Quantifying Societal Stress: Forecasting Historical London Mortality using Hardship Sentiment and Crime Data with Natural Language Processing and Time-Series*

Sebastian Olsen
Amsterdam University College
sebolsen2@gmail.com

Jelke Bloem
University of Amsterdam
j.bloem@uva.nl

Abstract

We study links between societal stress - quantified from 18th–19th century Old Bailey trial records - and weekly mortality in historical London. Using MacBERTh-based hardship sentiment and time-series analyses (CCF, VAR/IRF, and a Temporal Fusion Transformer, TFT), we find robust lead–lag associations. Hardship sentiment shows its strongest predictive contribution at a 5–6 week lead for mortality in the TFT, while mortality increases precede higher conviction rates in the courts. Results align with Epidemic Psychology and suggest that text-derived stress markers can improve forecasting of public-health relevant mortality fluctuations.

1 Introduction

"Great fears of the sickenesse here in the City, it being said that two or three houses are already shut up. God preserve us all!" This entry from Samuel Pepys' diary, written on Sunday 30 April (1665), encapsulates the dread that gripped London during the Great Plague. Accounts such as these show how epidemics are not purely biological phenomena, but events that have a widespread impact on social order. The model of Epidemic Psychology put forward by the sociologist Strong (1990) outlines how societies can be caught up in what he calls a "maelstrom" of impact, altering social structures.

This study investigates whether indicators of societal stress, expressed through historical crime records, relate to fluctuations in total weekly mortality. Such fluctuations are often indicative of public-health crises, but also reflect broader societal vulnerabilities. Researchers such as Alcabes (2009) have explored the social impact of epidemics, with studies primarily focusing on their sociological impacts (Chu et al., 2020; Muthuswamy, 2024). Modern quantitative studies exist, with Kastalskiy et al. (2021) investigating the impact of societal stress on the COVID-19 pandemic. Their research finds that countries able to manage stress can avoid strong second disease waves. These results highlight the potential for investigating similar dynamics in historical contexts. However, there is a distinct lack of quantitative research utilising high-quality historical data. We aim to better understand how societal stress has influenced mortality trends and fill this gap in historical quantitative studies.

Our primary research question is: "How are patterns of crime, judicial outcomes, and the sentiment expressed in criminal trial accounts related to weekly mortality fluctuations in London, reflecting societal stress during periods of crisis?". Our work builds on previous studies that link text sentiment to real-world measures (Bentley et al., 2014) and on quantitative legal history and sentiment analysis (Zhao et al., 2022; Pan et al., 2021).

Research Context and Motivation Modern studies (e.g., Kruspe et al. 2020; Wang et al. 2020) show that public sentiment can shift almost in real time - lockdowns and policy announcements produce immediate mood changes on social media. Such rapid shifts reflect how heightened societal stress and widespread hardship can impair decision-making, fuel policy overreactions, and undermine social stability (FeldmanHall et al., 2015). Historical accounts (for example Bishop de Belsunce on the Great Plague of Marseille; Devaux 2012) suggest comparable social dynamics may have occurred in earlier epidemics. By quantitatively analysing trial texts alongside weekly mortality, we

*Code for this study is available at: https://github.com/Seb-Olsen/ranlp25-hardship-mortality

Proceedings of the First Workshop NLP and Language Models for Digital Humanities associated with RANLP 2025, pages 112–119, Varna, Bulgaria, Sep 11, 2025.
https://doi.org/10.26615/978-954-452-106-6-010

ask whether spikes in archival hardship language precede or follow mortality surges and whether trial sentiment echoes Strong's "maelstrom" of fear-driven social collapse (Strong, 1990). Applying NLP and time-series forecasting gives us an empirical way to test these dynamics in historical data.

Contributions (1) We construct a hardship sentiment proxy from historical trial texts using MacBERTh and ABSA-style embedding similarity; (2) We quantify lead–lag structure with CCF, VAR/IRF, and predictive tests, separating contemporaneous from delayed associations; (3) We benchmark forecasting with TFT against naïve and seasonal baselines and interpret drivers via variable importance; (4) We connect mortality dynamics with judicial responses, providing quantitative evidence consistent with Epidemic Psychology.

