Proceedings of the Eighteenth International AAAI Conference on Web and Social Media (ICWSM 2024)

Analyzing Mentions of Death in Covid-19 Tweets

Divya Mani Adhikari[1], Muhammad Imran[2], Umair W. Qazi[2], Ingmar Weber[1]

[1]Interdisciplinary Institute for Societal Computing (I2SC), Saarland University, Saarbrücken, Germany
[2] Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
diad00001@uni-saarland.de, mimran@hbku.edu.qa, uqazi@hbku.edu.qa, iweber@cs.uni-saarland.de

Abstract

Many researchers have analyzed the potential of using tweets for epidemiology in general and for nowcasting COVID-19 trends in specific. Here, we focus on a subset of tweets that mention a *personal, COVID-related death*. We show that focusing on this set improves the correlation with official death statistics in six countries, while also picking up on mortality trends specific to different age groups and socio-economic groups. Furthermore, qualitative analysis reveals how politicized many of the mentioned deaths are. To help others reproduce and build on our work, we release a dataset of annotated tweets for academic research.

Introduction

During the COVID-19 pandemic, millions of people shared personal updates on social media, thereby providing an up-to-date account of the epidemiological situation. Due to its popularity and (former) ease of access, Twitter (now X) in particular has been used in efforts to nowcast epidemiological statistics and to study public opinion on, say, lock-down measures. We contribute to such efforts by providing an in-depth analysis of tweets mentioning a personal death close to the author. By using data from six countries, we show that focusing on this subset of death-related tweets greatly improves the correlation between tweet counts and official statistics. Furthermore, we show that these tweets contain interesting narratives, worth analyzing in their own right.

While our work is far from the first to look at the relationship between COVID-related tweets and COVID statistics, previous work did not attempt to identify mentions of individual deaths. For example, Cheng et al. (2021) observed a temporal lag between the rise in the number of COVID-19-related tweets and officially reported deaths 6–27 days later. Similarly, Turiel, Fernandez-Reyes, and Aste (2021) found evidence that COVID-19-related tweet intensity per region can forecast the number of deaths, one month later. Both Sarker et al. (2020) and Mackey et al. (2020) studied self-reported COVID-19 symptoms and as well as recovery progression and compared them with those reported in studies conducted in clinical settings. Conceptually, the work by Paul, Dredze, and Broniatowski (2014) and Broniatowski, Paul, and Dredze (2013) is related to ours as they show that for monitoring *flu* activity using tweets processed by a first-person classifier is helpful as it separates tweets discussing the flu in general from first-person reports of the flu.

Here, we analyze mentions of personally related COVID-19 deaths on Twitter. We do so using both regular expression (regex) based filtering as well as custom classifiers for identifying relevant tweets. Our main contributions are:

- We show that filtering for mentions of personal death improves the correlation with official statistics, compared to general COVID-related filtering.
- For the US, we find evidence for not only correlating with overall deaths but also the male-to-female ratio of deaths.
- We describe patterns of shifting Android-to-iOS ratios in tweets mentioning deaths, hinting at shifts in the socio-economic status of people being affected.
- We describe the dominant narratives in tweets mentioning personal COVID-related deaths.
- We release a dataset[*] of tweets mentioning COVID-19 deaths, together with the code to reproduce our results.

Methodology

Dataset The initial collection of tweets based on multilingual COVID-related keywords started in early 2020(Imran, Qazi, and Ofli 2022). These tweets were geotagged using Nominatim[†] based on the user-specified location from the user profile, the textual content of the tweet, or geo-coordinates provided in the tweet. We removed countries with high misclassification by Nominatim such as Colombia (e.g. "Denver, CO" classified as Colombia). The collection was further filtered for English tweets.

Further filtering was applied via regular expressions such as *my * passed*, *my * died*, *my * succumbed*, *lost * battle* to get the tweets that likely mentioned a personally related COVID death. The filtered dataset contained about 1.2 Million tweets from early January 2020 to the end of March 2021 – the end of the underlying data collection. In the following, we use tweets from March 2020 to the end of March 2021, unless specified otherwise.

