Media Sentiment and Government Healthcare Investment:
A Time-Series Analysis of GDELT News Data

By

Lige Zhang

Thesis submitted in partial fulfillment of

Bachelor of Business Administration in Enterprise Systems with Honours

St. Francis Xavier University

Antigonish, Nova Scotia

Thesis Advisor: Dr. David Mattie, PhD

Date: April, 2026

Acknowledgements

I would like to express my sincere gratitude to my thesis advisor, Dr. David Mattie, for his guidance, patience, and encouragement throughout the development of this research. His insightful feedback and continuous support were invaluable during the research and writing process.

I would also like to thank the faculty members at St. Francis Xavier University, particularly those in the Gerald Schwartz School of Business, for their support and guidance throughout my studies. Their teaching and encouragement greatly contributed to the completion of this thesis.

Finally, I am grateful to St. Francis Xavier University for providing an academic environment that made this research possible.

Abstract

Media play an important role in public decision-making, but empirical research on whether media tone influences actual government investment in public healthcare remains limited. This study uses news data from the Global Database of Events, Language, and Tone (GDELT) and data on current health expenditure expressed as a percentage of GDP across countries to examine the relationship between media tone and government public healthcare investment across a 20 years time-series. Using linear mixed-effects models with lagged media variables and controls for economic conditions and major public health events, we analyze convergent evidence from two independent GDELT datasets spanning 2004–2024. We find no significant media effect on healthcare expenditure. Instead, a long-term upward trend in spending and COVID-19 case burden emerge as the primary predictors, suggesting that healthcare investment decisions are driven by policy trajectories and direct public health threats rather than media framing.

List of Abbreviations Used

AIC – Akaike Information Criterion

BIC – Bayesian Information Criterion

COVID-19 – Coronavirus Disease 2019

GDP – Gross Domestic Product

GDELT – Global Database of Events, Language, and Tone

SD – Standard Deviation

Table of Contents

Table of Figures

List of Tables

Introduction

The increase in partisan media has led to a willingness for politicians to echo the messages found in these news sources (Prior, 2013) resulting in the potential for our politicians to develop echo chambers in which they draw insights and form opinions (Iyengar & Hahn, 2009). An increasing lack of balanced journalism has the potential to influence the landscape of government investment in important policy decisions that affect all citizens.

This research identifies relationships between the tone and sentiment of healthcare-related media coverage and government investment in healthcare programs over a twenty-year period. We hypothesize that negative healthcare coverage in partisan media sources precedes increases in healthcare investment.

Using mainstream media data and healthcare investment data over time, we quantify a potential relationship between the content of these media sources and healthcare investment. We also assess the effect size of media as it relates to funding decisions.

Government healthcare investment impacts people in different countries. It influences the quality of medical services, social determinants of health, healthcare burden, and individual lifestyles. A healthy population improves the development of the economy and society in Canada (Martin et al., 2018). Over the past two decades, healthcare spending has increased steadily, but the share of healthcare spending in GDP has fluctuated significantly. This elicits series of questions regarding whether media shapes healthcare policy outcomes, whether there is a temporal component in a healthcare policy processes, and whether sustained media attention ("media buzz") produces measurable effects over time.

With an evolving media landscape, and evidence that news coverage shows a significant impact on public decision-making in other countries (Okechukwu, 2024), it is important to understand government's relationship with media as there is a limited amount of research about the connection between media coverage and government healthcare investment in Canada.

The results quantify the associations between media coverage and healthcare investment decisions and identify temporal patterns in how politicians respond to media pressure. They expand our understanding of the interaction mechanisms between media coverage and policy, contribute insights into media strategy and policy-making processes in the healthcare sector, and provide insights into healthcare investment policy under the lens of big data.

The analysis also considers other potential influencing factors. These include the level of healthcare resources, such as hospital bed capacity, and major public health events, such as the COVID-19 pandemic. These factors may provide a more complete understanding of the interactions among media coverage, policy decisions, and healthcare investment.

Literature Review

This literature review examines existing research on media sentiment across different contexts, including studies on the role of media sentiment in shaping public healthcare policy, research focusing on media effects in healthcare contexts without a direct policy focus, evidence linking media sentiment to investment decision-making, and studies that leverage content analysis methods to capture trends within mainstream media sources. It highlights key findings and gaps that motivate this study.

Media Sentiment and Public Healthcare Policy

Existing research shows that media tone and sentiment significantly influence public perceptions of healthcare policy and reform through framing and emotional cues. Daw et al. (2014) use qualitative discourse analysis of media coverage from 1990 to 2010 to explore how various actors framed prescription drug financing and reform proposals. They find that media coverage was largely shaped by politicians and stakeholders, focused on costs, and framed pharmacare as competing with existing programs. This may reduce public support and make the reform harder to advance. Soroka, Maioni, and Martin (2013) use both micro- and macro-level analyses to examine how media framing and other factors shape the differences between individual and collective, retrospective and prospective, and evaluations within Canada's healthcare system. They reveal that media content may be a driver of the widening gap between retrospective and prospective attitudes, positioning mass media as an important factor influencing the healthcare policy process. Fredheim (2021) uses qualitative media text analysis to examine how media framing of a contentious health policy issue was shaped by various actors before and after the policy shift, and suggest that, media effects on policy processes may be dynamic and stage-dependent under specific policy and institutional contexts. Miller et al. (2012) identify and analyze 1,704 nursing home–related articles from LexisNexis, using content analysis and descriptive statistics to assess tone, themes, prominence, and key actors over time, and find that media coverage dominated by negative narratives poses challenges to healthcare reform in the United States, whereas it also strengthens public scrutiny and encourages regulatory reform. Apriliyanti, Utomo, and Purwanto (2022) examine the relationship between Indonesia's government health interventions and media tone based on a content analysis of online news and interviews with political advisors, as well as public health experts who were part of the core policymaking group. They find that during the COVID-19 pandemic, media may symbolically strengthen science–policy interactions, but government decisions mainly driven by economic considerations, resulting in limited policy responses and reduced public trust.

Media Sentiment in Healthcare Without a Policy Focus

Existing studies have examined media sentiment in healthcare contexts that are not directly related to public healthcare policy. Walker et al. (2021) use content and frame analysis to code negative media representations to examine how negative media representations of the National Health Service (NHS) fluctuate and intensify over time and suggest that negative news repetition builds perceptions of service inadequacy. Hart, Chinn, and Soroka (2020) use computer-assisted content analytic approaches to analyze early U.S. news coverage of COVID-19 and find that politicized and polarized reporting may have deepened partisan divides in public attitudes toward the pandemic. Carson, Martin, and Ratcliff (2020) apply a large-scale dataset combined with content analysis of media coverage and party press releases to show how negative campaign influence the election outcome. Niu et al. (2021) employ event study methodology to quantify the impact of Canadian policy announcements on Twitter sentiment toward COVID-19 NPIs (Non-Pharmaceutical Interventions). Lawson (2022) explores how journalists used open government data to report on the NHS winter crisis by Content analysis, thematic analysis, and journalist interviews.

Media Sentiment and Investment Decisions

Existing studies show that media sentiment has been linked to a range of investment decisions. Hsu et al. (2024) find that media sentiment can lead to overinvestment and influence investment efficiency by using the Google Cloud Natural Language API to analyse 106,484 released news in

China from 2007 to 2017. Liao, Wang, and Wu (2021) examine the link between media sentiment and firm acquisition decisions by using the CSS sentiment score of news from RavenPack, a leading global news database. Dang et al. (2019) use RavenPack News Analytics, to capture the extent of news coverage and the tone of the news and examine the impact of the media on firms' leverage adjustments. Baloria and Heese (2018) use quasi-natural variation in local Fox News availability to show how media slant alters firms' reaction in the U.S.

Content analysis of mainstream media

The Global Database of Events, Language, and Tone (GDELT), supported by Google Jigsaw, is a large-scale database that tracks global news coverage and media tone (GDELT Project, n.d.). It has been widely used to analyze patterns in mainstream media and their broader implications. Hu et al. (2023) build media inclination, represented by Goldstein score and attention indexes to link media tone to outward foreign direct investment (OFDI) behaviors at the firm level in China. Beasant (2025) uses media buzz data and compare it with energy investment figures to examine how government energy policies reflected by public opinion and media narratives or other influences. Bourgeois et al. (2019) Uses mainstream media events and mentions to build a source event interaction matrix and applies collaborative-filtering personalization to predict coverage and extract latent outlet preferences. Alipour et al. (2024) focus on how events are mentioned over time across different countries, to explain the imbalance in online news spreading. Korkmaz et al. (2016) develops a multi-source forecasting model using social media, news, and political-economic data to predict daily civil unrest events in six Latin American countries. Krieg et al. (2022) predict terrorist attacks in the United States by using a feature representation method and utilising localized news data to evaluate machine learning models. Consoli, Tiozzo Pezzoli and Tosetti (2022) deploy economic news information within a probabilistic forecasting framework with autoregressive recurrent networks (DeepAR) to forecast the Italian 10-year interest rate spread. Hu, Yuan and Liu (2025) construct a Geopolitical Relation Index, and a Comparative Relation Index to quantify shifts in Asia-Pacific countries' hedging strategies toward China or the US.

Important gaps exist despite the existing studies on media sentiment across different contexts. Few studies directly examine its impact on public healthcare policy outcomes, and there is limited evidence on whether media sentiment influences government healthcare expenditure over time. This study aims to address this gap by examining the relationship between media sentiment and public healthcare expenditure.

Data & Methods

Data Sources and Collection

This study adopts a quantitative approach using regression models with a time-series dataset to analyze GDELT data from 2005-2024. GDELT is a project which analyzes events, emotions, and other factors by tracking global news sources in 15-minute intervals and provides an open platform for data analysis (GDELT Project, n.d.). This study extracts healthcare-related news coverage from GDELT media sources and uses GDELT's tone metrics as sentiment measures, to establish regression models of media coverage volume and sentiment about healthcare.

During the data extraction phase, we identify notable differences in the structure and content of GDELT data before and after 2017. We employ two data collection approaches based on GDELT DOC 2.0 and GDELT Events. While GDELT DOC 2.0 offers advanced multilingual article-level search capabilities, its reliable coverage is limited to post-2017 data (GDELT Project, 2017).