2 Related Work

This research is situated at the intersection of historical NLP, time-series forecasting, and public health. The application of transformer-based models to time-series forecasting is a burgeoning field. The Temporal Fusion Transformer (TFT) (Lim et al., 2021) has shown state-of-the-art performance in various domains like retail and finance, but its application to historical or epidemiological data is less common. Makhonza et al. (2024) successfully applied TFT to modern mortality forecasting, demonstrating its potential. Our work extends this by applying TFT to a uniquely challenging historical dataset, characterised by noise, reporting lags, and non-stationarity.

Furthermore, the use of NLP to extract public health indicators is a well-established practice, with systematic reviews highlighting its application to infectious disease surveillance (Agbehadji and Awad, 2023) and population mental health monitoring (Gkotsis et al., 2024). Specific applications are diverse: foundational studies used sentiment analysis to predict flu outbreaks (Pan et al., 2021) and measure public reaction to health policies (Zhao et al., 2022), while more recent work employs sophisticated transformer models for detailed emotion classification (Cui et al., 2024) and leverages the latest generation of LLMs to analyse health discussions on social media (Zhang et al., 2024). Our study is novel in its application of these techniques to a large-scale historical corpus to create a sentiment-based proxy for societal stress. By linking this proxy to administrative records on mortality and crime, we create a new, quantitative lens through which to study historical societal dynamics.

3 Methodology

3.1 Data Sources

Our primary text source is the Old Bailey Sessions Papers XML corpus (Hitchcock et al., 2023), covering 1678–1849. This corpus was chosen for its detailed trial accounts and precise weekly dating, which mitigates the publication lag found in other historical text corpora and is crucial for time-series analysis. Crucially, its focus on London allowed for direct geographical and temporal alignment with our mortality dataset. The source boasts a 99.99% accurate transcription rate (Hitchcock and Turkel, 2016). We restricted our analysis to 1719–1829 to maximise data utility. We programmatically extracted structured metadata – `offenceCategory`, `verdictCategory`, and `punishmentCategory` – from the records (Hitchcock and Shoemaker, 2006), yielding a total of 138,078 trial observations.

For mortality data, we used the London Bills of Mortality dataset (Smith et al., 2020), which provides weekly mortality counts from 1644–1849. To stabilise variance and mitigate extreme spikes, we log-transformed the weekly death counts, creating the `log_deaths` variable. Izdebski et al. (2022) have previously used such mortality data as a reliable indicator of epidemic impact in pre-modern periods.

3.2 Sentiment Analysis

A key challenge was the linguistic complexity of historical texts and the lack of temporal generalization ability of language models (Verkijk et al., 2025). We leveraged MacBERTh, a BERT model adapted for historical English (1450-1950) (Devlin et al., 2019; Manjavacas and Fonteyn, 2022). Its pre-training on historical corpora provides robustness to spelling variations and helps mitigate semantic drift (Kutuzov and Giulianelli, 2020), making it highly suitable for fine-grained semantic tasks. Recent work has specifically demonstrated the value of using MacBERTh for aspect-based sentiment analysis (ABSA) in literary-historical contexts (Dejaeghere et al., 2024).

Informed by this precedent, we adopted an ABSA approach to quantify hardship within the trial narratives. We generated embeddings for each trial's text and computed reference embeddings for a curated list of hardship-related terms (Table 1). The resulting hardship sentiment score was obtained by cosine similarity between trial embeddings and the hardship reference vectors.

Terms for hardship sentiment		
poor	poverty	necessity
distress	hardship	starve
desperate	ruin	beggar
vagrant	hunger	want

Table 1. Terms used to generate reference embeddings indicative of hardship sentiment.

3.3 Validation and Robustness of the Hardship Measure

We qualitatively validated our hardship measure by auditing a stratified sample of trials. High-scoring texts clearly contained hardship narratives. For instance, one trial described a prisoner who begged a judge for mercy, stating: "What a sad thing will this be for my Wife, who has not a Farthing in the World." In contrast, low-scoring trials were typically terse and factual. To test robustness, we compared our MacBERTh-based scores to a simple keyword-frequency baseline; correlations and lead–lag patterns were similar. This triangulation supports the validity of our construct, though further multi-annotator evaluation remains a priority for future work.