[*]https://zenodo.org/doi/10.5281/zenodo.10839648
[†]https://nominatim.org/

Country	Total regex filtered death tweets	Total classifier filtered death tweets	Distinct users	Total official deaths	Pearson correlation between tweet counts and official deaths		
					TBCOV	Regex	Classifier
Australia†	1,063	494	392	910 (ABS)	0.081	**0.274**	0.256
Canada†	3,052	1,386	1,143	22,960 (JHU)	0.614*	**0.703****	0.676*
India†	1,737	952	857	162,927 (JHU)	-0.28	0.53	**0.724****
Italy†	421	303	289	109,317 (CPD)	0.391	**0.673***	0.555*
United Kingdom‡	9,636	5,219	3,995	152,704 (UKHSA)	0.697***	0.775***	**0.788*****
United States‡	41,211	25,058	18,854	563,035 (CDC)	0.029	0.771***	**0.79*****

Table 1: Basic data statistics and Pearson correlations with official COVID death statistics for the countries we looked at from March 2020 to March 2021. The regex-filtered tweet counts are after the deduplication steps (†Monthly, ‡Weekly, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Data annotation To annotate if the tweet contained a mention of COVID-related death, we randomly sampled 3k tweets, as well as 15k tweets from the top 15 countries with stratified sampling.[‡] We used Appen[§] as the platform for annotation. $1,790$ annotators participated and were asked to answer the following questions: (i) Does the tweet refer to the COVID-related death of one or more individuals personally known to the tweet's author?; (ii) What is the relationship between the tweet's author and the victim mentioned?; and (iii) Relative to the time of the tweet, when did the mentioned death occur?

Checking for bots Due to the intense discussions surrounding COVID-19, we were worried about bot involvement. To assess this, on Jan 31, 2023, we compared the percentage of active, deleted, and suspended Twitter users in our dataset (79.5%, 13.1%, and 7.4%, respectively) with that in a sample of tweets from April 2020/March 2021 provided by the Twitter Stream Grab project[¶] (74.3%/82.2%, 18.3%/10.9%, and 7.4%/6.9%, respectively). Given that (i) suspension is often linked to bot activity (Pierri et al. 2023), and that (ii) the suspension percentages are similar in our dataset on a generic sample, this suggests that overall bot activity is not unusually prominent in our sample.

(Near) duplicate removal To remove (near) duplicates, e.g., caused by "share this" buttons on news articles, as well as by other semi-automated means, we applied the following pipeline: i) Remove all retweets. ii) Remove tweets containing a URL but w/o attached media. iii) Use MinhashLSH (Broder et al. 1998) to remove any near-duplicates in the remaining collection. We used the *datasketch* (Zhu et al. 2023) library with parameters *threshold=0.8*, *shingle_size=4* and *num_perms=256* for the Minhash LSH. After deduplication, our dataset narrows down to 69k tweets (see Table 2).

Training the classifier To further remove false positives from the regex-filtered dataset, we used the annotations to train a binary classifier to detect tweets containing a mention of personal COVID-19-related deaths. We fine-tuned the COVID-Twitter-BERT v2 model (Müller, Salathé, and Kummervold 2023), which was trained on 97M tweets related to COVID-19. After finetuning for two epochs, we obtained a 10-fold cross-validated average F1-score of 0.81 ($\sigma = 0.04$)[¶]. In the end, the (single-annotator)-(model) agreement ($\kappa = .53$) exceeded the inter-annotator ($\kappa = .42$) agreement in terms Cohen's Kappa. The classifier classified about 40k tweets (out of 69k) as having a mention of a personally related COVID-19 death. To avoid sparsity, we only kept countries for which at least 20 tweets per month were available (ignoring March 2020 and allowing for two months having less than 20 tweets). This filtering left us with a total of six countries, namely: Australia, Canada, India, Italy, the United Kingdom, and the United States. We decided against training a classifier for the author-victim relationship and the relative time label due to highly imbalanced labels.

‡Data annotation was done before sub-setting to six countries.

§https://appen.com/

¶https://archive.org/details/twitterstream

¶Additional details about the hyperparameters and model training are available on the dataset and model release page.

