay process, where a proportion of zombie papers are quickly identified and retracted, while others persist in the literature, resistant to retraction pressures. The negative logistic model is well suited to capture this dynamic, as it allows for an initial high decay rate that slows over time, reflecting diminishing returns as only the most entrenched zombie papers remain in circulation.

Our model is therefore based on a negative logistic decay framework, where the decay of zombie papers is driven by retraction rates derived from the AFT-Weibull model (discussed in Section 3) and a carrying capacity representing unretracted papers that remain in the literature. The use of population dynamics in scientific modeling is well-established for understanding systemic behaviors such as knowledge growth, obsolescence, and correction processes (see, e.g., Bettencourt et al. (2008) and Vitanov & Ausloos (2012)). By simulating these processes, we can evaluate the impact of editorial interventions on the persistence of retracted papers.

Let $Z(t)$ represent the number of zombie papers at time $t$, $r$ the retraction rate—derived as the inverse of the expected time to retraction $T$ (from Equation (2) estimated using the AFT-Weibull model)—and $K$ the carrying capacity, which accounts for papers that will likely never be retracted due to limited scrutiny or missed detection, or be retracted after a very long time.

The differential equation governing the decay of zombie papers is given by:

$$\frac{dZ}{dt} = -r \times (Z - K) \tag{3}$$

where $(Z - K)$ represents the number of retractable zombie papers at any given time, with $r$ being the intensity of retraction efforts. The carrying capacity $K$ sets a lower bound for the persistent population, accounting for practical limitations like unreported or undefined misconduct, unsystematic reviews, or insufficient editorial resources (Smith (2000)).

As previously discussed, the retraction rate $r$ is determined by the expected time to retraction $T$, which is calculated using the covariates from the AFT-Weibull model. Specifically:

$$r = \frac{1}{T} \tag{4}$$

where $T$ can be derived from the AFT model under baseline conditions or adjusted to reflect different policy interventions. By simulating changes in retraction rates resulting from policies such as enhanced data transparency, replicability and reproducibility, as well as plagiarism detection, we model their effects on the long-term decay of zombie papers in the scientific body of knowledge.

To account for the long-term persistence of zombie papers, we define the carrying capacity $K$ based on observed time-to-retraction data from our dataset, reflecting that a small proportion of papers take an extended time to be scrutinized (or never be scrutinized). Specifically, we calculate the 95th percentile of time to retraction (in days) and determine the proportion of observations that exceed this threshold. This approach estimates $K$ as 5% of the initial number of zombie papers, $Z_0$, as follows:

$$K = 0.05 \times Z_0 \tag{5}$$

The carrying capacity, $K$, ensures that over time the number of zombie papers asymptotically approaches this limit, simulating a scenario where some papers take an extended time to be retracted, effectively impacting science as if they were never retracted.

Substituting expressions (4) and (5) into Equation (3), we obtain the final form of the ZPDD model:

$$\frac{dZ}{dt} = -\frac{1}{T} \times (Z - (0.05 \times Z_0)) \tag{6}$$

In the following section, we simulate the dynamics of the zombie paper population over a period of 5000 days, equivalent to roughly 13.7 years. This extended timeframe is chosen to capture the long-term effects of retraction dynamics and policy interventions, allowing us to observe how the population of zombie papers approaches the carrying capacity over a realistic editorial timescale. Given that retraction processes in academia often span multiple years, a 5000-day horizon provides a comprehensive view of both the initial impact and the diminishing returns of retraction efforts as the population stabilizes.

### 4.2 Editorial policy interventions and zombies

To explore how the scientific community might implement mechanisms to strengthen research integrity and mitigate the impact of retracted papers on scientific knowledge, we use our ZPDD model (Equation (6)) to assess a range of (non-exhaustive) potential editorial policy interventions and their counterfactual impacts on the evolution of the zombie paper population. For each policy intervention, we simulate a change in specific covariates from the baseline (i.e., no policy change) as follows:

1. Adjust the relevant covariates according to the specific policy intervention
2. Recalculate $T$ in Equation (6) based on the adjusted covariates, yielding $T_{policy}$
3. Define the retraction rate under the policy intervention as: $r_{policy} = \frac{1}{T_{policy}}$
4. Solve the ZPDD model using $r = r_{policy}$.

This framework allows us to evaluate the impact of specific policies on the persistence and decay of zombie papers within the scientific record. These editorial interventions are listed in Table 3 and address critical forms of research misconduct, including enhancing data transparency, improving reproducibility and replicability, and strengthening plagiarism detection (as highlighted by COPE). Each intervention is modeled to simulate a 30% reduction in its targeted issue, providing a concrete illustration of its potential impact on the zombie paper population. While these policies are widely recognized as pivotal for safeguarding research integrity (e.g., Miguel et al. (2014), Nosek et al. (2015), Christensen & Miguel (2018), Pérignon et al. (2024)), their implementation across the scientific ecosystem remains incomplete. To identify the unique contributions of each intervention, the simulation of single-policy strategies provides valuable insights. These findings can inform actionable editorial recommendations, such as:

- Mandating public sharing of raw datasets and analysis codes at the submission stage to enhance transparency.
- Requiring pre-registration of study designs to ensure methodological rigor and mitigate selective reporting.
- Implementing advanced plagiarism detection tools during peer review to strengthen originality checks.

Table 3: Editorial policy interventions tools

| Editorial policy interventions | Proxy variables | Changes | Meanings |
|---|---|---|---|
| **Individual (targeted) interventions** | | | |
| Increase data transparency | *Data* | ↓ Data issues by 30% | Strengthen requirements for data accessibility and transparency to minimize data-related retractions. |
| Improve replicability and reproducibility | *Result* | ↓ Results issues by 30% | Implement guidelines to improve replicability and reproducibility of research findings, reducing issues related to results. |
| Improve plagiarism detection | *Plagiarism* | ↓ Plagiarism issues by 30% | Enhance plagiarism detection during peer review to decrease instances of content duplication and improve integrity. |

Note: This table presents the simulated editorial policy interventions, including the associated variable names, implemented changes, and descriptions of each intervention's intended impact.

- Encouraging the publication of replication studies to validate key findings.[10]
- Creating dedicated editorial committees to efficiently handle misconduct allegations.

Combining interventions—such as pairing data-sharing mandates with pre-registration requirements—can amplify their effectiveness. Integrated approaches have the potential to significantly reduce zombie paper persistence, expediting their removal from the scientific record.

The effectiveness of these editorial interventions is evaluated across multiple dimensions, including publication year, geographical region, and journal field. We examine five time periods—1980, 1990, 2000, 2010, and 2020—to provide insights into the evolution of retraction practices. To facilitate interpretation, geographical zones have been grouped into four regions: America (North and South), Europe (Northern, Southern, Western, Central, and Eastern), Asia and Oceania, and Africa and the Middle East (see Table A.3 in Appendix A, excluding Central America). For journal fields, we have simplified the classification into four major domains: "Quantitative and Empirical Research", "Life and Applied Sciences", "Humanities and Social Sciences", and "Interdisciplinary and Multidisciplinary" (see Table A.2 in Appendix A, excluding Geography). Each domain group represents distinctive research characteristics that influence retraction dynamics.

Figure 5 and the attached table illustrate the decay of zombie papers across different time periods, along with key metrics such as the number of zombies remaining after 1,000 days, as well as the average and peak decay speed (calculated as $\frac{dZ}{dt} = \frac{Z_{t+1}-Z_t}{t_{t+1}-t_t}$) for each considered year.[11] The figure reveals a clear trend of increasing efficiency in zombie paper decay

[10] Reproducibility rates in fields like economics and finance remain surprisingly low, underscoring the need for such measures (Chang & Li (2017), Gertler et al. (2018), Herbert et al. (2021)).

[11] In research contexts, a period of 1,000 days (approximately 3 years) is a meaningful timeframe for assessing the retraction of problematic papers, balancing timely correction with maintaining scientific integrity.

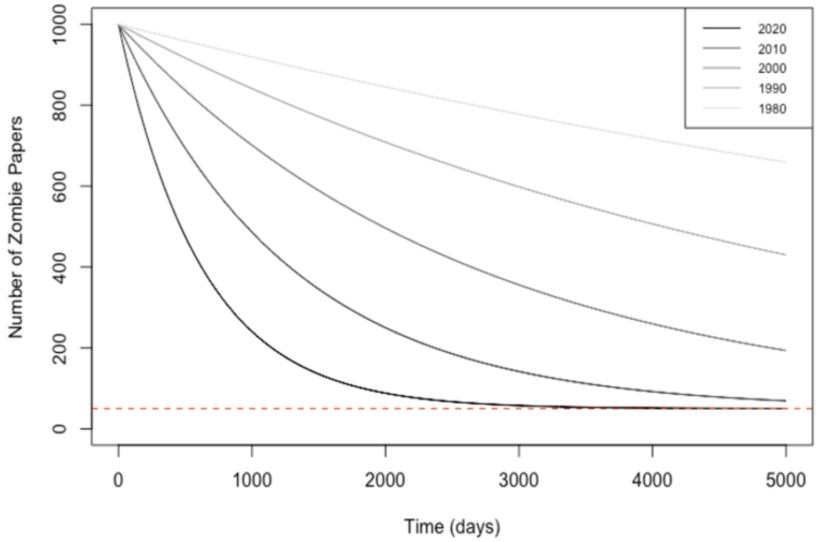
over the years, with 2020 showing the most rapid and sustained decline. This is evidenced by the significantly higher decay speed in 2020, indicating that recent editorial practices have become more effective in curtailing the persistence of zombie papers in the academic literature.

Figure 6 and the accompanying table illustrate the effects of various policy interventions—each targeting a 30% reduction in a specific retraction issue—on the decay of zombie papers in 2020. The modeled interventions include enhancing data transparency, improving reproducibility and replicability, and strengthening plagiarism detection. Each intervention is compared to a baseline scenario (without policy changes) across key metrics: time to reach 50% and 75% reduction, maximum decay rate, and survival rate at 1,000 days. The results show that improving data transparency is the most effective intervention, achieving the fastest reduction in paper counts and the highest peak decay rate, followed by strengthening plagiarism detection and enhancing reproducibility and replicability. Each intervention accelerates the decay of zombie papers compared to the baseline, underscoring the potential of targeted policy changes to reduce the persistence of problematic publications. These higher decay rates and lower survival rates at 1,000 days reflect the effectiveness of these interventions in fostering a more responsive and rigorous editorial environment, emphasizing the role of policy in enhancing academic integrity and minimizing the prevalence of zombie papers.