While GDELT Events offers longer historical coverage, it differs from GDELT DOC 2.0 in many aspects, such as data structure and themes. Given the differences across the datasets, it is not appropriate to standardize and harmonize the data collected from these two periods. To ensure consistency, the analysis focuses on daily average media tone and average intensity from 2017 to 2024 across four countries, namely Canada, China, the United States, and the United Kingdom, and based on five healthcare-related themes: WB_625_HEALTH_ECONOMICS_AND_FINANCE, WB_1287_HEALTH_INSURANCE, WB_1291_PUBLIC_HEALTH_INSURANCE, WB_1293_UNIVERSAL_HEALTH_INSURANCE, and WB_631_HUMAN_RESOURCES_FOR_HEALTH. These themes were selected because they are healthcare-related categories defined by the World Bank. These themes mainly relate to health economics and finance, health insurance and the allocation of healthcare workforce resources. Overall, they reflect policy areas that are closely linked to healthcare financing and healthcare investment. As a result, they are directly relevant to the focus of this study on healthcare investment. These data cover news media on daily basis. Therefore, we calculate average values per country, theme, and year to obtain aggregated annual measures of media tone and volume intensity. We also extract media sentiment data from GDELT 1.0 for Canada between 2004 and 2023, using keyword-based theme reconstruction to approximate GDELT 2.0 theme categories. This dataset serves as an independent validation sample to examine whether media tone was associated with health expenditure during normal (pre-COVID) periods.

To examine government healthcare investment, this study further uses health expenditure data from the World Bank (World Bank, n.d.-a). Current health expenditure (% of GDP) is obtained from the World Bank, using the indicator SH.XPD.CHEX.GD.ZS. This indicator measures current health expenditure as a percentage of gross domestic product (GDP) for each country. For this World Bank healthcare expenditure data, Canada and the United Kingdom are covered from 2017 to 2023, while data for the United States and China are available from 2017 to 2022. In addition, Population, total and GDP (Current US$) from world bank are included as control variables to account for scale effects across countries and to prevent population-related structural and economic differences from confounding the estimated media effects (World Bank, n.d.-c, n.d.-b).

Several indicators related to public health conditions are also collected from Our World in Data, including COVID-19 cases and deaths, hospital beds per capita, and life expectancy (Our World in Data, n.d.-b, n.d.-a, n.d.-c). These are used to control for differences in the severity of the epidemic, the supply of medical resources, and the basic level of population health.

Data Preparation

After data collection, we conduct several data preparation steps prior to model estimation. We merge GDELT 2.0 media data with World Bank indicators, COVID-19 data, hospital beds data, and life expectancy data at the country–year level, yielding n = 130 observations across four countries and five health themes over the period 2017–2023, with 7 years of observations for Canada and 6 years for the other countries. We then construct lagged media variables, including one-year (t−1) and two-year (t−2) lags of average tone and average intensity at the country–theme level. Observations with missing values for the dependent variable (expenditure) are also removed, resulting in n = 110 analytic observations after applying the lag structure.

In order to facilitate feature comparison, all continuous predictors are standardized (mean = 0, SD = 1). In addition, Missing values, primarily in the COVID-19 variables and hospital bed capacity measures, are imputed using country-specific mean substitution implemented via group-wise calculation.

Analytical Framework - Linear Mixed-Effects Models

We employ linear mixed effects models to account for the hierarchical structure of the data, in which repeated observations are nested within countries, and to accommodate unbalanced data and missing values without requiring listwise deletion, thereby preserving statistical power. Initially, we estimated a baseline model without lagged independent variables as a starting point. However, it did not show significant effects. We subsequently analyzed lead-lag relationships, an important pattern in high-frequency data (Huth & Abergel, 2014), between media coverage patterns and government investment decisions. Healthcare systems and budget structures vary across countries and change minimally over time (Hanson et al., 2022). To account for the nested structure of the data, a random intercept is included for Country, allowing for baseline variation in the outcome across different nations. In this model, we defin the dependent variable as government healthcare expenditure and the main explanatory variables as average media tone and media intensity. The model also includes two control variables: population and year. These are used to adjust for country size and capture overall time trends. Moreover, one-year (t−1) and two-year (t−2) lagged values of media tone and intensity are established to consider possible time lags in how media influence policy. T−1 lags represent the one-year effects of media coverage on healthcare expenditure. The t−2 lags represent the effects in the two-year term (see Figure 7).

Building on this initial specification, we then employed linear mixed-effects models (lme4 package in R) with random intercepts for country and theme, constructing a series of nested specifications that progressively expanded from baseline media-only models to full models incorporating economic controls, pandemic indicators, and temporal trends (Bates et al., 2003). Accordingly, the model can be formally expressed as follows:

$$y_{ijt} = \beta_0 + \sum_{k=1}^{K} \beta_k X_{k,ij(t-\ell)} + u_j + v_i + \varepsilon_{ijt}, \qquad \ell \in \{1,2\}$$

Where:

y_{ijt}: government healthcare expenditure with country j, theme i, year t.

β_0: the fixed intercept.

$X_{k,ij(t-\ell)}$: explanatory variables (i.e. media tone and intensity).

u_j: country-level random intercept.

v_i: theme random intercept

ε_{ijt}: error term.

$\ell \in \{1,2\}$: one-year and two-year lags.

In Models 1–3, we fit baseline media-only models, testing tone, intensity, and their joint effects with random intercepts for country to evaluate whether media variables independently predict expenditure. In Model 4, we add theme as an additional random effect to examine whether health-related themes differentially predict expenditure. In Model 5, Year is included as a fixed effect to account for long-term temporal trends in healthcare spending. In Model 6, an interaction term between tone and country is included to examine whether the effect of media tone differs across countries. In Models 7–8, we introduce lagged economic and pandemic control variables, including GDP, population, and COVID-19 deaths and cases, to examine whether media effects survive after accounting for economic and pandemic-related factors. In Model 9, we include lagged media tone and lagged COVID-19 deaths and cases, while retaining a country-level random intercept to assess whether media effects persist after accounting for delayed pandemic influences. In Model 10, we incorporate current media variables (tone and intensity) alongside lagged GDP, controlling for country-level heterogeneity to evaluate the role of economic conditions. In Model 11, we specify a more comprehensive model including lagged media tone, lagged GDP and population, and lagged COVID-19 deaths and cases, with random intercepts for both country and theme. In Models 12 and 13, COVID-19 indicators are introduced in two forms: deaths and cases in the same year (CovidNow) and one-year lagged deaths and cases (CovidLag), to distinguish immediate from delayed pandemic effects on health expenditure. In Models 14–18, we examine combined media and COVID specifications, pairing lagged media variables with both current and lagged COVID indicators, with interaction terms introduced in selected models. In Models 19–21, we specify full temporal models combining lagged media variables, current COVID-19 indicators, and Year trends to capture short-term pandemic shocks and longer-term expenditure trajectories.

Table 1 presents the specifications of all models. It summarizes the inclusion of key explanatory and lagged variables, as well as control variables and random effects across different model specifications.

Table 1 Model Specifications

	Explanatory Variables							Lagged Variables							Random Effects	
Model	Tone	Intensity	GDP	Population	COVID Deaths	COVID Cases	Year	Lag Tone	Lag Intensity	Lag GDP	Lag Population	Lag COVID Deaths	Lag COVID Cases	Interaction	Country	Theme
M1	✓														✓	
M2		✓													✓	
M3	✓	✓													✓	
M4	✓	✓													✓	✓
M5	✓	✓					✓								✓	✓
M6	✓	✓												Tone × Country		✓
M7								✓		✓	✓	✓	✓			
M8								✓		✓	✓	✓	✓		✓	
M9								✓				✓	✓		✓	
M10	✓	✓								✓					✓	
M11								✓		✓	✓	✓	✓		✓	✓
M12					✓	✓									✓	✓
M13												✓	✓		✓	✓
M14	✓	✓			✓	✓									✓	✓
M15								✓	✓			✓	✓		✓	✓
M16					✓	✓		✓							✓	✓
M17								✓				✓	✓		✓	✓
M18					✓			✓						Lagged Tone × Covid Deaths	✓	✓
M19					✓	✓	✓	✓							✓	✓
M20					✓	✓	✓	✓	✓						✓	✓
M21					✓	✓	✓		✓						✓	✓

Model Selection

Model selection is conducted by comparing all specifications using Akaike Information Criterion (AIC), while Bayesian Information Criterion (BIC) and log-likelihood values are included for reference. Both AIC and BIC consider the number of parameters when comparing models, which helps avoid selecting overly complex specifications. AIC emphasizes goodness of fit, whereas BIC more strongly penalizes model complexity and typically favors simpler models. Models with lower AIC values are preferred, indicating a better balance between model fit and complexity.

Robustness Checks

As a robustness check, we replicate the analysis using GDELT 1.0 data for Canada (2004–2023; n = 18 after lag construction) to examine whether the estimated media sentiment effects persist during the pre-COVID period. We then compare the findings from GDELT 1.0 and GDELT 2.0 to assess the consistency of results across measurement systems and time periods.

Software

The data analysis process is carried out through R (V4.4.0; R Core Team 2025). It is a programming language and environment for statistical analysis and data visualization. R helps to maintain a consistent analysis workflow including data preprocessing, model construction and result visualization. In this environment, different software packages are used to support different stages of the data analysis. Among them, Data preprocessing and variable construction are conducted using packages of tidyverse, dplyr, tidyr, and reshape2 (Wickham, 2010, 2016; Wickham, François, et al., 2014; Wickham, Vaughan, et al., 2014). Linear mixed-effects models are estimated using the lme4 package and the nlme package (Bates et al., 2003; Pinheiro et al., 1999). Statistical tests for the mixed-effects models are conducted using the lmerTest package (Kuznetsova et al., 2013). Model checks are performed using the performance package (Lüdecke et al., 2019). Model results are summarized and visualized using the sjPlot package (Lüdecke, 2013). The effects of predictors were examined using the effects package (Fox et al., 2003). Figures are created using the ggplot2 package (Wickham et al., 2007). The psych package is used for correlation analysis with significance testing (Revelle, 2007). The writexl package is used to export analytical results to Excel files (Ooms, 2017).