3.4 Time-Series Analysis and Forecasting

To explore the temporal dynamics between our variables, we employed several time-series analyses. Cross-Correlation Functions (CCF) were used to examine lead-lag relationships (Shumway and Stoffer, 2000). We then performed Granger causality tests to assess whether one time series could be used to forecast another (Granger, 1969). We tested for lags from 1 to 12 weeks.

Finally, to model the system's response to shocks, we fitted a Vector Autoregression (VAR) model. A VAR model expresses each variable as a linear function of its own past values and the past values of all other variables in the system (Sims, 1980). From the fitted VAR model, we generated Impulse Response Functions (IRFs), which trace the dynamic effect of a one-time, one-standard-deviation shock in one variable on the future trajectory of another variable, holding other shocks constant. This allows for a deeper analysis of the bivariate relationships.

For the primary forecasting task, we implemented the Temporal Fusion Transformer (TFT) (Lim et al., 2021), a neural network architecture with built-in interpretability. We optimised hyperparameters using the Optuna framework, aiming to minimise Symmetric Mean Absolute Percentage Error (SMAPE). The model was trained using `QuantileLoss` at the 0.1, 0.5, and 0.9 quantiles to estimate prediction intervals. We evaluated its performance against three standard baselines: a Naïve (last-value) forecast, a Seasonal Naïve (same-week-last-year) forecast, and a Historical Average forecast.

4 Results

The final DataFrame contains 5,768 weekly rows and 28 features (aggregated `hardship_sentiment`, crime proportions, conviction/punishment rates, trial counts, a `year_end_spike` indicator, and mortality variables). Trial-level hardship scores excluded 5,549 very-short trials (<30 words), i.e., $\approx 4.02\%$, to improve reliability (see Figure 1).

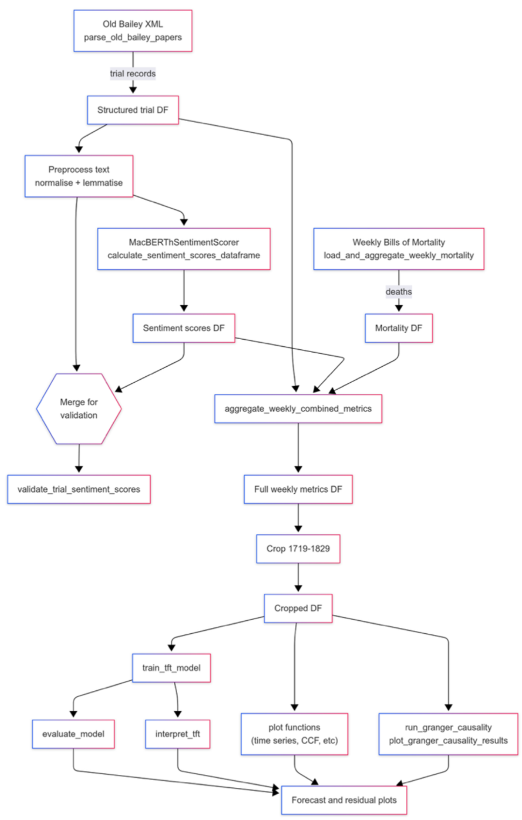


Figure 1. Pipeline for sentiment analysis and mortality data forecasting.

4.1 Exploratory Data Analysis

The weekly death count (Figure 2) shows a gradual decrease over the period, likely due to public health improvements (Porter, 1991). The prominent 'jitters' are artefacts of historical reporting, where parishes aggregated death counts at year-end (Figure 3). The conviction rate (Figure 4) was steady around 60%, but spiked to 97% between 1791–1793, possibly reflecting fears following the French Revolution (Eastwood, 1995). Across the series mean weekly deaths decline from ≈ 500 in the early 1700s to ≈ 420 by 1829, motivating the log-transform used throughout. The dataset therefore combines long-run decline with irregular, year-end aggregation spikes that we explicitly model.

