

Faculty of Economics and Business Administration, ICADE

Does the Press Favour Its Own? Home-Country Bias in Pharmaceutical Media Coverage

Author: Claudia Montero Meléndez

Director: Carlos Bellón Núñez-Mera

MADRID | June 2026

Abstract

Financial news, like any journalistic output, may reflect some degree of author and publisher subjectivity. Home-country bias may emerge from editorial and audience incentives, denser local sourcing networks, regulatory and cultural context, or informational proximity to the "home" firm. This thesis tests for such bias by analysing the coverage of Johnson & Johnson (US) and AstraZeneca (UK) in flagship financial newspapers from both countries between 2000 and 2025.

The study analyses whether the domestic firm receives more coverage and more favourable sentiment through Natural Language Processing techniques, with sentiment adjusted for each outlet's baseline tone. Sentiment is compared across countries, individual newspapers, article sections and specific event windows, with particular attention to the Covid-19 vaccine period. The results suggest statistically significant evidence consistent with home-country bias, particularly in the body of articles, although newspaper type, article section and event context strongly affect sentiment and make the pattern less uniform.

Overall, the thesis shows that media bias should not be interpreted only as explicit positive coverage of domestic firms. It can also appear as greater visibility, softer negative framing, or more favourable narrative development within comparable news contexts.

Key words: Media bias; home-country bias; financial journalism; pharmaceutical industry; sentiment analysis; Natural Language Processing; VADER; AstraZeneca; Johnson & Johnson; Covid-19

Resumen:

Las noticias financieras, como cualquier producto periodístico, pueden reflejar cierto grado de subjetividad por parte del autor y del medio que las publica. El sesgo de país de origen puede surgir de incentivos editoriales y de audiencia, redes locales de fuentes más densas, el contexto regulatorio y cultural, o la proximidad informativa hacia la empresa "doméstica". Este trabajo analiza la posible existencia de este sesgo mediante el estudio de la cobertura de Johnson & Johnson (US) y AstraZeneca (UK) en periódicos financieros de referencia de ambos países entre 2000 y 2025.

El estudio examina si la empresa doméstica recibe mayor cobertura y un sentimiento más favorable mediante técnicas de Procesamiento del Lenguaje Natural, ajustando el sentimiento según el tono base de cada medio. El sentimiento se compara entre países, periódicos individuales, secciones de artículos y ventanas temporales específicas, prestando especial atención al periodo de las vacunas contra la COVID-19. Los resultados sugieren evidencia estadísticamente significativa consistente con la existencia de sesgo de país de origen, aunque el tipo de periódico, la sección del artículo y el contexto del evento afectan de forma significativa al sentimiento y hacen que el patrón no sea uniforme. En conjunto, el trabajo muestra que el sesgo mediático no debe interpretarse únicamente como una cobertura explícitamente positiva de las empresas domésticas. También puede manifestarse como una mayor visibilidad, un encuadre negativo más moderado o desarrollo narrativo más favorable en contextos informativos comparables.

Palabras clave: sesgo mediático; sesgo de país de origen; periodismo financiero; industria farmacéutica; análisis de sentimiento; Procesamiento del Lenguaje Natural; VADER; AstraZeneca; Johnson & Johnson; Covid-19

Table of contents:

Section I: About the project 10
1. Motivation and research questions 10
2. Objectives and expected contributions 11
3. Thesis roadmap 12
Section II: Media Bias 13
1. Home-country media bias 13
2. Media sentiment, attention, and asset prices 14
3. Using Difference-in-Differences Event-Study Methods in Finance 16
Section III: Methodology 20
1. Research design and comparative logic 20
2. Data sources and sample construction 21
3. Sentiment measurement: VADER 22
4. Entity-level sentiment attribution 23
5. Newspaper-level sentiment normalization 25
6. Difference-in-Differences 26
Section IV: The Pharma Industry & Context (2000 to 2025) 28
1. Pharma landscape and relevance for U.S. and U.K. 28
2. Timeline of major firm and industry events (2000-2025) 29
3. Firm profiles and news: J&J (U.S.) and AstraZeneca (U.K.) 31
Johnson & Johnson 31
AstraZeneca 31
Section V: Exploratory Data Analysis 34
1. Basic Exploratory Data Analysis 34
2. Coverage Intensity Analysis: Country-bias 36
3. Exploratory Analysis of Newspaper Tone 37
Section VI: Sentiment Analysis 41
1. Sentiment Analysis Comparison 41
2. Robustness consideration: single-company articles 69
3. Overall interpretation of sentiment analysis 71
Section VII: Conclusions 73
1. Answer to research questions 73
2. Implications for media, investors and firms 75
3. Limitations and future research 76
Bibliography 78

Appendix 81

1. Reproducibility package: Python code 81

Preprocessing of data 81

List of figures:

Figure 1. Distribution of articles by country of newspaper source

Figure 2. Distribution of articles by company mentioned

Figure 3. Distribution of articles by year period (2000–2025)

Figure 4. Distribution of articles by type of article section

Figure 5. Company mentioned by country of newspaper source

Figure 6. Company mentioned in title by country of newspaper source

Figure 7. Average sentiment z-score by newspaper

Figure 8. Adjusted title sentiment by country and company

Figure 9. Adjusted snippet sentiment by country and company

Figure 10. Adjusted body sentiment by country and company

Figure 11. Newspaper-adjusted body sentiment by newspaper and company

Figure 12. Newspaper-adjusted body sentiment by newspaper and company (Business, Economy and Finance articles only)

Figure 13. Newspaper-adjusted body sentiment by newspaper and company (Frontpage and Main sections only)

Figure 14. Newspaper-adjusted body sentiment by newspaper and company (Health articles only)

Figure 15. Adjusted body sentiment by country and company: Vaccine development phase (March 2020 – March 2021)

Figure 16. Adjusted body sentiment by country and company: AstraZeneca vaccine authorization (December 2020)

Figure 17. Adjusted body sentiment by country and company: Johnson & Johnson vaccine authorization (February 2021)

Figure 18. Adjusted body sentiment by country and company: AstraZeneca vaccine suspension (March 2021)

Figure 19. Adjusted body sentiment by country and company: Johnson & Johnson vaccine suspension (April 2021)

Figure 20. Adjusted title sentiment by country and company: Johnson & Johnson vaccine suspension (April 2021)

List of tables:

Table 1. Key events timeline for Johnson & Johnson (2000–2025)

Table 2. Key events timeline for AstraZeneca (2000–2025)

Table 3. Average sentiment by newspaper without adjustment

Table 4. Average raw sentiment by country of newspaper source

Table 5. Adjusted sentiment by country and company (title, snippet and body)

Table 6. Difference-in-Differences Regression Results for Adjusted Body Sentiment

Table 7. Adjusted sentiment by newspaper and company

Table 8. Adjusted sentiment by country and company: Vaccine development phase (March 2020 – March 2021)

Table 9. Adjusted sentiment by country and company: AstraZeneca vaccine authorization (December 2020)

Table 10. Adjusted sentiment by country and company: Johnson & Johnson vaccine authorization (February 2021)

Table 11. Adjusted sentiment by country and company: AstraZeneca vaccine suspension (March 2021)

Table 12. Adjusted sentiment by country and company: Johnson & Johnson vaccine suspension (April 2021)

Table 13. Difference-in-Differences Regression Results for Adjusted Body Sentiment

Table 14. Single-Company Robustness Check of Adjusted Sentiment by Company

List of acronyms:

AZ — AstraZeneca

J&J — Johnson & Johnson

UK — United Kingdom

US — United States

Covid-19 — Coronavirus Disease 2019

DiD — Difference-in-Differences

NLP — Natural Language Processing

VADER — Valence Aware Dictionary and sEntiment Reasoner

WSJ — The Wall Street Journal

CDC — Centers for Disease Control and Prevention

EMA — European Medicines Agency

FDA — Food and Drug Administration

MHRA — Medicines and Healthcare products Regulatory Agency

Section I: About the project

1. Motivation and research questions

Media coverage plays a central role in shaping how firms are perceived by investors, consumers, regulators and the general public. This is particularly relevant in the pharmaceutical industry, where companies are not only evaluated as economic actors, but also as providers of products with direct implications for public health. In this context, newspapers may influence reputational dynamics by deciding how much attention to give to a firm, which events to emphasize and what tone to use when describing its actions.

This thesis focuses on the possibility of home-country media bias. The central idea is that newspapers may cover domestic firms differently from foreign firms, either because domestic companies are more relevant to their audience, because journalists have stronger informational links with domestic institutions, or because national narratives shape editorial framing. Such bias may appear through greater coverage intensity, more favourable sentiment, or softer treatment during negative events.

The pharmaceutical sector provides a particularly useful setting for this analysis. AstraZeneca and Johnson & Johnson are both large multinational healthcare companies, but they are strongly associated with different national contexts: AstraZeneca with the United Kingdom and Johnson & Johnson with the United States. Both firms also became highly visible during the Covid-19 pandemic due to their vaccine development, regulatory approvals and later safety-related concerns. This makes them suitable for comparing how British and American newspapers frame domestic and foreign pharmaceutical firms under similar industry conditions.

The main research question of this thesis is: Do British and American newspapers display home-country bias when covering AstraZeneca and Johnson & Johnson?

This general question is divided into the following sub-questions:

- Do newspapers publish more articles about the company associated with their own country?
- Do domestic companies receive more favourable sentiment than foreign companies after adjusting for each newspaper's baseline tone?
- Does home-country bias appear differently in titles, snippets and article bodies?
- Does the pattern vary across individual newspapers and article sections?
- Do specific event windows, during the Covid-19 vaccine period, reinforce or weaken the evidence of home-country bias?

2. Objectives and expected contributions

The main objective of this thesis is to evaluate whether home-country media bias is present in the coverage of AstraZeneca and Johnson & Johnson by British and American newspapers.

The specific objectives are:

1. To build a structured dataset of newspaper articles covering AstraZeneca and Johnson & Johnson in UK and US newspapers from 2000 to 2025, and to assess asymmetries in the number of articles per newspaper and firm over time.
2. To measure and compare media coverage and sentiment across countries, accounting for their source and article sections, using VADER sentiment analysis, entity-level attribution and newspaper-level sentiment adjustment.

3. To assess whether the observed patterns are consistent with home-country bias, including the analysis of selected event windows, especially during the Covid-19 vaccine period as a central event in the industry.

3. Thesis roadmap

This thesis is structured in seven sections.

Following this introduction, Section II reviews the relevant literature on media bias, sentiment analysis and event-study methods. Section III describes the methodology, covering the research design, sentiment measurement approach and the Difference-in-Differences method. Section IV provides context on the pharmaceutical industry and the two firms analyzed. Section V describes Data Exploratory Analysis. Section VI presents sentiment analysis results across countries, newspapers, article sections and event windows. Section VII concludes with answers to the research questions, implications and limitations.

Section II: Media Bias

1. Home-country media bias

Media bias has long been a central topic in political economy and communication studies, referring to systematic deviations in coverage that favor particular actors, perspectives, or interests (Gentzkow & Shapiro, 2010). Within this broader concept, home-country bias refers to the tendency of media outlets to disproportionately cover and/or positively frame domestic entities relative to foreign counterparts.

The existence of home-country bias can be explained through several complementary mechanisms:

Demand-side explanation

Media organizations cater to the preferences of their audience, which is typically more interested in domestic firms and issues (Mullainathan & Shleifer, 2005). Readers derive greater utility from information that is geographically, culturally, and economically proximate, leading news outlets to allocate more attention to domestic firms.

Supply-side factors

Journalists tend to have denser information networks within their home country, facilitating access to sources, corporate insiders, and regulatory bodies. This informational proximity reduces reporting costs and increases the likelihood of domestic coverage (Coval & Moskowitz, 1999). Additionally, institutional and regulatory environments may implicitly favor domestic narratives, especially in sectors with strong national relevance such as healthcare or pharmaceuticals.

Organizational and editorial incentives

Media outlets often depend on domestic advertising markets and corporate relationships, potentially creating subtle incentives to portray domestic firms more favorably (Dyck & Zingales, 2003). Even absent explicit pressure shared cultural and institutional frameworks can shape journalists' framing choices.

Empirical evidence of home-country bias has been documented across multiple domains. While much of the literature originates in finance (e.g., investor home bias), related evidence exists in publishing and information production. For instance, Zhu (2021) shows that the *New England Journal of Medicine* publishes a disproportionately high share of U.S.-based papers (65.44%) compared to other journals. Importantly, Zhu (2021) also finds that U.S.-based articles exhibit lower average research impact, measured through citation counts, than non-U.S. articles. This suggests that home-country bias may manifest not only in quantity of output but also in quality thresholds, implying preferential selection mechanisms.

Translating these insights to financial journalism, home-country bias may appear through:

- Coverage intensity: more articles about domestic firms and greater prominence of domestic corporate events
- Framing and tone: more favorable sentiment

This thesis builds on this framework by examining whether such bias exists in financial newspapers when covering Johnson & Johnson (U.S.) and AstraZeneca (U.K.).

2. Media sentiment, attention, and asset prices

Beyond coverage volume, the tone and sentiment of news play a crucial role in financial markets. A large body of literature demonstrates that media content affects investor beliefs, attention allocation, and ultimately asset prices.

Media as an intermediary

Traditional finance assumes that prices reflect all available information (Fama, 1970). However, behavioral finance highlights frictions in information processing. Media acts as a key intermediary, transforming complex corporate information into digestible narratives (Tetlock, 2007). This transformation is not neutral, since framing, word choice, or emphasis influence how investors interpret information.

Tetlock (2007) provides seminal evidence that negative media sentiment predicts downward pressure on stock prices, particularly in the short term. This effect is attributed to both information dissemination and behavioral biases such as overreaction.

Investor´s limited processing capacity

Attention is a scarce resource. Investors cannot process all available information, so they rely on salient signals, such as media coverage, to guide decisions (Barber & Odean, 2008). Firms receiving more media attention experience:

- Higher trading volume
- Increased price volatility
- Faster incorporation of information into prices

In this context, media bias can distort attention allocation, potentially leading investors to overweight domestic firms relative to foreign ones.

Tone, sentiment, and market outcomes

Recent advances in Natural Language Processing (NLP) have enabled systematic measurement of media tone. Studies such as Loughran and McDonald (2011) show that domain-specific sentiment dictionaries significantly improve the analysis of financial texts.

Empirical findings consistently indicate that positive tone is associated with abnormal returns, negative tone predicts lower returns and higher volatility and media pessimism can amplify market downturns (Tetlock, Saar-Tsechansky, & Macskassy, 2008).

Importantly, sentiment effects are asymmetric, since negative news tends to have a stronger impact than positive news, consistent with loss aversion.

If domestic media systematically portrays home firms more positively or allocates more attention to them, this may generate reputational advantages for domestic firms, mispricing or delayed correction for foreign firms and cross-country differences in market efficiency.

This thesis contributes to this literature by jointly analyzing coverage intensity and sentiment, allowing for a comprehensive test of media bias effects.

3. Using Difference-in-Differences Event-Study Methods in Finance

General Framework

Difference-in-Differences is a widely used empirical strategy in economics and finance to compare changes in an outcome variable between a treated group and a control group before and after a specific event or treatment. The intuition behind the method is that, if both groups would have followed similar trends in the absence of the treatment, any divergence after the event can be interpreted as the effect of the treatment. This identifying condition is known as the parallel trends assumption (Angrist & Pischke, 2009).

In its simplest form, the Difference-in-Differences estimator compares the evolution of an outcome before and after an event across two groups. One group is exposed to the

treatment, while the other acts as a control group. The difference between both changes is interpreted as the estimated treatment effect. This approach is especially useful when randomized experiments are not possible but comparable groups can be observed over time.

Event-study methods are closely related and are particularly common in finance. An event study examines how an outcome variable behaves around a specific event date, usually by comparing observed outcomes with a benchmark or expected value. In financial research, event studies are often used to estimate abnormal returns around corporate announcements, regulatory changes, earnings releases, mergers, accidents or other relevant shocks (MacKinlay, 1997). By observing the dynamics before and after an event, this methodology allows researchers to assess whether markets, media coverage, or other outcomes react systematically to new information.

In the context of media and financial markets, Difference-in-Differences and event-study approaches are useful because news coverage is often concentrated around identifiable events. Regulatory decisions, litigation announcements, vaccine approvals, safety concerns, earnings reports, or major corporate decisions may trigger sharp changes in both media attention and sentiment. Comparing how different newspapers cover domestic and foreign firms around these events can help identify whether coverage patterns differ systematically.

Application to this thesis

In this thesis, the logic of Difference-in-Differences and event-study methods is used as a comparative framework to study potential home-country media bias. The analysis

compares the coverage of AstraZeneca, a company associated with the United Kingdom, and Johnson & Johnson, a company associated with the United States, across British and American newspapers.

However, it is important to clarify that this study does not have a traditional control group in the strict experimental or Difference-in-Differences sense. The dataset consists of observational news articles, and there is no group of newspapers or companies that is completely unaffected by the events being studied. Both AstraZeneca and Johnson & Johnson are exposed to pharmaceutical industry news, Covid-19 developments, regulatory decisions, and broader market conditions. Therefore, the analysis should not be interpreted as a causal DiD design with a clearly treated and untreated group.

Instead, the study follows comparative control logic. The main comparison is between domestic and foreign coverage. British newspapers are expected to be more closely connected to AstraZeneca, while American newspapers are expected to be more closely connected to Johnson & Johnson. If home-country bias is present, British newspapers would be expected to cover AstraZeneca more intensively or more favorably than Johnson & Johnson, while American newspapers would be expected to show relatively more favorable coverage towards Johnson & Johnson.

The empirical analysis therefore compares domestic and foreign coverage across several dimensions, including article volume, publishing dates, and sentiment. This comparison does not prove causality, but it allows the study to identify whether the observed patterns are consistent with the expectations of home-country bias.

The event-study component is particularly relevant for the Covid-19 vaccine period. Both AstraZeneca and Johnson & Johnson were directly involved in vaccine development and faced intense media, regulatory and political scrutiny. This makes the pandemic period a useful setting for comparing how newspapers from both countries framed nationally associated and foreign pharmaceutical firms under similar global conditions.

In particular, the analysis pays attention to specific events such as vaccine development announcements, regulatory approvals, safety concerns, and temporary pauses or restrictions. These events generated high levels of media attention and provide suitable moments to observe whether coverage intensity and tone differ depending on the nationality of the company and the country of the newspaper.

However, the results should be interpreted cautiously. Differences in coverage may reflect real differences in the events affecting each company, not only media bias. For this reason, the empirical analysis is interpreted as evidence of patterns consistent with home-country bias, rather than as definitive causal proof.

Section III: Methodology

1. Research design and comparative logic

This thesis follows a comparative observational design aimed at identifying systematic differences in the media treatment of AstraZeneca (UK) and Johnson & Johnson (US) across British and American newspapers between 2000 and 2025. The central analytical question is whether each newspaper treats its nationally associated company more favourably than the foreign comparator, after accounting for each outlet's baseline tone.

As discussed in Section II, the design follows a comparative logic inspired by Difference-in-Differences and event-study methods, without constituting a strict causal identification strategy. The results are therefore interpreted as descriptive evidence of patterns consistent with home-country bias rather than as definitive causal proof.

The core identification strategy relies on a within-newspaper comparison. Each newspaper's own average sentiment serves as a baseline, allowing the analysis to ask not whether US newspapers are more positive than UK newspapers in absolute terms, but whether each outlet treats its domestically associated company more favourably than would be expected given its usual tone. This adjustment is essential because newspapers differ substantially in their general writing style, some of them being systematically more positive than others, and this difference must be separated from any company-specific or country-specific bias.

The analysis proceeds across four levels of comparison: country, individual newspaper, article section, and specific event windows during the Covid-19 pandemic.

2. Data sources and sample construction

News articles were retrieved from Factiva, a professional news database that provides access to full-text articles from major international newspapers. The search strategy involved downloading all articles mentioning AstraZeneca or Johnson & Johnson published in eight British and American newspapers between 2000 and 2025. The UK newspapers included in the sample are The Guardian, The Times, The Sunday Times, The Daily Telegraph and City AM. The US newspapers are The Wall Street Journal, The Washington Post and Barron's. These outlets were selected because they are flagship financial and general-interest newspapers in their respective countries, with substantial coverage of corporate and pharmaceutical news over the full sample period.

After downloading more than 14,000 articles, the dataset was cleaned by retaining only the variables relevant to the analysis: source name, title, snippet, body, publication date and section. A country variable was added to distinguish between UK and US outlets. A year variable was also derived from the publication date to allow temporal segmentation of the analysis.