Results

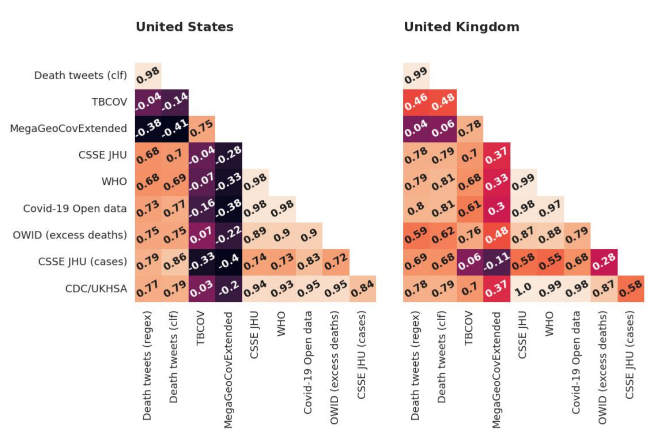

Figure 1: Pearson correlations for the United States and the United Kingdom between different tweet counts and COVID-19 death and confirmed case counts.

Example Tweet	Remarks
Today was a good day #happyday	Tweets w/o COVID-related terms were filtered out during the initial collection.
The number of Covid-related cases in our town is increasing daily. :(#covid19	Tweets w/o a mention of a COVID-related death were filtered out by the classifier.
User reposted *My father died yesterday due to Covid.*	Retweets were removed since we are interested in personally related COVID-19 deaths.
Rory Kinnear: My sister died of coronavirus https://theguardian.com/commentisfree/2020/...	Tweets with (non-media) URLs were removed as most of them quoted and pointed to news articles.
User 1 *My aunt lost her battle with Covid yesterday.* **User 2** *My aunt lsot her battle with covid yesterday.*	Near-duplicate tweets, possibly due to bots, were removed using MinhashLSH as a precaution.
My brother died yesterday due to Covid.	Such tweets remained at the end of our data preprocessing pipeline.
@politician Because of you my brother died of covid	Tweets in response threads were also kept.

Table 2: Examples of different types of tweets. All except the first one are covid-related. But only the last two are examples of unique, personal reports of covid deaths.

Time-series analysis To evaluate whether *death-specific* COVID-19 tweets are better for monitoring mortality trends compared to *all* COVID-related tweets, we correlated counts for different variants of filtered tweets with different official statistics for COVID-related deaths in the six countries (see Table 1). For the US we used the official data from the CDC (National Center for Health Statistics 2023)†, for the UK, the UK Health Security Agency (UKHSA)(UK Health Security Agency 2023), for Italy, the Civil Protection Department (CPD) (Civil Protection Department 2023), and, for Australia, the Australian Bureau of Statistics (Australian Bureau of Statistics 2023). For India and Canada, we use data from the Center for Systems Science and Engineering at Johns Hopkins University (CSSE JHU)(Dong, Du, and Gardner 2020). For the US and the UK, we also used aggregators such as the World Health Organization (WHO) (World Health Organization 2023), and Google Covid-19 Open data (Wahltinez et al. 2022). For excess mortality (US and UK), we used the Our World in Data (OWID) estimates which are based on the Human Mortality Database (Barbieri et al. 2015) and World Mortality Dataset (Karlinsky and Kobak 2021). To see how the regex/classifier filtering affects the correlation, we also used two general geotagged COVID-19 tweets datasets: TBCOV (Imran, Qazi, and Ofli 2022) and MegaGeoCovExtended (Lamsal 2023). For the US and the UK, we computed weekly correlations whereas for the remaining countries, we computed monthly correlations due to data sparsity.

We did not apply any "lags" to the tweets as, based on the data annotations, most tweets that explicitly mention the (relative) time of death are made on the same or the next day. We also do not apply such shifts to the official data as, by the time of our analysis, initial reporting lags had been fixed.

We observed a higher correlation between regex-filtered tweets and official deaths than with general COVID-related tweets (TBCOV and MegaGeoCovExtended). Moreover, using the classifier (clf) on the regex-filtered tweets to further remove false positives often further improved the correlations (see Table 1). Detailed correlation results for the US and UK can be found in Fig. 1. Fig. 2 shows the corresponding time series.