Figures 7 and 8 along with their accompanying tables break down the decay trends of zombie papers across specific regions: America and Europe in Figure 7, and Asia, Oceania, Africa, and the Middle East in Figure 8. In America and Europe, zombie paper reduction under the baseline scenario progresses at a relatively slower rate; however, targeted editorial interventions significantly accelerate this decay. For instance, in America, the implementation of data transparency reduces the time needed to reach a 50% reduction from 645 days to 570 days. Similar acceleration is observed with interventions focused on reproducibility and plagiarism detection. In Europe, the baseline decay rate is already comparatively higher than in other regions, and all interventions further enhance this process, with data transparency yielding the most pronounced effect, shortening the time to 50% reduction from 197 days to 174 days. In contrast, baseline decay rates in Asia, Oceania, Africa, and the Middle East are slower than those observed in Europe but faster than those in America. Here, editorial interventions, particularly reproducibility and plagiarism detection improvements, show notable impacts, albeit to a slightly lesser extent than in Europe. The accompanying tables underscore these findings by presenting key metrics, including time to 50% and 75% decay, maximum decay rates, and survival rates at 1,000 days for each region and intervention type. These metrics illustrate a consistent reduction in decay time across all interventions, with minor regional variations in maximum decay rates and long-term survival rates. Overall, while each intervention contributes to the reduction of zombie papers, their relative effectiveness varies by region. Data transparency consistently emerges as the most impactful intervention across regions, demonstrating its broad utility in accelerating the decay of zombie papers.

Figures 9 and 10 disentangle the effect of editorial policy interventions across research fields,

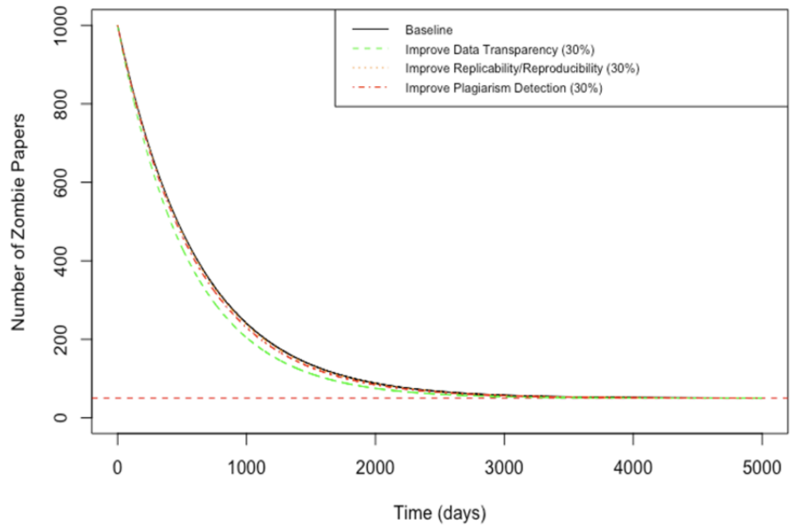
Figure 5: Decay of zombie papers over time




| Years | Number of Zombies (at 1,000 days) | Decay Speed | |
|---|---|---|---|
| | | Average | Peak |
| 1980 | 919 | 0.068 | 0.084 |
| 1990 | 840 | 0.113 | 0.174 |
| 2000 | 701 | 0.161 | 0.358 |
| 2010 | 486 | 0.186 | 0.739 |
| 2020 | 240 | 0.189 | 1.523 |

Note: This figure shows the decay trend of zombie papers over time, including characteristics such as the number of zombie papers remaining after 1000 days and the average and peak decay rates for the years 1980, 1990, 2000, 2010, and 2020.

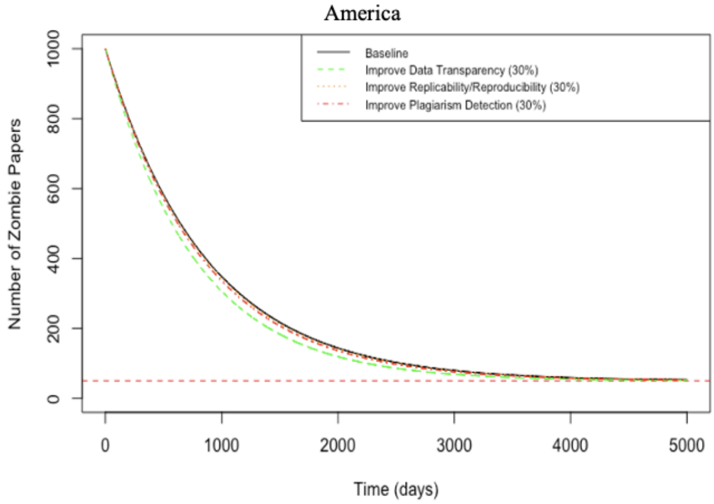
Figure 6: Decay of zombie papers across editorial policy interventions



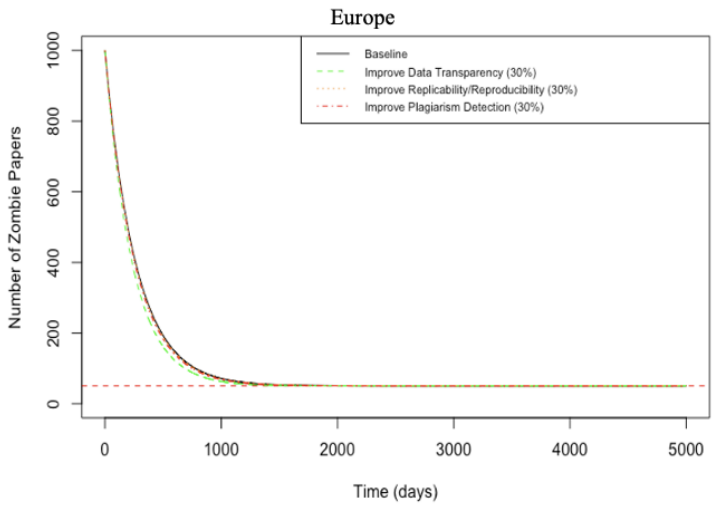

| Metrics | Baseline | Data Transparency (30%) | Reproducibility & Replicability (30%) | Plagiarism (30%) |
|---|---|---|---|---|
| Time to 50% Reduction (Days) | 465 | 411 | 460 | 449 |
| Time to 75% Reduction (Days) | 971 | 858 | 960 | 937 |
| Max Decay Rate (Papers/Day) | 1.523 | 1.724 | 1.541 | 1.578 |
| Survival Rate at 1,000 Days (%) | 24.07 | 20.44 | 23.73 | 23.00 |

Note: This figure illustrates the decay trend of zombie papers, along with characteristics under various editorial policy interventions, including improvements in data transparency, reproducibility, replicability, and plagiarism detection.

Figure 7: Decay of zombie papers across editorial policy interventions in America and Europe



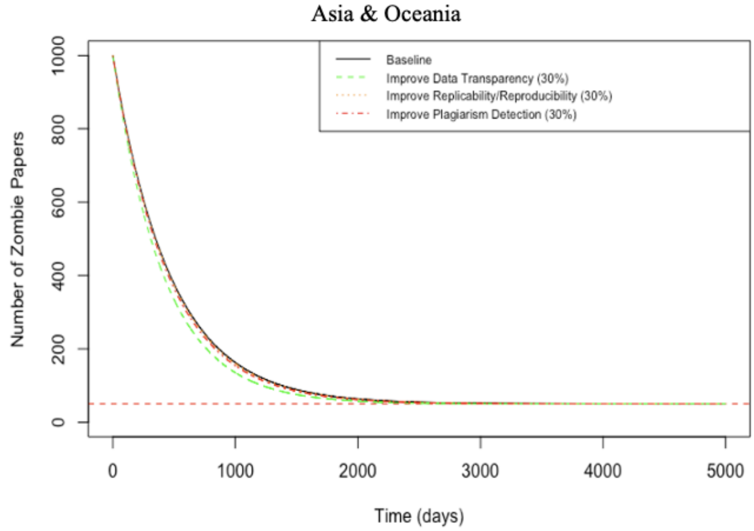

| Metrics | Baseline | Data Transparency (30%) | Reproducibility & Replicability (30%) | Plagiarism (30%) |
|---|---|---|---|---|
| Time to 50% Reduction (Days) | 645 | 570 | 638 | 623 |
| Time to 75% Reduction (Days) | 1,345 | 1,189 | 1,330 | 1,298 |
| Max Decay Rate (Papers/Day) | 1.099 | 1.244 | 1.112 | 1.139 |
| Survival Rate at 1,000 Days (%) | 34.83 | 30.61 | 34.44 | 33.60 |



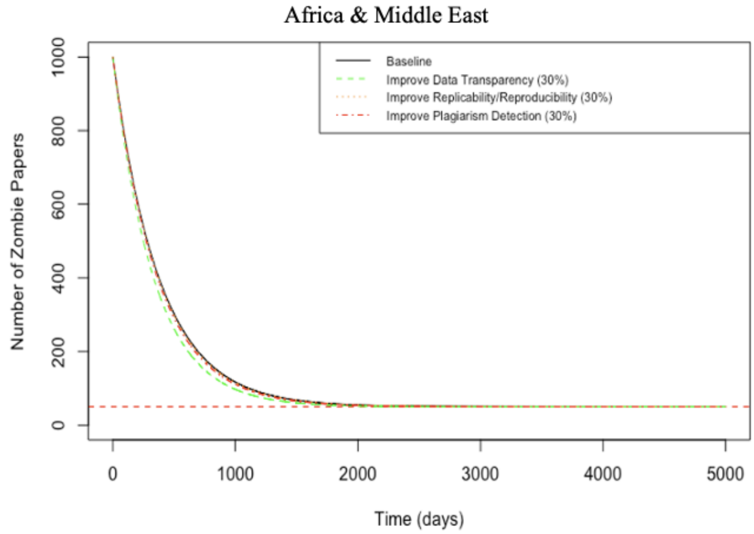

| Metrics | Baseline | Data Transparency (30%) | Reproducibility & Replicability (30%) | Plagiarism (30%) |
|---|---|---|---|---|
| Time to 50% Reduction (Days) | 197 | 174 | 195 | 190 |
| Time to 75% Reduction (Days) | 411 | 364 | 407 | 397 |
| Max Decay Rate (Papers/Day) | 3.591 | 4.063 | 3.632 | 3.721 |
| Survival Rate at 1,000 Days (%) | 7.150 | 6.306 | 7.060 | 6.875 |

Note: This figure illustrates the decay trend of zombie papers, with characteristics shown for America (top) and Europe (bottom) under various editorial policy interventions, including enhancements in data transparency, reproducibility, replicability, and plagiarism detection.