Each article was then classified according to which companies it mentioned. Using Python, a function based on regular expressions searched for variations of both company names across the title, snippet and body of each article. The search patterns covered common abbreviations and alternative spellings, including "johnson", "janssen", "j&j" for Johnson & Johnson, and "astrazeneca", "astra zeneca" and "az" for AstraZeneca. Articles were assigned a numerical code: 1 if only Johnson & Johnson was mentioned, 2 if only AstraZeneca was mentioned, 3 if both companies were mentioned, and 0 if neither was found. This classification is the basis for the entity-level sentiment attribution later described in Section III.4.

Finally, articles were categorized by section type using a dictionary-based mapping applied to the section column of the dataset. Section names were cleaned and normalized to ensure consistency across newspapers, and mapped into broader categories such as Business / Economy / Finance, Frontpage / Main sections, and Health / Science / Environment. This categorization was later used as a control dimension in the sentiment comparison.

3. Sentiment measurement: VADER

Sentiment analysis was conducted using VADER (Valence Aware Dictionary and sEntiment Reasoner), a lexicon- and rule-based Natural Language Processing tool developed by Hutto and Gilbert (2014). VADER was selected for this study for three reasons. First, it is specifically designed to handle short, emotionally expressive texts and performs well on news content without requiring a manually labelled training dataset. Second, it is fully transparent: the sentiment scores derive from a validated lexicon with publicly available valence ratings, making the results reproducible and interpretable. Third, its computational efficiency makes it suitable for processing a dataset of nearly 14,000 articles.

VADER works by comparing the words in a text against a predefined sentiment lexicon in which each item carries an associated valence score indicating positive, negative or neutral emotional content. Words associated with success, approval, effectiveness, growth or strong financial performance increase the positive sentiment score, while words related to safety concerns, lawsuits, regulatory risk, treatment failure or reputational damage increase the negative sentiment score. Beyond individual words, VADER accounts for punctuation, capitalization, intensifiers and negations, so that expressions

such as "very effective", "not safe" or "major concerns" are scored differently from the same words considered in isolation.

The main output used in this study is the compound score, a normalized weighted composite measure ranging from −1 to +1. Values closer to +1 indicate a more positive tone, values closer to −1 indicate a more negative tone, and values close to 0 indicate a neutral or mixed tone. Following the standard interpretation of VADER, values above 0.05 are classified as positive, values below −0.05 as negative, and values between −0.05 and 0.05 as neutral.

VADER was applied separately to three parts of each article: the title, the snippet and the body. This distinction is relevant because each part of a news article performs a different editorial function. The title captures the most visible framing of the article and is subject to space and visibility constraints. The snippet provides a short summary that often reflects editorial priorities. The body contains the full narrative development of the story, where the newspaper has more space to frame the company, contextualize events and shape the reader's interpretation. By calculating sentiment separately for these three components, the analysis can identify whether the tone differs between the headline and the full article, which is itself a relevant dimension of editorial framing.

After obtaining the three component scores, a general sentiment measure is calculated as the simple average of the title, snippet and body compound scores.

4. Entity-level sentiment attribution

A key methodological challenge in sentiment analysis applied to news articles is that VADER produces a sentiment score for the full text, without specifying which entity that sentiment refers to. This is straightforward when an article mentions only one of the two

companies: in that case, the sentiment of the title, snippet and body can be directly attributed to that firm, since it is the only corporate subject of the article.

However, when both AstraZeneca and Johnson & Johnson appear in the same article, direct attribution becomes problematic. An article may simultaneously discuss positive regulatory developments for AstraZeneca and negative safety concerns for Johnson & Johnson, or vice versa. Assigning a single article-level sentiment score to both companies would obscure these differences and introduce noise into the comparison.

To address this, articles mentioning both companies (classified with code 3 in the dataset) were treated separately using an entity-level attribution procedure. The text was divided into individual sentences using standard sentence tokenization. VADER was then applied only to the sentences in which each company appeared, based on the same regular expression patterns used for the initial classification. This produced separate sentiment scores for AstraZeneca and Johnson & Johnson within the same article, based on the specific sentences in which each firm was mentioned.

This sentence-level approach reduces the risk of attributing the tone associated with one company to the other, and improves the interpretability of the sentiment scores in articles where both firms are discussed. It shifts the analysis from article-level sentiment to entity-level sentiment attribution for the subset of articles containing both companies, making the comparison more precise, particularly during periods such as the COVID-19 vaccine phase when both firms were frequently discussed together in the same article.

The final sentiment dataset therefore combines two types of attribution: direct article-level scores for single-company articles, and sentence-level entity scores for dual-company articles. Both are expressed as compound scores on the same −1 to +1 scale, making them directly comparable across the full sample.

5. Newspaper-level sentiment normalization

A further methodological issue identified in the exploratory analysis is that newspapers differ substantially in their baseline tone. Some outlets use systematically more positive language across all their coverage, while others adopt a more critical or restrained writing style. If this variation is ignored, raw sentiment comparisons across newspapers or countries may reflect general editorial style rather than company-specific or country-specific bias.

An exploratory z-score analysis confirmed the presence of this pattern in the sample. This difference in outlet type, rather than national origin alone, would contribute to a gap between US and UK raw sentiment averages.

To address this, each article's sentiment score was normalized at the newspaper level by subtracting the average sentiment of the newspaper in which it was published:

$$Adjusted\ Sentiment = Article\ Sentiment - Newspaper\ Average\ Sentiment$$

This transformation sets the average adjusted sentiment of each newspaper close to zero, so that each article is evaluated relative to the normal tone of its own outlet rather than relative to the full dataset. A positive adjusted sentiment score indicates that the article is more positive than that newspaper's usual coverage, while a negative adjusted score indicates that it is more negative than the outlet's own average tone.

This adjustment is central to the interpretation of home-country bias in this study. Since AstraZeneca is associated with the United Kingdom and Johnson & Johnson with the United States, the relevant question is not whether US newspapers are more positive overall, but whether each newspaper treats its nationally associated company more favourably than would be expected given its own baseline. The newspaper-adjusted

sentiment scores are the primary outcome variable used in all comparisons presented in Section VI.

6. Difference-in-Differences

To assess whether the observed domestic-versus-foreign sentiment differences are statistically significant, a Difference-in-Differences estimator is computed following a standard two-way regression approach. The intuition behind the estimator is as follows. If home-country bias is present, AstraZeneca should receive relatively more favourable sentiment in UK newspapers than in US newspapers, and Johnson & Johnson should receive relatively more favourable sentiment in US newspapers than in UK newspapers. The DiD estimator captures this double difference:

$$DiD = (AZ_UK - AZ_US) - (JJ_UK - JJ_US)$$

A positive and statistically significant DiD estimate indicates that the domestic-versus-foreign sentiment gap is larger than would be expected by chance, providing evidence consistent with home-country bias.

In practice, this estimator is obtained from the following OLS regression:

$$Sentiment_adjusted = \beta 0 + \beta 1 \cdot is_AZ + \beta 2 \cdot is_UK + \beta 3 \cdot (is_AZ \times is_UK) + \varepsilon$$

where *is_AZ* is a binary indicator equal to 1 for AstraZeneca articles and 0 for Johnson & Johnson articles, *is_UK* is a binary indicator equal to 1 for articles published in UK newspapers and 0 for US newspapers, and *is_AZ* × *is_UK* is the interaction term between the two. The coefficient β3 on the interaction term is the Difference-in-Differences estimator. A positive and significant β3 indicates that AstraZeneca receives relatively more favourable adjusted sentiment in UK newspapers than in US newspapers, beyond what Johnson & Johnson receives in UK relative to US newspapers.

The regression is estimated on the full sample using newspaper-adjusted general sentiment as the outcome variable, and replicated using body sentiment, which is the component where home-country bias is most visible in the descriptive analysis. The same specification is then applied to each event window identified in the Covid-19 vaccine period, allowing the DiD estimate to vary across phases of the pandemic narrative. To ensure that the statistical inference is valid despite differences in sample size and sentiment variability across newspapers and article types, heteroskedasticity-robust standard errors are used in all regressions

It is important to note that this regression does not include newspaper fixed effects, since the outcome variable is already adjusted at the newspaper level. The within-newspaper normalization described in Section III.5 performs a function analogous to a newspaper fixed effect by removing each outlet's average tone before the regression is estimated.

Section IV: The Pharma Industry & Context (2000 to 2025)

1. Pharma landscape and relevance for U.S. and U.K.

The pharmaceutical industry represents one of the most research-intensive and globally integrated sectors, characterized by high fixed costs, long development cycles, and strong regulatory oversight. R&D investment, patent protection, and regulatory approval processes jointly shape firm behavior and market outcomes (DiMasi, Grabowski, & Hansen, 2016).

From a geographical perspective, the United States and the United Kingdom occupy central positions in the global pharmaceutical ecosystem. The U.S. dominates in terms of market size and regulatory influence through the Food and Drug Administration (U.S. General Services Administration, 2026), while the U.K. plays a critical role as a hub for scientific research, clinical trials, and multinational headquarters.

The U.S. pharmaceutical market accounts for a disproportionately large share of global revenues, driven by higher drug prices and a market-oriented healthcare system. In contrast, the U.K. operates under a publicly funded system, where pricing and reimbursement are subject to centralized negotiation, often through institutions such as NICE (Reinhardt, Hussey, & Anderson, 2004). These structural differences create distinct informational environments, potentially influencing how media outlets frame pharmaceutical firms.

The industry is also characterized by information asymmetry and uncertainty, particularly in drug development, where outcomes are highly uncertain and binary (success/failure of

trials). As a result, news, especially those regarding clinical trials, regulatory approvals, or safety concerns, plays a critical role in shaping investor expectations.

Moreover, pharmaceutical firms are particularly exposed to reputational dynamics, given the societal importance of their products. Media coverage can influence not only financial performance but also public trust, regulatory scrutiny, and risk of litigation.

Within this context, large multinational firms such as Johnson & Johnson and AstraZeneca operate across both markets, but remain strongly associated with their home countries. This makes them ideal candidates for analyzing home-country media bias, as they are comparable in size and global reach but embedded in different national narratives.

2. Timeline of major firm and industry events (2000-2025)

The period from 2000 to 2025 has been marked by significant events that have shaped both the pharmaceutical industry and media coverage patterns. These events provide natural shocks that are particularly suitable for event-study and sentiment analysis.

Early 2000s: Patent Cliffs and Industry Consolidation

The early 2000s were characterized by the so-called "patent cliff", where major blockbuster drugs lost patent protection, leading to revenue declines and increased competition from generics (Grabowski & Kyle, 2007). This period also saw increased merger and acquisition activity as firms sought to maintain pipelines and scale economies.

Mid-2000s to 2010s: Regulatory Scrutiny and Safety Concerns

Several high-profile drug withdrawals and safety concerns increased regulatory scrutiny. Notably, cases such as Vioxx highlighted the importance of post-market surveillance and

significantly affected investor perceptions of pharmaceutical risk (Topol, 2004). These events intensified media attention on safety, ethics, and corporate responsibility.

2010s: Rise of Biologics and Innovation Shift

The 2010s marked a transition toward biologics, oncology, and personalized medicine, increasing both the complexity and potential profitability of drug development (Scannell et al., 2012). During this period, firms like AstraZeneca repositioned their pipelines toward oncology, while diversified firms like Johnson & Johnson leveraged broader healthcare portfolios.

Covid-19 Pandemic (2020–2022)

The Covid-19 pandemic represents a major structural break for the industry. Pharmaceutical firms became central actors in global crisis management, particularly through vaccine development. AstraZeneca, in collaboration with Oxford University, developed one of the first widely distributed vaccines, while Johnson & Johnson introduced a single-dose alternative.

The pandemic dramatically increased media attention, public scrutiny and political involvement in pharmaceutical activities. It also created sharp fluctuations in sentiment, driven by vaccine efficacy results, safety concerns (e.g., rare side effects), and regulatory decisions.

Post-COVID Period (2022–2025): litigation, innovation, and market Adjustment

The post-pandemic period has been marked by ongoing litigation, strategic restructuring and spin-offs and a renewed focus on innovation and pipeline productivity.

3. Firm profiles and news: J&J (U.S.) and AstraZeneca (U.K.)

Johnson & Johnson

Johnson & Johnson (J&J) is one of the largest healthcare conglomerates globally, operating across pharmaceuticals, medical devices, and consumer health products. Its diversified structure distinguishes it from more specialized pharmaceutical firms and affects both its risk profile and media coverage.

The firm has historically maintained a strong reputation but has also faced significant controversies. Litigation related to opioid distribution and talc-based products has generated substantial negative media attention and financial liabilities. These events provide clear instances of negative sentiment shocks, which are particularly relevant for textual analysis.

Table 1. Key events timeline for Johnson & Johnson (2000–2025)

Johnson & Johnson			
Date	**News**	**Event Type**	**Expected Sentiment**
15/01/2010	Initial wave of McNeil (Tylenol) recalls begins	Product safety	Negative
30/09/2010	Large-scale Tylenol recall escalation	Product safety	Strong Negative
04/11/2013	$2.2B DOJ settlement (Risperdal marketing)	Litigation	Negative
14/12/2018	Reuters investigation links talc to cancer	Litigation/Reputation	Strong Negative
26/08/2019	Oklahoma opioid ruling against J&J	Litigation	Negative
30/03/2020	COVID-19 vaccine development announced	R&D	Positive
27/02/2021	FDA authorizes J&J COVID vaccine	Regulatory	Positive
13/04/2021	Vaccine paused due to blood clot concerns	Safety	Strong Negative
04/04/2023	Kenvue (consumer division) spinoff announced	Corporate strategy	Mixed
31/07/2023	Talc-related bankruptcy strategy rejected	Litigation	Negative

Source: Own elaboration from news articles

AstraZeneca

AstraZeneca is a U.K.-based global biopharmaceutical company with a strong focus on innovation, particularly in oncology, cardiovascular diseases, and rare conditions. Over the past two decades, the firm has undergone significant strategic transformation, shifting

from a traditional pharmaceutical model toward a research-intensive, innovation-driven approach.

Table 2. Key events timeline for AstraZeneca (2000–2025)

AstraZeneca			
Date	**Event**	**Event Type**	**Expected Sentiment**
05/04/2001	Loss of exclusivity pressures (Losec decline)	Patent/Revenue	Negative
01/07/2006	FDA-related safety scrutiny affecting AZ drugs	Regulation	Negative
06/01/2012	Strategic shift toward oncology announced	Strategy	Positive
02/05/2014	Pfizer takeover bid for AstraZeneca revealed	M&A	Mixed
19/05/2014	AstraZeneca rejects Pfizer offer	M&A	Positive (domestic framing)
17/08/2018	Strong oncology results (Tagrisso success)	Earnings/R&D	Positive
30/04/2020	Oxford–AstraZeneca vaccine partnership	R&D	Strong Positive
30/12/2020	UK approves AstraZeneca COVID vaccine	Regulatory	Positive
07/03/2021	Multiple countries suspend AZ vaccine	Safety	Negative
07/04/2021	EMA confirms clot risk link	Safety	Strong Negative
29/07/2022	Strong earnings driven by oncology	Earnings	Positive
2023–2024	Continued oncology and rare disease	Strategy	Positive

Source: Own elaboration from news articles

Among the set of identified events, particular attention is given to those related to the Covid-19 vaccine. Both AstraZeneca (U.K.) and Johnson & Johnson (U.S.) were directly involved in vaccine development and faced similar regulatory and reputational challenges. Hence, these events constitute a particularly suitable setting to test for media bias, as they are largely exogenous, highly salient, and objectively verifiable developments.

The approval, distribution, and subsequent safety concerns associated with Covid-19 vaccines were widely documented by regulatory agencies and international institutions, reducing ambiguity regarding the factual content of the news. At the same time, these events generated intense and heterogeneous media coverage across countries, particularly following the emergence of side-effect concerns and the temporary suspension of the AstraZeneca vaccine in several European countries in March 2021.

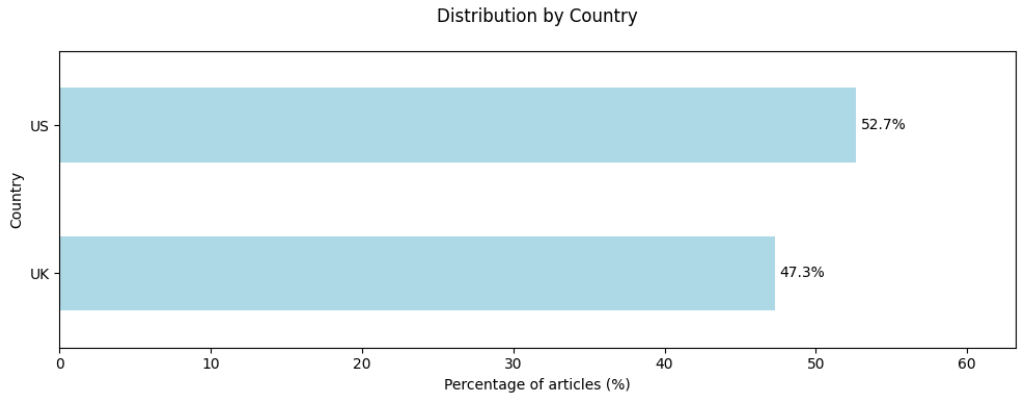
This combination of high informational clarity and strong media attention makes vaccine-related events especially appropriate for identifying potential home-country bias. In other words, differences in coverage tone or intensity across domestic and foreign newspapers can be more credibly interpreted as evidence of bias rather than differences in underlying fundamentals.

Section V: Exploratory Data Analysis

1. Basic Exploratory Data Analysis

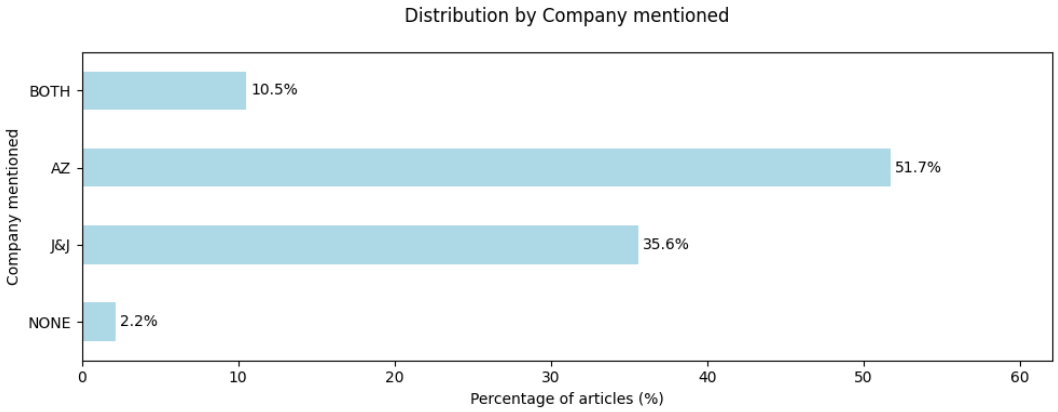
We begin with an exploratory analysis of the distribution of the main dataset variables in order to assess the representativeness and balance of the sample.