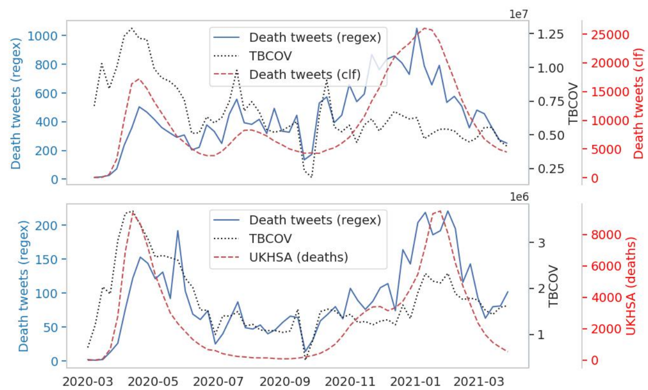

Figure 2: Classifier-filtered tweets and TBCOV against the official deaths from CDC (top) and the UKHSA (bottom).

Gender-based time-series analysis We also wanted to know if death-specific tweets can provide a signal for the trends in *gender-disaggregated* COVID deaths in the US and UK. Due to data sparsity in the annotations, we used a regex-based approach to assign a gender to the victim, such as counting mentions of "mom" or "sister" as a female death. We computed the correlation between the weekly gender disaggregated tweets for the US and the UK against the official weekly deaths by gender for the US (CDC, (National Center for Health Statistics 2023)) and the UK (ONS, (Office for National Statistics 2023)). We ignored March 2020 due to its weeks having less than 20 tweets for the male and female series. For the US, we found a correlation of 0.52 ($p < 0.001$) suggesting a weak signal. For the UK, a correlation of -0.01 ($p = 0.94$) suggested no relationship.

Age group-based time-series analysis Similar to the gender analysis, we used regex-based matching to map terms such as "grandmother", "grandpa" vs. "mother" or "dad" into different generation categories. We computed

†We followed the CDC's Sun-Sat week definition for the US, but otherwise used a Mon-Sun definition.

the weekly correlation between the ratio for the *parents*-to-*grandparents* tweets and the ratio of deaths in the 65+ vs. 35-64 age category (45-64 for the UK). There was only a negligible correlation of 0.138 ($p = 0.33$) for the US and a correlation of 0.169 ($p = 0.23$) for the UK.

Device usage as a socio-economic indicator Fatehkia et al. (2020) showed that the device type used to access social media provides a signal for relative wealth differences where, in general, Apple usage is linked to higher wealth compared to Android usage. Moreover, McGowan and Bambra (2022) showed that the mortality rates for COVID-19 were higher in socio-economically disadvantaged regions in the US. Assuming that, due to homophily and the structure of social networks, the socio-economic status of the tweet author is correlated to the socio-economic status of the mentioned victim. We use the tweet metadata to analyze the devices used, Apple vs. Android, during COVID-19 to look for relationships in who was dying during a given period.

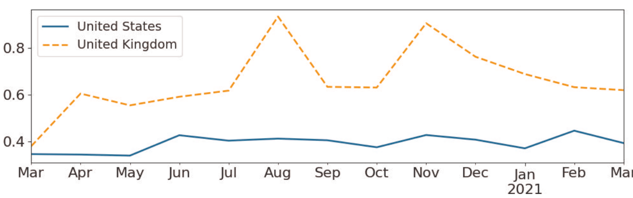

Figure 3: Ratio of Android-to-Apple devices used in tweets mentioning COVID-19 deaths for the US and the UK.

Interestingly, the US trends in Fig. 3 agree with results by Dukhovnov and Barbieri (2021) who found that during the first wave (Mar-May 2020) the most socio-economically advantaged counties (i.e. lowest Android usage) were highly affected. This pattern reversed after June, when we also observe a relative increase in Android usage in tweets mentioning deaths. Unfortunately, for the UK we could not find directly applicable statistics to validate the observed trends.