Figure 8: Decay of zombie papers across editorial policy interventions in Asia, Oceania, Africa and Middle East




| Metrics | Baseline | Data Transparency (30%) | Reproducibility & Replicability (30%) | Plagiarism (30%) |
|---|---|---|---|---|
| Time to 50% Reduction (Days) | 350 | 350 | 309 | 338 |
| Time to 75% Reduction (Days) | 730 | 730 | 722 | 704 |
| Max Decay Rate (Papers/Day) | 2.026 | 2.026 | 2.048 | 2.099 |
| Survival Rate at 1,000 Days (%) | 16.23 | 16.23 | 15.96 | 15.39 |




| Metrics | Baseline | Data Transparency (30%) | Reproducibility & Replicability (30%) | Plagiarism (30%) |
|---|---|---|---|---|
| Time to 50% Reduction (Days) | 281 | 249 | 278 | 272 |
| Time to 75% Reduction (Days) | 587 | 519 | 580 | 566 |
| Max Decay Rate (Papers/Day) | 2.519 | 2.850 | 2.547 | 2.610 |
| Survival Rate at 1,000 Days (%) | 11.67 | 9.708 | 11.47 | 11.06 |

Note: This figure illustrates the decay trend of zombie papers, with characteristics shown for Asia and Oceania (top) and Africa and Middle East (bottom) under various editorial policy interventions, including enhancements in data transparency, reproducibility, replicability, and plagiarism detection.

specifically “Quantitative and Empirical Research”, “Life and Applied Sciences” (in Figure 9), and “Humanities and Social Sciences”, as well as “Interdisciplinary and Multidisciplinary Domains” (in Figure 10).

In “Quantitative and Empirical Research”, the baseline decay rate is moderate, with zombie papers reaching a 50% reduction in 1,619 days. Data transparency significantly accelerates this, bringing the time down to 1,431 days, indicating the importance of transparent data practices in fields heavily reliant on empirical evidence. Maximum decay rate increases from 0.438 to 0.496 papers per day with this intervention, while the survival rate at 1,000 days decreases from 64.87% to 61.34%. Reproducibility and plagiarism detection also have beneficial effects but to a lesser extent, suggesting that transparency addresses a core issue in quantitative disciplines by reducing opacity in data use and enhancing accountability. For “Life and Applied Sciences”, decay rates are faster, with a baseline time to 50% reduction at 1,224 days, reflecting the field’s rapid research turnover. Data transparency further accelerates this to 1,081 days, with the maximum decay rate rising from 0.579 to 0.656 papers per day, and the survival rate at 1,000 days falling from 56.58% to 52.60%. These results highlight the crucial role of data integrity in a fast-moving field where outdated or flawed studies quickly lose relevance. The interventions amplify this natural decay, making zombie papers less viable as the field advances.

In “Humanities and Social Sciences”, the decay is considerably slower, with a baseline time to 50% reduction of 2,204 days, underscoring the persistence of papers in these areas. This slow decay may result from the longevity of theoretical and qualitative contributions, coupled with less frequent replication. Data transparency reduces this decay time to 1,948 days, with the maximum decay rate increasing from 0.322 to 0.364 papers per day and the survival rate at 1,000 days slightly dropping from 72.68% to 69.73%. This relatively limited impact reflects the challenge of enforcing rapid decay in a field less dependent on data-driven research, though transparency and reproducibility still provide noticeable improvements. “Interdisciplinary and Multidisciplinary Domains” display faster decay rates than “Humanities and Social Sciences”, with a baseline time of 834 days to reach a 50% reduction. Data transparency again proves highly effective, lowering this to 737 days, increasing the maximum decay rate from 0.851 to 0.962 papers per day, and reducing the survival rate at 1,000 days from 43.76% to 39.45%. The faster baseline decay suggests a higher likelihood of cross-verification and scrutiny from multiple disciplines, making zombie papers less sustainable. The interventions reinforce this fast decay, indicating that the interdisciplinary nature, involving diverse perspectives, makes these papers more resilient to obsolescence, with transparency amplifying this effect.

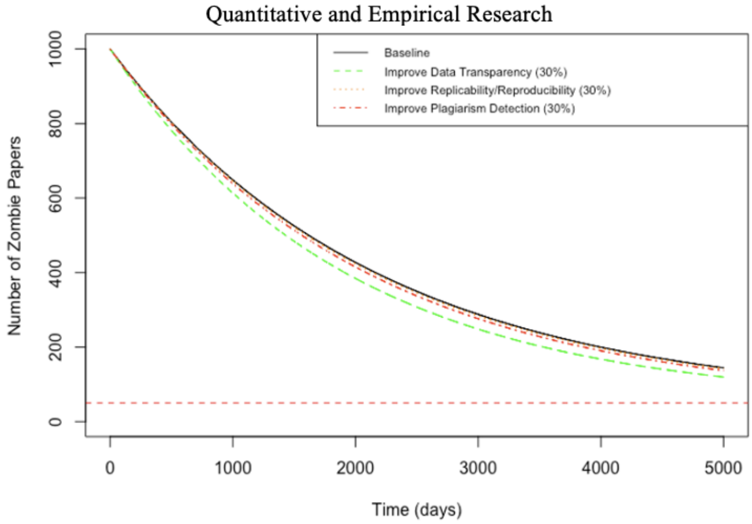
In summary, data transparency consistently emerges as the most impactful intervention across all fields, though its effectiveness varies by discipline. Quantitative, empirical, and applied sciences, where data integrity is foundational, respond most strongly to this intervention, benefiting from increased decay rates and reduced survival of zombie papers. “Humanities and Social Sciences”, with their slower pace of knowledge turnover, also benefit, but decay rates increase more modestly, reflecting traditional norms less focused on data scrutiny. Interdisciplinary fields naturally exhibit rapid decay, with interventions like data transparency reinforcing this resilience by supporting cross-disciplinary validation. Reproducibility and plagiarism detection also positively impact decay rates across fields but have a less pronounced effect than data transparency. This suggests that foundational improvements in data practices could yield substantial reductions in zombie papers, particularly in fields where empirical data is central to research validity.

### 4.3 Optimizing editorial policy interventions to minimize zombie paper persistence

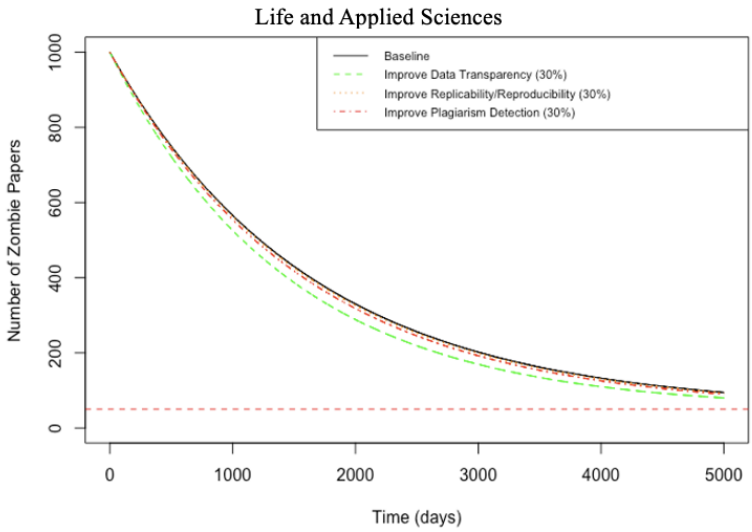
To provide an effective mix of editorial policy interventions aimed at minimizing zombie paper persistence, we develop a mathematical optimization framework that determines optimal intervention levels across key retraction issues: data transparency, replicability and reproducibility, and plagiarism detection. This framework is grounded in the ZPDD model, where the retraction rate is inversely related to the predicted retraction time $T$, which itself decreases as editorial interventions increase. The objective of the optimization is to minimize the number of zombie papers at a specific future time horizon (e.g., 1,000 days) by reducing $T$ through well-targeted policy interventions. To prevent unrealistically high levels of intervention, a penalty term is included in the optimization, designed to balance the trade-off between intervention intensity and practical feasibility. The penalty, which grows quadratically with each intervention level, discourages excessive intervention efforts unless they yield substantial reductions in the zombie paper population. This formulation ensures that the resulting intervention levels are both impactful and resource-efficient. By simulating and quantifying the effects of various policy mixes, this optimization provides editorial boards with actionable insights to enhance research integrity and reduce zombie paper persistence (see Appendix C.2 for the full mathematical formulation).

Figure 11 presents the optimization results across three critical dimensions. The top-left panel illustrates the relationship between editorial policy interventions and penalties. As the penalty parameter ($\theta$) increases, the optimal levels for each intervention (Data Transparency, Results Replicability/Reproducibility, and Plagiarism Detection) decline. This trend demonstrates a trade-off: higher penalties discourage intensive interventions. Data Transparency consistently maintains the highest intervention level, followed by Plagiarism Detection and Results Reproducibility, emphasizing its key role in reducing zombie papers within this model. The top-right panel shows the impact of penalties on the zombie population at 1,000 days. As $\theta$ increases, the zombie population rises and approaches an asymptote. The optimal penalty point is marked at $\theta = 1$, where the zombie population reaches its minimum under the penalty constraint. Beyond this threshold, diminishing returns set in as further penalty increases lead to a higher zombie population due to scaled-back intervention levels. The bottom panel highlights the optimal intervention mix that achieves the minimum zombie population given the penalty constraint. The optimal mix comprises a Data Transparency intervention level of

Figure 9: Decay of zombie papers across editorial policy interventions by journal area



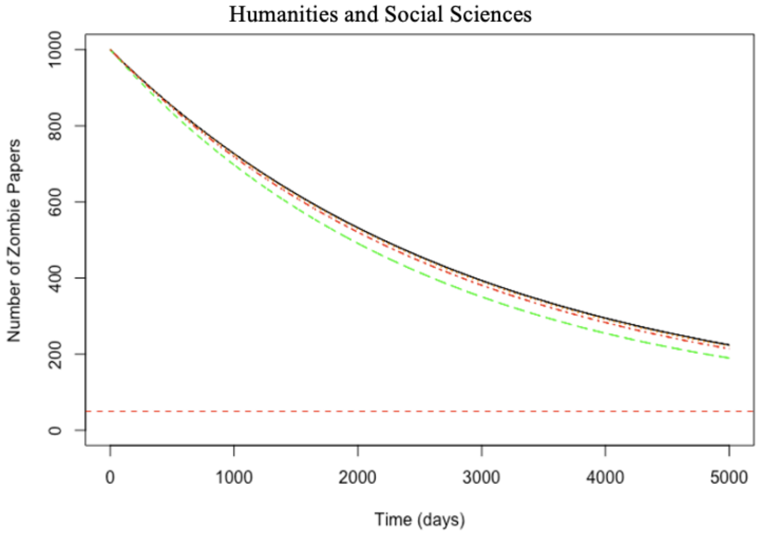

| Metrics | Baseline | Data Transparency (30%) | Reproducibility & Replicability (30%) | Plagiarism (30%) |
|---|---|---|---|---|
| Time to 50% Reduction (Days) | 1,619 | 1,431 | 1,601 | 1,562 |
| Time to 75% Reduction (Days) | 3,375 | 2,983 | 3,338 | 3,258 |
| Max Decay Rate (Papers/Day) | 0.438 | 0.496 | 0.443 | 0.454 |
| Survival Rate at 1,000 Days (%) | 64.87 | 61.34 | 64.56 | 63.88 |