Figure 1. Distribution of articles by country of newspaper source



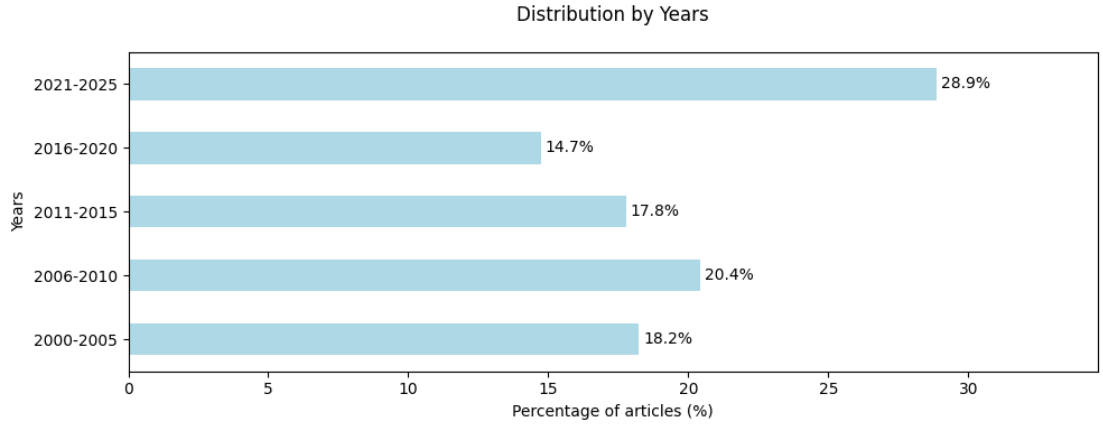
Source: Own elaboration from Python analysis based on Factiva data

Figure 2. Distribution of articles by company mentioned



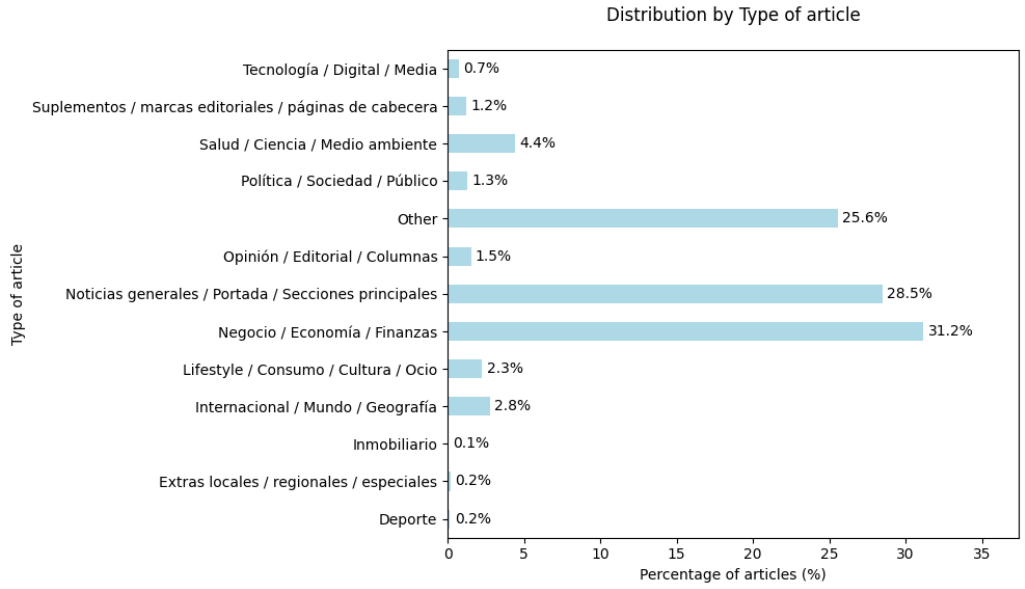
Source: Own elaboration from Python analysis based on Factiva data

Figure 3. Distribution of articles by year period (2000–2025)



Source: Own elaboration from Python analysis based on Factiva data

Figure 4. Distribution of articles by type of article section



Source: Own elaboration from Python analysis based on Factiva data

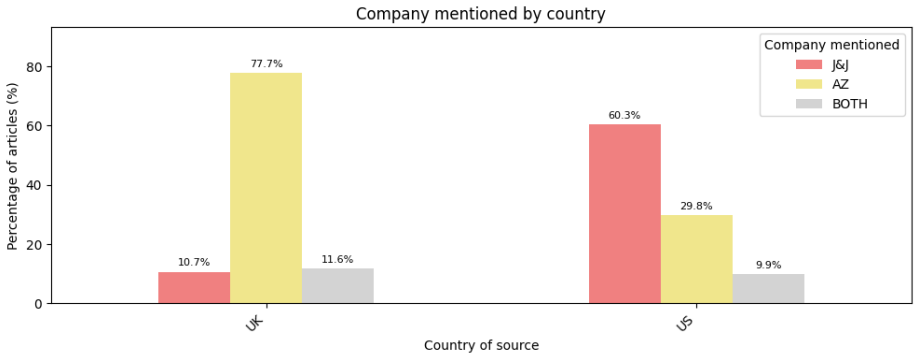
The descriptive statistics show that the number of articles is relatively balanced in terms of newspaper country. AstraZeneca is mentioned in a larger number of articles than Johnson & Johnson overall, and the highest concentration of news appears in the post-COVID-19 period, especially between 2021 and 2025. In terms of article type, the most

represented categories are business / economy / finance and front-page or main-section articles.

2. Coverage Intensity Analysis: Country-bias

This subsection examines whether there is an association between the volume of news coverage devoted to each company and the country of origin of the newspaper in which the articles were published.

Figure 5. Company mentioned by country of newspaper source

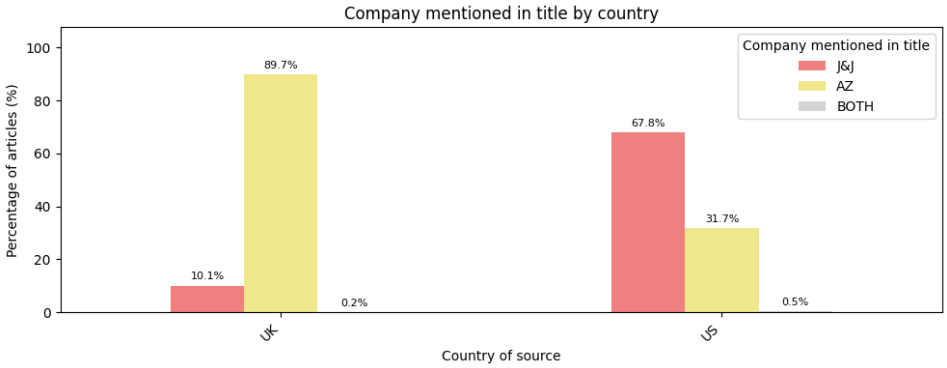


Source: Own elaboration from Python analysis based on Factiva data

The results show a clear predominance of mentions of the nationally associated company over the foreign company.

When the analysis is restricted to articles in which the company is mentioned in the title, the difference becomes even stronger: 89.7% of UK articles mention AstraZeneca in the title, compared with 67.8% of US articles mentioning Johnson & Johnson in the title. This suggests that home-country bias is reflected not only in overall article volume, but also in the prominence given to domestic firms in article headlines.

Figure 6. Company mentioned in title by country of newspaper source



Source: Own elaboration from Python analysis based on Factiva data

3. Exploratory Analysis of Newspaper Tone

As described in Section III, newspaper-adjusted sentiment scores are used throughout this analysis to control differences in baseline tone across outlets. This section first presents the exploratory z-score analysis that motivated this adjustment.

To explore this issue, z-scores were calculated for each article's sentiment score. The z-score measures how far an article's sentiment is from the overall sample mean, expressed in standard deviation units:

$$Z = \text{Article Sentiment} - \text{Sample Mean} \,/\, \text{Sample Standard Deviation}$$

The purpose of this step was not to classify articles as positive or negative according to VADER, but to compare the relative position of each newspaper within the sample. A positive average z-score indicates that a newspaper is more positive than the dataset average, while a negative average z-score indicates that it is less positive than the dataset average.

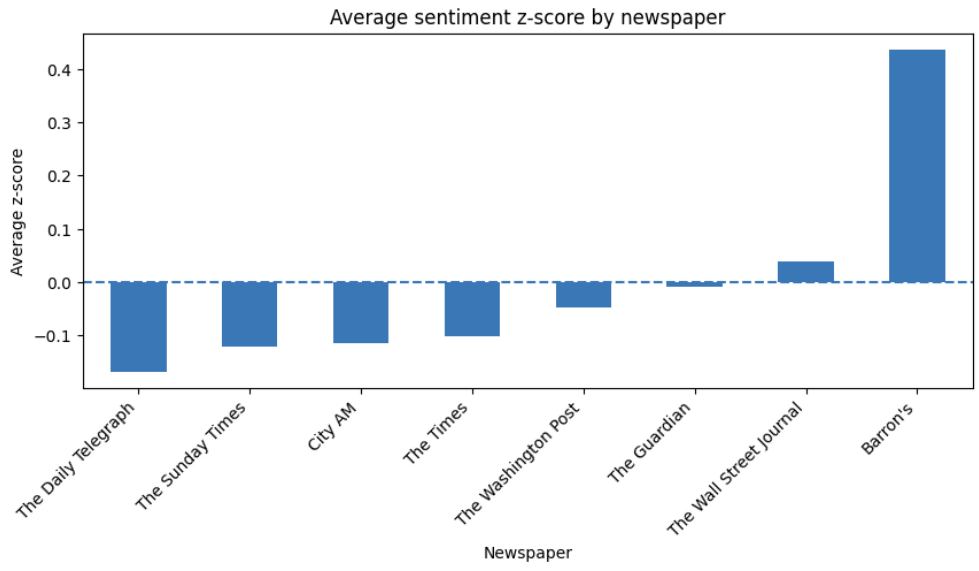
The exploratory results showed that, although all newspapers had positive average VADER sentiment scores in absolute terms, there were clear differences in their relative tone. UK newspapers in the sample generally appeared below the overall dataset average, while US newspapers were more heterogeneous and, in some cases, substantially more positive. In particular, Barron's displayed higher average sentiment scores than most other outlets.

Table 3. Average sentiment by newspaper without adjustment

	source_name	country	n_articles	avg_general_sentiment	avg_general_zscore	avg_title_zscore	avg_snippet_zscore	avg_body_zscore
2	The Daily Telegraph	UK	1877	0.112660	-0.168042	-0.117091	-0.152272	-0.104681
4	The Sunday Times	UK	178	0.131572	-0.121676	-0.098784	-0.020692	-0.140232
1	City AM	UK	579	0.134907	-0.113501	-0.044958	-0.035521	-0.140287
5	The Times	UK	2295	0.139583	-0.102036	-0.095261	-0.019868	-0.109718
7	The Washington Post	US	848	0.161980	-0.047127	-0.065768	-0.041430	-0.014775
3	The Guardian	UK	1623	0.177726	-0.008524	0.004574	-0.055002	0.027224
6	The Wall Street Journal	US	5334	0.196866	0.038403	0.057208	0.041077	0.004523
0	Barron's	US	1113	0.359279	0.436586	0.202354	0.234466	0.448065

Source: Own elaboration from Python analysis based on Factiva data

Figure 7. Average sentiment z-score by newspaper



Source: Own elaboration from Python analysis based on Factiva data

This finding is important because it suggests that a direct comparison between US and UK newspapers using raw sentiment scores could have been misleading. If some US newspapers are generally more positive and some UK newspapers are generally less positive, then a country-level difference in sentiment may reflect the newspapers' baseline tone rather than a specific bias towards Johnson & Johnson or AstraZeneca.

However, this result should not be interpreted as evidence that British newspapers are negative or that American newspapers are inherently positive. All average sentiment scores remained positive in absolute VADER terms. The z-score analysis only indicates that some newspapers are relatively more positive or less positive compared with the sample average.

Table 4. Average raw sentiment by country of newspaper source

	country	n_articles	avg_general_sentiment	avg_title_sentiment	avg_snippet_sentiment	avg_body_sentiment
0	UK	6552	0.140688	-0.006707	0.092091	0.336679
1	US	7295	0.217590	0.040947	0.166945	0.444880

Source: Own elaboration from Python analysis based on Factiva data

A further caveat is that newspaper type may influence these results. Barron's and The Wall Street Journal are strongly oriented towards financial and business coverage. Their articles may therefore emphasize investment performance, stock prices, revenues, earnings, market expectations or corporate growth. This type of vocabulary can generate more positive sentiment scores in VADER. Consequently, the observed difference between US and UK newspapers should not be interpreted as purely national; it may also reflect the financial orientation of specific outlets and the type of articles included in the sample.

To address this, as discussed in Section III.5, sentiment scores were normalized at the newspaper level. This transformation sets the average adjusted sentiment of each newspaper close to zero, so that all subsequent comparisons evaluate each company relative to the normal tone of its own outlet rather than relative to the full dataset. The newspaper-adjusted sentiment scores are used as the primary outcome variable throughout the remainder of this section.

Section VI: Sentiment Analysis

1. Sentiment Analysis Comparison

The purpose of this section is to compare the sentiment assigned to AstraZeneca and Johnson & Johnson across countries, newspapers, article sections and specific event windows. The analysis uses newspaper-adjusted sentiment scores, meaning that each article is evaluated relative to the average tone of the newspaper in which it was published. Therefore, a positive adjusted score does not necessarily mean that an article is positive in absolute terms, but rather that it is more positive than that newspaper's usual tone. Similarly, a negative adjusted score indicates that the article is more negative than the outlet's own average.

Country-level sentiment comparison

At the country level, results show that the evidence of home-country bias depends on the part of the article being analyzed. While headline sentiment is relatively similar across countries, snippet and especially body sentiment reveal clearer domestic-versus-foreign differences.

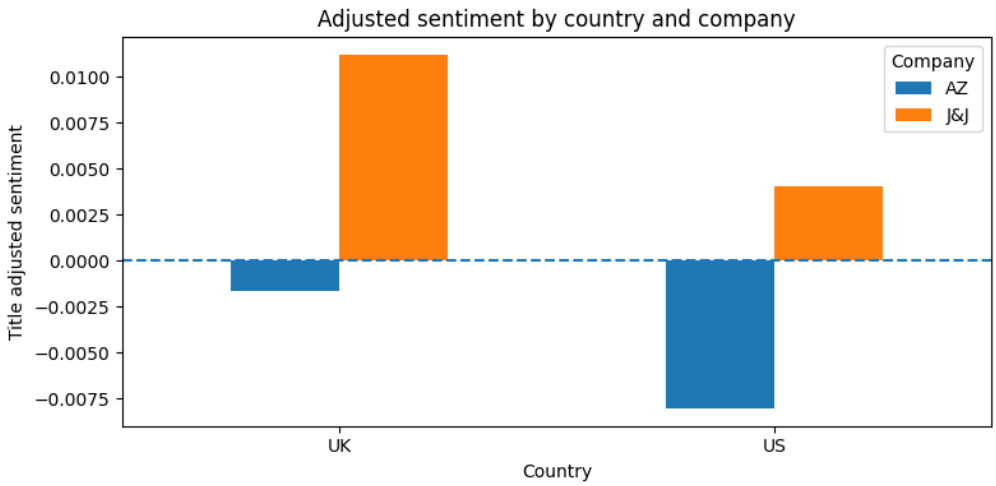
Table 5. Adjusted sentiment by country and company (title, snippet and body)

	company	country	n_articles	avg_title_adjusted	avg_snippet_adjusted	avg_body_adjusted
0	AZ	UK	5837	-0.001635	0.004713	0.015439
1	AZ	US	2781	-0.008020	-0.000252	-0.035254
2	J&J	UK	1457	0.011169	-0.028927	-0.062686
3	J&J	US	4924	0.004027	0.000131	0.019789

Source: Own elaboration from Python analysis based on Factiva data

At the title level, there is little evidence of country-level bias. AstraZeneca receives slightly negative adjusted headline sentiment in both UK and US newspapers, with values of -0.0016 and -0.0080 respectively. Johnson & Johnson receives slightly positive adjusted headline sentiment in both countries, with 0.0112 in UK newspapers and 0.0040 in US newspapers. This suggests that headlines do not strongly differentiate between domestic and foreign firms. Instead, headline tone appears to be relatively stable across countries and may be influenced more by headline-writing conventions or company-specific events than by the nationality of the newspaper.

Figure 8. Adjusted title sentiment by country and company

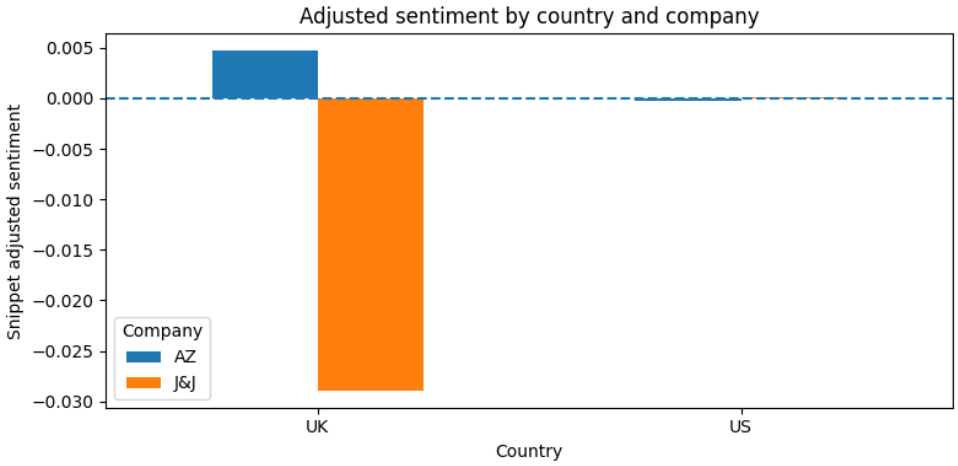


Source: Own elaboration from Python analysis based on Factiva data

The snippet-level results show a more visible domestic-versus-foreign pattern. AstraZeneca receives positive adjusted snippet sentiment in UK newspapers, at 0.0047, while its score in US newspapers is slightly negative, at -0.0003. Johnson & Johnson shows the opposite pattern: it receives a clearly negative adjusted snippet score in UK newspapers, at -0.0289, while its US score is almost neutral, at 0.0001. This suggests that

article summaries already contain some evidence of domestic favourability: UK newspapers summarize AstraZeneca coverage more favourably than US newspapers do, while US newspapers summarize Johnson & Johnson coverage less negatively than UK newspapers do.

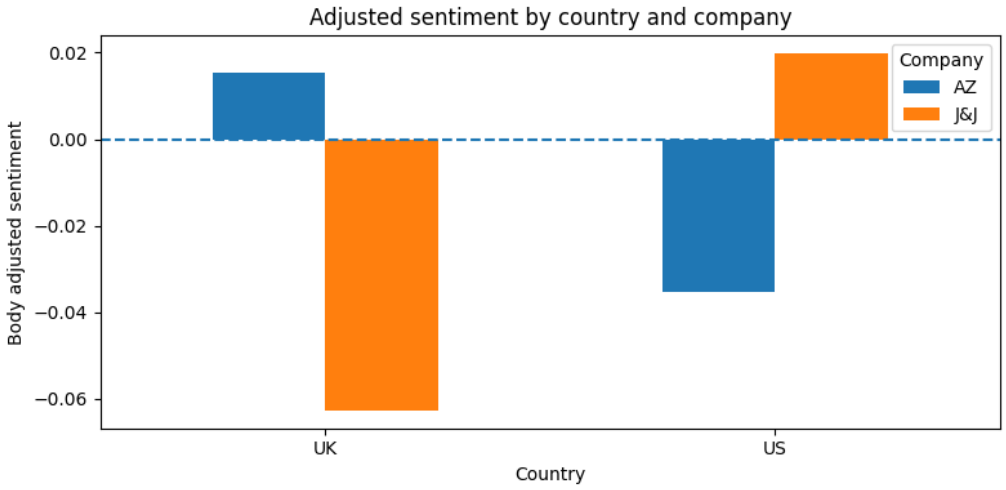
Figure 9. Adjusted snippet sentiment by country and company



Source: Own elaboration from Python analysis based on Factiva data

The clearest evidence appears in the body of the articles. AstraZeneca's adjusted body sentiment is positive in UK newspapers, at 0.0154, but negative in US newspapers, at -0.0353. Conversely, Johnson & Johnson receives negative adjusted body sentiment in UK newspapers, at -0.0627, but positive adjusted body sentiment in US newspapers, at 0.0198. This means that, in the full development of the article, each company is treated more favourably by newspapers from its own country.

Figure 10. Adjusted body sentiment by country and company



Source: Own elaboration from Python analysis based on Factiva data

This result is particularly important because the body contains the most complete narrative. The title and snippet are short and may be constrained by visibility, concision or editorial style, whereas the body allows newspapers to develop context, explanations, comparisons and evaluative framing. Therefore, the fact that body sentiment shows a clear domestic-versus-foreign pattern suggests that home-country bias is more visible in the detailed narrative construction of the article than in the headline.

Overall, the country-level results suggest that home-country bias is weak in headlines, more visible in snippets, and strongest in the body of the articles. In other words, the evidence does not suggest that newspapers necessarily signal bias through explicit headline tone. Instead, the domestic company tends to receive more favourable treatment in the full article text, where the newspaper has more space to frame the company, contextualize events and shape the reader's interpretation.

To assess whether the observed domestic-versus-foreign differences are statistically significant, a Difference-in-Differences regression was estimated on the full sample. The outcome variable is newspaper-adjusted body sentiment, which is the component where the descriptive analysis reveals the clearest home-country pattern. The DiD estimator, the coefficient on the interaction term between the AstraZeneca indicator and the UK newspaper indicator, is 0.1332 and is statistically significant at the 1 percent level ($p < 0.001$, $N = 14{,}870$). This indicates that AstraZeneca receives significantly more favourable adjusted body sentiment in UK newspapers relative to US newspapers, beyond what Johnson & Johnson receives in UK relative to US newspapers. The result is robust to the use of heteroskedasticity-consistent standard errors (HC3), with a 95% confidence interval of [0.087, 0.180].