Qualitative analysis Given that 56% of tweets from a combined US-UK death-related dataset start by mentioning, i.e. typically replying to, another user, we wondered if people were blaming the government, or potentially others for a lack of caution. For this exploratory analysis, we preprocessed the tweets by removing hashtags, emoji, numbers, URLs, and stopwords, and computed the most frequent bigrams in the combined data for the US and UK (Fig. 4). As a proof-of-concept study, we sampled and inspected 50 tweets for each of the following bigrams: "nursing home", "death certificate" and "wear mask".

For "nursing homes", most of the tweets mentioned that an elderly family member died alone in the nursing home due to COVID restrictions. Many of the tweets also blamed a political figure, such as Donald Trump, for the death. The bigram "death certificate" was mostly used in two contexts: (i) in claims that the death certificate falsely listed COVID-19, suggesting that the government tried to keep the case count high, and (ii) in reports of *missing* COVID-19 on the death certificate, despite testing positive, suggesting that the government was trying to keep the case count low. Tweets

Figure 4: Top 100 bigrams for the US-UK combined dataset.

mentioning "wear mask" were mostly supportive of masks, sometimes suggesting that a family member died because someone else refused to wear a mask. Overall, there appeared to be lots of politicization of family deaths, with empathetic tweets such as "@user sorry for your loss. my dad also died of covid-19" in the minority.

Discussion and Conclusion

While our analysis was done in retrospect, we advocate for applying classifier-based filtering in any epidemiological *nowcasting* setting. Beyond strengthening the correlations with official statistics, our filtering strategy could improve temporal stability. For example, Fig. 2 shows a decent fit between the unfiltered TBCOV and CDC death counts until about Oct. 2020. However, after COVID had lost its "novelty", tweet counts no longer closely followed the occurrence of death, whereas the death-specific tweets continued to track the occurrence. This observation mirrors learnings from the failure of Google Flu Trends (Lazer et al. 2014).

Limitations Our analysis is only done on *English* content on *Twitter* and the applicability to other settings needs to be demonstrated. Double-counting is another challenge, as someone's deceased mother could be someone else's aunt. Our results show that despite this limitation, the trends in the filtered timeline are closer to the official data. Finally, both the classifier as well as the Nominatim-based geocoding are noisy. We believe that an improvement here would only further strengthen the overall results. Similarly, broadening the initial regex used for filtering could increase the overall dataset, helping to address sparsity challenges.

Ethical considerations At the individual level, data on specific deaths could be misused for, say, scams. ("I'm your grandpa's illegitimate daughter."). For this reason, we will only share our data with academic researchers upon reasonable request. At the societal level, if a government were to use signals from social media for, say, allocating aid, this could disadvantage communities with reduced social media usage as their suffering would remain hidden. Furthermore, reliance on these signals could attract bots and astroturfing, again favoring better-resourced actors.

Conclusion Our work shows that more data is not always better and that focusing on tweets explicitly mentioning a COVID death is useful both for (i) monitoring trends in COVID mortality in real-time, as well as (ii) picking up on themes in discussion of such deaths. While arguably "obvious" in hindsight, we are unaware of other work exploring this simple yet effective strategy. We acknowledge that applying our insight to X with its more restricted access might be difficult. However, we hope that together with the data and code* we release, our insights can still be transferred to other platforms such as, say, comments on YouTube.

Acknowledgments

We would like to acknowledge Kareem Darwish for their help with creating the data annotation task.

References

Australian Bureau of Statistics. 2023. COVID-19 Mortality in Australia: Deaths registered until 30 November 2023. https://www.abs.gov.au/articles/covid-19-mortality-australia-deaths-registered-until-30-november-2023. [Date accessed 20 Dec. 2023].

Barbieri, M.; Wilmoth, J. R.; Shkolnikov, V. M.; Glei, D.; Jasilionis, D.; Jdanov, D.; Boe, C.; Riffe, T.; Grigoriev, P.; and Winant, C. 2015. Data Resource Profile: The Human Mortality Database (HMD). *International Journal of Epidemiology*, 44(5): 1549–1556.

Broder, A. Z.; Charikar, M.; Frieze, A. M.; and Mitzenmacher, M. 1998. Min-Wise Independent Permutations (Extended Abstract). In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, 327–336. New York, NY, USA: Association for Computing Machinery. ISBN 0897919629.