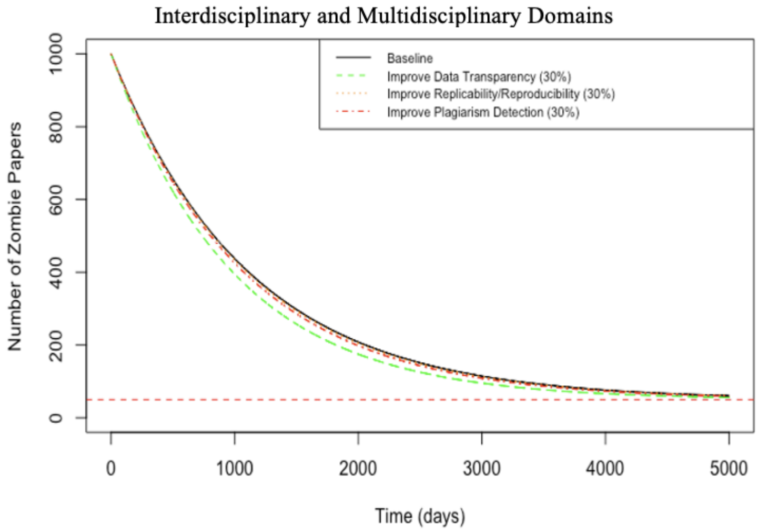

| Metrics | Baseline | Data Transparency (30%) | Reproducibility & Replicability (30%) | Plagiarism (30%) |
|---|---|---|---|---|
| Time to 50% Reduction (Days) | 1,224 | 1,081 | 1,210 | 1,181 |
| Time to 75% Reduction (Days) | 2,552 | 2,255 | 2,523 | 2,463 |
| Max Decay Rate (Papers/Day) | 0.579 | 0.656 | 0.586 | 0.600 |
| Survival Rate at 1,000 Days (%) | 56.58 | 52.60 | 56.23 | 55.45 |

Note: This figure illustrates the decay trend of zombie papers, with characteristics shown for "Quantitative and Empirical Research" (top) and "Life and Applied Sciences" (bottom) under various editorial policy interventions, including enhancements in data transparency, reproducibility, replicability, and plagiarism detection.

Figure 10: Decay of zombie papers across editorial policy interventions by journal area (con't)



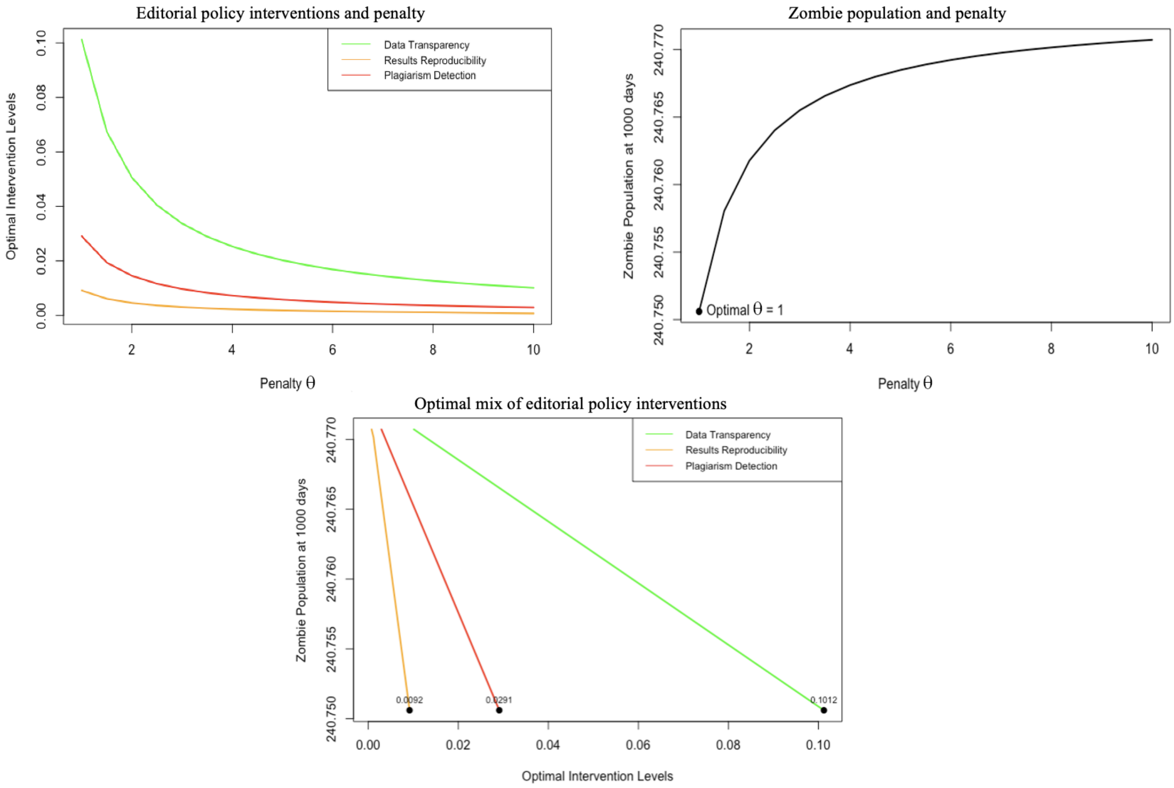

| Metrics | Baseline | Data Transparency (30%) | Reproducibility & Replicability (30%) | Plagiarism (30%) |
|---|---|---|---|---|
| Time to 50% Reduction (Days) | 2,204 | 1,948 | 2,179 | 2,127 |
| Time to 75% Reduction (Days) | 4,596 | 4,062 | 4,544 | 4,435 |
| Max Decay Rate (Papers/Day) | 0.322 | 0.364 | 0.325 | 0.333 |
| Survival Rate at 1,000 Days (%) | 72.68 | 69.73 | 72.42 | 71.85 |




| Metrics | Baseline | Data Transparency (30%) | Reproducibility & Replicability (30%) | Plagiarism (30%) |
|---|---|---|---|---|
| Time to 50% Reduction (Days) | 834 | 737 | 824 | 805 |
| Time to 75% Reduction (Days) | 1,738 | 1,536 | 1,719 | 1,678 |
| Max Decay Rate (Papers/Day) | 0.851 | 0.962 | 0.860 | 0.881 |
| Survival Rate at 1,000 Days (%) | 43.76 | 39.45 | 43.37 | 42.52 |

Note: This figure illustrates the decay trend of zombie papers, with characteristics shown for "Humanities and Social Sciences" (top) and "Interdisciplinary and Multidisciplinary Domains" (bottom) under various editorial policy interventions, including enhancements in data transparency, reproducibility, replicability, and plagiarism detection.

Figure 11: Optimal mix of editorial policy interventions for the decay of zombie papers at 1,000 days




Note: This figure reports the results of the optimization calculus showing (i) the editorial policy interventions with respect to penalty; (ii) the zombie population at 1,000 days with respect to penalty; and (iii) the optimal mix of editorial policy interventions with respect to the zombie population at 1,000 days.

0.1012 (equivalent to a 10.12% reduction), Plagiarism Detection at 0.0291 (2.91% reduction), and Results Reproducibility at 0.0092 (0.92% reduction). With this configuration, the minimum achievable zombie population at 1,000 days is approximately 240.75 papers. This figure effectively captures the balance between intervention intensity and penalty, demonstrating the optimal mix needed to minimize the zombie population over time while managing penalty trade-offs.

# 5 Discussion and editorial policy recommendations

Our findings underscore several key implications for editorial policies that can enhance the efficiency and transparency of the retraction process while reducing the persistence and influence of zombie papers. A central theme emerging from our results is the need for proactive strategies and systemic reforms to address both the persistence of flawed research and the disparities in retraction practices.

One key recommendation is the enhancement of data transparency. We find that open access to raw datasets and analysis codes significantly reduces the time required to detect errors or misconduct, facilitating faster retractions. To this end, academic journals should systematically mandate the submission of datasets, analysis codes, and supplementary materials at the time of manuscript submission. Establishing open-access repositories linked to published articles would further support transparency and replicability.

The results also highlight the importance of promoting reproducibility, replicability, and methodological rigor. The persistence of zombie papers reveals a lack of robust reproduction and replication efforts across disciplines. Policies encouraging pre-registration of study designs, adherence to reporting standards, and prioritization of reproducible and replication studies can address this gap. Academic journals could dedicate specific sections or platforms to reproducible and replication studies and null results, reinforcing the emphasis on methodological integrity.

Another point emphasized by our investigation is the need for advanced plagiarism detection and improved misconduct monitoring. Prolonged retraction timelines in cases of severe misconduct suggest that journals must invest in state-of-the-art plagiarism detection tools and provide targeted training for editors and reviewers. Specialized review boards for misconduct investigations could enhance the capacity to address complex cases efficiently.

The visibility of retraction notices is another crucial area for improvement. Low awareness of retracted articles contributes to their continued citation. Journals should ensure that retracted articles are prominently marked and linked to their retraction notices in citation databases and journal platforms. Clear notifications in reference indexes such as PubMed and CrossRef would further minimize the propagation of flawed research.

Geographic and institutional disparities in retraction practices also require attention. Our findings reveal inefficiencies in under-resourced regions, emphasizing the need for standardized global retraction guidelines. Capacity-building efforts and collaborative initiatives between journals and institutions can help bridge these gaps, ensuring equitable enforcement of research integrity standards.

A broader cultural shift is necessary to prioritize ethical practices in research. The current "publish or perish" culture often incentivizes misconduct and delays its detection. Institutions and journals should reward ethical behavior through recognition programs and dedicated funding while incorporating ethics training into researcher and editorial staff development. Finally, leveraging technology offers promising solutions for addressing the issue of zombie papers. Artificial intelligence-driven tools could monitor citation networks in real time, flag unusual citation patterns, and track the impact of retracted studies. Cross-disciplinary collaboration in editorial boards and oversight committees could further enhance the detection and prevention of research misconduct.

Adopting these measures can strengthen the integrity of the scientific record, mitigate the impact and persistence of zombie papers, and foster a research environment centered on transparency, accountability, and quality.

## 6 Conclusion

This paper addresses the critical issue of zombie papers, i.e., retracted or destined-for-retraction publications that continue to influence academic discourse despite their discredited status. These papers represent a significant challenge to the self-correcting mechanisms of science, as their flawed findings propagate through citations, distorting knowledge production and diverting future research efforts. Through a comprehensive analysis combining survival models and a novel theoretical framework—the Zombie Population Decay Dynamics (ZPDD) model—we shed light on the factors driving zombie paper persistence and propose actionable editorial strategies to mitigate their impact.

Our findings reveal that the time taken to retract papers varies significantly depending on the nature of the misconduct, with serious issues such as data falsification or fabrication requiring more extended investigation and resolution periods. Geographic and institutional disparities also play a crucial role, with systemic inefficiencies in under-resourced regions, such as parts of Eastern Europe, contributing to prolonged retraction processes. Furthermore, the analysis highlights that editorial policies, including journal characteristics, significantly influence the persistence of zombie papers. Subscription-based journals, for instance, demonstrate faster retraction times than open-access journals, likely due to stricter oversight and more robust editorial resources.