Table 6. Difference-in-Differences Regression Results for Adjusted Body Sentiment

```
================================================================================
DIFFERENCE-IN-DIFFERENCES RESULTS
Outcome: DiD estimator (is_AZ x is_UK interaction)
Positive coef = AZ treated more favourably in UK vs US than JJ
Significance: *** p<0.01  ** p<0.05  * p<0.1
================================================================================
                    Window                          Outcome  DiD coef  CI lower  CI upper  p-value     N Sig
               Full sample general_sentiment_adjusted_source    0.0173   -0.0122    0.0467   0.2513 14999
               Full sample    body_sentiment_adjusted_source    0.1332    0.0867    0.1796   0.0000 14870 ***
Dev & Approval (Mar20-Jun21) general_sentiment_adjusted_source    0.0365   -0.0124    0.0854   0.1432  4106
Dev & Approval (Mar20-Jun21)    body_sentiment_adjusted_source    0.1756    0.1000    0.2512   0.0000  4030 ***
         AZ release (Dec20) general_sentiment_adjusted_source    0.1048   -0.1267    0.3364   0.3750   325
         AZ release (Dec20)    body_sentiment_adjusted_source    0.1973   -0.1404    0.5350   0.2521   313
         JJ release (Feb21) general_sentiment_adjusted_source    0.1721    0.0582    0.2860   0.0031   771 ***
         JJ release (Feb21)    body_sentiment_adjusted_source    0.2946    0.1100    0.4792   0.0018   757 ***
      AZ suspension (Mar21) general_sentiment_adjusted_source    0.0355   -0.0819    0.1529   0.5531   703
      AZ suspension (Mar21)    body_sentiment_adjusted_source    0.1552   -0.0210    0.3315   0.0844   696   *
      JJ suspension (Apr21) general_sentiment_adjusted_source    0.0646   -0.0395    0.1687   0.2238   657
      JJ suspension (Apr21)    body_sentiment_adjusted_source    0.2633    0.1035    0.4230   0.0012   652 ***

MAIN RESULT (Full sample, body sentiment):
  DiD coefficient: 0.1332
  95% CI: [0.0867, 0.1796]
  p-value: 0.0 ***
  N: 14870
```

Source: Own elaboration from Python analysis based on Factiva data

When the same specification is estimated using general sentiment — the average of title, snippet and body scores — the DiD coefficient is 0.0173 and is not statistically significant ($p = 0.251$). This result is consistent with the descriptive finding that home-country bias is concentrated in the body of articles rather than in headlines or summaries, and confirms that aggregate sentiment averages can mask the pattern by diluting the body signal with less informative title and snippet scores.

Newspaper-level sentiment comparison

The newspaper-level results reveal substantial heterogeneity across outlets. This is an important finding because it shows that national media bias cannot be interpreted only at the country level. Individual newspapers differ in their editorial focus, political positioning, audience, article mix and coverage style. Therefore, even within the same country, newspapers may frame AstraZeneca and Johnson & Johnson differently.

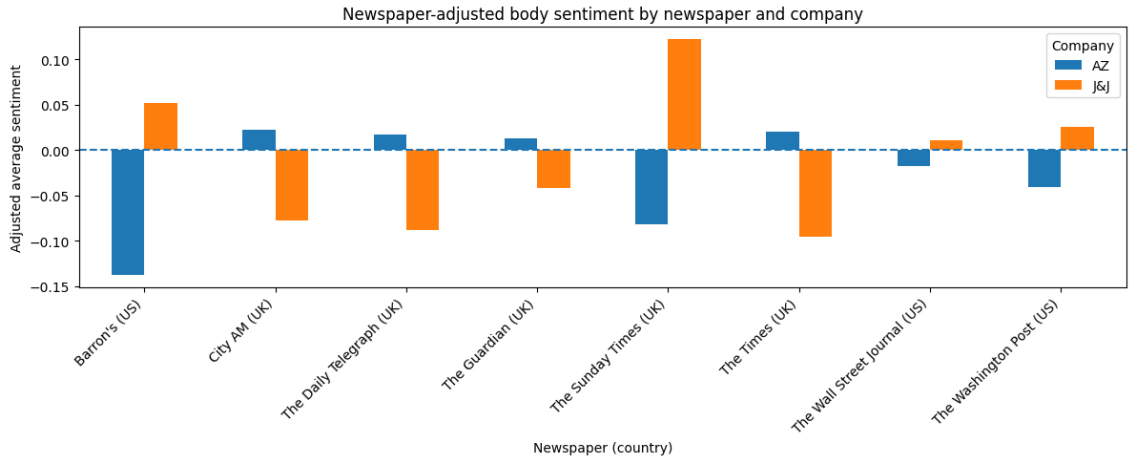
Table 7. Adjusted sentiment by newspaper and company

	source_name	company	country	n_articles	avg_adjusted_sentiment	avg_title_adjusted	avg_snippet_adjusted	avg_body_adjusted
0	Barron's	AZ	US	328	-0.057481	-0.017608	-0.032807	-0.138058
1	Barron's	J&J	US	867	0.021746	0.005532	0.010665	0.051811
2	City AM	AZ	UK	490	0.011754	-0.003664	0.016612	0.022732
3	City AM	J&J	UK	144	-0.039995	0.016614	-0.073501	-0.077351
4	The Daily Telegraph	AZ	UK	1708	0.001157	0.000704	-0.004569	0.016711
5	The Daily Telegraph	J&J	UK	326	-0.006060	-0.006625	0.036096	-0.088688
6	The Guardian	AZ	UK	1469	-0.015590	0.001827	0.019035	0.013246
7	The Guardian	J&J	UK	465	0.049251	-0.012412	-0.121766	-0.042318
8	The Sunday Times	AZ	UK	124	-0.043016	-0.005797	-0.036665	-0.081666
9	The Sunday Times	J&J	UK	87	0.061310	0.010281	0.063868	0.122499
10	The Times	AZ	UK	2046	0.005990	-0.005081	0.002696	0.020083
11	The Times	J&J	UK	435	-0.028171	0.036136	-0.016838	-0.095831
12	The Wall Street Journal	AZ	US	2073	-0.008922	-0.007672	0.005890	-0.017965
13	The Wall Street Journal	J&J	US	3444	0.005370	0.004253	-0.003354	0.010768
14	The Washington Post	AZ	US	380	-0.015488	-0.000451	-0.009628	-0.040729
15	The Washington Post	J&J	US	613	0.009601	0.000214	0.005090	0.024931

Source: Own elaboration from Python analysis based on Factiva data

Building on the country-level results, the newspaper-level comparison will focus on adjusted body sentiment. This is also the most appropriate component for analysing newspaper-level framing, because the body contains the full narrative of the article: the context provided, the explanations selected, the risks emphasized, and the way each company is presented to the reader.

Figure 11. Newspaper-adjusted body sentiment by newspaper and company



Source: Own elaboration from Python analysis based on Factiva data

In the US sample, all three newspapers treat Johnson & Johnson more favourably than AstraZeneca.

Barron's, The Wall Street Journal and The Washington Post all assign positive or less negative adjusted body sentiment to Johnson & Johnson, while AstraZeneca receives negative adjusted body sentiment in each of them. This is consistent with the home-country bias hypothesis, since Johnson & Johnson would represent the national company and AstraZeneca the foreign company.

Barron's and The Wall Street Journal show a similar pattern, although the difference is much stronger in Barron's. In Barron's, AstraZeneca records an adjusted body sentiment of -0.1381, while Johnson & Johnson records 0.0518. In The Wall Street Journal, the difference is smaller but follows the same direction: AstraZeneca records -0.0180, while Johnson & Johnson records 0.0108.

A plausible explanation is that both outlets are strongly business- and market-oriented. Barron's is explicitly focused on market analysis, stock analysis and financial commentary, while The Wall Street Journal is one of the main US business newspapers. Therefore, their coverage is likely to be shaped by the perspective of American investors, firms and markets. From this point of view, Johnson & Johnson may appear as a more familiar and institutionally relevant domestic company, whereas AstraZeneca may be framed as a more distant foreign corporate actor.

The Washington Post also follows the same direction, although it has a different editorial profile. AstraZeneca records an adjusted body sentiment of -0.0407, while Johnson & Johnson records 0.0249. This result is particularly relevant because The Washington Post is not primarily a financial newspaper. Unlike Barron's and The Wall Street Journal, it is a general-interest national newspaper with a stronger focus on politics, public affairs and national issues. Nevertheless, it still treats the US company more favourably at the body level. This suggests that the domestic-favourability pattern is not limited to financial newspapers, but may also appear in general-interest outlets when the domestic firm is more familiar and institutionally closer to the newspaper's audience.

The UK newspapers show a broadly consistent pattern, with one clear exception.

City AM, The Daily Telegraph, The Guardian and The Times all treat AstraZeneca more favourably than Johnson & Johnson. In all four outlets, AstraZeneca records positive adjusted sentiment, while Johnson & Johnson records negative adjusted sentiment. The pattern is especially clear in The Times, where AstraZeneca records 0.0201 compared with -0.0953 for Johnson & Johnson, and in The Daily Telegraph, where AstraZeneca records 0.0167 compared with -0.0869 for Johnson & Johnson.

This pattern is consistent with the home-country bias hypothesis, since AstraZeneca is the UK-associated company and Johnson & Johnson is the foreign comparator. This interpretation can apply across different types of newspapers. For example, City AM's business and finance orientation makes it likely to frame AstraZeneca as a major UK corporate actor; The Times and The Daily Telegraph may reflect a more institutional or national-market framing; and The Guardian shows that domestic familiarity can also appear in a newspaper that is usually more critical of large corporations. Therefore, the UK results again suggest that home-country familiarity is not limited to one specific editorial ideology, but appears across different types of British newspapers.

The main exception in the UK sample is The Sunday Times. Unlike the other British newspapers, it treats Johnson & Johnson much more favourably than AstraZeneca. AstraZeneca records a negative adjusted sentiment of -0.0817, while Johnson & Johnson records a strongly positive value of 0.1225. This is the clearest case where the expected home-country pattern does not appear.

A plausible explanation is the specific editorial identity of The Sunday Times. News UK (n.d.) describes The Sunday Times as a newspaper that "sets the news agenda" and "holds the powerful to account", which suggests a more adversarial and accountability-oriented

style of journalism. This may make the outlet more willing to scrutinize powerful domestic actors rather than frame them favourably. It is also important not to assume that The Sunday Times should behave in the same way as The Times. UK government documentation notes that The Times and The Sunday Times have taken different editorial lines and pursued different investigative areas, despite common ownership (Department for Digital, Culture, Media & Sport, 2019). Therefore, the divergence between The Times and The Sunday Times is not necessarily inconsistent. Rather, it reinforces the idea that home-country bias can be moderated or even overridden by outlet-specific editorial culture.

Taken together, the newspaper-level analysis suggests that home-country bias is clearer and more consistent than the aggregate country-level comparison initially suggested. In the US, all three newspapers favour Johnson & Johnson over AstraZeneca. In the UK, four out of five newspapers favour AstraZeneca over Johnson & Johnson. The only clear exception is The Sunday Times. These results suggest that home-country bias is not mechanical, but it does appear as a strong pattern across most outlets once each newspaper is analysed individually.

Sentiment comparison by article section

Not all articles published by the same newspaper perform the same editorial function. A front-page article, a business report and a health-related piece may cover the same company from very different perspectives. As a result, the content and purpose of the article can strongly affect its sentiment. For example, business articles are more likely to discuss firms through market performance, revenues, strategy or investment expectations,

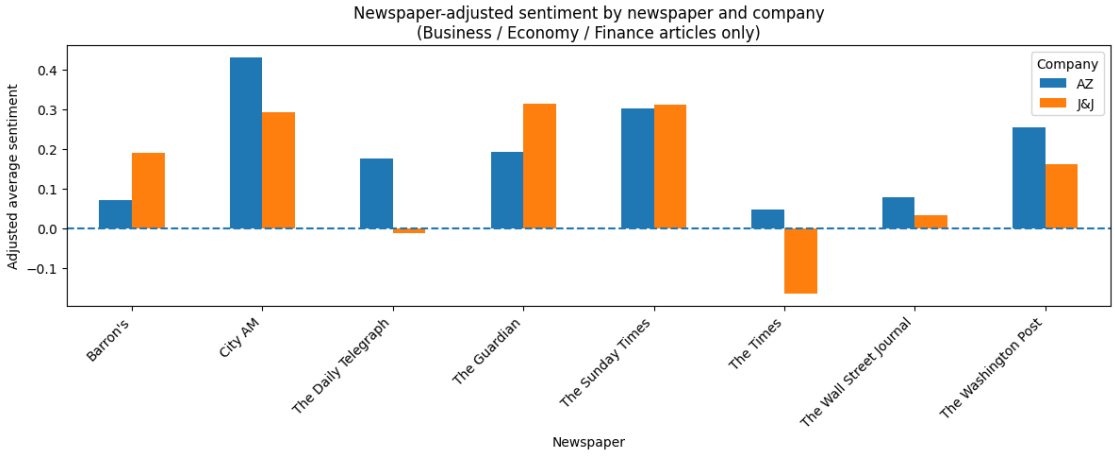
while front-page and main-section articles are more likely to focus on high-salience events, controversies or reputational risks.

For this reason, the analysis was replicated across the most relevant article sections identified in the dataset. In the exploratory analysis, Business / Economy / Finance and Frontpage / Main sections were among the categories with the largest number of articles, making them especially important for understanding the overall sentiment results. In addition, Health articles are included because, although they are less numerous and less evenly distributed, they are particularly relevant in a pharmaceutical-industry context.

The results show that article section has a strong effect on sentiment.

The **Business, Economy and Finance** sections shows the most positive sentiment overall.

Figure 12. Newspaper-adjusted body sentiment by newspaper and company (Business, Economy and Finance articles only)



Source: Own elaboration from Python analysis based on Factiva data

In this category, most newspaper-company pairs are above zero, meaning that articles in financial sections tend to treat both AstraZeneca and Johnson & Johnson more positively than each newspaper's average tone.

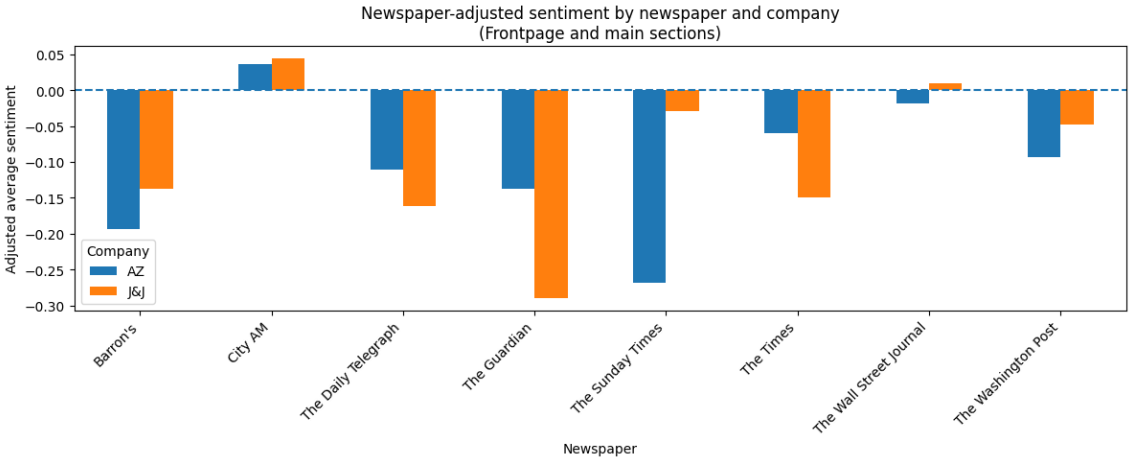
This is an important finding because it suggests that business coverage frames pharmaceutical firms mainly through corporate performance, market relevance, investment prospects, revenues, strategy, partnerships and business developments. As a result, the tone is generally more favourable than in sections where the same companies may appear in relation to public controversy, litigation, regulation or reputational risk.

The business-section results also show that home-country bias is less uniform in financial coverage. In some UK newspapers, such as City AM, The Daily Telegraph and The Times, AstraZeneca receives more favourable sentiment than Johnson & Johnson, which is consistent with the home-country bias hypothesis. However, this pattern is not universal. The Guardian and The Sunday Times treat Johnson & Johnson more favourably in this section, and in the US sample The Wall Street Journal and The Washington Post appear more favourable towards AstraZeneca than Johnson & Johnson.

Therefore, the main insight from the business section is not that home-country bias disappears, but that the section itself creates a more positive environment for both companies. When articles are framed as corporate, financial or market stories, sentiment becomes more favourable overall and the domestic-versus-foreign pattern becomes less consistent.

The **Frontpage and Main sections** show the opposite pattern. In this category, adjusted sentiment is generally negative for both companies.

Figure 13. Newspaper-adjusted body sentiment by newspaper and company (Frontpage and Main sections only)



Source: Own elaboration from Python analysis based on Factiva data

This pattern is intuitive. Articles placed in front-page or main-news sections usually correspond to stories with broader public relevance. These are more likely to involve high-impact events, public controversies, regulatory decisions, lawsuits, political reactions, safety concerns or reputational crises. Therefore, the negative tone observed in these sections should not automatically be interpreted as bias against one company. It may reflect the fact that only the most serious or conflict-driven corporate stories reach the most visible parts of the newspaper.

However, there are still relative differences between companies. In several US newspapers, Johnson & Johnson is treated less negatively than AstraZeneca, while in several UK newspapers AstraZeneca is treated less negatively than Johnson & Johnson. This indicates that home-country bias may still operate inside a generally negative news environment.

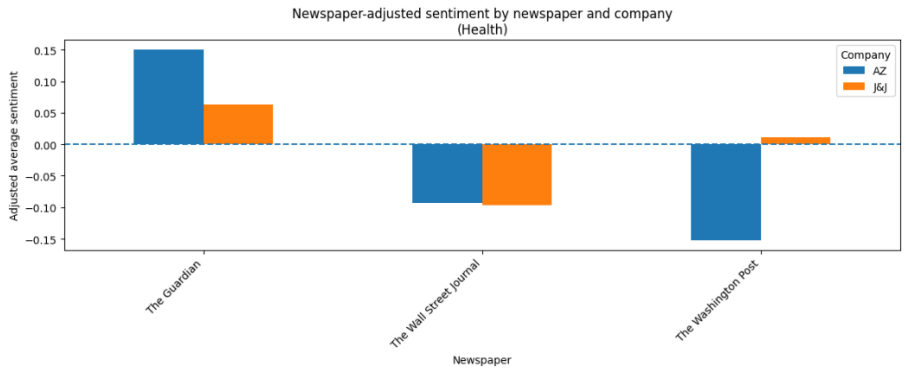
City AM is an important exception. In its frontpage and main-section articles, both AstraZeneca and Johnson & Johnson show positive adjusted sentiment. This is likely related to the nature of the outlet itself. Unlike general-interest newspapers, City AM is a business and finance newspaper, so even its more prominent sections are likely to maintain a corporate, market-oriented framing.

The Wall Street Journal also appears less negative than most outlets in this category, with Johnson & Johnson slightly positive and AstraZeneca only slightly negative. This may reflect a similar mechanism, although less strongly than in City AM.

This is an important nuance. Home-country bias does not necessarily mean that the domestic company receives positive coverage. In more conflict-oriented sections, the domestic company may simply receive less negative coverage than the foreign company. Therefore, the frontpage and main-section results suggest that section-level negativity and home-country bias can coexist.

The **Health** sections presents a more limited but still useful comparison.

Figure 14. Newspaper-adjusted body sentiment by newspaper and company (Health articles only)



Source: Own elaboration from Python analysis based on Factiva data

Health articles are available for fewer newspapers, so this section should be interpreted as complementary rather than as one of the main sources of evidence. However, the pattern is still informative because health coverage is especially relevant in the pharmaceutical industry.

In this section, the results provide some evidence consistent with home-country bias, although the pattern is not uniform. The Guardian favours AstraZeneca, with both companies receiving positive sentiment but AstraZeneca clearly higher. The Washington Post shows the opposite national pattern: AstraZeneca is clearly negative, while Johnson & Johnson is slightly positive. The Wall Street Journal is the main exception, as both companies receive similarly negative sentiment, suggesting a more generally cautious tone towards pharmaceutical health coverage rather than a clear domestic preference.

Overall, the section-level analysis shows that accounting for article section is essential. Sentiment changes substantially depending on whether the company appears in a business article, a frontpage/main-section article or a health-related article. Business coverage tends to create a more positive environment for both companies, while frontpage and main-section coverage is generally more negative because it concentrates higher-salience and more conflict-oriented stories. Health coverage is more mixed, depending on whether the article emphasizes medical contribution or public-health risk.

This matters especially because the newspapers in the sample are not homogeneous. Some outlets, such as City AM, Barron's and The Wall Street Journal, have a strong business or financial orientation, while others are broader general-interest newspapers. Without taking section into account, these structural differences between newspapers could be mistaken for sentiment bias.