Broniatowski, D. A.; Paul, M. J.; and Dredze, M. 2013. National and Local Influenza Surveillance through Twitter: An Analysis of the 2012-2013 Influenza Epidemic. *PLOS ONE*, 8(12): e83672. Publisher: Public Library of Science.

Cheng, I. K.; Heyl, J.; Lad, N.; Facini, G.; and Grout, Z. 2021. Evaluation of Twitter data for an emerging crisis: an application to the first wave of COVID-19 in the UK. *Scientific Reports*, 11(1): 19009. Number: 1 Publisher: Nature Publishing Group.

Civil Protection Department. 2023. Italian COVID-19 data. https://github.com/pcm-dpc/COVID-19/blob/master/README_EN.md. [Date accessed 20 Dec. 2023].

Dong, E.; Du, H.; and Gardner, L. 2020. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20(5): 533–534. Publisher: Elsevier.

Dukhovnov, D.; and Barbieri, M. 2021. County-level socio-economic disparities in COVID-19 mortality in the USA. *International Journal of Epidemiology*, 51(2): 418–428.

Fatehkia, M.; Tingzon, I.; Orden, A.; Sy, S.; Sekara, V.; Garcia-Herranz, M.; and Weber, I. 2020. Mapping socioeconomic indicators using social media advertising data. *EPJ Data Science*, 9(1): 22.

Imran, M.; Qazi, U.; and Ofli, F. 2022. TBCOV: Two Billion Multilingual COVID-19 Tweets with Sentiment, Entity, Geo, and Gender Labels. *Data*, 7(1).

Karlinsky, A.; and Kobak, D. 2021. Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. *eLife*, 10: e69336. Publisher: eLife Sciences Publications, Ltd.

Lamsal, R. 2023. MegaGeoCOV Extended.

Lazer, D.; Kennedy, R.; King, G.; and Vespignani, A. 2014. The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 343(6176): 1203–1205.

Mackey, T.; Purushothaman, V.; Li, J.; Shah, N.; Nali, M.; Bardier, C.; Liang, B.; Cai, M.; and Cuomo, R. 2020. Machine Learning to Detect Self-Reporting of Symptoms, Testing Access, and Recovery Associated With COVID-19 on Twitter: Retrospective Big Data Infoveillance Study. *JMIR Public Health and Surveillance*, 6(2): e19509.

McGowan, V. J.; and Bambra, C. 2022. COVID-19 mortality and deprivation: pandemic, syndemic, and endemic health inequalities. *The Lancet Public Health*, 7(11): e966–e975. Publisher: Elsevier.

Müller, M.; Salathé, M.; and Kummervold, P. E. 2023. COVID-Twitter-BERT: A natural language processing model to analyse COVID-19 content on Twitter. *Front. Artif. Intell.*, 6: 1023281.

National Center for Health Statistics. 2023. Provisional COVID-19 Deaths by Week, Sex, and Age. https://data.cdc.gov/d/vsak-wrfu. [Date accessed 20 Dec. 2023].

Office for National Statistics. 2023. Deaths registered weekly in England and Wales, provisional. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/weeklyprovisionalfiguresondeathsregisteredinenglandandwales. [Date accessed 20 Dec. 2023].

Paul, M. J.; Dredze, M.; and Broniatowski, D. 2014. Twitter improves influenza forecasting. *PLoS Curr.*, 6.

Pierri, F.; Luceri, L.; Chen, E.; and Ferrara, E. 2023. How does Twitter account moderation work? Dynamics of account creation and suspension on Twitter during major geopolitical events. *EPJ Data Science*, 12(1): 43.

Sarker, A.; Lakamana, S.; Hogg-Bremer, W.; Xie, A.; Al-Garadi, M. A.; and Yang, Y.-C. 2020. Self-reported COVID-19 symptoms on Twitter: an analysis and a research resource. *Journal of the American Medical Informatics Association: JAMIA*, 27(8): 1310–1315.

Turiel, J.; Fernandez-Reyes, D.; and Aste, T. 2021. Wisdom of crowds detects COVID-19 severity ahead of officially available data. *Scientific Reports*, 11(1): 13678.