Based on these findings, our study proposes several policy recommendations to accelerate retractions and reduce the influence of flawed research. Mandating greater data transparency, such as requiring the public sharing of raw datasets and analysis codes, can enhance the replicability of research and facilitate the identification of errors. Similarly, promoting reproducibility through pre-registration of study designs and the publication of replication studies can mitigate the propagation of flawed research. Enhancing plagiarism detection with advanced tools during the peer review process and increasing the accessibility and visibility of retraction notices are also critical for minimizing the impact of zombie papers. Addressing institutional and geographic disparities through standardized global practices and targeted initiatives for under-resourced regions is essential for ensuring equitable enforcement of research integrity standards worldwide.

In addition to these practical measures, our paper underscores the need for a cultural shift in the academic community to prioritize quality and transparency over quantity. This would involve creating or reinforcing an environment where transparency, reproducibility, and accountability are integral to the publication process. Leveraging technological advancements, such as artificial intelligence for misconduct detection and automated citation monitoring, can further streamline retraction processes and reduce the influence of zombie papers.

Although our paper lays a strong foundation for understanding the dynamics of zombie paper persistence, it also opens avenues for future research. Specifically, it would be interesting to build on our findings to explore the long-term effects of specific editorial interventions on retraction dynamics and examine how policy changes and technologies such as artificial intelligence may influence the behavior of researchers and institutions. Additionally, assessing zombie papers' economic and reputational costs on the broader scientific community could provide a deeper understanding of their systemic impact. Finally, evaluating the effectiveness of coordinated global efforts to standardize editorial practices and ensure equitable access to resources for addressing retractions would be a promising research avenue. Overall, addressing zombie paper persistence requires a multifaceted approach that combines dedicated editorial policies, technological innovations, and global cooperation.

## References


Anderson, R. G., Greene, W. H., McCullough, B. D. & Vinod, H. D. (2008), 'The role of data/code archives in the future of economic research', *Journal of Economic Methodology* **15**(1), 99–119.

Azoulay, P., Furman, J. L., Krieger & Murray, F. (2015), 'Retractions', *Review of Economics and Statistics* **97**(5), 1118–1136.

Beall, J. (2013), 'Predatory publishing is just one of the consequences of gold open access', *Learned Publishing,* **26**(2), 79–84.

Bergh, D. D., Sharp, B. M. & Li, M. (2017), 'Tests for identifying "red flags" in empirical findings: Demonstration and recommendations for authors, reviewers, and editors', *Academy of Management Learning & Education* **16**(1), 110–124.

Bettencourt, L. M. A., Kaiser, D. I., Kaur, J., Castillo-Chàvez, C. & Wojick, D. E. (2008), 'Population modeling of the emergence and development of scientific fields', *Scientometrics* **75**(3), 495–518.

Bozeman, B. & Boardman, C. (2014), *Research collaboration and team science: A state-of-the-art review and agenda*, Springer.

Brier, G. W. (1950), 'Verification of forecasts expressed in terms of probability', *Monthly Weather Review* **78**(1), 1–3.

Bruch, E. & Atwell, J. (2015), 'Agent-based models in empirical social research', *Sociological Methods & Research* **44**(2), 186–221.

Budd, J. M., Sievert, M. & Schultz, T. R. (1998), 'Phenomena of retraction: reasons for retraction and citations to the publications', *Jama* **280**(3), 296–297.

Bushee, B. J., Core, J. E., Guay, W. & Hamm, S. J. (2010), 'The role of the business press as an information intermediary', *Journal of Accounting Research* **48**(1), 1–19.

Chang, A. C. & Li, P. (2017), 'A preanalysis plan to replicate sixty economics research papers that worked half of the time', *American Economic Review* **107**(5), 60–64.

Christensen, G. & Miguel, E. (2018), 'Transparency, reproducibility, and the credibility of economics research', *Journal of Economic Literature* **56**(3), 920–980.

COPE (2022), 'COPE Retraction guidelines. Version 2. November 2019', *Science Editor and Publisher* **6**(2), 148–154.

Cox, A., Craig, R. & Tourish, D. (2018), 'Retraction statements and research malpractice in economics', *Research Policy* **47**(5), 924–935.

Cox, D. R. (1972), 'Regression models and life-tables', *Journal of the Royal Statistical Society: Series B (Methodological)* **34**(2), 187–202.

Cummings, J. N. & Kiesler, S. (2005), 'Collaborative research across disciplinary and organizational boundaries', *Social Studies of Science* **35**(5), 703–722.

Drake, M. S., Guest, N. M. & Twedt, B. J. (2014), 'The media and mispricing: The role of the business press in the pricing of accounting information', *The Accounting Review* **89**(5), 1673–1701.

Dunkler, D., Ploner, M., Schemper, M. & Heinze, G. (2018), 'Weighted Cox regression using the R package coxphw', *Journal of Statistical Software* **84**, 1–26.

Fanelli, D. (2009), 'How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data', *PLOS One* **4**(5), e5738.

Fang, F. C., Steen, R. G. & Casadevall, A. (2012), 'Misconduct accounts for the majority of retracted scientific publications', *Proceedings of the National Academy of Sciences* **109**(42), 17028–17033.

Furman, J. L., Jensen, K. & Murray, F. (2012), 'Governing knowledge in the scientific community: Exploring the role of retractions in biomedicine', *Research Policy* **41**(2), 276–290.

Gertler, P., Galiani, S. & Romero, M. (2018), 'How to make replication the norm', *Nature* **554**(7693), 417–419.

Graf, E., Schmoor, C., Sauerbrei, W. & Schumacher, M. (1999), 'Assessment and comparison of prognostic classification schemes for survival data', *Statistics in Medicine* **18**(17-18), 2529–2545.

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L. & Rosati, R. A. (1982), 'Evaluating the yield of medical tests', *Jama* **247**(18), 2543–2546.

Herbert, S., Kingi, H., Stanchi, F. & Vilhubern, L. (2021), The Reproducibility of Economics Research: A Case Study, Working papers 853, Banque de France.

Hesselmann, F., Graf, V., Schmidt, M. & Reinhart, M. (2017), 'The visibility of scientific misconduct: A review of the literature on retracted journal articles', *Current Sociology* **65**(6), 814–845.

Horbach, S. P. & Halffman, W. (2017), 'The ghosts of hela: How cell line misidentification contaminates the scientific literature', *PLOS One* **12**(10), e0186281.

Horbach, S. P. & Halffman, W. (2019), 'The ability of different peer review procedures to flag problematic publications', *Scientometrics* **118**(1), 339–373.

Houghton, F. (2022), 'Keep calm and carry on: moral panic, predatory publishers, peer review, and the emperor's new clothes', *Journal of the Medical Library Association: JMLA* **110**(2), 233.

Kaplan, E. L. & Meier, P. (1958), 'Nonparametric estimation from incomplete observations', *Journal of the American Statistical Association* **53**(282), 457–481.

Karabag, S. F. & Berggren, C. (2012), 'Retraction, Dishonesty and Plagiarism: Analysis of a Crucial Issue for Academic Publishing, and the Inadequate Responses from Leading Journals in Economics and Management Disciplines', *Journal of Applied Economics and Business Research* **2**(3), 172–183.

LaFollette, M. C. (2000), 'The evolution of the "scientific misconduct" issue: an historical overview (44535c)', *Proceedings of the Society for Experimental Biology and Medicine* **224**(4), 211–215.

Le Maux, B., Necker, S. & Rocaboy, Y. (2019), 'Cheat or perish? A theory of scientific customs', *Research Policy* **48**(9), 103792.

Mehra, M. R., Ruschitzka, F. & Patel, A. N. (2020), 'Retraction-Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis', *The Lancet* **395**(10240), 1820.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., Glennerster, R., Green, D. P., Humphreys, M., Imbens, G. et al. (2014), 'Promoting transparency in social science research', *Science* **343**(6166), 30–31.

Moher, D., Shamseer, L., Cobey, K. D., Lalu, M. M., Galipeau, J., Avey, M. T., Ahmadzai, N., Alabousi, M., Barbeau, P., Beck, A. et al. (2017), 'Stop this waste of people, animals and money', *Nature* **549**(7670), 23–25.

Necker, S. (2014), 'Scientific misbehavior in economics', *Research Policy* **43**(10), 1747–1759.

Necker, S. (2016), 'Why do scientists cheat? Insights from behavioral economics', *Review of Social Economy* **74**(1), 98–108.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G. et al. (2015), 'Promoting an open research culture', *Science* **348**(6242), 1422–1425.

Pérignon, C., Akmansoy, O., Hurlin, C., Dreber, A., Holzmeister, F., Huber, J., Johannesson, M., Kirchler, M., Menkveld, A. J., Razen, M. et al. (2024), 'Computational reproducibility in finance: Evidence from 1,000 tests', *The Review of Financial Studies* **37**(11), 3558–3593.

Schemper, M., Wakounig, S. & Heinze, G. (2009), 'The estimation of average hazard ratios by weighted Cox regression', *Statistics in Medicine* **28**(19), 2473–2489.

Smith, R. (2000), 'What is research misconduct?', *Proceedings-royal college of physicians of edinburgh* **30**(1; SUPP/7), 4–8.

Steen, R. G. (2011), 'Retractions in the scientific literature: is the incidence of research fraud increasing?', *Journal of Medical Ethics* **37**(4), 249–253.

Steen, R. G., Casadevall, A. & Fang, F. C. (2013), 'Why Has the Number of Scientific Retractions Increased?', *PLOS one* **8**(7), 1–9.

Steneck, N. H. (2006), 'Fostering integrity in research: Definitions, current knowledge, and future directions', *Science and Engineering Ethics* **12**, 53–74.

Tennant, J. P., Waldner, F., Jacques, D. C., Masuzzo, P., Collister, L. B. & Hartgerink, C. H. (2016), 'The academic, economic and societal impacts of open access: an evidence-based review', *F1000Research* **5**.

Therneau, T. M., Grambsch, P. M. & Fleming, T. R. (1990), 'Martingale-based residuals for survival models', *Biometrika* **77**(1), 147–160.

Tilman, D., Isbell, F. & Cowles, J. M. (2014), 'Biodiversity and ecosystem functioning', *Annual Review of Ecology, Evolution, and Systematics* **45**(1), 471–493.

Vitanov, N. & Ausloos, M. (2012), 'Knowledge epidemics and population dynamics models for describing idea diffusion', *Understanding Complex Systems* .

Wagner, C. S., Whetsell, T. A. & Leydesdorff, L. (2017), 'Growth of international collaboration in science: revisiting six specialties', *Scientometrics* **110**, 1633–1652.

Wei, L.-J. (1992), 'The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis', *Statistics in Medicine* **11**(14-15), 1871–1879.