However, once section is taken into account, the results still show a clear pattern consistent with home-country bias. In several sections, the domestic company is treated more favourably than the foreign comparator, even when the overall tone of that section is positive or negative for both firms. This is particularly important in the Frontpage and Main sections: although sentiment is generally negative, Johnson & Johnson tends to be less negative in US newspapers, while AstraZeneca tends to be less negative in UK newspapers. Therefore, home-country bias does not necessarily appear as positive coverage, but often as relatively softer treatment within the same editorial context.

Section composition shapes the general level of sentiment, but domestic-versus-foreign differences remain visible inside specific sections. For this reason, article section should be treated as a key control variable in the interpretation of media sentiment. It allows the analysis to distinguish between differences caused by news type and differences that are more consistent with home-country framing.

Event-window and period-specific sentiment analysis

To complement the aggregate comparison, the analysis was replicated around specific periods and events identified in the companies timelines. The objective of this step is to observe whether the sentiment patterns found in the full sample remain stable during specific periods of high media attention were news should have been more objective: vaccine releases, safety concerns, takeovers, etc.

This approach follows the event-study logic discussed in the methodology section, but it should not be interpreted as a causal event study in the strict econometric sense. The analysis does not compare a fully treated group with a fully untreated control group.

Instead, event windows are used as a descriptive framework to examine how sentiment varies across domestic and foreign newspapers when coverage is concentrated around relevant corporate events for AstraZeneca and J&J.

i. The Covid19 Pandemic

To provide some context, as mentioned in Section III, both companies followed comparable but not identical COVID-19 vaccine timelines. Johnson & Johnson announced the identification of its lead COVID-19 vaccine candidate on 30 March 2020, while AstraZeneca announced its partnership with the University of Oxford one month later, on 30 April 2020. However, AstraZeneca reached regulatory authorization earlier in the United Kingdom: the Oxford/AstraZeneca vaccine was authorized by the UK medicines regulator on 30 December 2020, whereas Johnson & Johnson's single-dose vaccine received Emergency Use Authorization from the FDA on 27 February 2021.

The two vaccines were also affected by safety-related pauses during 2021. In the case of AstraZeneca, the first suspension occurred on 7 March 2021, when Austria paused the use of one vaccine batch as a precaution while investigating reports of blood-clotting events. Several other countries later suspended or restricted the vaccine during March, before the European Medicines Agency stated that the vaccine's benefits continued to outweigh the risks, while acknowledging a possible link to very rare blood clots. Johnson & Johnson faced a similar issue shortly afterwards: on 13 April 2021, the FDA and CDC recommended a temporary pause in the use of the Janssen vaccine in the United States while investigating rare blood-clotting cases.

This timeline is important because it shows that both companies experienced a transition from positive development and approval news to safety-related uncertainty. Therefore, the event-window analysis should distinguish between the initial vaccine-development and authorization phase, where sentiment may be more favourable, and the later safety-concern phase, where coverage is more likely to become negative or risk-oriented.

- **From the announcement of vaccine development to its release:**

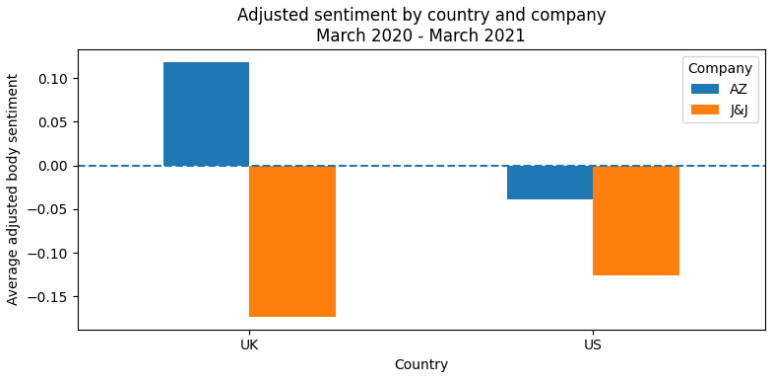
The first period analyzed covers the Covid-19 vaccine development and early approval phase, from March 2020 to June 2021. This period is broader than a single event window and should therefore be interpreted as a phase of sustained vaccine-related attention.

Table 8. Adjusted sentiment by country and company: Vaccine development phase (March 2020 – March 2021)

	company	country	n_articles	avg_adjusted_sentiment	avg_title_adjusted	avg_snippet_adjusted	avg_body_adjusted
0	AZ	UK	1100	0.062082	0.034993	0.081515	0.118132
1	AZ	US	444	0.012143	0.015335	0.029120	-0.038489
2	J&J	UK	261	-0.040100	0.065780	-0.082330	-0.172969
3	J&J	US	302	-0.021501	-0.014205	0.065164	-0.125221

Source: Own elaboration from Python analysis based on Factiva data

Figure 15. Adjusted body sentiment by country and company: Vaccine development phase (March 2020 – March 2021)



Source: Own elaboration from Python analysis based on Factiva data

The results show that AstraZeneca occupied a particularly strong position during this phase, especially in UK newspapers. This is consistent with the fact that the Oxford/AstraZeneca vaccine became closely associated with British scientific capacity and national vaccination strategy. In this period, UK newspapers treated AstraZeneca more favourably than Johnson & Johnson, which is consistent with the home-country bias hypothesis.

In US newspapers, however, the pattern is less symmetrical. Johnson & Johnson does not receive the same clear positive advantage that AstraZeneca receives in UK newspapers. This suggests that the early COVID-vaccine period was not simply a balanced case of each country favouring its own company. Instead, the broader narrative appears to have been shaped by the earlier and more internationally visible role of the Oxford/AstraZeneca vaccine.

This first period therefore provides evidence of home-country bias, but mainly in the UK sample. It also shows that positive sentiment during the pandemic was not distributed evenly across firms. Timing mattered: the company that reached a major vaccine milestone earlier received more favourable coverage during this phase.

- **Release of AZ vaccine:**

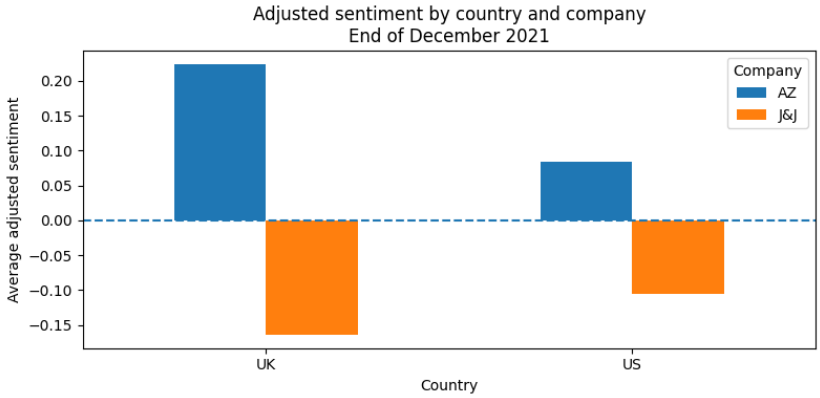
The UK authorization of the Oxford/AstraZeneca vaccine the 30th of December 2020 is one of the most relevant positive events for the analysis because it combined scientific, regulatory and national dimensions. The event was not only a corporate milestone for AstraZeneca, but also a symbolic achievement for the UK, given the association between with Oxford University.

Table 9. Adjusted sentiment by country and company: AstraZeneca vaccine authorization (December 2020)

	company	country	n_articles	avg_adjusted_sentiment	avg_title_adjusted	avg_snippet_adjusted	avg_body_adjusted
0	AZ	UK	201	0.141592	0.109423	0.131033	0.223792
1	AZ	US	42	0.080999	0.102458	0.013157	0.084378
2	J&J	UK	61	0.014364	0.147825	0.126335	-0.163798
3	J&J	US	21	0.058584	-0.137732	-0.128718	-0.105888

Source: Own elaboration from Python analysis based on Factiva data

Figure 16. Adjusted body sentiment by country and company: AstraZeneca vaccine authorization (December 2020)



Source: Own elaboration from Python analysis based on Factiva data

The results show a clearly favourable reaction towards AstraZeneca in both UK and US newspapers, but the effect is stronger in the UK. This suggests that the approval of the vaccine was interpreted positively beyond the British press, but that UK newspapers framed the event especially favourably. In US newspapers, the AstraZeneca release appears more distant. For UK outlets, the AstraZeneca release is a domestic scientific and regulatory milestone; for US outlets, it is more likely to be treated as an international pharmaceutical development.

The comparison with Johnson & Johnson reinforces this interpretation. During this window, Johnson & Johnson is not the central company in the news cycle and receives much weaker sentiment than AstraZeneca. This indicates that the favourable tone observed in this window is not simply a general vaccine optimism effect, but is specifically linked to AstraZeneca's regulatory milestone.

Overall, the AstraZeneca release window shows how positive corporate events can activate domestic favourability while also attracting favourable international coverage. In this case, AstraZeneca benefits from both effects: the objective importance of the vaccine authorization at that time and its strong association with the UK.

- **Release of J&J vaccine:**

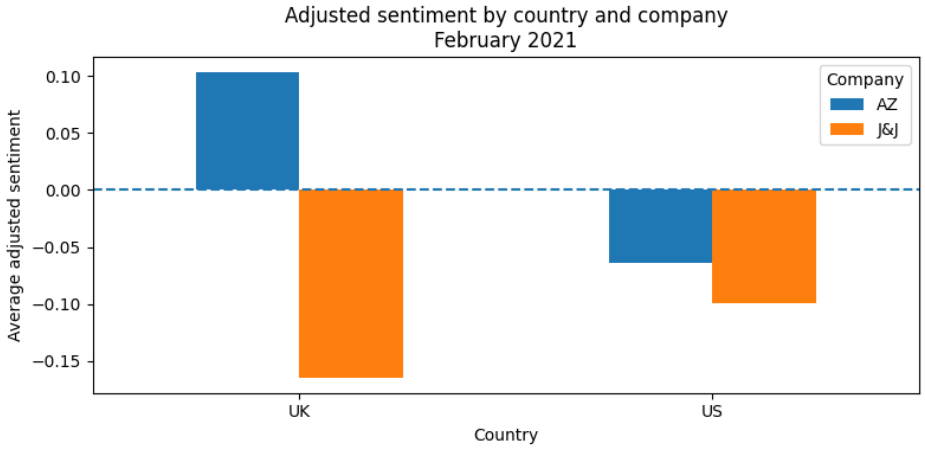
The Johnson & Johnson release window captures the reaction around the FDA authorization of the Janssen single-dose vaccine the 27th February 2021. This event is comparable to the AstraZeneca release window because it also represents a major regulatory milestone. However, the results suggest that the media response was not as clearly positive as in the AstraZeneca release period.

Table 10. Adjusted sentiment by country and company: Johnson & Johnson vaccine authorization (February 2021)

	company	country	n_articles	avg_adjusted_sentiment	avg_title_adjusted	avg_snippet_adjusted	avg_body_adjusted
0	AZ	UK	1136	0.052748	0.034741	0.067200	0.103275
1	AZ	US	463	-0.000359	0.011014	0.025117	-0.063451
2	J&J	UK	276	-0.033793	0.073668	-0.034088	-0.164472
3	J&J	US	333	-0.003812	0.007404	0.091369	-0.098992

Source: Own elaboration from Python analysis based on Factiva data

Figure 17. Adjusted body sentiment by country and company: Johnson & Johnson vaccine authorization (February 2021)



Source: Own elaboration from Python analysis based on Factiva data

The most visible pattern is that UK newspapers still treat AstraZeneca much more favourably than Johnson & Johnson, which remains negative. This suggests that, even during the Johnson & Johnson authorization window, British newspapers continue to favor AstraZeneca, even though US newspapers started to attribute a more negative sentiment to AZ already. By this time, AstraZeneca's international coverage was beginning to lose part of the highly positive tone due to safety concerns.

In US media, Johnson & Johnson is treated less negatively than in UK newspapers, which is consistent with a softer domestic treatment. However, the company does not receive a strong positive sentiment effect despite the FDA authorization. This may be because the Johnson & Johnson vaccine arrived later in the pandemic narrative, when media coverage was already moving beyond the initial optimism of vaccine development and approval. By February 2021, attention was increasingly focused on rollout logistics, comparative efficacy, supply constraints and emerging safety concerns.

In other words, the window does not show a simple replacement of AstraZeneca by Johnson & Johnson. Rather, it shows a more cautious media environment overall, where AstraZeneca remains strongly supported in the UK but less so in the US, and Johnson & Johnson benefits from domestic proximity without achieving a clearly positive sentiment advantage.

- **Suspension of AZ vaccine:**

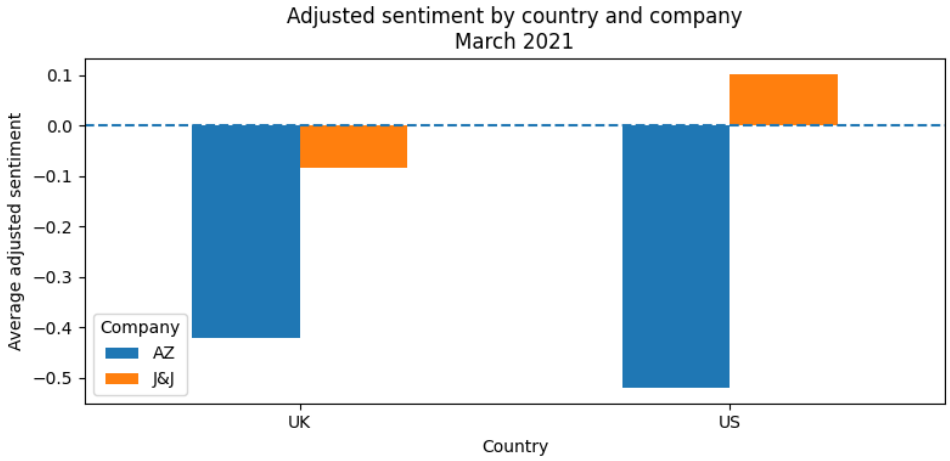
The AstraZeneca suspension window captures the shift from positive vaccine-development coverage to safety-related concern. This is a crucial subperiod because it tests whether home-country bias remains visible when the domestic company is affected by negative news.

Table 11. Adjusted sentiment by country and company: AstraZeneca vaccine suspension (March 2021)

	company	country	n_articles	avg_adjusted_sentiment	avg_title_adjusted	avg_snippet_adjusted	avg_body_adjusted
0	AZ	UK	208	-0.269621	-0.048988	-0.278109	-0.421002
1	AZ	US	84	-0.270955	-0.090834	-0.230079	-0.519547
2	J&J	UK	57	0.001314	0.095761	-0.031091	-0.084256
3	J&J	US	80	0.152483	0.145902	0.277041	0.101533

Source: Own elaboration from Python analysis based on Factiva data

Figure 18. Adjusted body sentiment by country and company — AstraZeneca vaccine suspension (March 2021)



Source: Own elaboration from Python analysis based on Factiva data

The results show a clear deterioration in sentiment towards AstraZeneca in both UK and US newspapers.

However, the decline is not equally strong across countries. AstraZeneca receives negative adjusted sentiment in the UK, but the negative value is more pronounced in US newspapers. This suggests that the suspension damaged AstraZeneca's media tone internationally, but that UK newspapers treated the company less harshly than US newspapers did. Hence, we can see that, even when the event itself is negative, the domestic press does not necessarily become positive; instead, the bias may appear as relatively softer treatment.

The comparison with Johnson & Johnson reinforces this interpretation. In UK newspapers, Johnson & Johnson remains negative during this window, while in US newspapers it receives positive adjusted sentiment.

Overall, the AstraZeneca suspension window shows that negative safety events can strongly reduce sentiment, but they do not eliminate the domestic-versus-foreign pattern.

- **Suspension of J&J vaccine:**

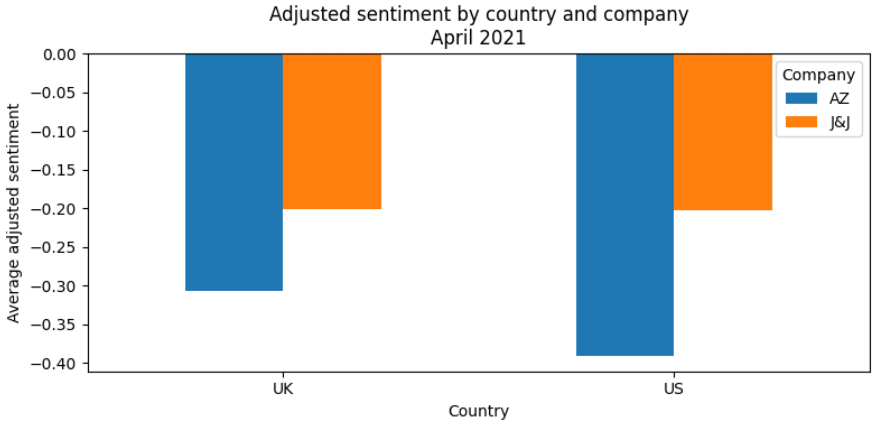
The Johnson & Johnson suspension window captures the period in which the FDA and CDC recommended a temporary pause in the use of the Janssen vaccine due to rare blood-clotting concerns. This event takes place after the main vaccine approval phase, when the media narrative had already shifted from vaccine development and authorization to safety, side effects and public-health risk.

Table 12. Adjusted sentiment by country and company: Johnson & Johnson vaccine suspension (April 2021)

	company	country	n_articles	avg_adjusted_sentiment	avg_title_adjusted	avg_snippet_adjusted	avg_body_adjusted
0	AZ	UK	291	-0.187073	-0.008414	-0.167936	-0.306323
1	AZ	US	120	-0.197822	-0.053450	-0.181963	-0.391361
2	J&J	UK	116	-0.105645	-0.073683	-0.150107	-0.201304
3	J&J	US	160	-0.177316	-0.091576	-0.232867	-0.202136

Source: Own elaboration from Python analysis based on Factiva data

Figure 19. Adjusted body sentiment by country and company: Johnson & Johnson vaccine suspension (April 2021)



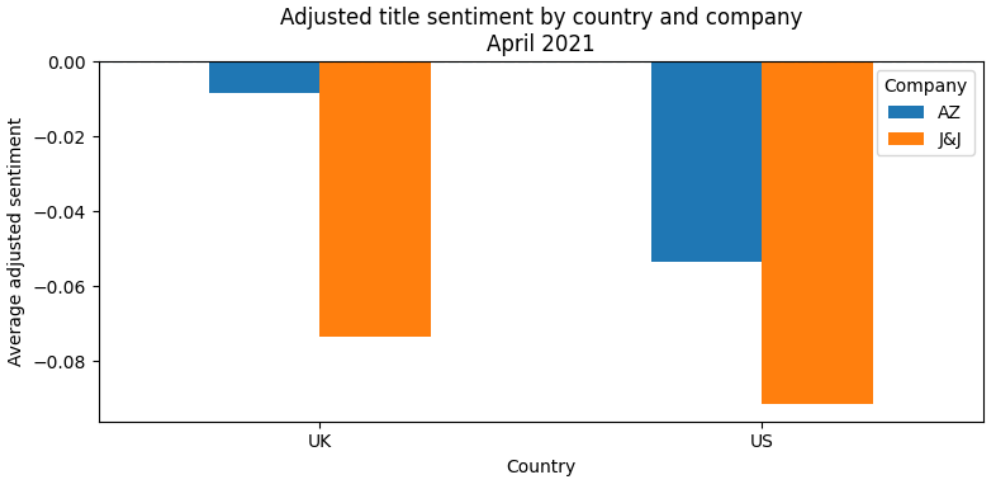
Source: Own elaboration from Python analysis based on Factiva data

The results show that sentiment becomes negative across all company-country pairs. This is an important finding because it suggests that, by April 2021, vaccine coverage was no longer dominated by the positive narrative of scientific progress or regulatory approval. Instead, the news cycle was increasingly focused on the potential risks associated with vaccines. In this context, both AstraZeneca and Johnson & Johnson are framed more negatively than in previous windows.

However, the results still show some relative differences. AstraZeneca remains more negative in US newspapers than in UK newspapers, suggesting that the UK press continues to treat the UK-associated company somewhat less harshly. Johnson & Johnson also receives negative sentiment in both countries, but the comparison does not show the same strong positive domestic effect that appeared in earlier approval-related windows.

The title-level results are also informative:

Figure 20. Adjusted title sentiment by country and company: Johnson & Johnson vaccine suspension (April 2021)



Source: Own elaboration from Python analysis based on Factiva data

Titles are negative for both companies, but UK headlines appear less severe overall than US headlines. This may suggest a more cautious or less alarmist headline treatment in the UK sample. However, this pattern may also reflect the composition of the newspapers included in each country. The US sample contains strongly business- and market-oriented outlets, such as Barron's and The Wall Street Journal, where vaccine safety concerns may be framed partly through corporate risk, investor uncertainty and potential stock-market implications. Therefore, the stronger negativity in US coverage may reflect both national framing and the financial nature of some outlets.