Results

Descriptive Statistics and Exploratory Analysis

We collected 160 observations over four countries (Canada, the United States, the United Kingdom, and China), with five health-related media themes, and eight years from 2017 to 2024. However, the dependent variable, healthcare expenditure as a percentage of GDP, had missing values in 30 observations. Incomplete observations were removed, leaving 130 complete observations in the model. After introducing lagged variables by country and theme, the sample size decreased further. Lagged models require data from the previous year. For example, in Canada, under the theme "Economics Finance", the average tone in 2017 was −0.438. When a one-year lag was applied, this value (−0.438) became the 2018 lagged observation for the same series. Therefore, the first year in each country-theme series could not be included in the one-year lag model. As a result, the final sample size was reduced to 110 observations.

The correlation matrix (Figure 1) shows multiple statistically significant relationships among the healthcare expenditure and the independent variables. Expenditure is positively correlated with COVID-19 deaths ($r = 0.70$, $p < 0.001$) and COVID-19 cases ($r = 0.61$, $p < 0.001$), illustrating that pandemic severity is closely associated with increases in health spending. Expenditure has a positive correlation with media intensity (r = 0.53, p < 0.001). In contrast, expenditure is negatively correlated with average media tone ($r = -0.40$, $p < 0.001$), suggesting that more negative media sentiment is associated with higher health expenditure. Hospital beds per capita also has a strong negative correlation with expenditure ($r = -0.70$, $p < 0.001$), which appears to be contrary to intuition and does not match common expectations. Overall, based on the correlation matrix, we found that pandemic-related variables show the strongest associations with healthcare expenditure, while media tone and intensity exhibit more moderate relationships. These descriptive patterns provide initial guidance for subsequent multivariate modelling.

It is worth noting that the correlation matrix shows potential multicollinearity concerns. Specifically, Covid-19 deaths and Covid-19 cases are highly correlated ($r = 0.98$), indicating an almost perfect linear relationship. In addition, population and hospital beds per capita are strongly correlated ($r = 0.99$), suggesting these variables are essentially redundant. This phenomenon shows there is overlapping information among the related variables, and including them together in the same model may limit the generalizability of the model. As a result, we did not include hospital beds per capita in the subsequent models, since no new information is added by knowing hospital beds due to its correlation with population. Moreover, the identification of multicollinearity provided further support for excluding Covid-19 deaths in the final model selection process to achieve a more parsimonious specification.

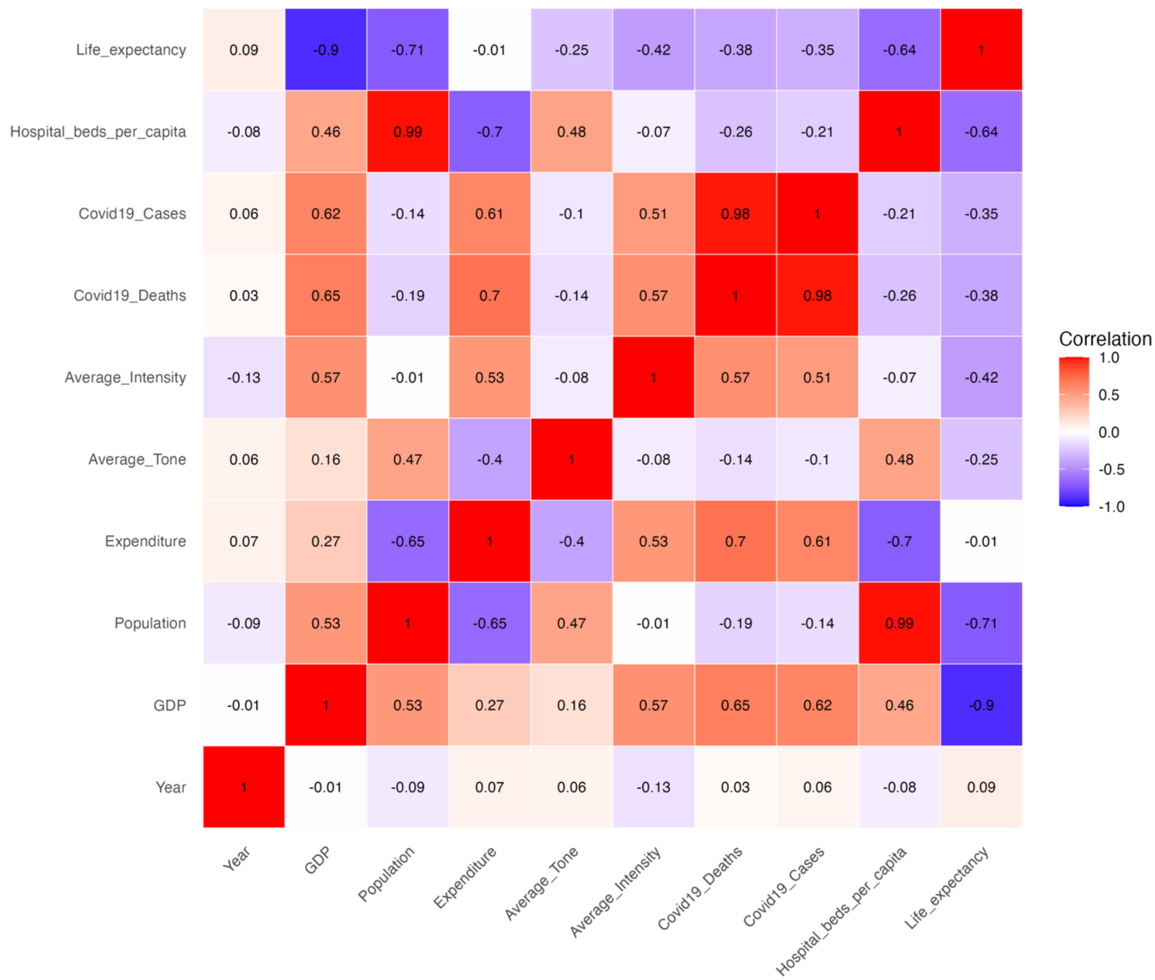

Figure 1 Correlation Heatmap of Healthcare Expenditure and Independent Variables

The exploratory plots provide additional insight into temporal trends and cross-national variation in healthcare expenditure, as well as differences in media tone and intensity across countries and themes.

Figure 2 illustrates the trends in healthcare expenditure as a percentage of GDP across the four countries from 2017 to 2023. Overall, healthcare expenditure remained relatively stable across countries from 2017 to 2019, increased sharply in 2020, and then declined after 2021. There are also some differences between countries: The United States consistently remained at the highest level, while China stayed at a relatively lower level. Canada and the United Kingdom were in the middle range and showed similar trends over time. The general spike in 2020 likely reflects the impact of the COVID-19 pandemic.

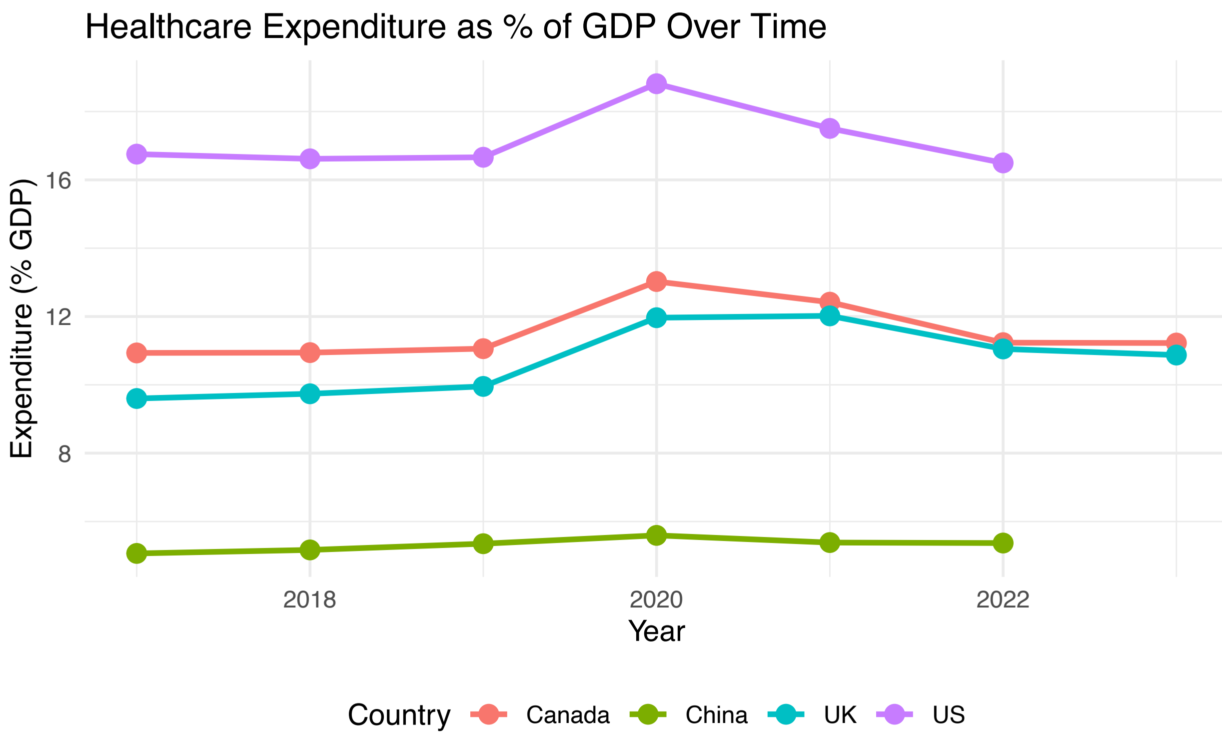

Figure 2 Healthcare Expenditure as % of GDP Over Time

Figure 3 shows the relationship between healthcare expenditure as a share of GDP and average media tone. Overall, there is no clear linear relationship between healthcare expenditure and media tone, and the fitted lines for each country show only a very slight downward trend. It is also noted that the data points for the United States, Canada, and the United Kingdom are mainly concentrated in the negative range (Average Tone < 0), while observations for China are mostly located in the positive range (Average Tone > 0). Compared with other countries, China has more health-related news with positive media sentiment. These differences indicate potential heterogeneity at the country level, the warranting consideration of country-specific intercepts in later model specifications.