Wienke, A. (2010), *Frailty models in survival analysis*, Chapman and Hall/CRC.

Xu, H., Ding, Y., Zhang, C. & Tan, B. C. (2023), 'Too official to be effective: An empirical examination of unofficial information channel and continued use of retracted articles', *Research Policy* **52**(7), 104815.

# Appendix

## A Dataset and variables

Table A.1: Zombie papers database

| Variables | Range/Details | Key Insights & Occurrences |
|---|---|---|
| Publication | 1923-2023 | 2021 (2,554) |
| Retraction | 1960-2023 | 2023 (4,197) |
| Journals | 5,755 | Plos One (629) |
| Publishers | 1,281 | Springer (4,235) |
| Authors | 24,318 | Ali Nazari (100) |
| Countries | 164 | China (11,086) |
| Institutions | 82,246 | Fin. Univ. Under Russian Gov. (49) |
| Paywalled | Subscription / Open-access | 821 / 24,659 |
| Reasons of retraction | Data issues | 11,556 |
| | Results issues | 8,411 |
| | Authorship issues | 1,882 |
| | Peer review/editorial issues | 7,030 |
| | Duplication/plagiarism issues | 6,913 |
| | Referencing issues | 2,787 |
| | Withdraw | 1,125 |
| | Miscommunication | 1,983 |
| | Non-reportable | 2,572 |
| | Various ethical violation | 4,251 |
| Investigation by | Publishers/Journals | 10,013 |
| | Company/Institution/Third Party | 6,197 |

Note: This table reports the main variables of our zombie papers database sourced from the Retraction Watch Database (RWD), along with key insights and occurrences. The reasons for retraction have been simplified by grouping certain categories together according to the RWD taxonomy. For the full list of retraction reasons, please refer to: https://retractionwatch.com/retraction-watch-database-user-guide/retraction-watch-database-user-guide-appendix-b-reasons/.

Table A.2: Main research domains of journals

| Variables | Main domains | Occurrence | Groups |
|---|---|---|---|
| Area 1 | Computer Science; Data Science, Information Systems; Robotics | 2834 (11.1%) | Quantitative and Empirical Research |
| Area 2 | Pharmacology, Toxicology and Pharmaceutics | 1031 (4.0%) | Life Sciences and Applied Sciences |
| Area 3 | Medicine | 4357 (17.1%) | Life Sciences and Applied Sciences |
| Area 4 | Engineering | 1370 (5.3%) | Quantitative and Empirical Research |
| Area 5 | Biochemistry, Genetics and Molecular Biology | 5004 (19.6%) | Life Sciences and Applied Sciences |
| Area 6 | Chemistry engineering; Chemistry; Material Sciences | 2273 (8.9%) | Interdisciplinary and Multidisciplinary |
| Area 7 | Physics | 692 (2.7%) | Quantitative and Empirical Research |
| Area 8 | Neuroscience | 507 (1.8%) | Interdisciplinary and Multidisciplinary |
| Area 9 | Earth and Planetary Sciences; Environmental Science; Ecology; Energy | 1313 (5.1%) | Interdisciplinary and Multidisciplinary |
| Area 10 | Arts and Humanities; History | 456 (1.7%) | Humanities and Social Sciences |
| Area 11 | Immunology; Microbiology; Virology | 618 (2.4%) | Life Sciences and Applied Sciences |
| Area 12 | Business; Management; Accounting | 460 (1.8%) | Quantitative and Empirical Research |
| Area 13 | Sociology; Anthropology; Ethnology | 255 (1.0%) | Humanities and Social Sciences |
| Area 14 | Multidisciplinary | 1324 (5.1%) | Interdisciplinary and Multidisciplinary |
| Area 15 | Psychology | 385 (1.5%) | Humanities and Social Sciences |
| Area 16 | Mathematics | 411 (1.6%) | Quantitative and Empirical Research |
| Area 17 | Agricultural and Biological Sciences | 855 (3.5%) | Life Sciences and Applied Sciences |
| Area 18 | Economics; Finance; Econometrics | 341 (1.3%) | Quantitative and Empirical Research |
| Area 19 | Dentistry | 102 (0.4%) | Life Sciences and Applied Sciences |
| Area 20 | Veterinary | 67 (0.2%) | Life Sciences and Applied Sciences |
| Area 21 | Education | 453 (1.7%) | Humanities and Social Sciences |
| Area 22 | Law; Administrative Science; Military Science | 302 (1.1%) | Humanities and Social Sciences |
| Area 23 | Geography | 6 (0.02%) | X |
| Area 24 | Political Science | 64 (0.2%) | Humanities and Social Sciences |

Note: This table reports the main research domains of journals based on Web of Science and Scimago categories. The occurrences of each domain in the database are also provided.

Table A.3: Main geographical locations of authors

| Main Locations | Occurrence |
|---|---|
| Asia | 15411 (60.4%) |
| Oceania | 357 (1.4%) |
| Middle East | 1430 (5.6%) |
| Africa | 363 (1.4%) |
| North America | 2593 (10.1%) |
| Central America | 196 (0.7%) |
| South America | 308 (1.2%) |
| North Europe | 212 (0.8%) |
| Central Europe | 598 (2.3%) |
| South Europe | 339 (1.3%) |
| West Europe | 1295 (5.0%) |
| East Europe | 2360 (9.2%) |

Note: This table reports the main geographical locations of the corresponding (or first) author. The occurrences of each location in the database are also provided.

Table A.4: Top 10 largest publishers

| Top publishers (2022) | Number of journals | Occurrences |
|---|---|---|
| Springer | 3763 | 4,235 (16.6%) |
| Taylor & Francis | 2912 | 1,452 (5.6%) |
| Elsevier | 2674 | 3,717 (14.5%) |
| Wiley | 1691 | 1,750 (6.8%) |
| SAGE | 1208 | 723 (2.8%) |
| OMICS | 705 | 7 (0.02%) |
| De Gruyter | 513 | 97 (0.38%) |
| Oxford University Press | 500 | 65 (0.25%) |
| InderScience | 472 | 0 (0.00%) |
| Brill | 461 | 3 (0.01%) |

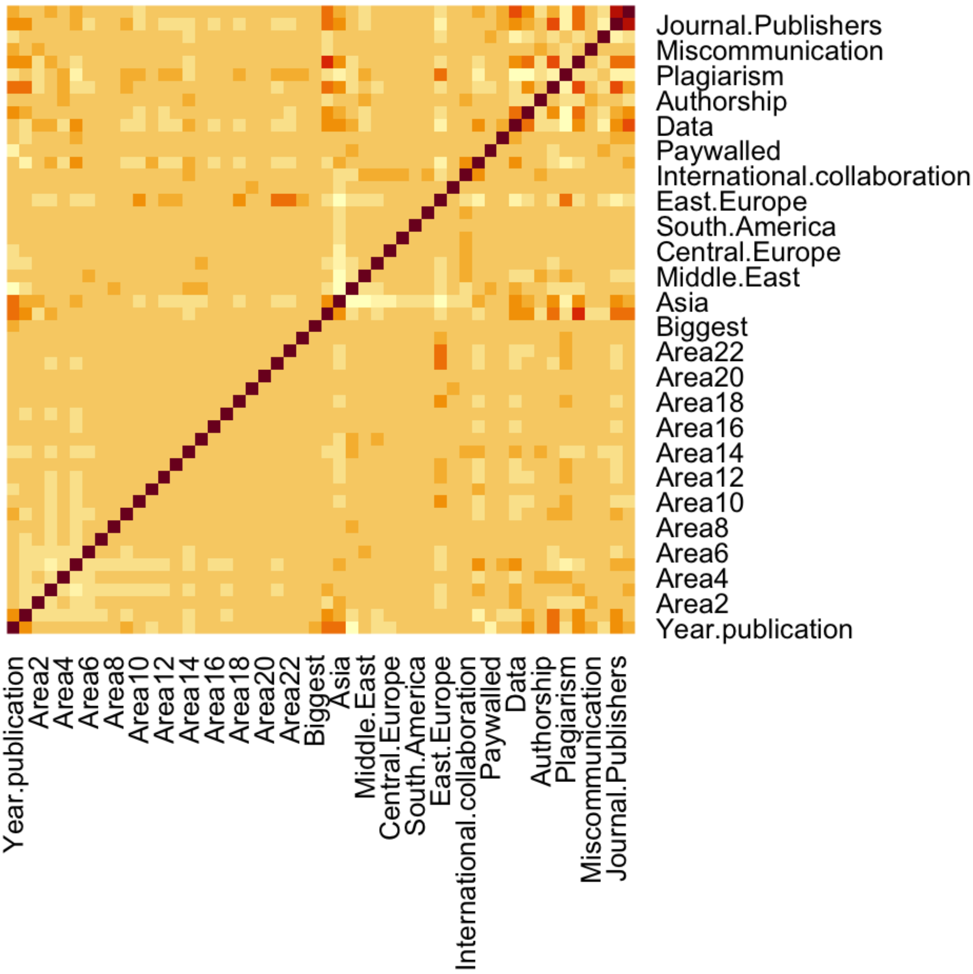
Note: This table lists the top 10 largest publishers by the number of managed journals. The occurrences of each publisher in the database are also provided.

Table A.5: List of endogenous and exogenous variables

| Categories | Variable names | Data types | Descriptions |
|---|---|---|---|
| Time to retraction (endogenous) | Time | Numeric | Difference in days between retraction date and publication date |
| Time period | Year | Numeric | Year of publication |
| Journal charateristics | Area | Binary (24 dummies) | Domain of the journals |
| | Paywalled | Binary | 1 = Subscription-based, 0 = Open-access |
| Authors information | Geo | Binary (12 dummies) | Geographical zone of corresponding (or first) authors |
| | Inter | Binary | 1 = collaboration among authors from various countries, 0 = no international collaboration |
| | Team | Numeric | Number of authors |
| Publisher characteristics | Biggest | Binary | 1 = Publisher is among the top 10 largest in the world by managed journals |
| | Predatory | Binary | 1 = Predatory publisher, 0 = Not predatory |
| Reasons of retractions | Data | Binary | Any concerns/issues about data, images, figures, and materials (including manipulation, falsification, fabrication, duplication, contamination, etc.) |
| | Result | Binary | Any concerns/issues about analyses, results, conclusions, methods, or text (including manipulation, fabrication, and duplication) |
| | Authorship | Binary | Any question/dispute about authorship or affiliation (including complaints from companies/institutions/third parties) |
| | Peer | Binary | Any concerns/issues with peer review due to journal error (including duplicate publication, rogue editor) |
| | Plagiarism | Binary | Any concerns about duplication, self-plagiarism, or plagiarism of any part of the text |
| | Reference | Binary | Any question/controversy/dispute over proper crediting of ideas, analyses, text, or data (including citing retracted works) |
| | Withdraw | Binary | Article retracted as part of the journal's regular process of updating guidelines or reviews (include journal/publisher transferred an article from one platform to another) |
| | Miscommunication | Binary | Issues related to miscommunication between authors, journals, publishers, or third parties (including nonpayment of fees) |
| | Ethics | Binary | Any issues regarding ethical violations by authors/institutions/third parties (including criminal proceedings, legal threats, etc.) |
| | None | Binary | No reasons reported |
| Investigation | Investigation1 | Binary | Investigation by Journal/Publisher of allegations |
| | Investigation2 | Binary | Investigation conducted by company, institution, or third party |

Note: This table presents the full list of variables used in our paper, including their names, types, and detailed descriptions.