The Difference-in-Differences specification was replicated across each event window to assess whether the domestic-versus-foreign pattern identified in the full sample remains statistically significant during specific periods of high media attention. The results are summarised in the table below.

Table 13. Difference-in-Differences Regression Results for Adjusted Body Sentiment

```
================================================================================
DIFFERENCE-IN-DIFFERENCES RESULTS
Outcome: DiD estimator (is_AZ x is_UK interaction)
Positive coef = AZ treated more favourably in UK vs US than JJ
Significance: *** p<0.01  ** p<0.05  * p<0.1
================================================================================
                    Window                          Outcome  DiD coef  CI lower  CI upper  p-value     N Sig
               Full sample general_sentiment_adjusted_source    0.0173   -0.0122    0.0467   0.2513 14999
               Full sample    body_sentiment_adjusted_source    0.1332    0.0867    0.1796   0.0000 14870 ***
Dev & Approval (Mar20-Jun21) general_sentiment_adjusted_source    0.0365   -0.0124    0.0854   0.1432  4106
Dev & Approval (Mar20-Jun21)    body_sentiment_adjusted_source    0.1756    0.1000    0.2512   0.0000  4030 ***
         AZ release (Dec20) general_sentiment_adjusted_source    0.1048   -0.1267    0.3364   0.3750   325
         AZ release (Dec20)    body_sentiment_adjusted_source    0.1973   -0.1404    0.5350   0.2521   313
         JJ release (Feb21) general_sentiment_adjusted_source    0.1721    0.0582    0.2860   0.0031   771 ***
         JJ release (Feb21)    body_sentiment_adjusted_source    0.2946    0.1100    0.4792   0.0018   757 ***
      AZ suspension (Mar21) general_sentiment_adjusted_source    0.0355   -0.0819    0.1529   0.5531   703
      AZ suspension (Mar21)    body_sentiment_adjusted_source    0.1552   -0.0210    0.3315   0.0844   696   *
      JJ suspension (Apr21) general_sentiment_adjusted_source    0.0646   -0.0395    0.1687   0.2238   657
      JJ suspension (Apr21)    body_sentiment_adjusted_source    0.2633    0.1035    0.4230   0.0012   652 ***


MAIN RESULT (Full sample, body sentiment):
  DiD coefficient: 0.1332
  95% CI: [0.0867, 0.1796]
  p-value: 0.0 ***
  N: 14870
```

Source: Own elaboration from Python analysis based on Factiva data

During the vaccine development and approval phase (March 2020 to June 2021), the DiD estimate for body sentiment is 0.1756 and highly significant ($p < 0.001$, $N = 4{,}030$), suggesting that home-country bias was particularly strong during this period. This is consistent with the descriptive finding that AstraZeneca occupied a central position in the UK vaccine narrative during the development phase.

During the Johnson & Johnson vaccine release window (February 2021), the DiD estimate is 0.2946 for body sentiment ($p = 0.002$, $N = 757$), indicating that even during J&J's major regulatory milestone, UK newspapers continued to treat AstraZeneca significantly more favourably relative to how US newspapers treated Johnson & Johnson.

During the AstraZeneca suspension window (March 2021), the DiD estimate for body sentiment is 0.1552 and significant at the 10 percent level ($p = 0.084$, $N = 696$). Although weaker than in positive event windows, the result suggests that home-country bias persisted even when AstraZeneca was the subject of negative safety-related news, manifesting as relatively softer treatment in UK newspapers rather than positive coverage.

During the Johnson & Johnson suspension window (April 2021), the DiD estimate is the largest across all windows at 0.2633 for body sentiment ($p = 0.001$, $N = 652$). This suggests that when Johnson & Johnson faced its own safety crisis, UK newspapers treated AstraZeneca significantly more favourably relative to how US newspapers treated Johnson & Johnson.

The AstraZeneca release window (December 2020) does not yield a statistically significant DiD estimate for either outcome variable, likely due to the relatively small

sample size available within that narrow event window (N = 313–325). The differences observed in that period should therefore be interpreted with caution.

Taken together, the event-window DiD results suggest that home-country bias is not only a full-sample phenomenon but is statistically detectable across multiple specific periods.

Results suggest that home-country bias is strongest during positive or achievement-oriented events. However, this favourable treatment weakens during safety-related windows. In those cases, domestic bias does not disappear completely, but it is more likely to appear as relatively less negative coverage rather than as positive sentiment.

Therefore, the event-window analysis refines the interpretation of home-country bias. Bias should not be understood as a constant tendency to praise domestic firms. Instead, it operates conditionally, which means that it is stronger when the event is positive and weaker when the event is negative.

2. Robustness consideration: single-company articles

One potential limitation of article-level sentiment analysis is that some articles mention both AstraZeneca and Johnson & Johnson. Although entity-level sentiment attribution was applied in those cases, complex narratives in which both companies are involved may have led to wrong attribution, and, therefore, misleading results.

For this reason, an additional robustness check consists of replicating the same calculations using only articles that mention one company. This subset reduces ambiguity because the sentiment of the article is more directly attributable to either AstraZeneca or Johnson & Johnson. If the same domestic-versus-foreign patterns remain visible in the single-company sample or become even stronger, the results would be more robust.

Table 14. Single-Company Robustness Check of Adjusted Sentiment by Company

	company	country	n_articles	avg_title_adjusted	avg_snippet_adjusted	avg_body_adjusted
0	AZ	UK	5076	-0.000907	0.000888	0.007454
1	AZ	US	2088	-0.010742	-0.004688	-0.012357
2	J&J	UK	696	0.006616	-0.006476	-0.054362
3	J&J	US	4231	0.005301	0.002314	0.006098

Source: Own elaboration from Python analysis based on Factiva data

The robustness check confirms that the domestic-versus-foreign pattern identified in the full sample is not driven by ambiguous sentiment attribution in dual-company articles. When the analysis is restricted to articles mentioning only one company, the results remain consistent with the home-country bias hypothesis across all three sentiment components.

In terms of body sentiment, which is the component where the pattern is most visible, AstraZeneca receives positive adjusted sentiment in UK newspapers (+0.007) but negative adjusted sentiment in US newspapers (-0.012). Johnson & Johnson shows the opposite pattern: it receives clearly negative adjusted body sentiment in UK newspapers (-0.054) but positive adjusted body sentiment in US newspapers (+0.006). These differences are directionally identical to those observed in the full sample, and the gap between domestic and foreign treatment is of comparable magnitude.

At the title level, differences are small and close to zero across all groups, which is also consistent with the full-sample finding that home-country bias is weak in headlines. At the snippet level, the pattern is present but more modest than in the body.

Overall, the single-company robustness check strengthens the main findings. The fact that the domestic-versus-foreign sentiment pattern remains visible, and even becomes more pronounced in some cases, when dual-company articles are excluded suggests that the entity-level attribution procedure applied to those articles did not introduce systematic error into the results. The home-country bias pattern appears to be a genuine feature of the data rather than a methodological artefact.

3. Overall interpretation of sentiment analysis

Overall, the sentiment analysis provides evidence consistent with home-country media bias, although the pattern is not uniform across all dimensions.

The clearest evidence appears when sentiment is analysed at a more detailed level, in the body of the articles. This suggests that home-country bias is not always visible in headline tone, which briefly summarizes an event in a few words, or aggregate averages, but becomes clearer when the full narrative and editorial context are considered. At the country level, article bodies reveal a clear domestic-versus-foreign pattern. UK newspapers tend to treat AstraZeneca more favourably than Johnson & Johnson, while US newspapers tend to do the opposite.

The newspaper-level analysis reinforces this conclusion. In the US sample, all three newspapers treat J&J more favourably than AZ. In the UK sample, most newspapers treat AZ more favourably than J&J, with The Sunday Times as the main exception as an outlet more dedicated to investigative journalism. This shows that home-country bias is not mechanical, but it appears as a strong pattern across most outlets once each newspaper is analysed individually.

The section-level analysis adds an important finding. Sentiment changes substantially depending on whether the pharma articles appear in a business section, a frontpage, or a health-related section. Business coverage tends to generate a more positive environment for both companies, while frontpage and main-section coverage is generally more negative. However, even after accounting for section, the domestic company is often treated more positively than the foreign comparator. This means that article section affects the general level of sentiment, but does not eliminate the evidence of home-country bias.

The event-window analysis further shows that bias depends on the type of event. During positive vaccine-related milestones, domestic favourability becomes more visible. During safety-related events, sentiment becomes more negative overall, but domestic companies may still receive relatively softer treatment. Home-country bias therefore result in news which can be amplified, weakened or partially hidden depending on the nature of the event.

Taken together, the results suggest that media bias should not be interpreted only as openly positive coverage of domestic firms. It may also appear as greater visibility, less negative framing, more favourable contextualization or stronger narrative legitimacy. The analysis does not prove causality, since the dataset is observational and both companies were exposed to different events. Nevertheless, the findings show systematic differences in the treatment of AstraZeneca and Johnson & Johnson across newspapers countries of origin. This provides meaningful descriptive evidence that home-country bias is present in financial and general-interest newspaper coverage of pharmaceutical firms.

Section VII: Conclusions

1. Answer to research questions

This thesis set out to examine whether British and American newspapers display home-country bias when covering AstraZeneca and Johnson & Johnson. The evidence gathered across five dimensions of analysis, including coverage intensity, country-level sentiment, newspaper-level sentiment, article section, and event windows, consistently points in the same direction, although with important nuances.

Do newspapers publish more articles about the company associated with their own country?

Yes. UK newspapers publish a substantially higher share of articles mentioning AstraZeneca, while US newspapers publish proportionally more articles mentioning Johnson & Johnson. This asymmetry is even more pronounced when the analysis is restricted to articles in which the company name appears in the title, suggesting that domestic firms not only receive more coverage but also more prominent placement.

Do domestic companies receive more favourable sentiment than foreign companies after adjusting for each newspaper's baseline tone?

Yes, particularly in the body of articles. At the country level, UK newspapers assign more positive adjusted body sentiment to AstraZeneca than to Johnson & Johnson, while US newspapers do the opposite. The Difference-in-Differences estimator for body sentiment across the full sample is 0.133 and is statistically significant at the 1 percent level ($p <$

0.001, N = 14,870), confirming that the domestic-versus-foreign pattern is not a random fluctuation.

Does home-country bias appear differently in titles, snippets and article bodies?

Yes. The pattern is weakest in titles, more visible in snippets, and strongest in the body of articles. This finding suggests that home-country bias is not primarily expressed through explicit headline framing, but through the way newspapers develop context, select which aspects to emphasize and construct the narrative in the full article text.

Does the pattern vary across individual newspapers and article sections?

Yes, substantially. At the newspaper level, all three US newspapers treat Johnson & Johnson more favourably than AstraZeneca, and four out of five UK newspapers treat AstraZeneca more favourably than Johnson & Johnson. The Sunday Times is the main exception, likely due to its more adversarial editorial culture. At the section level, business and finance articles generate a more positive environment for both companies, while frontpage and main-section articles are generally more negative. However, even within sections, the domestic company tends to receive relatively better treatment than the foreign comparator.

Do specific event windows reinforce or weaken the evidence of home-country bias?

Both. During positive events, such as vaccine development announcements and regulatory approvals, domestic favourability is particularly strong and statistically significant. During safety-related events, the overall sentiment becomes negative for both companies, but the domestic company tends to receive softer treatment than the foreign

comparator. The DiD estimates remain significant across most event windows, with the largest estimate observed during the Johnson & Johnson suspension period (DiD = 0.263, p = 0.001), when UK newspapers continued to treat AstraZeneca more favourably even as both companies faced safety concerns simultaneously.

2. Implications for media, investors and firms

The findings of this thesis carry several practical implications for different audiences.

For media readers and investors, the results suggest that financial and general-interest newspaper coverage of pharmaceutical firms is not neutral with respect to the national origin of the company being covered. Readers who rely on a single national source of information may be systematically exposed to a more favourable framing of domestically associated firms and a more critical framing of foreign competitors. Investors making decisions based on media narratives should therefore be aware that sentiment signals extracted from news articles may be partially shaped by home-country framing rather than by underlying corporate fundamentals alone. Cross-national news aggregation can therefore help mitigate this source of bias.

For firms operating across national markets, the results suggest that reputational dynamics in international media are asymmetric. A foreign firm entering a new market may face structurally less favourable media coverage than a domestically associated competitor, even under comparable corporate conditions. This asymmetry may be particularly pronounced during high-salience events, where media framing has the greatest potential to influence public trust and investor sentiment. Foreign firms may therefore benefit from

investing in local communication strategies and media relationships to partially offset the informational disadvantage implied by home-country bias.

For researchers and journalists, the findings reinforce the importance of cross-national comparative approaches in media analysis.

Future research could extend the methodology to other industries, other country pairs, or longer time horizons to assess whether the patterns identified here are specific to the pharmaceutical sector or reflect a broader feature of financial journalism.

3. Limitations and future research

This thesis has several limitations that should be acknowledged when interpreting the results.

First, the analysis relies on VADER, a general-purpose lexicon-based sentiment tool. Although VADER performs well on news content and has been widely used in financial text analysis, it was not specifically developed for pharmaceutical journalism. Some domain-specific terms may not be optimally scored by the VADER lexicon. Future research could compare VADER results with those obtained using transformer-based models such as FinBERT, which may capture more nuanced sentiment in specialised financial texts.

Second, the Difference-in-Differences specification used in this study does not constitute a causal identification strategy in the strict econometric sense, as mentioned throughout the study. There is no fully untreated control group, and both companies were exposed to many of the same industry-level shocks throughout the sample period. The results are

therefore interpreted as descriptive evidence of patterns consistent with home-country bias rather than as causal estimates. However, a more rigorous causal design with clearer treatment-control distinction is difficult to construct in observational media data.

Third, the sample is limited to eight newspapers from two countries. Although the selected outlets include different types of newspapers, ranging from broadsheet and financial press to more popular national newspapers, they do not capture the full media ecosystem. Moreover, the media landscapes of the UK and the US are not perfectly equivalent, meaning that newspapers with a similar role or reputation in one country may not be fully comparable to those in the other. Therefore, the results should be interpreted as evidence from a selected group of influential national newspapers rather than as a representation of all media coverage.

Finally, this study focuses exclusively on coverage intensity and sentiment as a measure of media bias. However, there are other dimensions such as article placement or narrative structure through which home-country bias can operate. A multimodal analysis combining sentiment with these other dimensions would provide a richer and more complete picture of how national identity shapes financial journalism.

Bibliography

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention on the buying behavior of individual and institutional investors. *Review of Financial Studies*, 21(2), 785–818.

Coval, J. D., & Moskowitz, T. J. (1999). Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance*, 54(6), 2045–2073.

Department for Digital, Culture, Media & Sport. (2019). The Times and The Sunday Times application: Secretary of State decision. GOV.UK. https://assets.publishing.service.gov.uk/media/5ff6f242d3bf7f65d066a8cc/Final_MT_Letter_to_Parties__Times_S_Times_undertakings__SoS_decision_-_Google_Docs_Redacted_V2.pdf

DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. Journal of Health Economics, 47, 20–33.

Dyck, A., & Zingales, L. (2003). The media and asset prices. *Working Paper*, Harvard Business School.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417.

Grabowski, H., & Kyle, M. (2007). Generic competition and market exclusivity periods in pharmaceuticals. *Managerial and Decision Economics*, 28(4–5), 491–502.

Gentzkow, M., & Shapiro, J. M. (2010). What drives media slant? Evidence from U.S. daily newspapers. *Econometrica*, 78(1), 35–71.

Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216–225. https://doi.org/10.1609/icwsm.v8i1.14550

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35–65.

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature*, 35(1), 13–39.

Mullainathan, S., & Shleifer, A. (2005). The market for news. *American Economic Review*, 95(4), 1031–1053.

News UK. (n.d.). The Sunday Times. Retrieved May 20, 2026, from https://www.news.co.uk/our-brands/the-sunday-times/

Reinhardt, U. E., Hussey, P. S., & Anderson, G. F. (2004). U.S. health care spending in an international context. Health Affairs, 23(3), 10–25.

Scannell, J. W., Blanckley, A., Boldon, H., & Warrington, B. (2012). Diagnosing the decline in pharmaceutical R&D efficiency. *Nature Reviews Drug Discovery*, 11(3), 191–200.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139–1168.

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3), 1437–1467.

Topol, E. J. (2004). Failing the public health—Rofecoxib, Merck, and the FDA. *New England Journal of Medicine*, 351(17), 1707–1709.

Zhu, H. (2021). Home country bias in academic publishing: A case study of the New England Journal of Medicine. *Learned Publishing*, 34(4), 578–584.