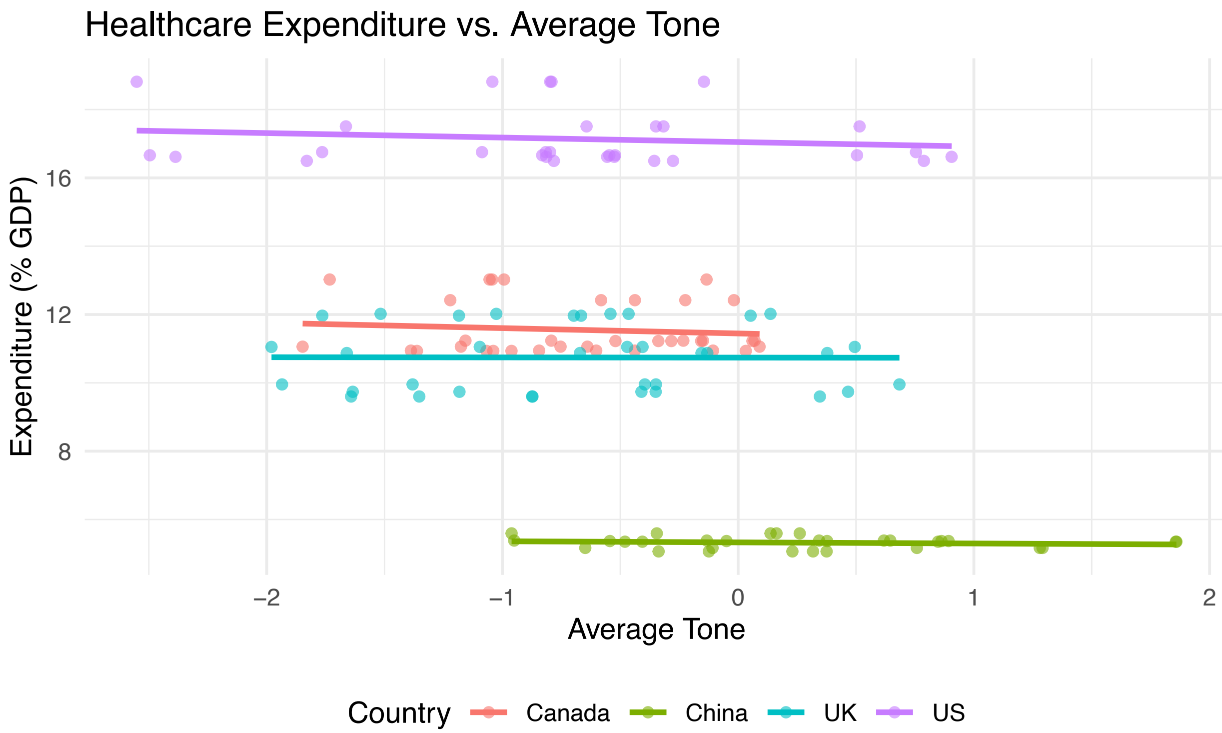

Figure 3 Healthcare Expenditure vs. Average Tone

Figure 4 shows the relationship between healthcare expenditure as a share of GDP and average media intensity. For the United States, healthcare expenditure remains relatively flat regardless of media intensity. In the United Kingdom, there is a slight increase in expenditure as media intensity rises, but this effect is not statistically significant. Overall, healthcare expenditure levels are similar for the United Kingdom and Canada, while the United States has higher expenditure and China has lower expenditure. Media coverage appears relatively higher in the United States, but healthcare expenditure as a share of GDP remains relatively stable.

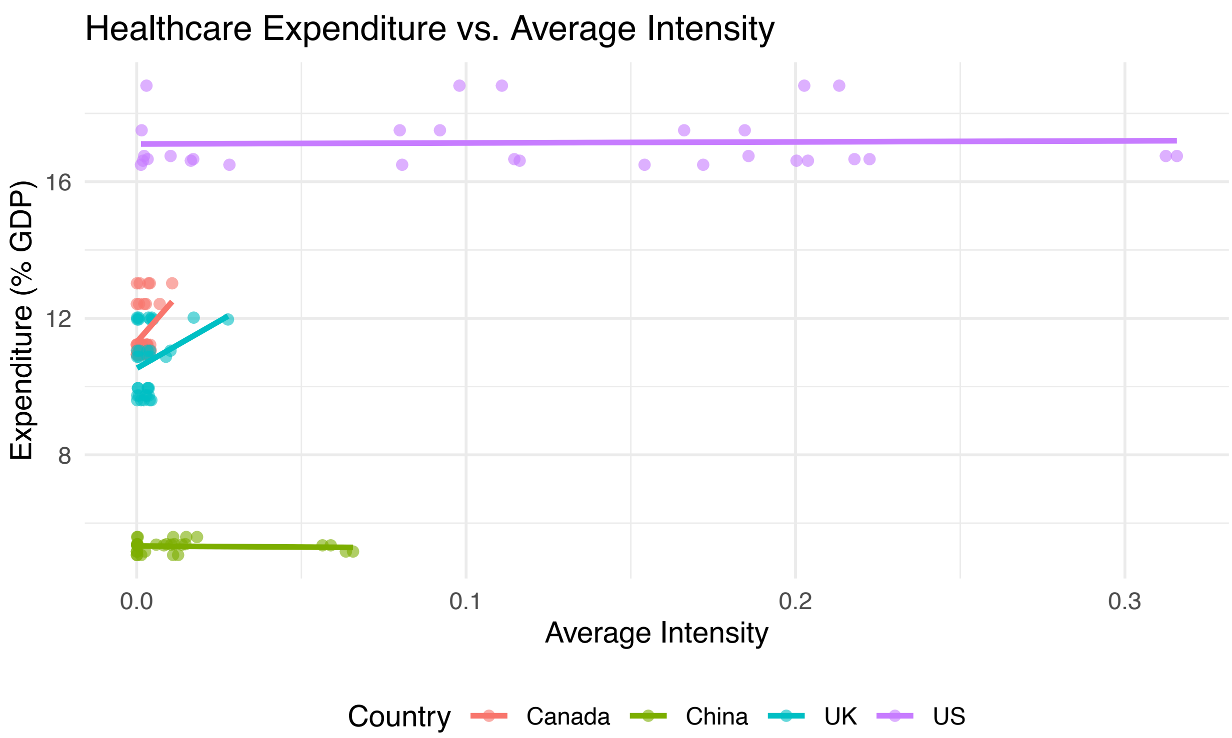

Figure 4 Healthcare Expenditure vs. Average Intensity

Figure 5 compares the relationship between healthcare expenditure as a share of GDP and average media tone across different health-related themes. Under the themes of "Economics Finance", "Human Resources", "Insurance", and "Public Insurance", healthcare expenditure and media tone generally showcase a modest negative association. By contrast, the fitted line for the theme of "Universal Insurance" is relatively flat. At the same time, data points within each theme are widely dispersed, suggesting substantial variation across countries and years and that the observed relationships are not uniform. Private Insurance and healthcare economics within media sources exhibiting a largely negative tone in Canada, UK, and US.

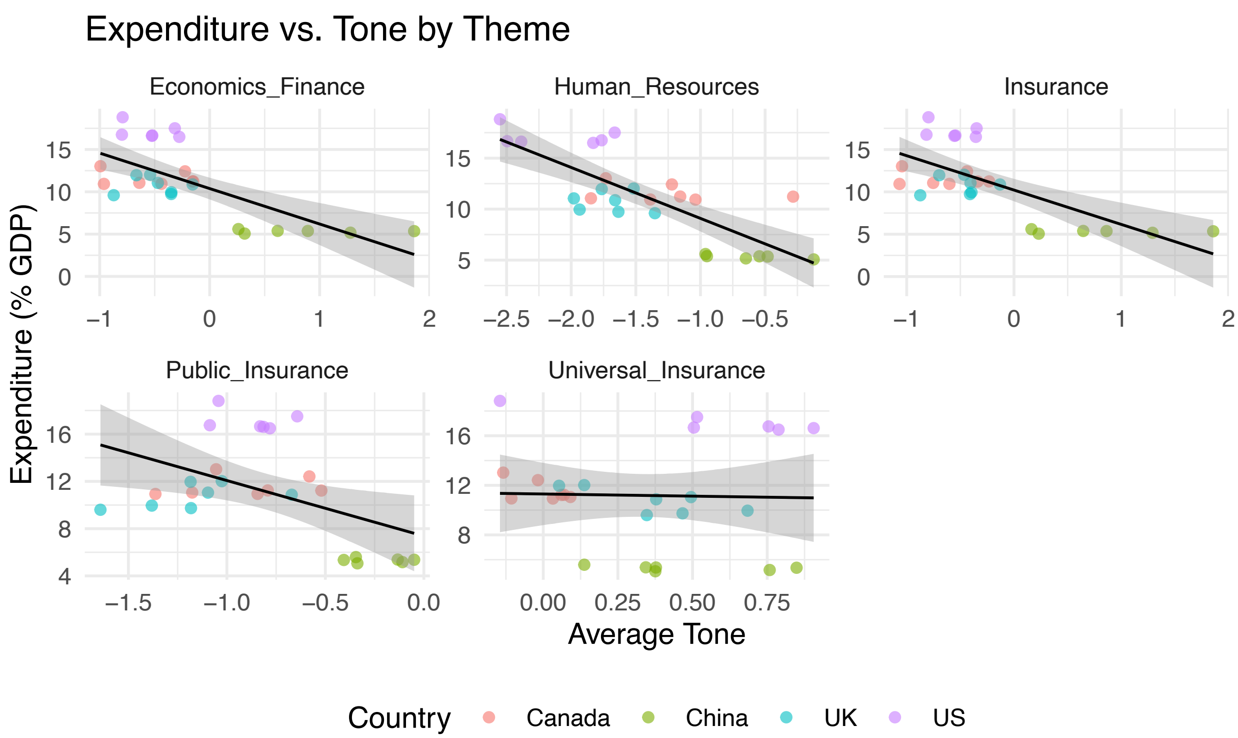

Figure 5 Expenditure vs. Tone by Theme

Baseline Media Sentiment Models (M1–M6)

In the baseline stage, we examined the fixed effects of the media variables in the current year. From Model 1 to Model 3, random intercepts were specified for country to account for clustering at the country level. Model 1 included only average tone, with a coefficient of β = −0.027 (p = 0.183). Model 2 included only Average Intensity, with a coefficient of β = 0.021 (p = 0.425). Model 3 included both average tone and Average Intensity. In this joint model, the coefficient for average tone was β = −0.028 (p = 0.170), and the coefficient for Average Intensity was β = 0.022 (p = 0.389). According to all three models, neither media variable was statistically significant, indicating an insignificant association with Expenditure. In terms of model fit, the AIC values ranged from −24.93 to −23.67, with only small differences. This indicates that including media tone and intensity did not meaningfully improve model fit.