UK Health Security Agency. 2023. UKHSA data dashboard - COVID-19. https://api.ukhsa-dashboard.data.gov.uk/. [Date accessed 20 Dec. 2023].

Wahltinez, O.; et al. 2022. COVID-19 Open-Data a global-scale spatially granular meta-dataset for coronavirus disease.

World Health Organization. 2023. WHO Coronavirus (COVID-19) dashboard > Data [Dashboard]. https://data.who.int/dashboards/covid19/deaths?n=c. [Date accessed 20 Dec. 2023].

Zhu, E.; Markovtsev, V.; Astafiev, A.; Ha, C.; Łukasiewicz, W.; Foster, A.; Sinusoidal36; Oriekhov, A.; Halliwell, J.; JonR; Mann, K.; Joshi, K.; Rosenthal, M. J.; TianHuan, Q.; Ibraimoski, S.; Thakur, S.; Ortolani, S.; Titusz; Letal, V.; Bentley, Z.; fpug; hguhlich; long2ice; oisincar; and Assa, R. 2023. ekzhu/datasketch: v1.6.4.

Ethics checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, see the Limitations section. Another artifact, that could not be mentioned in the paper but, that could be in the dataset is multiple mentions of the death of the same person by different users for example a public figure which could cause an overestimation of the number of deaths based on the tweets.
   (e) Did you describe the limitations of your work? Yes, see the Limitations section.
   (f) Did you discuss any potential negative societal impacts of your work? Yes, see the Ethical Considerations section.
   (g) Did you discuss any potential misuse of your work? Yes, see the Ethical Considerations section.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, in the Ethical Considerations section we mention that the dataset and the model will be released with access control to academic researchers on a reasonable request.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? NA
   (b) Have you provided justifications for all theoretical results? NA
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
   (e) Did you address potential biases or limitations in your theoretical framework? NA
   (f) Have you related your theoretical results to the existing literature in social science? NA
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? No, due to space constraints. These details are available on the dataset and model release page on Zenodo. See: https://doi.org/10.5281/zenodo.10839649
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? Yes, see the Training the classifier section.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? No, due to space constraints. These details are available on the dataset and model release page on Zenodo. See: https://doi.org/10.5281/zenodo.10839649
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes, see the Training the classifier section.
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? No, due to space constraints. One possible "cost" of misclassification could be the overestimation of deaths by the classifier for a certain region. If the policymakers were to use this for decision-making, it could deprive other regions of the necessary resources.
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes
   (b) Did you mention the license of the assets? No, since the dataset isn't publicly available.
   (c) Did you include any new assets in the supplemental material or as a URL? Yes, we include the dataset and the code to reproduce the experiment as supplemental material.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? No, since we are analyzing publicly available Twitter data. The sheer volume of the tweets also makes it difficult and using the data of only users who consented could introduce bias to the data.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes, due to the possibility of PII in the tweets, we only release the dataset on reasonable request from academic researchers.

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles/)? No, not in the paper but we have made our dataset available on Zenodo following the FAIR guidelines. See: https://doi.org/10.5281/zenodo.10839649

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. Communications of the ACM, 64(12): 86–92)? Yes, we created a Datasheet for datasets which is available along with the released dataset on Zenodo. See: https://doi.org/10.5281/zenodo.10839649

6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...

(a) Did you include the full text of instructions given to participants and screenshots? Yes, we include a summary of the instructions in the Dataset section and the full instructions are available on the dataset release page. See: https://doi.org/10.5281/zenodo.10839649

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? No, not in the paper due to space constraints. The description of the instructions for annotation to the human annotators on Appen started with: 'CONTENT WARNING: The task includes going through mentions of death, and might cause emotional discomfort.'. We feel that this warning adequately describes potential risks to the annotators.

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? No, due to space constraints in the paper. The annotators were paid approximately $15 per hour. We spent a total of $116 and $471.38 for the 3k and 15k annotation jobs on Appen respectively.

(d) Did you discuss how data is stored, shared, and deidentified? Yes, some in the Ethical Considerations section. Appen anonymized the annotator's data before we accessed it. The dataset is stored in the Zenodo repository and will be only shared with academic researchers on a reasonable request.