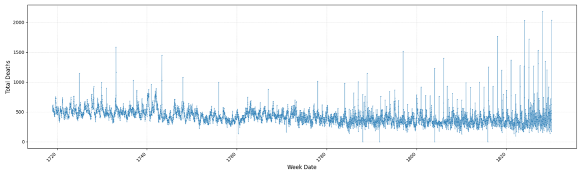


Figure 2. Weekly Death Rates (1719–1829).

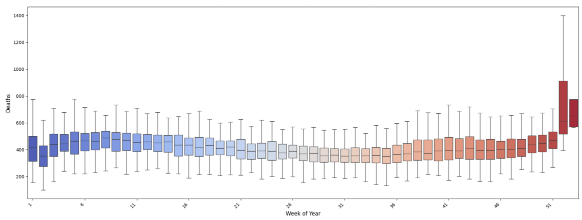


Figure 3. Distribution of Deaths by Week (1719–1829).

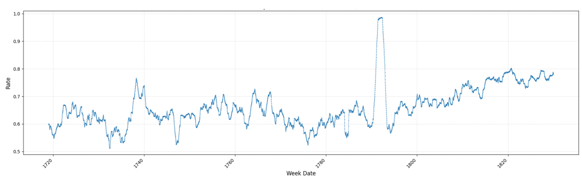
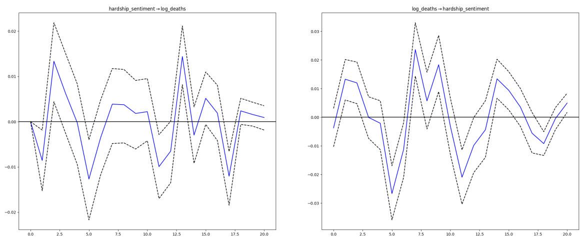


Figure 4. Weekly Conviction Rate (1719–1829).

Analyses of trial and verdict counts around 1788–1792 show that the conviction-rate spike was not driven by fewer trials, suggesting a genuine increase in conviction likelihood during that period.

4.2 Shock Analysis with VAR/IRF

VAR/IRF analysis shows that an unexpected shock in hardship sentiment leads to a statistically significant increase in the growth rate of `log_deaths` about two weeks later, followed by oscillatory effects (Figure 5a). Conversely, a shock in `log_deaths` prompts smaller but statistically significant responses in hardship sentiment (Figure 5b).



(a) Hardship → Deaths IRF (b) Deaths → Hardship IRF

Figure 5. Impulse response functions (95% CIs).

4.3 Lead–lag associations and predictive tests

The cross-correlation plot (Figure 6) shows a modest peak when `hardship_sentiment` leads `log_deaths` by 4–6 weeks, and a stronger, significant correlation when deaths lead sentiment by 17–20 weeks. A five-year rolling correlation (Figure 7) reveals a highly dynamic relationship, with a strong positive peak (>+0.6) around the 1740 "Great Frost" (Engler et al., 2013).

Granger causality tests (Figure 8) for `hardship_sentiment` causing `log_deaths` were not statistically significant. However, tests show that `log_deaths` significantly Granger-causes the conviction rate, particularly with a lag of two or more weeks (Figure 9).

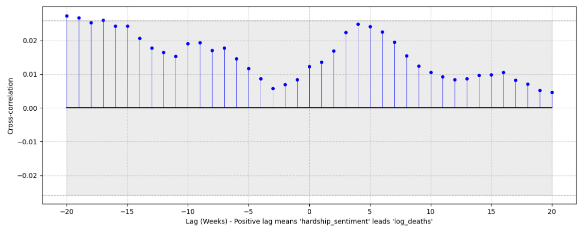


Figure 6. Cross-correlation of hardship and deaths.

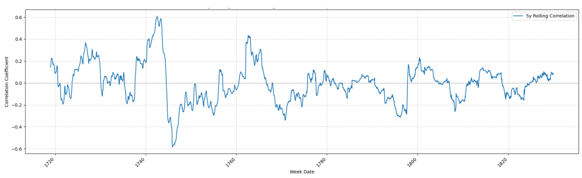


Figure 7. Five-year rolling correlation.