Based on Models 1–3, we added an additional random intercept, Theme_Short, in Model 4 to examine whether systematic differences exist across healthcare themes. However, the estimated variance for Theme was zero, suggesting that under the current data structure, there were no meaningful differences in expenditure across themes.

In Model 5, Year was added as a fixed effect based on the previous model to control for time trends. The results show that Year was statistically significant ($\beta = 0.048$, $p = 0.006$), which is the strongest predictor in the model. In addition, after adding the year variation, the AIC decreased from −21.67 (Model 4) to −27.13. This shows a clear model improvement. We also compared Model 5 and Model 4 by using the likelihood ratio test. The result ($\chi^2 = 7.46$, $df = 1$, $p = 0.006$) suggests that the model is significantly improved when including the time variable.

In Model 6, we further included the interaction term between Average Tone and Country (Tone × Country) to examine whether the effect of media tone on healthcare expenditure is different at the country level. The results show that all Country coefficients were statistically significant (all $p < 0.001$), indicating clear differences in baseline expenditure levels across countries. However, none of the Tone × Country interaction terms were statistically significant (all $p > 0.50$). This suggests that the association between media tone and expenditure does not vary across countries.

Lagged and COVID-Controlled Models (M7–M18)

In the baseline models, we did not find a significant association between media variables and healthcare expenditure. However, previous literature suggests that media influence on policy decisions may exist with time lags rather than occurring immediately (McCombs & Shaw, 1972; S. N. Soroka, 2002; Walgrave & Van Aelst, 2006). Therefore, we subsequently introduced one-year lagged media variables to examine whether media tone and intensity have effects on government healthcare expenditure through a delayed mechanism. Meanwhile, the COVID-19 pandemic during 2020–2022 may represent a potential confound. During this period, healthcare expenditure had a substantial fluctuation (see Figure 2), and media tone and intensity were also likely influenced by the development of the pandemic. To avoid the biased estimation caused by COVID-related factors, we include COVID-19 Cases and COVID-19 Deaths as the direct impact of the pandemic on healthcare expenditure in the subsequent models. For example, in Models 12 and 16, we included current COVID-19 deaths and COVID-19 cases. The results show a improvement in model fit, with the AIC decreasing to −50.94 (Model 12) compared to the baseline models (AIC around −24). This indicates that COVID-related variables significantly improved explanatory power. By contrast, in Models 13 and 17, we introduced the one-year lagged COVID variables. However, model fit did not show further improvement. This suggests that the effect of the pandemic on healthcare expenditure is more likely to be immediate rather than delayed.

Across the remaining model specifications, only in Model 10, which did not include COVID controls, the coefficients of Average Tone approached marginal significance ($p < 0.18$), though it still did not reach conventional significance levels. Once COVID-related variables were introduced, all media coefficients were non-significant (all media $p > 0.18$). This pattern held regardless of whether media variables were entered in their unlagged form or with one-year lags.

Best-Fitting Model: M19

Among all 21 model specifications (see Table 2), Model 19 demonstrated the best model fit (AIC = −70.40; BIC = −48.79), substantially outperforming the previous models.

Table 2 Model Comparison

Model	AIC	BIC	LogLik
M1: Tone	−24.93	−14.13	16.46
M2: Intensity	−23.78	−12.98	15.89
M3: Tone+Intensity	−23.67	−10.17	16.84
M4: +Theme RE	−21.67	−5.47	16.84
M5: +Year	−27.13	−8.23	20.57
M6: Tone×Country	−43.06	−13.35	32.53
M7: LinearExtra	−6.28	12.62	10.14
M8: ExtraxCountry	−22.03	−0.43	19.02
M9: CovidxCountry	−22.74	−6.53	17.37
M10: Tone+Int+GDP	−21.80	−5.60	16.90
M11: LagFull+Theme	−20.03	4.27	19.02
M12: CovidNow	−50.94	−34.74	31.47
M13: CovidLag	−22.50	−6.30	17.25
M14: MediaLag+CovidNow	−47.13	−25.52	31.56
M15: MediaLag+CovidLag	−18.74	2.86	17.37
M16: ToneLag+CovidNow	−49.07	−30.17	31.53
M17: ToneLag+CovidLag	−20.74	−1.83	17.37
M18: ToneLag×CovidNow	−47.44	−28.54	30.72
M19: ToneLag+CovidNow+Year	−70.40	−48.79	43.20
M20: ToneLag+IntLag+CovidNow+Year	−68.40	−44.09	43.20
M21: IntensityLag+CovidNow+Year	−70.16	−48.55	43.08

In Model 20 and Model 21, we added media variables with a 1-year lag. Model 20 added one-year lagged media tone and intensity to Model 19. However, neither Average Tone Lag1 (p = 0.623) nor Average Intensity Lag1 (p = 0.990) reached statistical significance. Model 21 included only Average Intensity Lag1 (p = 0.980), which was not significant as well. Compared with

Model 19, these two models have nearly identical fit (AIC = −68.40 and −70.16, respectively). Neither specification provided meaningful improvement.

Based on the result, we finally selected Model 19 as the best-fitting model. This model includes Average Tone Lag1, Covid19 Deaths, Covid19 Cases, and Year as fixed effects, with two random intercepts, Country and Theme_Short. In this model, Year is the strongest predictor ($\beta = 0.073$, $p < 0.001$). It means that as the year goes on, health expenditure increases. Covid19 Cases is also statistically significant ($\beta = -0.195$, $p = 0.043$). In contrast, Covid19 Deaths is not statistically significant ($\beta = 0.022$, $p = 0.857$). One possible reason is the very strong correlation between Covid19 Deaths and Covid19 Cases ($r = 0.98$), making it difficult for deaths to explain additional variation once cases are included. The lagged media tone variable, Average Tone Lag1, is also not significant ($\beta = -0.008$, $p = 0.623$), suggesting that media tone does not have a clear effect on healthcare expenditure in the model.

To further examine the contribution of Year, we compared Model 16 and Model 19 by using a likelihood ratio test. Compared with Model 19, Model 16 was identical in specification except that it excludes the Year variable. The test results showed that the model fit was improved when adding the Year variable ($\chi^2 = 23.33$, $df = 1$, $p < 0.001$). The AIC also decreased from −49.07 to −70.40, indicating a substantial improvement. This finding indicates that Year plays an important role in explaining changes in healthcare expenditure.

From the random effects results, the standard deviation of the Country random intercept is 1.14, indicating that the baseline healthcare expenditure is different across countries. However, the variance for Theme remained zero. This suggests that there are no baseline differences in expenditure across healthcare themes. The residual standard deviation is 0.14. This means that after accounting for the fixed effects and country-level differences, the remaining unexplained variation is relatively small.

Best-Performing Model: M22

In Model 19, a singular fit was observed. In this model, the random intercept variance for Theme Short was zero, which means this random effect did not provide an additional contribution. In addition, the two COVID-related variables were nearly collinear ($r = 0.98$). To improve model stability and parsimony, we removed Covid19 Deaths and the Theme Short random intercept, and constructed a parsimonious Model 22. Among all model specifications, Model 22 demonstrated the best model fit (AIC = −74.36; BIC = −58.2), substantially outperforming the previous models, including Model 19 (AIC = −70.40). In this model, Year is still the strongest predictor ($\beta = 0.073$, $p < 0.001$). Covid19 Cases is also highly statistically significant ($\beta = -0.178$, $p < .001$). In contrast, the lagged media tone variable, Average Tone Lag1, is not statistically significant ($\beta = -0.008$, $p = 0.613$) (see Table 3).

Table 3 Fixed Effects Estimates for Model 22

Variable	Estimate	Std. Error	df	*t*-value	Pr(>|t|)
Intercept	0.008	0.573	3.99	0.014	9.90×10^{-1}
Average Tone Lag1	−0.008	0.016	106.04	−0.508	6.13×10^{-1}
Covid19 Cases	−0.178	0.023	106.22	−7.864	3.28×10^{-12}
Year	0.073	0.014	106.01	5.109	1.44×10^{-6}

Diagnostic plots for Model 22 (Figure 6) illustrates that the model performs reasonably well. The first plot (Residuals vs Fitted) shows how the model's prediction errors change across different predicted values. In this plot, most points are scattered around the zero line without forming a clear curve or pattern. This means that while the model makes smaller or larger errors depending on the predicted values, model performs reasonably well across the range of fitted values. The Normal Q–Q plot compares the distribution of the residuals with a theoretical normal distribution. In this figure, most of the points lie close to the reference line, suggesting that the residuals are approximately normally distributed. The Scale-Location plot examines whether the size of the residuals changes across different fitted values. In this figure, the points appear relatively evenly spread, suggesting that the size of the errors remains fairly consistent across the range of predicted values. The Residuals vs Leverage plot examines whether any observations have an unusually large influence on the model. From the figure, there are no clearly extreme or outstanding points. The observations with higher leverage values tend to have relatively small residuals. This indicates that no single observation exerts disproportionate influence on the model results.