Figure A.1: Heatmap of variable correlations




Note: This figure depicts the correlation matrix for the variables included in our model. The intensity of the color represents the strength of the correlation, with darker shades indicating stronger relationships.

# B Robustness checks and sensitivity analysis

This section presents the robustness checks and sensitivity analysis of our approach as discussed in the main body of the paper. Specifically, we conduct: (i) a test of proportional hazards; (ii) model comparisons; and (iii) a bootstrap variant of the AFT-Weibull estimation.

## B.1 Tests of proportional hazards assumption

The Cox regression model relies on the proportional hazards (PH) assumption, which states that covariates have a multiplicative and time-invariant effect on the hazard. Violations of this assumption can result in biased inferences. Table B.1 reports the results of the Schoenfeld residual test, which evaluates the correlation between scaled Schoenfeld residuals and time. Non-significant results indicate that the PH assumption holds, while significant results suggest a violation, implying time-dependent covariate effects.

We reject the null hypothesis of proportional hazards both globally and for individual covariates, except for "Sociology; Anthropology; Ethnology" and "Multidisciplinary" domains.

Table B.1: Test for proportional hazards assumption

| Variables | Chi-square statistic | P-values |
|---|---|---|
| **Time period** | | |
| Year publication | 783.3 | 0.000*** |
| **Journal characteristics** | | |
| Computer Science; Data Science, Information Systems; Robotics | 243.1 | 0.000*** |
| Pharmacology, Toxicology and Pharmaceutics | 1.626 | 0.202 |
| Medicine | 17.42 | 0.000*** |
| Engineering | 20.94 | 0.000*** |
| Biochemistry, Genetics and Molecular Biology | 75.95 | 0.000*** |
| Chemistry engineering; Chemistry; Material Sciences | 148.1 | 0.000*** |
| Physics | 28.10 | 0.000*** |
| Neuroscience | 13.92 | 0.000*** |
| Earth and Planetary Sciences; Environmental Science; Ecology; Energy | 72.12 | 0.000*** |
| Arts and Humanities; History | 16.19 | 0.000*** |
| Immunology; Microbiology; Virology | 18.69 | 0.000*** |
| Business; Management; Accounting | 20.89 | 0.000*** |
| Sociology; Anthropology; Ethnology | 0.439 | 0.507 |
| Multidisciplinary | 2.267 | 0.132 |
| Psychology | 25.02 | 0.000*** |
| Mathematics | 15.91 | 0.000*** |
| Agricultural and Biological Sciences | 43.23 | 0.000*** |
| Economics; Finance; Econometrics | 33.36 | 0.000*** |
| Dentistry | 8.037 | 0.000*** |
| Veterinary | 6.252 | 0.012** |
| Education | 73.14 | 0.000*** |
| Law; Administrative Science; Military Science | 103.5 | 0.000*** |
| Political Science | 25.43 | 0.000*** |
| Paywalled | 14.63 | 0.000*** |
| **Authors information** | | |
| Asia | 106.2 | 0.000*** |
| Oceania | 21.08 | 0.000*** |
| Middle East | 24.08 | 0.000*** |
| Africa | 6.895 | 0.008*** |
| North America | 68.26 | 0.000*** |
| South America | 29.29 | 0.000*** |
| North Europe | 8.389 | 0.003*** |
| Central Europe | 38.95 | 0.000*** |
| South Europe | 21.08 | 0.000*** |
| West Europe | 109.3 | 0.000** |
| East Europe | 546.1 | 0.000*** |
| International collaboration | 40.66 | 0.000*** |
| log(number authors +1) | 11.44 | 0.000*** |
| **Publishers characteristics** | | |
| Top 10 publishers | 0.004 (0.001) | 0.000*** |
| Predatory | -0.277 (0.0019) | 0.000*** |
| **Reasons of retraction** | | |
| Data | 359.2 | 0.000*** |
| Results | 474.9 | 0.000*** |
| Authorship | 9.231 | 0.002*** |
| Peer | 314.5 | 0.000** |
| Plagiarism | 102.0 | 0.000*** |
| References | 851.9 | 0.000*** |
| Miscommunication | 11.36 | 0.078 |
| Ethics | 16.92 | 0.000*** |
| None | 1415.8 | 0.000*** |
| **Investigation** | | |
| Journal/Publishers | 1022.7 | 0.000*** |
| Company/Institutions/Third Party | 530.1 | 0.000*** |
| **Global** | **4231** | **0.000*** ** |

Note: This table reports the results of the Schoenfeld residual test for the proportional hazards assumption. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively, and signify rejection of the proportional hazards assumption.

## B.2 Model comparisons

In Table B.2, we compare the performance of the various models employed in our analysis. The evaluation is based on several key metrics, including: (i) Bayesian Information Criterion (BIC), which measures the trade-off between model fit and complexity; (ii) Concordance Index (C-index), which assesses the model's ability to correctly rank survival times; (iii) Brier Score, which quantifies the accuracy of probabilistic predictions; (iv) Integrated Absolute Error (IAE), and (v) Integrated Squared Error (ISE), both of which measure the overall prediction error across the survival time. Based on the results, the AFT-Weibull model demonstrates the best overall performance, with the lowest BIC and strong predictive accuracy as reflected by the C-index. Additionally, the model consistently shows lower error rates in terms of Brier Score, IAE, and ISE, further confirming its robustness in modeling the survival time of zombie papers.

Table B.2: Model performance comparison

| | BIC | C-index | Brier | IAE | ISE |
|---|---|---|---|---|---|
| **Proportional Hazards** | | | | | |
| Cox | 436243 | 0.671 | 0.637 | 179 | 10.4 |
| Cox frailty | 437788 | 0.290 | 0.703 | 353 | 50.8 |
| **Non-Propotional Hazards** | | | | | |
| Cox weighted | 420336 | 0.538 | 0.602 | 230 | 30.8 |
| AFT-Weibull | 371891 | 0.674 | 0.564 | 135 | 5.31 |
| AFT-Log Normal | 382087 | 0.669 | 0.568 | 288 | 28 |
| AFT-Gaussian | 387401 | 0.670 | 0.613 | 179 | 44.3 |
| Random Survival Forest | X | 0.311 | 0.581 | 180 | 11.4 |

Note: This table compares the performance of the models using (i) Bayesian Information Criterion (BIC), (ii) Concordance Index (C-index), (iii) Brier Score, (iv) Integrated Absolute Error (IAE), and (v) Integrated Squared Error (ISE).

## B.3 Bootstrap AFT-Weibull estimation

To account for potential estimation uncertainty, we implemented a bootstrap-based AFT-Weibull model. Table B.3 presents the results derived from 5,000 replications, reinforcing the robustness of our main findings.

Table B.3: Boostrap results of AFT-Weibull model

| | Total | 2-years | 2 to 5-years | >5-years |
|---|---|---|---|---|
| Baseline (intercept) | 151.7*** | 26.28*** | 26.03*** | 37.17*** |
| **Time period** | | | | |
| Year publication | 0.930*** | 0.989*** | 0.990*** | 0.985*** |
| **Journal characteristics** | | | | |
| Computer Science; Data Science, Information Systems; Robotics | 1.656 | 0.828 | 1.544** | 1.102 |
| Pharmacology, Toxicology and Pharmaceutics | 1.283 | 0.722 | 1.391* | 1.168** |
| Medicine | 1.252 | 0.669 | 1.408* | 1.117** |
| Engineering | 1.506 | 0.759 | 1.364 | 1.147** |
| Biochemistry, Genetics and Molecular Biology | 1.466 | 0.730 | 1.379* | 1.182*** |
| Chemistry engineering; Chemistry; Material Sciences | 1.410 | 0.666 | 1.339 | 1.190*** |
| Physics | 1.074 | 0.659 | 1.355 | 1.165** |
| Neuroscience | 1.347 | 0.676 | 1.402* | 1.208*** |
| Earth and Planetary Sciences; Environmental Science; Ecology; Energy | 0.933 | 0.556 | 1.416* | 1.100 |
| Arts and Humanities; History | 1.484 | 0.811 | 1.447* | 1.113 |
| Immunology; Microbiology; Virology | 1.185 | 0.646 | 1.334 | 1.183** |
| Business; Management; Accounting | 1.482 | 0.708 | 1.484** | 1.096 |
| Sociology; Anthropology; Ethnology | 1.404 | 0.693 | 1.465* | 1.084 |
| Multidisciplinary | 1.574 | 0.736 | 1.390* | 1.193*** |
| Psychology | 1.339 | 0.700 | 1.422* | 1.218*** |
| Mathematics | 1.258 | 0.699 | 1.510** | 1.073 |
| Agricultural and Biological Sciences | 0.909 | 0.635 | 1.351 | 1.191*** |
| Economics; Finance; Econometrics | 1.365 | 0.824 | 1.488** | 1.108 |
| Dentistry | 1.499 | 0.767 | 1.327 | 1.209*** |
| Veterinary | 1.215 | 0.723 | 1.483* | 1.171 |
| Education | 1.291 | 0.784 | 1.387* | 1.063 |
| Law; Administrative Science; Military Science | 1.502 | 0.825 | 1.446* | 1.071 |
| Political Science | 1.684 | 0.807 | 1.373 | 1.008 |
| Paywalled | 0.705*** | 1.088** | 0.938*** | 0.919*** |
| **Authors information** | | | | |
| Asia | 1.122*** | 1.113* | 1.003 | 1.037 |
| Oceania | 0.909** | 0.985 | 0.943 | 1.095** |
| Middle East | 1.291 | 1.154** | 1.025 | 1.114*** |
| Africa | 0.902 | 1.089 | 0.982 | 0.990 |
| North America | 0.933 | 1.025 | 0.976 | 1.048* |
| South America | 0.803*** | 0.843** | 1.042 | 1.014 |
| North Europe | 0.762*** | 1.083 | 1.048 | 1.013 |
| Central Europe | 0.838*** | 0.946 | 0.954 | 1.000 |
| South Europe | 1.044 | 0.969 | 1.012 | 1.037 |
| West Europe | 0.944 | 0.970 | 0.996 | 1.036 |
| East Europe | 1.717*** | 1.478*** | 1.112*** | 0.983 |
| International collaboration | 0.996 | 1.006 | 1.007 | 1.012 |
| log(number authors +1) | 1.086*** | 1.055*** | 1.023*** | 1.011 |
| **Publishers characteristics** | | | | |
| Top 10 publishers | 1.046*** | 1.019 | 0.991* | 1.036*** |
| Predatory | 0.762*** | 0.963* | 0.948*** | 0.957*** |
| **Reasons of retraction** | | | | |
| Data | 1.510*** | 1.103*** | 1.047*** | 1.043*** |
| Results | 1.037*** | 1.058*** | 1.019*** | 0.985* |
| Authorship | 1.025 | 1.042 | 0.995 | 1.026* |
| Peer | 0.968* | 0.969 | 0.983* | 1.038*** |
| Plagiarism | 1.125*** | 0.918*** | 1.019** | 1.041*** |
| References | 0.834*** | 1.139*** | 0.903*** | 0.945** |
| Miscommunication | 1.032* | 1.023 | 1.037*** | 0.982 |
| Ethics | 1.001 | 1.042** | 0.999 | 0.990 |
| None | 0.643*** | 0.545*** | 1.006 | 0.993 |
| **Investigation** | | | | |
| Journal/Publishers | 1.160*** | 1.182*** | 0.993 | 1.006 |
| Company/Institutions/Third Party | 0.999 | 0.977 | 0.998 | 0.987 |

Note: This table reports the estimated coefficients (in exponential form) from the bootstrap AFT-Weibull model applied to the full sample, as well as to early, mid, and late stages of retraction. The results are based on 5,000 bootstrap replications. Statistical significance is denoted by ***, **, and * for the 1%, 5%, and 10% levels, respectively.