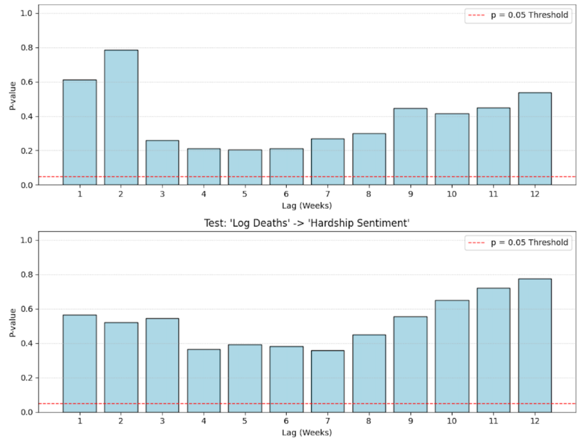


Figure 8. Granger Test: Hardship → Deaths.

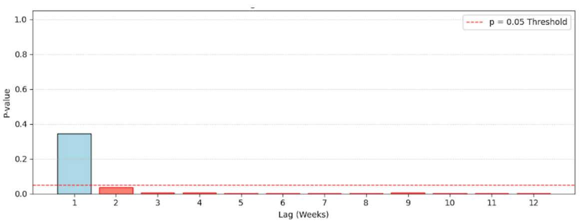


Figure 9. Granger Test: Deaths → Conviction Rate.

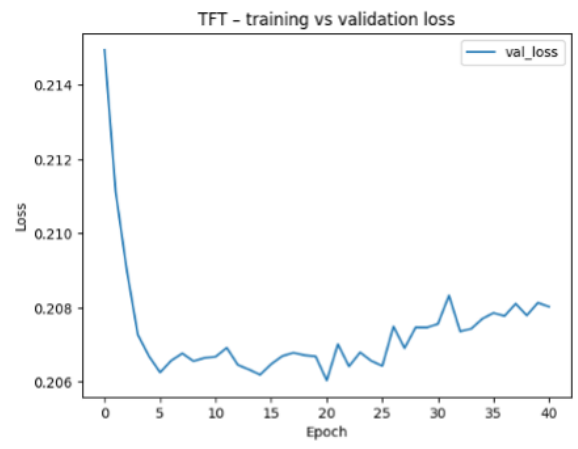
4.4 TFT Model Performance and Error Analysis

The TFT model outperformed all baselines (Table 2), achieving an MAE of 115.9. An error analysis (Figure 10) reveals the TFT's robustness. The validation loss curve (Figure 10a) shows that early stopping at epoch 20 prevented overfitting. The forecast visualization (Figure 10b) shows that the TFT model successfully captures the overall mortality trend and year-end spikes, unlike the baselines. The Seasonal Naïve model performs worst (SMAPE: 43.8%), likely because it overfits to past seasonal patterns and is brittle to the irregular shocks that drive historical mortality.

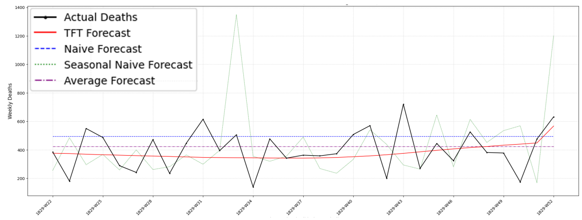
Model	MAE	MSE	SMAPE (%)	RMSE
TFT	115.9	20396.4	30.9	142.8
Naive	136.5	28921.7	34.3	170.1
Seasonal Naive	191.5	68216.0	43.8	261.2
Average	118.4	20562.2	31.0	143.4

Table 2. Forecasting performance of TFT and baselines.

The model's interpretability features highlight which signals drive forecasts. Among decoder variables, the `year_end_spike` indicator was most influential ($\approx 45\%$; Figure 11a). For encoder variables, hardship sentiment with a six-week lag was the strongest predictor ($\approx 16\%$; Figure 11b).