Appendix

1. Reproducibility package: Python code

Preprocessing of data

```
# 1. Importing dataset

import pandas as pd

# Establish path
DATA = "/content/AZ_J&J.xlsx"

# Basic cleaning in excel (columns & formatting, no nule values)

df = pd.read_excel(DATA)

# Table to visualize columns
columns_table = pd.DataFrame({
    "column_name": df.columns,
    "index": range(len(df.columns))
})

print(columns_table)

# %%
# 2. Including company named (already in database but no accuracy)

import re

# Establish search patterns and create function to detect them in the title, snippet or body

jj_pattern = r"\b(johnson's?|janssen|johnson\s*&?\s*johnson|j&j|johnson)\b"
az_pattern = r"\b(astra[\s-]*zeneca|az)\b"    #validate that it does not create false positives "az"

jj_regex = re.compile(jj_pattern, re.IGNORECASE)
az_regex = re.compile(az_pattern, re.IGNORECASE)

def detect_company_code(row):
    text = " ".join([
        str(row.get("title", "")),
        str(row.get("snippet", "")),
        str(row.get("body", ""))
    ])

    jj = bool(jj_regex.search(text))
    az = bool(az_regex.search(text))

    # Codify depending on which company appears
```

```
    if jj and az:
        return 3
    elif jj:
        return 1
    elif az:
        return 2
    else:
        return 0

# Apply to dataframe

df["company_mentioned"] = df.apply(detect_company_code, axis=1)

# Print the heading of the result

print(df[["title", "company_mentioned"]].head())

# %%
# Including a columns to refer to mentions in the title

def detect_company_title_code(row):
    text = str(row.get("title", ""))

    jj = bool(jj_regex.search(text))
    az = bool(az_regex.search(text))

    if jj and az:
        return 3
    elif jj:
        return 1
    elif az:
        return 2
    else:
        return 0

df["company_mentioned_title"] = df.apply(detect_company_title_code, axis=1)

print(df[["title", "company_mentioned", "company_mentioned_title"]].head())

# %%
# 3. Dividing by type of article (using column section)

import unicodedata

# Create function to ensure similar names will be considered the same
def clean_section(x):
    if pd.isna(x):
        return ""
    x = str(x).strip()  # Eliminar espacios extra alrededor
    x = re.sub(r"\s+", "", x)  # No spaces
    return x.lower()  # Lower case
```

```
# Define the groups (used GPT-5.4 to categorize them accordingly using column section)

group_patterns = {
    "Negocio / Economía / Finanzas": [
        "business", "markets", "marketplace", "financial", "finance", "economy", "earnings",
        "personalfinance", "money", "money&investing", "invest", "investing", "wsjpro",
        "proprivatemarkets", "creditmarkets", "marketwatch", "marketweek",
"worldstockmarkets",
        "europemarkets", "asiamarkets", "options", "mutualfunds"
    ],
    "Noticias generales / Portada / Secciones principales": [
        "news", "daily", "pageone", "frontsection", "asection", "a-section", "uapageone",
        "leadingthenews", "business;frontpage", "news;frontpage", "what’snews",
        "breakingnews", "generalnews"
    ],
    "Internacional / Mundo / Geografía": [
        "world", "worldnews", "international", "foreign", "overseasnews", "u.s.", "us", "u.s.news",
"usnews",
        "uknews", "australianews", "europe", "asia", "latinamerica", "newyork", "newyorkregion"
    ],
    "Política / Sociedad / Público": [
        "politics", "politics&policy", "national", "national-politics", "national-economy", "national-
enterprise",
        "society", "education", "law", "investigative"
    ],
    "Salud / Ciencia / Medio ambiente": [
        "health", "personalhealth", "healthandwellness", "health&family", "science",
"sciencejournal",
        "environment", "health-environment-science", "inthelab", "coronavirus"
    ],
    "Tecnología / Digital / Media": [
        "tech", "technology", "techcenter", "technologyjournal", "technologyweek",
"businesstechnology",
        "asiatechnology", "technology&media", "media", "media&marketing",
"marketing&media", "the-switch"
    ],
    "Opinión / Editorial / Columnas": [
        "opinion", "editorial", "editorial-opinion", "oped", "commentisfree", "letters",
"letterstotheeditor",
        "columnist", "heardonthestreet", "features;opinion", "editorial;opinioncolumns"
    ],
    "Lifestyle / Consumo / Cultura / Ocio": [
        "life", "lifestyle", "lifeandstyle", "food", "travel", "books", "bookworld", "music",
        "tvweek", "leisure&arts", "theguide", "escape", "atleisure", "homenews", "style", "home",
"yourmoney"
    ],
    "Deporte": [
        "sport", "sports", "guardiansportpages"
    ],
    "Inmobiliario": [
        "realestate", "newyorkrealestate", "localliving"
    ],
```

```
    "Suplementos / marcas editoriales / páginas de cabecera": [
        "guardianfinancialpages", "guardiancitypages", "guardianhomepages",
"guardianfeaturespages",
        "guardianinternationalpages", "guardiansocietypages", "guardianssciencepages",
        "guardianmoneypages", "guardianleaderpages", "barronsmagazine", "barron'scover",
"guardianweekly"
    ],
    "Extras locales / regionales / especiales": [
        "d.c.extra", "annearundelextra", "montgomeryextra", "fairfaxextra",
"southernmarylandextra",
        "saturdayextra", "questnews", "special", "preview", "review", "year-endreview"
    ]
}

# Define function to asign article type

def assign_group(section):
    s = clean_section(section)

    if s == "":
        return "Other"  # If there is no value or 0

    for group, patterns in group_patterns.items():
        for pattern in patterns:
            if re.search(pattern, s):
                return group

    return "Other"  # If it does not belong to any group

# Create new column

df["article_type"] = df["section"].apply(assign_group)

# Print the heading of the result

print(df[["section", "article_type"]].head())

# Download an excel file to check

data_preprocessed = "/content/data_preprocessed.xlsx"
df.to_excel(data_preprocessed, index=False)


# %%
# Distribution graphics

import matplotlib.pyplot as plt

# %%
df['publication_date'] = pd.to_datetime(df['year'], unit='ns', errors='coerce')

# Extract real year
```

```
df['year'] = df['publication_date'].dt.year

df['year'] = pd.to_numeric(df['year'], errors='coerce')

# Group years in 5-year periods
df['year_group'] = pd.cut(
    df['year'],
    bins=[1999, 2005, 2010, 2015, 2020, 2025],
    labels=['2000-2005', '2006-2010', '2011-2015', '2016-2020', '2021-2025']
)

# %%
columns_to_plot = ['country', 'company_mentioned', 'source_name', 'year_group', 'article_type']

axis_labels = {
    'country': 'Country',
    'company_mentioned': 'Company mentioned',
    'source_name': 'Source',
    'year_group': 'Years',
    'article_type': 'Type of article'
}

company_labels = {
    0: 'NONE',
    1: 'J&J',
    2: 'AZ',
    3: 'BOTH'
}

for col in columns_to_plot:
    percentages = df[col].value_counts(normalize=True).sort_index() * 100

    if col == 'company_mentioned':
        percentages = percentages.rename(index=company_labels)

    ax = percentages.plot(
        kind='barh',
        figsize=(10, max(4, len(percentages) * 0.45)),
        color='lightblue'
    )

    ax.bar_label(
        ax.containers[0],
        labels=[f'{v:.1f}%' for v in percentages.values],
        padding=3
    )

    plt.title(f'Distribution by {axis_labels[col]}', pad=20)
    plt.xlabel('Percentage of articles (%)')
    plt.ylabel(axis_labels[col])

    plt.xlim(0, percentages.max() * 1.20)
```

```
    plt.tight_layout()
    plt.show()

# %%
# Crossing variables for coverage analysis

labels = {
    0: 'NONE',
    1: 'J&J',
    2: 'AZ',
    3: 'BOTH'
}

ct = pd.crosstab(df['country'], df['company_mentioned'])
ct = ct.rename(columns=labels)

# No representar NONE
ct = ct.drop(columns='NONE', errors='ignore')

# Convertir a porcentaje por país
ct_percent = ct.div(ct.sum(axis=1), axis=0) * 100

# Colores por compañía
colors = {
    'J&J': 'lightcoral',
    'AZ': 'khaki',
    'BOTH': 'lightgray'
}

ax = ct_percent.plot(
    kind='bar',
    figsize=(10, 4),
    color=[colors[col] for col in ct_percent.columns]
)

# Etiquetas de porcentaje encima de cada barra
for container in ax.containers:
    ax.bar_label(
        container,
        labels=[f'{v:.1f}%' if v > 0 else '' for v in container.datavalues],
        padding=3,
        fontsize=8
    )

plt.title('Company mentioned by country')
plt.xlabel('Country of source')
plt.ylabel('Percentage of articles (%)')
plt.xticks(rotation=45, ha='right')
plt.legend(title='Company mentioned')

plt.ylim(0, ct_percent.max().max() * 1.20)
```

```
plt.tight_layout()
plt.show()

# %%
# Crossing variables for coverage analysis

labels = {
    0: 'NONE',
    1: 'J&J',
    2: 'AZ',
    3: 'BOTH'
}

ct = pd.crosstab(df['country'], df['company_mentioned_title'])
ct = ct.rename(columns=labels)

# No representar NONE
ct = ct.drop(columns='NONE', errors='ignore')

# Convertir a porcentaje por país
ct_percent = ct.div(ct.sum(axis=1), axis=0) * 100

# Colores por compañía
colors = {
    'J&J': 'lightcoral',
    'AZ': 'khaki',
    'BOTH': 'lightgray'
}

ax = ct_percent.plot(
    kind='bar',
    figsize=(10, 4),
    color=[colors[col] for col in ct_percent.columns]
)

# Etiquetas de porcentaje encima de cada barra
for container in ax.containers:
    ax.bar_label(
        container,
        labels=[f'{v:.1f}%' if v > 0 else '' for v in container.datavalues],
        padding=3,
        fontsize=8
    )

plt.title('Company mentioned in title by country')
plt.xlabel('Country of source')
plt.ylabel('Percentage of articles (%)')
plt.xticks(rotation=45, ha='right')
plt.legend(title='Company mentioned in title')

plt.ylim(0, ct_percent.max().max() * 1.20)
```

```
plt.tight_layout()
plt.show()

# %%
#Other crossings for further analysis

df_filtered = df[df['article_type'] == 'Negocio / Economía / Finanzas']

pd.crosstab(df_filtered['country'], df_filtered['article_type']).plot(
    kind='bar',
    figsize=(10, 4),
    color='lightblue',
    legend=False
)

plt.title('Business and Finance Articles by Source')
plt.xlabel('Country')
plt.ylabel('Number of articles')
plt.xticks(rotation=45, ha='right')
plt.show()

df_filtered = df[df['article_type'] == 'Negocio / Economía / Finanzas']

pd.crosstab(df_filtered['source_name'], df_filtered['article_type']).plot(
    kind='bar',
    figsize=(10, 4),
    color='lightblue',
    legend=False
)

plt.title('Business and Finance Articles by Source')
plt.xlabel('Source')
plt.ylabel('Number of articles')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Data Analysis

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import re

# %%
# Importing dataset

# Establish path
DATA = "/content/data_preprocessed.xlsx"

# Basic cleaning in excel (columns y formatting, no nule values)
```

```
df = pd.read_excel(DATA)

# Table to visualize columns
columns_table = pd.DataFrame({
    "column_name": df.columns,
    "index": range(len(df.columns))
})

print(columns_table)

# %%
# Sentiment analysis using Vader (install and import)

!pip -q install vaderSentiment

from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

analyzer = SentimentIntensityAnalyzer()

# %%
# Preprocessing of data for sentiment analysis

def preprocess_text(text):
    if isinstance(text, str):
        return text.lower()
    return str(text)

df['cleaned_title'] = df['title'].apply(preprocess_text)
df['cleaned_snippet'] = df['snippet'].apply(preprocess_text)
df['cleaned_body'] = df['body'].apply(preprocess_text)

# %% [markdown]
# Analysis of newspaper z-scores

# %%
def sentiment_analyzer_scores(sentence):
    return analyzer.polarity_scores(sentence)

df['title_sentiment'] = df['cleaned_title'].apply(
    lambda x: sentiment_analyzer_scores(x)['compound']
)

df['snippet_sentiment'] = df['cleaned_snippet'].apply(
    lambda x: sentiment_analyzer_scores(x)['compound']
)

df['body_sentiment'] = df['cleaned_body'].apply(
    lambda x: sentiment_analyzer_scores(x)['compound']
)

df['general_sentiment'] = df[
```

```
    ['title_sentiment', 'snippet_sentiment', 'body_sentiment']
].mean(axis=1)

# %%
sentiment_cols = [
    'title_sentiment',
    'snippet_sentiment',
    'body_sentiment',
    'general_sentiment'
]

for col in sentiment_cols:
    df[col] = pd.to_numeric(df[col], errors='coerce')

df['title_sentiment_zscore'] = (
    df['title_sentiment'] - df['title_sentiment'].mean()
) / df['title_sentiment'].std()

df['snippet_sentiment_zscore'] = (
    df['snippet_sentiment'] - df['snippet_sentiment'].mean()
) / df['snippet_sentiment'].std()

df['body_sentiment_zscore'] = (
    df['body_sentiment'] - df['body_sentiment'].mean()
) / df['body_sentiment'].std()

df['general_sentiment_zscore'] = (
    df['general_sentiment'] - df['general_sentiment'].mean()
) / df['general_sentiment'].std()

# %%
zscore_by_source = df.groupby(['source_name', 'country']).agg(
    n_articles=('general_sentiment_zscore', 'count'),
    avg_general_sentiment=('general_sentiment', 'mean'),
    avg_general_zscore=('general_sentiment_zscore', 'mean'),
    avg_title_zscore=('title_sentiment_zscore', 'mean'),
    avg_snippet_zscore=('snippet_sentiment_zscore', 'mean'),
    avg_body_zscore=('body_sentiment_zscore', 'mean')
).reset_index()

zscore_by_source = zscore_by_source.sort_values(
    by='avg_general_zscore',
    ascending=True
)

zscore_by_source

# %%
# Representing these z-scores graphically

zscore_by_source.plot(
    x='source_name',
```

```
    y='avg_general_zscore',
    kind='bar',
    figsize=(10, 4),
    legend=False
)

plt.axhline(0, linestyle='--')
plt.title('Average sentiment z-score by newspaper')
plt.xlabel('Newspaper')
plt.ylabel('Average z-score')
plt.xticks(rotation=45, ha='right')
plt.show()

# %%
# Checking positivity or negativity of countries

positivity_country = df.groupby('country').agg(
    n_articles=('general_sentiment', 'count'),
    avg_general_sentiment=('general_sentiment', 'mean'),
    avg_title_sentiment=('title_sentiment', 'mean'),
    avg_snippet_sentiment=('snippet_sentiment', 'mean'),
    avg_body_sentiment=('body_sentiment', 'mean')
).reset_index()

positivity_country

# %%

# %%
# Articles where only one company is mentioned
# 1 = J&J
# 2 = AZ

df_12 = df[df['company_mentioned'].isin([1, 2])].copy()

# Articles where both companies are mentioned
# 3 = BOTH

df_both = df[df['company_mentioned'].isin([3])].copy()

# %% [markdown]
# Analysis of 1&2

# %%
company_map = {1: 'J&J', 2: 'AZ'}
df_12['company'] = df_12['company_mentioned'].map(company_map)

# %% [markdown]
# Analysis of 3

# %%
# Patterns for J&J and AstraZeneca
```

```
jj_pattern = r"\b(johnson's?|johnson's?|janssen|johnson\s*&?\s*johnson|j&j|johnson)\b"
az_pattern = r"\b(astra[\s-]*zeneca|astrazeneca|az)\b"

def company_sentiment_from_sentences(text, company_pattern):
    if pd.isna(text):
        return np.nan

    text = str(text)

    sentences = re.split(r'(?<=[.!?])\s+', text)

    company_sentences = [
        sentence for sentence in sentences
        if re.search(company_pattern, sentence, flags=re.IGNORECASE)
    ]

    if len(company_sentences) == 0:
        return np.nan

    scores = [
        analyzer.polarity_scores(sentence)['compound']
        for sentence in company_sentences
    ]

    return np.mean(scores)

# %%
# J&J sentiment by part of article

df_both['jj_title_sentiment'] = df_both['cleaned_title'].apply(
    lambda x: company_sentiment_from_sentences(x, jj_pattern)
)

df_both['jj_snippet_sentiment'] = df_both['cleaned_snippet'].apply(
    lambda x: company_sentiment_from_sentences(x, jj_pattern)
)

df_both['jj_body_sentiment'] = df_both['cleaned_body'].apply(
    lambda x: company_sentiment_from_sentences(x, jj_pattern)
)

# AZ sentiment by part of article

df_both['az_title_sentiment'] = df_both['cleaned_title'].apply(
    lambda x: company_sentiment_from_sentences(x, az_pattern)
)

df_both['az_snippet_sentiment'] = df_both['cleaned_snippet'].apply(
    lambda x: company_sentiment_from_sentences(x, az_pattern)
)
```

```
df_both['az_body_sentiment'] = df_both['cleaned_body'].apply(
    lambda x: company_sentiment_from_sentences(x, az_pattern)
)

# General sentiment by company

df_both['jj_general_sentiment'] = df_both[
    ['jj_title_sentiment', 'jj_snippet_sentiment', 'jj_body_sentiment']
].mean(axis=1)

df_both['az_general_sentiment'] = df_both[
    ['az_title_sentiment', 'az_snippet_sentiment', 'az_body_sentiment']
].mean(axis=1)

# %%
# Making sure df_both sentiment columns are numeric

both_sentiment_cols = [
    'jj_title_sentiment',
    'jj_snippet_sentiment',
    'jj_body_sentiment',
    'jj_general_sentiment',
    'az_title_sentiment',
    'az_snippet_sentiment',
    'az_body_sentiment',
    'az_general_sentiment'
]

for col in both_sentiment_cols:
    df_both[col] = pd.to_numeric(df_both[col], errors='coerce')

# %%
base_cols = [
    'source_name', 'country', 'publication_date', 'year',
    'section', 'article_type', 'company',
    'title_sentiment', 'snippet_sentiment',
    'body_sentiment', 'general_sentiment'
]

# %%
# Articles of 1 company
df_12_long = df_12[base_cols].copy()

# %%
# Articles from both divide into 2 lines
df_both_jj = df_both.copy()
df_both_jj['company'] = 'J&J'
df_both_jj['title_sentiment'] = df_both_jj['jj_title_sentiment']
df_both_jj['snippet_sentiment'] = df_both_jj['jj_snippet_sentiment']
df_both_jj['body_sentiment'] = df_both_jj['jj_body_sentiment']
df_both_jj['general_sentiment'] = df_both_jj[
    ['title_sentiment', 'snippet_sentiment', 'body_sentiment']
```

```
].mean(axis=1)

df_both_az = df_both.copy()
df_both_az['company'] = 'AZ'
df_both_az['title_sentiment'] = df_both_az['az_title_sentiment']
df_both_az['snippet_sentiment'] = df_both_az['az_snippet_sentiment']
df_both_az['body_sentiment'] = df_both_az['az_body_sentiment']
df_both_az['general_sentiment'] = df_both_az[
    ['title_sentiment', 'snippet_sentiment', 'body_sentiment']
].mean(axis=1)

# %%
df_sentiment_final = pd.concat(
    [
        df_12_long,
        df_both_jj[base_cols],
        df_both_az[base_cols]
    ],
    ignore_index=True
)

sentiment_cols_final = [
    'title_sentiment',
    'snippet_sentiment',
    'body_sentiment',
    'general_sentiment'
]

for col in sentiment_cols_final:
    df_sentiment_final[col] = pd.to_numeric(df_sentiment_final[col], errors='coerce')

df_sentiment_final.head()

# %%
# Recalculate newspaper baseline using the final analysis dataset
df_sentiment_final['source_avg_title_sentiment_final'] = (
    df_sentiment_final.groupby('source_name')['title_sentiment'].transform('mean')
)

df_sentiment_final['source_avg_snippet_sentiment_final'] = (
    df_sentiment_final.groupby('source_name')['snippet_sentiment'].transform('mean')
)

df_sentiment_final['source_avg_body_sentiment_final'] = (
    df_sentiment_final.groupby('source_name')['body_sentiment'].transform('mean')
)

df_sentiment_final['source_avg_general_sentiment_final'] = (
    df_sentiment_final.groupby('source_name')['general_sentiment'].transform('mean')
)

# Adjust sentiment relative to the final newspaper baseline
```

```
df_sentiment_final['title_sentiment_adjusted_source'] = (
    df_sentiment_final['title_sentiment'] -
    df_sentiment_final['source_avg_title_sentiment_final']
)

df_sentiment_final['snippet_sentiment_adjusted_source'] = (
    df_sentiment_final['snippet_sentiment'] -
    df_sentiment_final['source_avg_snippet_sentiment_final']
)

df_sentiment_final['body_sentiment_adjusted_source'] = (
    df_sentiment_final['body_sentiment'] -
    df_sentiment_final['source_avg_body_sentiment_final']
)

df_sentiment_final['general_sentiment_adjusted_source'] = (
    df_sentiment_final['general_sentiment'] -
    df_sentiment_final['source_avg_general_sentiment_final']
)

# %%
check_adjustment = df_sentiment_final.groupby('source_name').agg(
    mean_title_adjusted=('title_sentiment_adjusted_source', 'mean'),
    mean_snippet_adjusted=('snippet_sentiment_adjusted_source', 'mean'),
    mean_body_adjusted=('body_sentiment_adjusted_source', 'mean'),
    mean_general_adjusted=('general_sentiment_adjusted_source', 'mean')
).reset_index()

check_adjustment

# %%
# DID Analysis

import statsmodels.formula.api as smf
from scipy import stats

df_did = df_sentiment_final.copy()

# Dummy by company: 1 = AZ, 0 = J&J
df_did['is_AZ'] = (df_did['company'] == 'AZ').astype(int)

# Dummy país: 1 = UK, 0 = US
df_did['is_UK'] = (df_did['country'] == 'UK').astype(int)

def run_did(data, outcome, label):
    try:
        model = smf.ols(
            f'{outcome} ~ is_AZ + is_UK + is_AZ:is_UK',
            data=data
        ).fit(cov_type='HC3')

        coef = model.params['is_AZ:is_UK']
```

```
        pval = model.pvalues['is_AZ:is_UK']