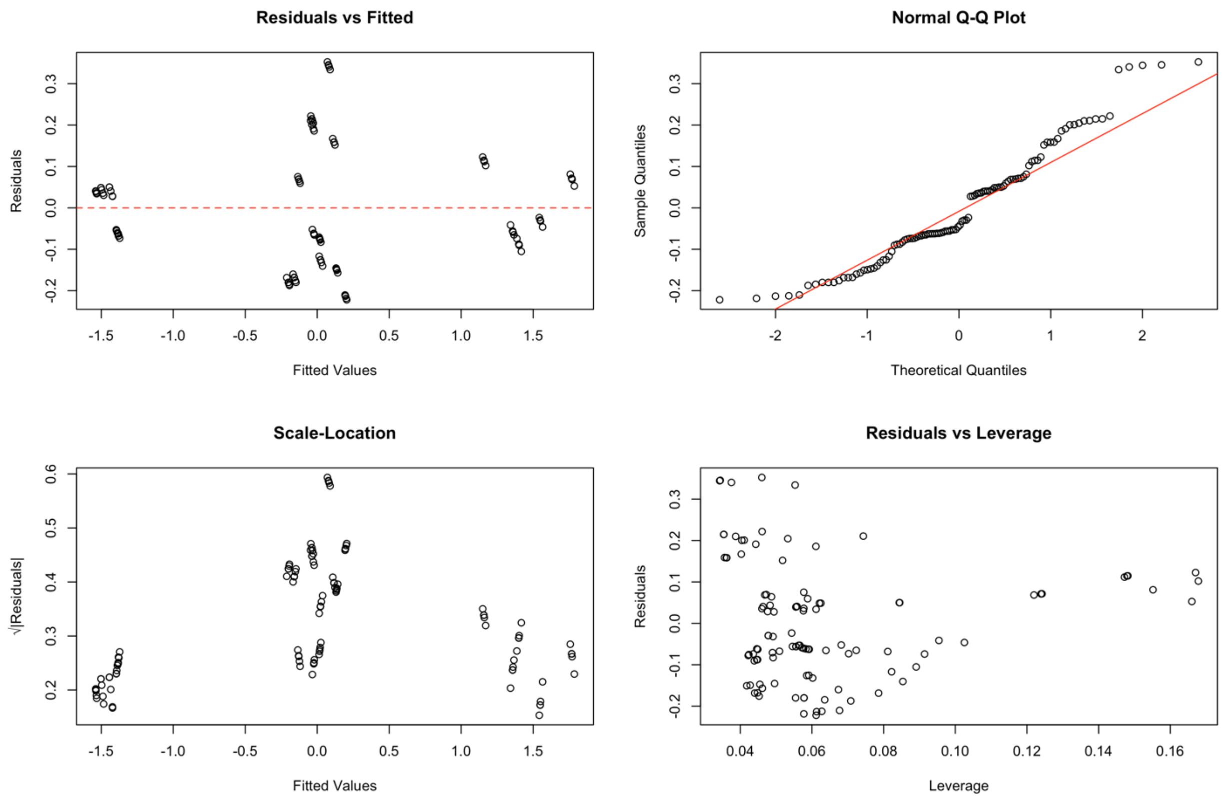

Figure 6 Diagnostic Plots for Model 22

Standardized Effects and Interpretation

To compare the relative effects of the predictor variables, we standardized the independent variables in Model 22 and re-estimated the model. The results show that the time variable, Year_std, remains the strongest predictor ($\beta = 0.299$, $p < 0.001$). Meanwhile, the pandemic case variable, Cases_std, also shows strong statistical significance ($\beta = -0.728$, $p < 0.001$), indicating that, even at the standardized scale, time trends and the epidemic's impact still have a significant effect on healthcare expenditure. The media tone variable, ToneLag1_std, lagged by one period, still does not reach statistical significance ($\beta = -0.033$, $p = 0.613$), consistent with the results before standardization.

In Model 22, the standard deviation of the Country random intercept is 1.15, and the residual standard deviation is 0.14. Specifically, the Country-specific random intercept estimates show that the United States (1.70) has a positive deviation, while China (−1.53) has a negative deviation. Canada (−0.03) and the UK (−0.14) are close to zero. This means that China's baseline healthcare spending is below average, while the United States' baseline healthcare spending is above average, and the United Kingdom and Canada's healthcare spending is close to average.

Reversal Effect Discovery and Resolution

In the initial lagged model analysis, we did not include pandemic-related variables. Instead, we introduced one-year and two-year lags of the media variables (n = 90) in the model. We found that one-year lagged media tone (Average_Tone_Lag1_scaled) was significantly negative ($\beta = -0.562$, $p < 0.001$), whereas two-year lagged media tone (Average_Tone_Lag2_scaled) was significantly positive ($\beta = 0.599$, $p < 0.001$). Moreover, the predictor: one-year lagged media intensity (Average_Intensity_Lag1_scaled) was significantly positive ($\beta = 0.861$, $p < 0.001$), while two-year lagged media intensity (Average_Intensity_Lag2_scaled) was significantly negative ($\beta = -0.783$, $p < 0.001$). All lagged media variables were highly statistically significant in this model (all $p < 0.001$). This reversal pattern is visually illustrated in Figure 7.

At first glance, these opposite-direction lagged effects appear to suggest a cyclical or delayed reversal mechanism in media influence. However, when COVID-19 variables were added in the subsequent models, this “reversal effect” disappeared. All media coefficients became non-significant (all $p > 0.08$). This suggests that the previously observed reversal effect was more probably driven by omitted COVID-related confounding rather than reflecting a genuine media dynamic.

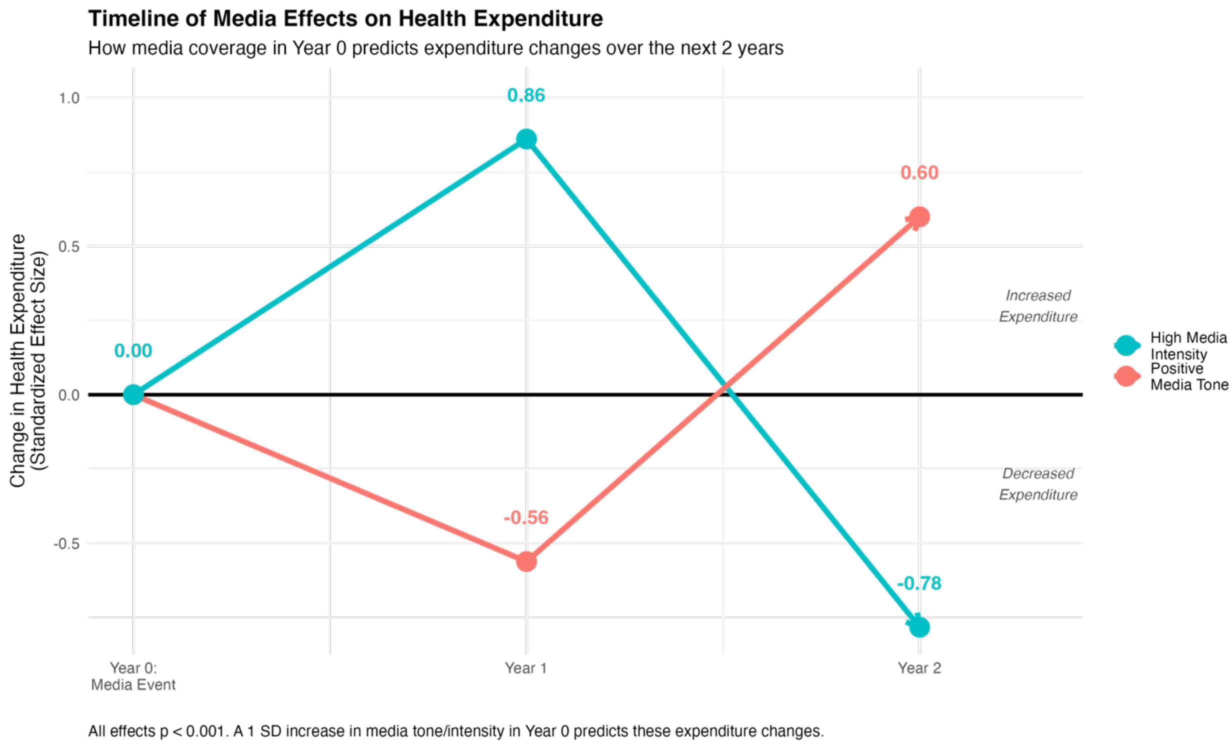

Figure 7 Reversal effects on Health Expenditure

Robustness check: GDELT 1.0 Canada-Only Analysis (2004–2023)

In the robustness check, we used historical GDELT 1.0 data and restricted the analysis to Canada (2004–2023). After introducing one- and two-year lagged variables, the sample size was reduced to n = 18 from 2006 to 2023. Following the stepwise inclusion of control variables, the final model (Healthcare_scaled ~ AvgSentiment_Lag1_scaled + AvgSentiment_Lag2_scaled + GDP_scaled + Population_scaled + Hospital_Beds_scaled + Year_scaled) was selected based on the AIC criterion (AIC = 19.79), significantly lower than the maximum AIC of 51.45 observed in the sentiment-only model.

The results show that neither the one-year lagged sentiment variable ($\beta = -0.056$, $p = 0.532$) nor the two-year lagged sentiment variable ($\beta = -0.150$, $p = 0.199$) reached statistical significance. It is worth noting that GDP was the dominant predictor in the GDELT 1.0 analysis ($\beta = 0.660$, $p =$

0.003). The overall model explained a large proportion of variance ($R^2 = 0.92$), which supports a high degree of fit. The results show that in the Canadian sample, changes in healthcare expenditure were strongly influenced by macroeconomic factors, with a very weak lagged effect from media sentiment.

By the robustness check, convergent evidence was found between GDELT 1 and 2: media effects replicate across two independent measurement systems were not significant, based on different time spans (approximately 20 years compared with 8 years), and different country samples (Canada alone versus four nations). In addition, no reversal effect appeared in the GDELT 1.0 data, further supporting COVID confounding as the explanation.

Summary of Findings

Across 21 model specifications and two independent datasets, media tone and intensity do not significantly predict government healthcare expenditure when appropriate controls are included. Meanwhile, the strongest predictors are an upward trend in spending (Year) and immediate COVID-19 case burden (crisis). This shows apparent media effects (including the reversal effect) were artifacts of omitted pandemic variables.

These results suggest that fiscal healthcare policy represents a boundary condition for agenda-setting theory. In other words, even if the media attempts to influence policy, it does not appear to affect healthcare spending. Media framing may shape public opinion, but it does not seem to change the structural and economic factors that determine actual budget decisions.

Discussion

Summary and Interpretation of Main Findings

In this section, we return to the core research question of this study: Does media tone or intensity influence government healthcare investment? Based on systematic tests across multiple model specifications and two independent datasets, the answer is clear: Once key control variables, such as time trends, macroeconomic factors, and pandemic-related shocks, are included, neither media tone nor media intensity significantly predicts government healthcare expenditure. Therefore, from an empirical perspective, we do not find evidence supporting the claim that media tone or intensity directly influences fiscal healthcare investment decisions.

The year variable plays an important role in these results. In all models, Year was the strongest predictor. This means that healthcare expenditure as a share of GDP increased steadily during the study period. This pattern reflects long-term structural trends rather than short-term ones. For example, population aging, higher rates of chronic diseases, medical technology development, and medical cost inflation may all have an influence on healthcare spending over time. This finding is consistent with existing research (Hartwig & Sturm, 2014). Therefore, the effect of the year mainly represents a structural growth trend in healthcare finance, not short-term media influence.