# C Zombies Population Decay Dynamics (ZPDD)

## C.1 Derivation and illustration of ZPDD

To model the evolution of the zombie paper population over time and the effect of retraction efforts on its dynamics, we choose a negative logistic decay model from ecology, adapted to our context with a retraction rate derived from the predicted time to retraction $T$ (from the AFT model). This model demonstrates how the population of zombie papers decays over time toward a carrying capacity $K$.

The model is defined as:

$$\frac{dZ}{dt} = -r \times (Z - K) \tag{C.1}$$

where $(Z - K)$ represents the number of retractable zombie papers at any given time, assuming that $K$ papers are resistant to retraction. The negative sign (unlike in a traditional logistic growth model) indicates decay rather than growth, as our goal is to reduce $Z$ over time.

Substituting $r = \frac{1}{T}$ into the model:

$$\frac{dZ}{dt} = -\frac{1}{T} \times (Z - K) \tag{C.2}$$

indicates that the rate of decay of $Z$ is proportional to the number of retractable zombie papers $(Z - K)$, with the effectiveness of retraction efforts represented by $\frac{1}{T}$. Smaller $T$ values (faster retraction) result in a higher rate of decay.

Rewriting the previous equation as:

$$\frac{dZ}{Z - K} = -\frac{1}{T} dt \tag{C.3}$$

and integrating both sides with respect to $Z$ and $t$:

$$\int \frac{1}{Z - K} dZ = \int -\frac{1}{T} dt \tag{C.4}$$

yields

$$\ln \mid Z - K \mid = -\frac{t}{T} + C \tag{C.5}$$

where $C$ is the constant of integration, allowing the solution to match initial conditions, such as the initial zombie population $Z_0$ at $t = 0$ in our context.

To solve for $Z$, we exponentiate both sides of the previous equation:

$$e^{\ln |Z-K|} = e^{-\frac{t}{T}+C} \tag{C.6}$$

Simplifying this expression, and letting $C' = e^C$, we have:

$$\mid Z - K \mid = C'.e^{-\frac{t}{T}} \tag{C.7}$$

Assuming the initial condition $Z(0) = Z_0$ at $t = 0$, we find:

$$Z_0 - K = C'.e^0 = C' \tag{C.8}$$

Substituting $C' = Z_0 - K$ back into Equation (C.7) gives the final solution describing the decay of zombie papers over time:

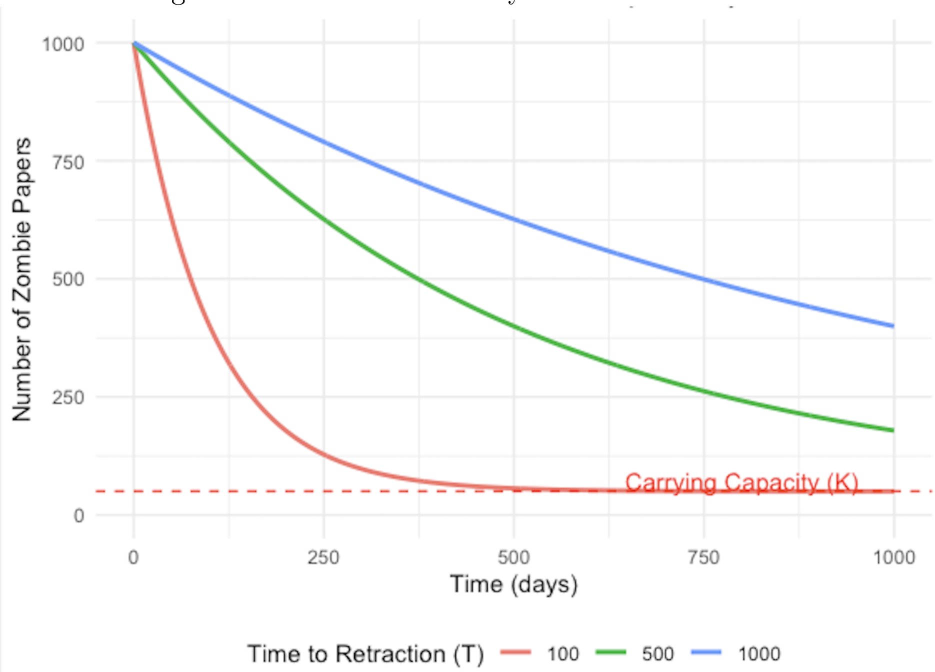
$$Z(t) = K + (Z_0 - K).e^{-\frac{t}{T}} \tag{C.9}$$

where $Z_0$ is the initial population of zombie papers, $T$ is the predicted time to retraction (from the AFT-Weibull model) controlling the rate of decay, and $K$ is the carrying capacity (the minimum achievable level of zombie papers).

Figure C.1 illustrates the theoretical decay of zombie papers over time for different time-to-retraction values $T$ (100, 500, and 1000 days) using the ZPDD model described above. These values simulate the impact of editorial policies on reducing $T$, which in turn increases the retraction rate $r = \frac{1}{T}$. A lower $T$ (e.g., $T = 100$ days, shown in red) represents a more effective editorial environment, resulting in a higher retraction rate and faster decay. In contrast, a higher $T$ (e.g., $T = 1000$ days, shown in blue) reflects a less effective editorial environment, leading to slower decay. The dashed red line at $K = 50$ represents the minimum population level, or "carrying capacity" of zombie papers—those that are highly resistant to retraction, even under strong editorial efforts. As $Z$ approaches $K$, the decay rate decreases, illustrating the diminishing returns in retraction efforts as only the most entrenched zombie papers remain.

This model provides a robust framework for understanding the decay of zombie papers and evaluating the impact of various editorial policy environments. The decay rate is directly influenced by the predicted time to retraction $T$, derived from the empirical AFT model, thus aligning the theoretical model with empirical findings. By adjusting $T$, we simulate the effects of different editorial policies on retraction dynamics, as discussed in Section 2.2. A lower $T$ represents more effective policy environments, resulting in faster decay of zombie papers, while higher values of $T$ indicate less effective policies, leading to slower decay. The model incorporates a minimum population threshold $K$, representing zombie papers that are resistant to retraction. As $Z$ (the population of zombie papers) approaches $K$, the decay rate further decreases, reflecting the diminishing returns often observed in retraction efforts for deeply entrenched papers. This decay dynamic aligns with principles in population dynamics, where populations asymptotically approach a limit, making the model intuitive and suitable for analyzing retraction challenges. By capturing both immediate and long-term effects of editorial policies, this model serves as a flexible tool for evaluating policy effectiveness and understanding the persistence of zombie papers over time.

Figure C.1: Theoretical decay of Zombies over time




Note: This figure presents the theoretical decay of zombie papers over time with respect to different time to retraction scenarios, respectively for $T = 100$, 500, and 1000 days. Initial population $Z_0 = 1000$ and carrying capacity $K = 0.05 \times Z_0$.

### C.2 Zombie paper persistence and optimal mix of editorial policy interventions

To effectively reduce the number of zombie papers by a specific future time horizon, we define an optimization problem where the objective function represents the zombie paper population $Z(T_{max})$ at the target time $T_{max}$. The aim is to minimize this population by adjusting the levels of editorial policy interventions that target key retraction issues, balancing intervention effectiveness with cost.

The objective function is defined as:

$$\min_{p_1,p_2,p_3} \int_0^{T_{\max}} Z(t)dt + \theta(p_1 + p_2 + p_3) \tag{C.10}$$

where $p_1$, $p_2$, and $p_3$ denote intervention levels aimed at improving data transparency, result reproducibility, and plagiarism detection, respectively. $T_{max}$ is the duration over which we seek to minimize the zombie paper population (1,000 days in our context), and $Z(t)$ is governed by ZPDD model. $\theta(p_1 + p_2 + p_3)$ is the quadratic penalty term that discourages high intervention levels unless they yield substantial reductions in $Z(T_{max})$, with $\theta$ representing the penalty weight. This quadratic penalty term rises steeply as each $p_i$ $(i = 1, 2, 3)$ approaches its upper limit, promoting resource-efficient intervention levels.

We express the retraction time $T$ as a function of the intervention levels $p_1$, $p_2$, and $p_3$:

$$T = T_0 - \alpha_1 p_1 - \alpha_2 p_2 - \alpha_3 p_3 \tag{C.11}$$

where $T_0$ is the baseline retraction time with no interventions, and $\alpha_1$, $\alpha_2$, and $\alpha_3$ are positive coefficients that quantify the effectiveness of each intervention in reducing $T$.

Each intervention level is constrained between 0 (no intervention) and 1 (maximum feasible intervention):

$$0 \leq p_1, p_2, p_3 \leq 1 \tag{C.12}$$

This constraint keeps interventions within realistic and feasible bounds.

To determine the optimal mix of interventions, we employ a numerical simulation approach. For each combination of intervention levels ($p_1$, $p_2$, and $p_3$), the model adjusts the retraction time $T$ according to the specified effectiveness parameters, then simulates the zombie paper population $Z(t)$ over the defined period. The objective is to minimize the zombie population at a specific future time point $T_{max}$, by iteratively refining the intervention levels. At each iteration, the model recalculates $T$ and $Z(t)$, while also incorporating a penalty term that discourages excessive interventions unless they yield substantial reductions in $Z(T_{max})$. This process continues until the objective function, defined as $Z(t_{max})$ plus the penalty term, reaches its minimum value. The resulting intervention levels represent the optimal, cost-effective policy mix for minimizing the zombie paper population at the target time horizon.