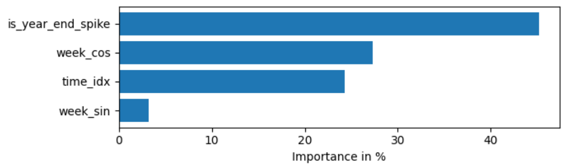


(a) TFT Validation Loss (early stop at epoch 20).

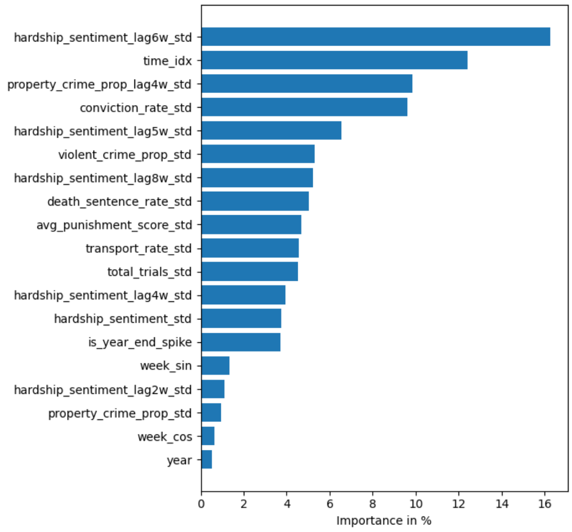


(b) Actual vs. TFT and Baseline Forecasts - Validation Period

Figure 10. TFT model performance and forecast visualization.



(a) TFT – Decoder Variables Importance.



(b) TFT – Encoder Variables Importance.

Figure 11. TFT model interpretability: feature importance.

5 Discussion

This research finds that patterns of crime, judicial outcomes, and sentiment in trial accounts are significantly related to weekly mortality in historical London. The finding that hardship sentiment lagged by six weeks is the top encoder variable for the TFT model strongly supports the hypothesis that hardship has a tangible, though delayed, relationship with mortality. This delay may be attributable to the time it takes for chronic stress and worsening living conditions to physically impact individuals, a conclusion supported by sociological research on unemployment and immune dysfunction (Milner et al., 2013; Balakin et al., 2025).

The relationship is multifaceted and bidirectional. While contemporaneous correlation is weak ($r = 0.01$), time-lagged analyses reveal a more complex story. The CCF, VAR/IRF, and TFT results all point to hardship sentiment having its greatest impact on mortality with an approximate six-week delay. Furthermore, mortality increases appear to precede rises in hardship sentiment by about 17-20 weeks, and Granger-cause a rise in conviction rates after a lag of two weeks. This indicates a bidirectional lead–lag pattern wherein public health crises influence public sentiment and judicial responses.

The finding that increased deaths Granger-cause higher conviction rates (Figure 9) raises complex questions about judicial objectivity, corroborating conclusions from Jedwab et al. (2021). Was the judicial system swayed by public fear, leading to a search for 'scapegoats' during periods of high mortality? Was it used by authorities to restore order after periods of high societal stress? These questions merit further research into the response of the legal system to drivers of societal stress.

These results contribute directly to Strong's (1990) concept of the "maelstrom", where disease outbreaks are followed by suspicion and moral controversy. As Strong notes, "friends, family and neighbours may be feared... the world may be turned upside down" (1990, p. 252). Our findings build on this by showing that hardship sentiment and judicial metrics can act as quantitative predictors of future weekly mortality. Methodologically, this offers a new way to track this historical effect and serves as a strong use case for applying historical language models like MacBERTh to extract meaningful, predictive signals from complex textual archives.

5.1 Limitations and Future Work

One key limitation is the quality of 18th-century data reporting. Spikes in weekly deaths are often artifacts of year-end data aggregation. While our `year_end_spike` feature helps mitigate this, future work could also benchmark our ABSA-style approach against other sentiment analysis techniques. A second challenge lies in the historical texts themselves, which contain spelling variations and semantic drift that are only partially mitigated by models like MacBERTh.

Another limitation is the use of total weekly mortality as a proxy for epidemic periods. This aggregate measure includes deaths from other causes potentially influenced by societal stress (e.g., malnutrition). Future work with more disaggregated, cause-specific death records could differentiate the impact of hardship more clearly. The findings of this study are specific to London from 1719-1829; generalizing these links to other contexts or modern data requires caution and further investigation to validate the relationships.