In the final model, we observe a negative association between COVID-19 case burden and healthcare expenditure. At first glance, this finding may seem counterintuitive, as one might expect that a more severe pandemic would lead to higher spending. However, this pattern likely reflects a reallocation of fiscal resources during the crisis period (De Foo et al., 2023). Governments may have prioritized emergency response measures, such as temporary facilities, testing programs, and short-term interventions, which may have limited the expansion of long-term baseline healthcare investment. Therefore, the negative relationship may capture short-term budget adjustments rather than a reduction in overall healthcare commitment. In addition, because COVID-19 cases and deaths are highly correlated, they contain largely overlapping information. For this reason, only the case variable was retained in the final specification to avoid collinearity concerns.

Why the Reversal Effect Was an Artifact

Initially, we observed a reversal effect in the lagged media models. We found highly significant effects (all $p < 0.001$), which at that stage appeared to indicate a meaningful dynamic relationship between media variables and healthcare expenditure. However, once we introduced COVID-19 controls into the model, these effects disappeared entirely, suggesting that the earlier findings were likely influenced by an omitted external shock rather than reflecting a stable media-policy mechanism.

This interpretation is further supported by the GDELT 1.0 evidence. In the 20-year Canadian dataset, which largely covers the pre-COVID period, the reversal effect did not appear. Despite the longer time span and the use of a different measurement system, no similar pattern was detected, indicating that the reversal was specific to the pandemic period rather than a general phenomenon.

From a methodological perspective, this process is very important. At the beginning, we found significant effects. But these effects disappeared after we systematically introduced additional controls. After using alternative specifications as robustness checks, the results became non-significant. This suggests that omitted variables played an important role and strengthens the credibility of our results.

Boundary Conditions for Agenda-Setting Theory

Agenda-setting theory suggests that media coverage influences policy priorities not by determining what the public thinks, but by shaping what issues the public and policymakers think about (McCombs & Shaw, 1972).

While agenda-setting theory successfully explains media influence in discretionary policy domains such as environmental regulation responding to climate coverage (Brulle et al., 2012), nuclear policy shifts following crisis reporting (Kepplinger & Lemke, 2016), or public concern and subsequent policy attention to energy investment in wind projects (Beasant, 2025), healthcare expenditure appears to be shaped mainly by long-term structural factors rather than media influence.

Healthcare policy may differ from energy, environmental, or other policies for several reasons. First, healthcare spending may be largely limited by existing programs, demographic pressures, and multi-year budget cycles. A large portion of health expenditure is tied to ongoing commitments such as hospital operations, insurance systems, and long-term care. These structural obligations potentially limit the possibility of short-term fiscal adjustments, even when media attention increases. Second, energy and environmental policies are often more discretionary. Governments can introduce new regulations, subsidies, or adjust investment programs in response to public debate or media coverage. In these domains, policy decisions may be more flexible and therefore more responsive to shifts in media coverage. Third, healthcare is experienced directly and continuously by citizens. Individuals directly interact with the health system through health insurance, hospitals, clinics or other health-related services. This phenomenon may potentially reduce the reliance on media as an information source to form their views about healthcare. This may also help explain why, in the existing literature, many studies examine media and health-related topics (such as public attitude or electoral behavior), while relatively few focus directly on media influence over fiscal healthcare policy decisions. This pattern is also consistent with the view that media effects on policy processes may be dynamic and stage-dependent (Fredheim, 2021).

The COVID-19 crisis context also helps explain our results. During the pandemic, governments might mainly respond to rising case numbers and death rates. This information was clear and direct. Policymakers did not appear to need to rely heavily on media coverage to understand the severity of the situation. When experiential information is overwhelming, the role of the media may become less important. In other words, when the crisis is immediate and obvious, policy decisions may be more likely to follow actual data rather than media framing. In addition, government responses during the pandemic may also be constrained by economic considerations, which may further limit the influence of media coverage on policy decisions (Apriliyanti et al., 2022).

Limitations

One important limitation concerns sample size and statistical power. The full sample includes n = 130 observations, which decreases to 110 after introducing lagged variables. In addition, the analysis covers only four countries. Although we tested multiple model specifications, the relatively limited number of observations may restrict the ability to detect small or subtle effects. If media influence does exist, such effects may require a larger sample size or a longer time series to be properly identified. Therefore, the null findings regarding media effects in this study should be interpreted with caution.

Another limitation concerns data granularity. This study uses annual data, which aggregates changes within each year. Policy responses and media attention may occur over shorter time periods. Media effects on policy may operate on quarterly or even monthly cycles. However, annual data may weaken or mask these short-term fluctuations. Therefore, if media effects mainly occur within shorter time units, the data used in this study may not be able to detect these changes.

Country selection also presents certain limitations. This study includes four countries with very different healthcare systems, including predominantly public universal systems, mixed systems, and more market-based structures. These countries were selected to increase variation in institutional context. However, this institutional diversity also introduces heterogeneity, which may not be fully captured by random intercepts. Structural characteristics of different healthcare systems may influence how governments respond to media attention. Therefore, the results may differ in countries with less structured or less institutionalized healthcare systems.

Another limitation concerns theme selection. This study includes only five healthcare-related themes. Although the model results show that the random intercept estimates for all five themes are zero, this does not necessarily mean that media influence is identical across topics. Different media themes may vary in issue importance, political sensitivity, and public attention. Therefore, using only five relatively broad themes may mask potential differences at more specific topic levels.

In addition, The tone and intensity scores in GDELT are generated by algorithms. Although this allows large-scale analysis, these measures may not capture all details of news coverage. Subtle differences in language or framing may not be reflected in the overall scores. As a result, possible effects on public opinion or government decisions may not be fully identified.

Another limitation is the measurement of healthcare expenditure as a percentage of GDP. This indicator is influenced by both the numerator (health spending) and the denominator (GDP), so changes in the ratio do not necessarily reflect health policy adjustments alone. For example, during economic downturns, GDP may decline even if healthcare spending remains stable. In such cases, the share of health expenditure in GDP may increase without any deliberate policy expansion. Therefore, fluctuations in this measure may partly reflect broader economic conditions rather than direct changes in healthcare investment. In addition, healthcare systems differ across countries. In some countries, most health spending is publicly financed, while in others private expenditure plays a larger role. As a result, national-level health expenditure does not fully represent government investment alone. Changes in private spending, such as out-of-pocket payments or insurance coverage, may also affect the overall ratio. This makes it more difficult to isolate the direct fiscal decisions made by governments.

Finally, the models in this study may be subject to omitted variable bias. Although we controlled for COVID-related factors, economic conditions, and time trends, other factors may still influence healthcare spending. For example, election cycles, changes in the ruling party, population aging, and technological advancements in medicine may all affect health expenditure. These structural and political factors were beyond the scope of the present study. However, they may play an important role in long-term healthcare finance decisions. Future research could incorporate additional political and demographic variables to better understand the determinants of government healthcare investment.

Conclusion

This study, based on four countries, two decades of data, two independent measurement systems, and multiple model specifications with systematic robustness checks, finds no evidence that media tone or intensity significantly influences government healthcare expenditure. After including key control variables, media variables did not show stable or significant effects across the main models. Instead, healthcare spending appears to be driven primarily by structural economic forces and direct public health threats rather than media framing.

The findings suggest that fiscal healthcare policy may represent a boundary condition for agenda-setting theory. Although media coverage may shape public attention and influence policy debates in more discretionary domains, healthcare finance decisions are embedded in long-term institutional arrangements and macroeconomic structures. In the GDELT 1.0 analysis, GDP emerged as the strongest predictor of healthcare expenditure, highlighting the role of economic fundamentals. In the GDELT 2.0 analysis, COVID-19 case burden played a central role, indicating that governments respond more strongly to direct and measurable crisis signals. Overall, when structural economic conditions or immediate public health threats dominate, media framing may have limited influence on government healthcare investment decisions.

In conclusion, this study contributes to a clearer understanding of the relationship between media dynamics and government healthcare investment, and it provides a foundation for future research exploring media effects in other policy domains.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During manuscript preparation, the author employed ChatGPT (OpenAI, GPT-5.3 model) solely for editorial support (e.g., language refinement, translation, and clarity improvement). The AI system did not contribute to the conceptualization, methodology, analysis, interpretation, or conclusions of the study. All scholarly judgments and substantive content are exclusively those of the author. The author reviewed and approved all AI-assisted edits and accepts full responsibility for the published work.

References

Alipour, S., Di Marco, N., Avalle, M., Etta, G., Cinelli, M., & Quattrociocchi, W. (2024). The drivers of global news spreading patterns. *Scientific Reports*, *14*(1), 1519. https://doi.org/10.1038/s41598-024-52076-6

Apriliyanti, I. D., Utomo, W. P., & Purwanto, E. A. (2022). Examining the policy narratives and the role of the media in policy responses to the COVID-19 crisis in Indonesia. *Journal of Asian Public Policy*, *15*(3), 541–557. https://doi.org/10.1080/17516234.2021.1954770

Baloria, V. P., & Heese, J. (2018). The effects of media slant on firm behavior. *Journal of Financial Economics*, *129*(1), 184–202. https://doi.org/10.1016/j.jfineco.2018.04.004

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2003). *lme4: Linear Mixed-Effects Models using "Eigen" and S4* (p. 1.1-38) [Dataset]. https://doi.org/10.32614/CRAN.package.lme4

Beasant, C. (2025). *Media's Relationship with Energy Investment: Analyzing the GDELT Data* [Honours thesis]. St. Francis Xavier University.

Bourgeois, D., Rappaz, J., & Aberer, K. (2019). *Selection Bias in News Coverage: Learning it, Fighting it*. https://doi.org/10.48550/ARXIV.1904.07536

Brulle, R. J., Carmichael, J., & Jenkins, J. C. (2012). Shifting public opinion on climate change: An empirical assessment of factors influencing concern over climate change in the U.S., 2002–2010. *Climatic Change*, *114*(2), 169–188. https://doi.org/10.1007/s10584-012-0403-y

Carson, A., Martin, A. J., & Ratcliff, S. (2020). Negative campaigning, issue salience and vote choice: Assessing the effects of the Australian Labor party's 2016 "Mediscare" campaign. *Journal of Elections, Public Opinion and Parties*, *30*(1), 83–104. https://doi.org/10.1080/17457289.2018.1563093

Consoli, S., Tiozzo Pezzoli, L., & Tosetti, E. (2022). Neural Forecasting of the Italian Sovereign Bond Market with Economic News. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(Supplement_2), S197–S224. https://doi.org/10.1111/rssa.12813

Dang, T. L., Dang, V. A., Moshirian, F., Nguyen, L., & Zhang, B. (2019). News media coverage and corporate leverage adjustments. *Journal of Banking & Finance*, *109*, 105666. https://doi.org/10.1016/j.jbankfin.2019.105666

Daw, J. R., Morgan, S. G., Collins, P. A., & Abelson, J. (2014). Framing Incremental Expansions to Public Health Insurance Systems: The Case of Canadian Pharmacare. *Journal of Health Politics, Policy and Law*, *39*(2), 295–330. https://doi.org/10.1215/03616878-2416229

De Foo, C., Verma, M., Tan, S. Y., Hamer, J., van der Mark, N., Pholpark, A., Hanvoravongchai, P., Cheh, P. L. J., Marthias, T., Mahendradhata, Y., & others. (2023). Health financing policies during the COVID-19 pandemic and implications for universal health care: A case study of 15 countries. *The Lancet Global Health*, *11*(12), e1964–e1977.