        ci_low = model.conf_int().loc['is_AZ:is_UK', 0]
        ci_high = model.conf_int().loc['is_AZ:is_UK', 1]
        n = int(model.nobs)

        if pval < 0.01:
            sig = '***'
        elif pval < 0.05:
            sig = '**'
        elif pval < 0.1:
            sig = '*'
        else:
            sig = ''

        return {
            'Window': label,
            'Outcome': outcome,
            'DiD coef': round(coef, 4),
            'CI lower': round(ci_low, 4),
            'CI upper': round(ci_high, 4),
            'p-value': round(pval, 4),
            'N': n,
            'Sig': sig
        }
    except Exception as e:
        return {
            'Window': label,
            'Outcome': outcome,
            'DiD coef': None,
            'CI lower': None,
            'CI upper': None,
            'p-value': None,
            'N': None,
            'Sig': f'Error: {e}'
        }

# Event study: Covid19

windows = {
    'Full sample': df_did,

    'Dev & Approval (Mar20-Jun21)': df_did[
        (df_did['publication_date'] >= '2020-03-30') &
        (df_did['publication_date'] <= '2021-06-30')
    ],

    'AZ release (Dec20)': df_did[
        (df_did['publication_date'] >= '2020-12-15') &
        (df_did['publication_date'] <= '2021-01-15')
    ],

    'JJ release (Feb21)': df_did[
```

```
        (df_did['publication_date'] >= '2021-02-01') &
        (df_did['publication_date'] <= '2021-03-15')
    ],

    'AZ suspension (Mar21)': df_did[
        (df_did['publication_date'] >= '2021-03-07') &
        (df_did['publication_date'] <= '2021-04-07')
    ],

    'JJ suspension (Apr21)': df_did[
        (df_did['publication_date'] >= '2021-04-13') &
        (df_did['publication_date'] <= '2021-05-13')
    ],
}

# Run DiD pfir each

outcomes = [
    'general_sentiment_adjusted_source',
    'body_sentiment_adjusted_source'
]

results = []

for window_name, window_data in windows.items():
    for outcome in outcomes:
        result = run_did(window_data, outcome, window_name)
        results.append(result)

df_did_results = pd.DataFrame(results)

print("=" * 80)
print("DIFFERENCE-IN-DIFFERENCES RESULTS")
print("Outcome: DiD estimator (is_AZ x is_UK interaction)")
print("Positive coef = AZ treated more favourably in UK vs US than JJ")
print("Significance: *** p<0.01  ** p<0.05  * p<0.1")
print("=" * 80)

print(df_did_results.to_string(index=False))

main = df_did_results[
    (df_did_results['Window'] == 'Full sample') &
    (df_did_results['Outcome'] == 'body_sentiment_adjusted_source')
].iloc[0]

print("\n")
print("MAIN RESULT (Full sample, body sentiment):")
print(f"  DiD coefficient: {main['DiD coef']}")
print(f"  95% CI: [{main['CI lower']}, {main['CI upper']}]")
print(f"  p-value: {main['p-value']} {main['Sig']}")
print(f"  N: {main['N']}")
```

```
# %% [markdown]
# Begginning analysis

# %%
# Checking home-country bias - full period

country_bias_adjusted = df_sentiment_final.groupby(
    ['company', 'country']
).agg(
    n_articles=('general_sentiment_adjusted_source', 'count'),
    avg_title_adjusted=('title_sentiment_adjusted_source', 'mean'),
    avg_snippet_adjusted=('snippet_sentiment_adjusted_source', 'mean'),
    avg_body_adjusted=('body_sentiment_adjusted_source', 'mean')
).reset_index()

country_bias_adjusted

# %%
import matplotlib.pyplot as plt

plot_country_adjusted = country_bias_adjusted.pivot_table(
    index='country',
    columns='company',
    values='avg_title_adjusted'
)

plot_country_adjusted.plot(
    kind='bar',
    figsize=(8, 4)
)

plt.title('Adjusted sentiment by country and company')
plt.xlabel('Country')
plt.ylabel('Title adjusted sentiment')
plt.xticks(rotation=0)
plt.axhline(0, linestyle='--')
plt.legend(title='Company')
plt.tight_layout()
plt.show()

# %%
import matplotlib.pyplot as plt

plot_country_adjusted = country_bias_adjusted.pivot_table(
    index='country',
    columns='company',
    values='avg_snippet_adjusted'
)

plot_country_adjusted.plot(
    kind='bar',
    figsize=(8, 4)
```

```
)

plt.title('Adjusted sentiment by country and company')
plt.xlabel('Country')
plt.ylabel('Snippet adjusted sentiment')
plt.xticks(rotation=0)
plt.axhline(0, linestyle='--')
plt.legend(title='Company')
plt.tight_layout()
plt.show()

# %%
import matplotlib.pyplot as plt

plot_country_adjusted = country_bias_adjusted.pivot_table(
    index='country',
    columns='company',
    values='avg_body_adjusted'
)

plot_country_adjusted.plot(
    kind='bar',
    figsize=(8, 4)
)

plt.title('Adjusted sentiment by country and company')
plt.xlabel('Country')
plt.ylabel('Body adjusted sentiment')
plt.xticks(rotation=0)
plt.axhline(0, linestyle='--')
plt.legend(title='Company')
plt.tight_layout()
plt.show()

# %%
# Checking adjusted bias by source_name, company, country

bias_by_source_company_adjusted = df_sentiment_final.groupby(
    ['source_name', 'company', 'country']
).agg(
    n_articles=('general_sentiment_adjusted_source', 'count'),
    avg_adjusted_sentiment=('general_sentiment_adjusted_source', 'mean'),
    avg_title_adjusted=('title_sentiment_adjusted_source', 'mean'),
    avg_snippet_adjusted=('snippet_sentiment_adjusted_source', 'mean'),
    avg_body_adjusted=('body_sentiment_adjusted_source', 'mean')
).reset_index()

bias_by_source_company_adjusted

# %%
# Including country to identify
bias_by_source_company_adjusted['newspaper_country'] = (
```

```
    bias_by_source_company_adjusted['source_name']
    + ' ('
    + bias_by_source_company_adjusted['country']
    + ')'
)

# Pivot table for adjusted sentiment
plot_adjusted = bias_by_source_company_adjusted.pivot_table(
    index='newspaper_country',
    columns='company',
    values='avg_body_adjusted'
)

plot_adjusted.plot(
    kind='bar',
    figsize=(12, 5)
)

plt.title('Newspaper-adjusted body sentiment by newspaper and company')
plt.xlabel('Newspaper (country)')
plt.ylabel('Adjusted average sentiment')
plt.xticks(rotation=45, ha='right')
plt.legend(title='Company')
plt.axhline(0, linestyle='--')
plt.tight_layout()
plt.show()

# %%

# %%
# Difference AZ vs J&J inside each newspaper, controlled for baseline tone

pivot_adjusted = bias_by_source_company_adjusted.pivot_table(
    index=['source_name', 'country'],
    columns='company',
    values='avg_adjusted_sentiment'
).reset_index()

pivot_adjusted['AZ_minus_JJ_adjusted'] = (
    pivot_adjusted['AZ'] - pivot_adjusted['J&J']
)

pivot_adjusted = pivot_adjusted.sort_values(
    by='AZ_minus_JJ_adjusted',
    ascending=True
)

pivot_adjusted

#AZ_minus_JJ_adjusted > 0  → el periódico trata mejor a AZ que a J&J
#AZ_minus_JJ_adjusted < 0  → el periódico trata mejor a J&J que a AZ
#AZ_minus_JJ_adjusted ≈ 0  → tratamiento parecido
```

```
# %%
# NOW WE WILL DO IT CONCENTRATING ON 3 SPECIFIC SECTIONS: health, business
and portada

# %%
df_filtered = df_sentiment_final[df_sentiment_final['article_type'] == 'Negocio / Economía /
Finanzas']

bias_business = df_filtered.groupby(
    ['source_name', 'company', 'country']
).agg(
    n_articles=('general_sentiment_adjusted_source', 'count'),
    avg_adjusted_sentiment=('general_sentiment_adjusted_source', 'mean'),
    avg_title_adjusted=('title_sentiment_adjusted_source', 'mean'),
    avg_snippet_adjusted=('snippet_sentiment_adjusted_source', 'mean'),
    avg_body_adjusted=('body_sentiment_adjusted_source', 'mean')
).reset_index()

plot_business = bias_business.pivot_table(
    index='source_name',
    columns='company',
    values='avg_body_adjusted'
)

plot_business.plot(
    kind='bar',
    figsize=(12, 5)
)

plt.title('Newspaper-adjusted sentiment by newspaper and company\n(Business / Economy /
Finance articles only)')
plt.xlabel('Newspaper')
plt.ylabel('Adjusted average sentiment')
plt.xticks(rotation=45, ha='right')
plt.legend(title='Company')
plt.axhline(0, linestyle='--')
plt.tight_layout()
plt.show()

# %%
df_filtered = df_sentiment_final[df_sentiment_final['article_type'] == 'Salud / Ciencia / Medio
ambiente']

bias_business = df_filtered.groupby(
    ['source_name', 'company', 'country']
).agg(
    n_articles=('general_sentiment_adjusted_source', 'count'),
    avg_adjusted_sentiment=('general_sentiment_adjusted_source', 'mean'),
    avg_title_adjusted=('title_sentiment_adjusted_source', 'mean'),
    avg_snippet_adjusted=('snippet_sentiment_adjusted_source', 'mean'),
    avg_body_adjusted=('body_sentiment_adjusted_source', 'mean')
```

```
).reset_index()

plot_business = bias_business.pivot_table(
    index='source_name',
    columns='company',
    values='avg_body_adjusted'
)

plot_business.plot(
    kind='bar',
    figsize=(12, 5)
)

plt.title('Newspaper-adjusted sentiment by newspaper and company\n(Health)')
plt.xlabel('Newspaper')
plt.ylabel('Adjusted average sentiment')
plt.xticks(rotation=45, ha='right')
plt.legend(title='Company')
plt.axhline(0, linestyle='--')
plt.tight_layout()
plt.show()

# %%
df_filtered = df_sentiment_final[df_sentiment_final['article_type'] == 'Noticias generales /
Portada / Secciones principales']

bias_business = df_filtered.groupby(
    ['source_name', 'company', 'country']
).agg(
    n_articles=('general_sentiment_adjusted_source', 'count'),
    avg_adjusted_sentiment=('general_sentiment_adjusted_source', 'mean'),
    avg_title_adjusted=('title_sentiment_adjusted_source', 'mean'),
    avg_snippet_adjusted=('snippet_sentiment_adjusted_source', 'mean'),
    avg_body_adjusted=('body_sentiment_adjusted_source', 'mean')
).reset_index()

plot_business = bias_business.pivot_table(
    index='source_name',
    columns='company',
    values='avg_body_adjusted'
)

plot_business.plot(
    kind='bar',
    figsize=(12, 5)
)

plt.title('Newspaper-adjusted sentiment by newspaper and company\n(Frontpage and main
sections)')
plt.xlabel('Newspaper')
plt.ylabel('Adjusted average sentiment')
plt.xticks(rotation=45, ha='right')
```

```
plt.legend(title='Company')
plt.axhline(0, linestyle='--')
plt.tight_layout()
plt.show()

# %% [markdown]
# Analyzing specific timings and events

# %%
# COVID 19

# %%
# Ensuring date format
df_sentiment_final['publication_date'] = pd.to_datetime(df_sentiment_final['publication_date'],
errors='coerce')

# Filter between 1st March 2020 and 1st March 2021
df_period = df_sentiment_final[
    (df_sentiment_final['publication_date'] >= '2020-03-01') &
    (df_sentiment_final['publication_date'] <= '2021-03-01')
].copy()

period_bias_adjusted = df_period.groupby(
    ['company', 'country']
).agg(
    n_articles=('general_sentiment_adjusted_source', 'count'),
    avg_adjusted_sentiment=('general_sentiment_adjusted_source', 'mean'),
    avg_title_adjusted=('title_sentiment_adjusted_source', 'mean'),
    avg_snippet_adjusted=('snippet_sentiment_adjusted_source', 'mean'),
    avg_body_adjusted=('body_sentiment_adjusted_source', 'mean')
).reset_index()

period_bias_adjusted

# %%
import matplotlib.pyplot as plt
import pandas as pd

# Pivot for plotting
plot_period_adjusted = period_bias_adjusted.pivot_table(
    index='country',
    columns='company',
    values='avg_body_adjusted'
)

# Plot
plot_period_adjusted.plot(
    kind='bar',
    figsize=(8, 4)
)

plt.title('Adjusted sentiment by country and company\nMarch 2020 - March 2021')
```

```
plt.xlabel('Country')
plt.ylabel('Average adjusted body sentiment')
plt.xticks(rotation=0)
plt.axhline(0, linestyle='--')
plt.legend(title='Company')
plt.tight_layout()
plt.show()

# %%
#AZ release

# Ensuring date format
df_sentiment_final['publication_date'] = pd.to_datetime(df_sentiment_final['publication_date'],
errors='coerce')

# Filter between 1st June 2020 and 1st March 2021
df_period = df_sentiment_final[
    (df_sentiment_final['publication_date'] >= '2020-12-15') &
    (df_sentiment_final['publication_date'] <= '2021-01-15')
].copy()

period_bias_adjusted = df_period.groupby(
    ['company', 'country']
).agg(
    n_articles=('general_sentiment_adjusted_source', 'count'),
    avg_adjusted_sentiment=('general_sentiment_adjusted_source', 'mean'),
    avg_title_adjusted=('title_sentiment_adjusted_source', 'mean'),
    avg_snippet_adjusted=('snippet_sentiment_adjusted_source', 'mean'),
    avg_body_adjusted=('body_sentiment_adjusted_source', 'mean')
).reset_index()

period_bias_adjusted

# %%
# Pivot for plotting
plot_period_adjusted = period_bias_adjusted.pivot_table(
    index='country',
    columns='company',
    values='avg_body_adjusted'
)

# Plot
plot_period_adjusted.plot(
    kind='bar',
    figsize=(8, 4)
)

plt.title('Adjusted sentiment by country and company\nEnd of December 2021')
plt.xlabel('Country')
plt.ylabel('Average adjusted sentiment')
plt.xticks(rotation=0)
plt.axhline(0, linestyle='--')
```

```
plt.legend(title='Company')
plt.tight_layout()
plt.show()

# %%
#AZ suspension

# Ensuring date format
df_sentiment_final['publication_date'] = pd.to_datetime(df_sentiment_final['publication_date'],
errors='coerce')

# Filter between 1st June 2020 and 1st March 2021
df_period = df_sentiment_final[
    (df_sentiment_final['publication_date'] >= '2021-03-01') &
    (df_sentiment_final['publication_date'] <= '2021-03-20')
].copy()

period_bias_adjusted = df_period.groupby(
    ['company', 'country']
).agg(
    n_articles=('general_sentiment_adjusted_source', 'count'),
    avg_adjusted_sentiment=('general_sentiment_adjusted_source', 'mean'),
    avg_title_adjusted=('title_sentiment_adjusted_source', 'mean'),
    avg_snippet_adjusted=('snippet_sentiment_adjusted_source', 'mean'),
    avg_body_adjusted=('body_sentiment_adjusted_source', 'mean')
).reset_index()

period_bias_adjusted

# %%
# Pivot for plotting
plot_period_adjusted = period_bias_adjusted.pivot_table(
    index='country',
    columns='company',
    values='avg_body_adjusted'
)

# Plot
plot_period_adjusted.plot(
    kind='bar',
    figsize=(8, 4)
)

plt.title('Adjusted sentiment by country and company\nMarch 2021')
plt.xlabel('Country')
plt.ylabel('Average adjusted sentiment')
plt.xticks(rotation=0)
plt.axhline(0, linestyle='--')
plt.legend(title='Company')
plt.tight_layout()
plt.show()
```

```
# %%
#J&J release

# Ensuring date format
df_sentiment_final['publication_date'] = pd.to_datetime(df_sentiment_final['publication_date'],
errors='coerce')

# Filter between 1st June 2020 and 1st March 2021
df_period = df_sentiment_final[
    (df_sentiment_final['publication_date'] >= '2020-02-20') &
    (df_sentiment_final['publication_date'] <= '2021-03-05')
].copy()

period_bias_adjusted = df_period.groupby(
    ['company', 'country']
).agg(
    n_articles=('general_sentiment_adjusted_source', 'count'),
    avg_adjusted_sentiment=('general_sentiment_adjusted_source', 'mean'),
    avg_title_adjusted=('title_sentiment_adjusted_source', 'mean'),
    avg_snippet_adjusted=('snippet_sentiment_adjusted_source', 'mean'),
    avg_body_adjusted=('body_sentiment_adjusted_source', 'mean')
).reset_index()

period_bias_adjusted

# %%
# Pivot for plotting
plot_period_adjusted = period_bias_adjusted.pivot_table(
    index='country',
    columns='company',
    values='avg_body_adjusted'
)

# Plot
plot_period_adjusted.plot(
    kind='bar',
    figsize=(8, 4)
)

plt.title('Adjusted sentiment by country and company\nFebruary 2021')
plt.xlabel('Country')
plt.ylabel('Average adjusted sentiment')
plt.xticks(rotation=0)
plt.axhline(0, linestyle='--')
plt.legend(title='Company')
plt.tight_layout()
plt.show()

# %%
#J&J suspension

# Ensuring date format
```

```
df_sentiment_final['publication_date'] = pd.to_datetime(df_sentiment_final['publication_date'],
errors='coerce')

# Filter between 1st June 2020 and 1st March 2021
df_period = df_sentiment_final[
    (df_sentiment_final['publication_date'] >= '2021-04-01') &
    (df_sentiment_final['publication_date'] <= '2021-04-20')
].copy()

period_bias_adjusted = df_period.groupby(
    ['company', 'country']
).agg(
    n_articles=('general_sentiment_adjusted_source', 'count'),
    avg_adjusted_sentiment=('general_sentiment_adjusted_source', 'mean'),
    avg_title_adjusted=('title_sentiment_adjusted_source', 'mean'),
    avg_snippet_adjusted=('snippet_sentiment_adjusted_source', 'mean'),
    avg_body_adjusted=('body_sentiment_adjusted_source', 'mean')
).reset_index()

period_bias_adjusted

# %%
# Pivot for plotting
plot_period_adjusted = period_bias_adjusted.pivot_table(
    index='country',
    columns='company',
    values='avg_body_adjusted'
)

# Plot
plot_period_adjusted.plot(
    kind='bar',
    figsize=(8, 4)
)

plt.title('Adjusted sentiment by country and company\nApril 2021')
plt.xlabel('Country')
plt.ylabel('Average adjusted sentiment')
plt.xticks(rotation=0)
plt.axhline(0, linestyle='--')
plt.legend(title='Company')
plt.tight_layout()
plt.show()

# %%
# Pivot for plotting
plot_period_adjusted = period_bias_adjusted.pivot_table(
    index='country',
    columns='company',
    values='avg_title_adjusted'
)
```

```
# Plot
plot_period_adjusted.plot(
    kind='bar',
    figsize=(8, 4)
)

plt.title('Adjusted title sentiment by country and company\nApril 2021')
plt.xlabel('Country')
plt.ylabel('Average adjusted sentiment')
plt.xticks(rotation=0)
plt.axhline(0, linestyle='--')
plt.legend(title='Company')
plt.tight_layout()
plt.show()

# %%
# OTHER EVENTS

# %%
#J&J 01/10/2010 NEGATIVE

# Ensuring date format
df_sentiment_final['publication_date'] = pd.to_datetime(df_sentiment_final['publication_date'],
errors='coerce')

# Filter between 1st June 2020 and 1st March 2021
df_period = df_sentiment_final[
    (df_sentiment_final['publication_date'] >= '2010-09-30') &
    (df_sentiment_final['publication_date'] <= '2010-10-15')
].copy()

period_bias_adjusted = df_period.groupby(
    ['company', 'country']
).agg(
    n_articles=('general_sentiment_adjusted_source', 'count'),
    avg_adjusted_sentiment=('general_sentiment_adjusted_source', 'mean'),
    avg_title_adjusted=('title_sentiment_adjusted_source', 'mean'),
    avg_snippet_adjusted=('snippet_sentiment_adjusted_source', 'mean'),
    avg_body_adjusted=('body_sentiment_adjusted_source', 'mean')
).reset_index()

period_bias_adjusted

# %%
# Pivot for plotting
plot_period_adjusted = period_bias_adjusted.pivot_table(
    index='country',
    columns='company',
    values='avg_adjusted_sentiment'
)

# Plot
```

```
plot_period_adjusted.plot(
    kind='bar',
    figsize=(8, 4)
)

plt.title('Adjusted sentiment by country and company\nOctober 2010')
plt.xlabel('Country')
plt.ylabel('Average adjusted sentiment')
plt.xticks(rotation=0)
plt.axhline(0, linestyle='--')
plt.legend(title='Company')
plt.tight_layout()
plt.show()

# %%
#AZ 19/05/2014 POSITIVE

# Ensuring date format
df_sentiment_final['publication_date'] = pd.to_datetime(df_sentiment_final['publication_date'],
errors='coerce')

# Filter between 1st June 2020 and 1st March 2021
df_period = df_sentiment_final[
    (df_sentiment_final['publication_date'] >= '2014-05-19') &
    (df_sentiment_final['publication_date'] <= '2014-05-25')
].copy()

period_bias_non_adjusted = df_period.groupby(
    ['company', 'country']
).agg(
    n_articles=('general_sentiment', 'count'),
    avg_sentiment=('general_sentiment', 'mean'),
    avg_title=('title_sentiment', 'mean'),
    avg_snippet=('snippet_sentiment', 'mean'),
    avg_body=('body_sentiment', 'mean')
).reset_index()

period_bias_non_adjusted

# %%
# Pivot for plotting
plot_period_adjusted = period_bias_adjusted.pivot_table(
    index='country',
    columns='company',
    values='avg_adjusted_sentiment'
)

# Plot
plot_period_adjusted.plot(
    kind='bar',
    figsize=(8, 4)
)
```

```
plt.title('Adjusted sentiment by country and company\nOctober 2010')
plt.xlabel('Country')
plt.ylabel('Average adjusted sentiment')
plt.xticks(rotation=0)
plt.axhline(0, linestyle='--')
plt.legend(title='Company')
plt.tight_layout()
plt.show()

# %%
# Robustness check (same process for articles where only 1 company is mentioned)
```

Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, CLAUDIA MONTERO MELÉNDEZ, estudiante del DOBLE GRADO DE ADMINISTRACIÓN DE EMPRESAS Y BUSINESS ANALYTICS (E2+ ANALYTICS) de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado " Does the Press Favour Its Own? Home-Country Bias in Pharmaceutical Media Coverage", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Referencias:** Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
2. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
3. **Interpretador de código:** Para realizar análisis de datos preliminares.
4. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
5. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
6. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
7. **Traductor:** Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente

de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 15/06/2026

Firma: Claudia Montero Meléndez