A final limitation concerns causal interpretation. While lead–lag structure is suggestive, our analyses cannot rule out unobserved confounders (e.g., climate, policing intensity). We therefore interpret findings as predictive associations only.

6 Conclusion

This study explored the temporal relationships between societal stress indicators and weekly mortality in 18th and 19th-century London. Our analysis revealed several key findings: (1) hardship sentiment, particularly with a 5-6 week lag, is a significant leading predictor of weekly mortality fluctuations; (2) patterns of crime also demonstrate predictive value; and (3) a bidirectional pattern is observed, with rising mortality improving forecasts of subsequent conviction rates.

The demonstrated link between text-derived hardship and mortality suggests the potential utility of similar sentiment analysis techniques on modern, large-scale textual data (e.g., social media, news reports) for early-warning systems of societal distress. By providing a quantitative lens on Epidemic Psychology, this work underscores the far-reaching societal impacts of public health crises and opens new avenues for both historical inquiry and contemporary policy.

References

Israel Eben Agbehadji and Alhassan Awad. 2023. A systematic review of machine learning and deep learning on social media for public health surveillance of infectious diseases. *Social Network Analysis and Mining*, 13(1):52.

Philip Alcabes. 2009. *Dread: How Fear and Fantasy Have Fueled Epidemics from the Black Plague to Avian Flu*. PublicAffairs.

E. Balakin, K. Yurku, M. Ivanov, A. Izotov, V. Nakhod, and V. Pustovoyt. 2025. Regulation of stress-induced immunosuppression in the context of neuroendocrine, cytokine, and cellular processes. *Biology*, 14(1):76.

R. Alexander Bentley, Alberto Acerbi, Paul Ormerod, and Vasileios Lampos. 2014. Books average previous decade of economic misery. *PLoS ONE*, 9(1):e83147.

Irene Y. Chu, Parisha Alam, Heidi J. Larson, and Li Lin. 2020. Social consequences of mass quarantine during epidemics: a systematic review with implications for the COVID-19 response. *Journal of Travel Medicine*, 27(7).

Yubao Cui, Xin Li, Shuai Ma, and Qi Zhang. 2024. Roberta-based multi-label emotion classification for public health social media data. *Applied Sciences*, 14(5):2075.

Tanne Dejaeghere, Pushpinder Singh, Els Lefever, and Julie Birkholz. 2024. Exploring aspect-based sentiment analysis methodologies for literary-historical research purposes. In *Proceedings of the 4th Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA)*.

Christian A. Devaux. 2012. Small oversights that led to the great plague of Marseille (1720–1723): Lessons from the past. *Infection, Genetics and Evolution*, 14:169–185.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.

David Eastwood. 1995. E. P. Thompson, Britain, and the French revolution. *History Workshop Journal*, 39:79–88.

Steven Engler, Franz Mauelshagen, Jürg Luterbacher, and Johannes P. Werner. 2013. The irish famine of 1740–1741: famine vulnerability and "climate migration". *Climate of the Past*, 9(3):1161–1179.

Oriel FeldmanHall, Candace M. Raio, Jonathan T. Kubota, Michael G. Seiler, and Elizabeth A. Phelps. 2015. The effects of social context and acute stress on decision making under uncertainty. *Psychological Science*, 26(12):1918–1926.

Georgios Gkotsis, Alison O'Mara-Eves, Stan Zammit, Samuel R Chamberlain, David Gunnell, Trevor Thompson, and James Thomas. 2024. Natural language processing for public mental health surveillance: A scoping review. *JMIR Mental Health*, 11:e48275.

Clive W. J. Granger. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438.

Tim Hitchcock and Robert Shoemaker. 2006. Digitising history from below: The Old Bailey proceedings online, 1674-1834. *History Compass*, 4(2):193–202.

Tim Hitchcock, Robert Shoemaker, Clive Emsley, Sharon Howard, and Jessica McLaughlin. 2023. The Old Bailey proceedings online, 1674–1913. Version 9.0.