Fox, J., Weisberg, S., Price, B., Friendly, M., & Hong, J. (2003). *effects: Effect Displays for Linear, Generalized Linear, and Other Models* (p. 4.2-4) [Dataset]. https://doi.org/10.32614/CRAN.package.effects

Fredheim, N. A. G. (2021). Life, death or drugs: Actor framing power on the news media coverage of health care policy. *European Journal of Communication*, *36*(3), 270–285. https://doi.org/10.1177/0267323120978721

GDELT Project. (n.d.). *The GDELT Project*. Retrieved January 17, 2026, from https://www.gdeltproject.org/

GDELT Project. (2017, January 20). *GDELT 2.0 API Debuts!* https://blog.gdeltproject.org/gdelt-doc-2-0-api-debuts/

Hanson, K., Brikci, N., Erlangga, D., Alebachew, A., De Allegri, M., Balabanova, D., Blecher, M., Cashin, C., Esperato, A., Hipgrave, D., Kalisa, I., Kurowski, C., Meng, Q., Morgan, D., Mtei, G., Nolte, E., Onoka, C., Powell-Jackson, T., Roland, M., … Wurie, H. (2022). The Lancet Global Health Commission on financing primary health care: Putting people at the centre. *The Lancet Global Health*, *10*(5), e715–e772. https://doi.org/10.1016/S2214-109X(22)00005-5

Hart, P. S., Chinn, S., & Soroka, S. (2020). Politicization and Polarization in COVID-19 News Coverage. *Science Communication*, *42*(5), 679–697. https://doi.org/10.1177/1075547020950735

Hartwig, J., & Sturm, J.-E. (2014). Robust determinants of health care expenditure growth. *Applied Economics*, *46*(36), 4455–4474. https://doi.org/10.1080/00036846.2014.964829

Hsu, Y., Hua, M., Liu, H., & Wang, Q. (2024). News sentiment and investment efficiency: Evidence from China. *European Financial Management*, *30*(3), 1587–1617. https://doi.org/10.1111/eufm.12454

Hu, B., Luan, Q., Meng, X., & Wang, K. (2023). Media Inclination and Outward Foreign Direct Investment: Evidence from Chinese Firms. *China & World Economy*, *31*(6), 134–155. https://doi.org/10.1111/cwe.12511

Hu, Q., Yuan, L., & Liu, B. (2025). Spatiotemporal evolution of hedging effects in Asia-Pacific countries amid Sino-US competition: Insights from massive event data. *PLOS ONE*, *20*(1), e0317308. https://doi.org/10.1371/journal.pone.0317308

Huth, N., & Abergel, F. (2014). High frequency lead/lag relationships—Empirical facts. *Journal of Empirical Finance*, *26*, 41–58. https://doi.org/10.1016/j.jempfin.2014.01.003

Iyengar, S., & Hahn, K. S. (2009). Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use. *Journal of Communication*, *59*(1), 19–39. https://doi.org/10.1111/j.1460-2466.2008.01402.x

Kepplinger, H. M., & Lemke, R. (2016). Instrumentalizing Fukushima: Comparing Media Coverage of Fukushima in Germany, France, the United Kingdom, and Switzerland. *Political Communication*, *33*(3), 351–373. https://doi.org/10.1080/10584609.2015.1022240

Korkmaz, G., Cadena, J., Kuhlman, C. J., Marathe, A., Vullikanti, A., & Ramakrishnan, N. (2016). Multi-source models for civil unrest forecasting. *Social Network Analysis and Mining*, *6*(1), 50. https://doi.org/10.1007/s13278-016-0355-8

Krieg, S. J., Smith, C. W., Chatterjee, R., & Chawla, N. V. (2022). Predicting terrorist attacks in the United States using localized news data. *PLOS ONE*, *17*(6), e0270681. https://doi.org/10.1371/journal.pone.0270681

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2013). *lmerTest: Tests in Linear Mixed Effects Models* (p. 3.2-0) [Dataset]. https://doi.org/10.32614/CRAN.package.lmerTest

Lawson, B. (2022). Realizing the benefits of open government data: Journalists' coverage of the NHS winter crisis, 2016–17. *The Information Society*, *38*(1), 25–35. https://doi.org/10.1080/01972243.2021.1998274

Liao, R., Wang, X., & Wu, G. (2021). The role of media in mergers and acquisitions. *Journal of International Financial Markets, Institutions and Money*, *74*, 101299. https://doi.org/10.1016/j.intfin.2021.101299

Lüdecke, D. (2013). *sjPlot: Data Visualization for Statistics in Social Science* (p. 2.9.0) [Dataset]. https://doi.org/10.32614/CRAN.package.sjPlot

Lüdecke, D., Makowski, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., Wiernik, B. M., & Thériault, R. (2019). *performance: Assessment of Regression Models Performance* (p. 0.15.3) [Dataset]. https://doi.org/10.32614/CRAN.package.performance

Martin, D., Miller, A. P., Quesnel-Vallée, A., Caron, N. R., Vissandjée, B., & Marchildon, G. P. (2018). Canada's universal health-care system: Achieving its potential. *The Lancet*, *391*(10131), 1718–1735. https://doi.org/10.1016/S0140-6736(18)30181-8

McCombs, M. E., & Shaw, D. L. (1972). The Agenda-Setting Function of Mass Media. *Public Opinion Quarterly*, *36*(2), 176. https://doi.org/10.1086/267990

Miller, E. A., Tyler, D. A., Rozanova, J., & Mor, V. (2012). National Newspaper Portrayal of U.S. Nursing Homes: Periodic Treatment of Topic and Tone. *The Milbank Quarterly*, *90*(4), 725–761. https://doi.org/10.1111/j.1468-0009.2012.00681.x

Niu, J., Rees, E., Ng, V., & Penn, G. (2021). Statistically Evaluating Social Media Sentiment Trends towards COVID-19 Non-Pharmaceutical Interventions with Event Studies. *Proceedings of the Sixth Social Media Mining for Health (#SMM4H) Workshop and Shared Task*, 1–6. https://doi.org/10.18653/v1/2021.smm4h-1.1

Okechukwu, C. (2024). Media Influence on Public Opinion and Political Decision-Making. *International Journal of Political Science Studies*, *1*(1), 13–24.

Ooms, J. (2017). *writexl: Export Data Frames to Excel "xlsx" Format* (p. 1.5.4) [Dataset]. https://doi.org/10.32614/CRAN.package.writexl

Our World in Data. (n.d.-a). *Cumulative confirmed COVID-19 cases and deaths, World*. Retrieved January 10, 2026, from https://ourworldindata.org/grapher/cumulative-deaths-and-cases-covid-19

Our World in Data. (n.d.-b). *Hospital beds per 1,000 people*. Retrieved January 10, 2026, from https://ourworldindata.org/grapher/hospital-beds-per-1000-people

Our World in Data. (n.d.-c). *Life expectancy*. Retrieved January 10, 2026, from https://ourworldindata.org/grapher/life-expectancy

Pinheiro, J., Bates, D., & R Core Team. (1999). *nlme: Linear and Nonlinear Mixed Effects Models* (p. 3.1-168) [Dataset]. https://doi.org/10.32614/CRAN.package.nlme

Prior, M. (2013). Media and Political Polarization. *Annual Review of Political Science*, *16*(1), 101–127. https://doi.org/10.1146/annurev-polisci-100711-135242

R Core Team. (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Revelle, W. (2007). *psych: Procedures for Psychological, Psychometric, and Personality Research* (p. 2.6.1) [Dataset]. https://doi.org/10.32614/CRAN.package.psych

Soroka, S., Maioni, A., & Martin, P. (2013). What Moves Public Opinion on Health Care? Individual Experiences, System Performance, and Media Framing. *Journal of Health Politics, Policy and Law*, *38*(5), 893–920. https://doi.org/10.1215/03616878-2334656

Walker, C., Hanna, P., & Raisborough, J. (2021). Representations of the National Health Service (NHS) in UK print media. *Health Promotion International*, *36*(1), 178–186. https://doi.org/10.1093/heapro/daaa044

Wickham, H. (2010). *reshape2: Flexibly Reshape Data: A Reboot of the Reshape Package* (p. 1.4.5) [Dataset]. https://doi.org/10.32614/CRAN.package.reshape2

Wickham, H. (2016). *tidyverse: Easily Install and Load the "Tidyverse"* (p. 2.0.0) [Dataset]. https://doi.org/10.32614/CRAN.package.tidyverse

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., & Van Den Brand, T. (2007). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics* (p. 4.0.1) [Dataset]. https://doi.org/10.32614/CRAN.package.ggplot2

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2014). *dplyr: A Grammar of Data Manipulation* (p. 1.2.0) [Dataset]. https://doi.org/10.32614/CRAN.package.dplyr

Wickham, H., Vaughan, D., & Girlich, M. (2014). *tidyr: Tidy Messy Data* (p. 1.3.2) [Dataset]. https://doi.org/10.32614/CRAN.package.tidyr

World Bank. (n.d.-a). *Current health expenditure (% of GDP)*. Retrieved November 9, 2025, from https://data360.worldbank.org/en/indicator/WB_WDI_SH_XPD_CHEX_GD_ZS

World Bank. (n.d.-b). *GDP (current US$)*. Retrieved November 9, 2025, from https://data.worldbank.org/indicator/NY.GDP.MKTP.CD

World Bank. (n.d.-c). *Population, total*. Retrieved November 9, 2025, from https://data.worldbank.org/indicator/SP.POP.TOTL