Tim Hitchcock and William J. Turkel. 2016. The Old Bailey proceedings, 1674–1913: Text mining for evidence of court behavior. *Law and History Review*, 34(4):929–955.

Adam Izdebski, Piotr Guzowski, Roman Poniat, Luca Masci, J. Palli, C. Vignola, M. Bauch, C. Cocozza, R. Fernandes, F. C. Ljungqvist, and 1 others. 2022. Palaeoecological data indicates land-use changes across Europe linked to spatial heterogeneity in mortality during the black death pandemic. *Nature Ecology & Evolution*, 6(3):297–306.

Remi Jedwab, Allan M. Khan, Jonathan Russ, and Esha D. Zaveri. 2021. Epidemics, pandemics, and social conflict: Lessons from the past and possible scenarios for COVID-19. *World Development*, 147:105629.

Igor A. Kastalskiy, Elena V. Pankratova, Evgeny M. Mirkes, Victor B. Kazantsev, and Alexander N. Gorban. 2021. Social stress drives the multi-wave dynamics of COVID-19 outbreaks. *Scientific Reports*, 11(1):2021.

Anna Kruspe, Matthias Häberle, Iona Kuhn, and Xiao Xiang Zhu. 2020. Cross-language sentiment analysis of European twitter messages during the COVID-19 pandemic. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*.

Andrey Kutuzov and Mario Giulianelli. 2020. UiO-UvA at SemEval-2020 task 1: Contextualised embeddings for lexical semantic change detection. ArXiv preprint arXiv:2005.00050.

Bryan Lim, Sercan Ö. Arık, Nicolas Loeff, and Tomas Pfister. 2021. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764.

B. Makhonza, N. Mogodi, and R. Mbuvha. 2024. Mortality forecasting using temporal fusion transformers. *SSRN Electronic Journal*.

Enrique Manjavacas and Lauren Fonteyn. 2022. Adapting vs. pre-training language models for historical languages. *Journal of Data Mining & Digital Humanities*.

Allison Milner, Andrew Page, and Anthony D. LaMontagne. 2013. Long-term unemployment and suicide: A systematic review and meta-analysis. *PLoS ONE*, 8(1):e51333.

V. V. Muthuswamy. 2024. Epidemics and society: A historical lens on public health and community resilience. *Journal of Natural Science, Biology and Medicine*, 15:58–71.

Wenju Pan, Ru Wang, Wenting Dai, Guiping Huang, Chang Hu, Wei Pan, and S. Liao. 2021. China public psychology analysis about COVID-19 under considering sina weibo data. *Frontiers in Psychology*, 12.

Roy Porter. 1991. Cleaning up the Great Wen: public health in eighteenth-century London. *Medical History*, 35(S11):61–75.

Robert H. Shumway and David S. Stoffer. 2000. *Time Series Analysis and Its Applications*, 1st edition. Springer, New York.

Christopher A. Sims. 1980. Macroeconomics and reality. *Econometrica*, 48(1):1–48.

R. S. Smith, R. D. Davenport, and G. N. Newton. 2020. London weekly bills of mortality, 1644-1849.

P. Strong. 1990. Epidemic psychology: a model. *Sociology of Health & Illness*, 12(3):249–259.

Stella Verkijk, Piek Vossen, and Pia Sommerauer. 2025. Language models lack temporal generalization and bigger is not better. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 20629–20637.

Shihan Wang, Marijn Schraagen, Erik Tjong Kim Sang, and Mehdi Dastani. 2020. Dutch general public reaction on governmental COVID-19 measures and announcements in twitter data.

Zian Zhang, Yixuan Sun, Martha Effting, Iris van der Vegt, and A. J. Ton Klein. 2024. Assessing the utility of ChatGPT for public health surveillance: a case study of social media analysis on heat-health discussions. *Frontiers in Public Health*, 12:1358607.

Shu Zhao, Lin Chen, Yang Liu, Ming Yu, and H. Han. 2022. Deriving anti-epidemic policy from public sentiment: A framework based on text analysis with microblog data. *PLoS ONE*, 17(8):e0270953.