Stringent containment and closure policies have been widely implemented by governments to prevent the transmission of COVID-19. Yet, such policies have significant impacts on people’s emotions and mental well-being. Here, we study the effects of pandemic containment policies on public sentiment in Singapore. We computed daily sentiment values scaled from −1 to 1, using high-frequency data of ~240,000 posts from highly followed public Facebook groups during January to November 2020. The lockdown in April saw a 0.1 unit rise in daily average sentiment, followed by a 0.2 unit increase with partially lifting of lockdown in June, and a 0.15 unit fall after further easing of restrictions in August. Regarding the impacts of specific containment measures, a 0.13 unit fall in sentiment was associated with travel restrictions, whereas a 0.18 unit rise was related to introducing a facial covering policy at the start of the pandemic. A 0.15 unit fall in sentiment was linked to restrictions on public events, post-lockdown. Virus infection, wearing masks, salary, and jobs were the chief concerns found in the posts. A 2 unit increase in these concerns occurred even when some restrictions were eased in August 2020. During pandemics, monitoring public sentiment and concerns through social media supports policymakers in multiple ways. First, the method given here is a near real-time scalable solution to study policy impacts. Second, it aids in data-driven and evidence-based revision of existing policies and implementation of similar policies in the future. Third, it identifies public concerns following policy changes, addressing which can increase trust in governments and improve public sentiment.

COVID-19 | containment policies | public sentiment | social media data | causal analysis

Since January 2020, governments across the globe have been using various policy instruments to inform and protect their residents during the COVID-19 pandemic. From lockdown to slow resumption of social and economic activities during various phases, policies are being designed and implemented by governments with varying impacts. The Oxford COVID-19 Government Response Tracker collects systematic information on policy measures that more than 180 countries have taken to tackle COVID-19 since January 2020. This tracker divides pandemic-related government policies into four major categories (i.e., healthcare, containment and closure, economic, and miscellaneous policies) (1). Among these categories, containment and closure policies are a common tool, which impose restrictions on people’s daily life in order to prevent disease transmission. Typical examples of these policies include stay-at-home orders, social distancing requirements, closing of schools, workplaces, and recreation venues, as well as travel restrictions. While containment policies are generally effective in reducing the number of COVID-19 cases (2), significant adverse impacts are seen on mental health and mental well-being. Knowing the impacts of these policy interventions on people’s mental well-being and sentiments is valuable for evidence-based policymaking under fast-changing pandemic conditions.

Analysis of social media data, such as through sentiment analysis, can help policymakers discern people’s sentiments and concerns, thereby increasing policymakers’ understanding of policy impacts on social phenomena (3, 4). In the context of the COVID-19 pandemic, several studies have used social media data to highlight public concerns (5–8) and gauge sentiments from user comments (9, 10). Additionally, some studies used social media data to examine stress, panic, and other psychological consequences (11–15), as well as the impacts of online misinformation (16, 17) and government outreach (18) during the pandemic. However, we did not find research about the impacts of government policies for handling the pandemic on public sentiment over time. We aim to add to the prior body of work in the following ways. First, our study extends prior research, which typically covered a few months. This duration may not be sufficient, considering the changes in policies as the pandemic continues for longer.

Second, we add to earlier studies that uncovered topics discussed from social media posts, by indicating how the topics (in our case concerns) changed in response to policy

Significance

For effective pandemic response, policymakers need tools that can assess policy impacts in near real-time. This requires policymakers to monitor changes in public well-being due to policy interventions. Particularly, containment measures affect people’s mental well-being, yet changes in public emotions and sentiments are challenging to assess. Our work provides a solution by using social media posts to compute salient concerns and daily public sentiment values as a proxy of mental well-being. We demonstrate how public sentiment and concerns are impacted by various containment policy sub-types. This approach provides key benefits of using a data-driven approach to identify public concerns and provides near real-time assessment of policy impacts by computing daily public sentiment based on postings on social media.

Author affiliations: Institute of Systems Science, National University of Singapore, 119615 Singapore; and Department of Information Systems and Analytics, National University of Singapore, 117417 Singapore

Author contributions: P.C.S. and A.K. designed research; P.C.S. and A.K. performed research; P.C.S. analyzed data; and P.C.S. and A.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

To whom correspondence may be addressed. Email: prakashc@nus.edu.sg.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2117292119/-/DCSupplemental.

Published May 3, 2022.
interventions over time, using robust methods. Third, we identify the effects of various types of pandemic policies, which was unexplored. For example, containment and closure policies comprise eight subtypes, whose implementation varies across the pandemic period (1).

In this article, using daily data collected from popular and public Facebook groups, we studied the impacts of containment policy subtypes on average daily sentiment and public concerns, reflected in ~240,000 posts in Singapore over a period of 11 mo from January to November 2020. Singapore is valuable to study, as it has been lauded for its handling of the COVID-19 pandemic (19). However, mental well-being issues were seen, with the “zero-case” containment approach (where the goal is to prevent even a single case of COVID-19) being adopted during our study period. We used robust methods to demonstrate the impact of containment policy interventions on public sentiment and on key concerns that emerged and sometimes persisted during the pandemic. First, using rigorous econometric and machine learning techniques, we estimated the causal relationships between containment policy changes and public sentiment observed on social media, while controlling for concurrent health care (e.g., facial covering mandate) and economic (e.g., income support) policies, as well as pandemic (e.g., daily case counts), and economic indicators (e.g., unemployment rate). Second, we identified various phase-specific and overarching public concerns expressed on social media using natural language processing (NLP) techniques, and their changes when policies changed. These concerns helped us to explain changes in public sentiment, as containment policies changed during the pandemic. This work is of direct relevance to policymakers because it presents a robust, scalable way to use social media data to study policy impacts and thereby implement evidence-based policies for pandemic management.

Results

To uncover the relationships between implementation of containment policies, daily sentiment, and public concerns, we employed a multistep approach. We first collected data from multiple sources and constructed our variables. Our social media data consisted of user posts from seven relevant and most highly-followed public Facebook pages in Singapore. These pages were identified from a social media benchmarking site (20). The pages were selected based on the number of active users, and to cover a mix of government, news, and community groups. The government pages comprised Gov.sg (417,285 followers) and the Prime Minister’s Facebook page (1,663,586 followers). We also included pages of three news agencies (i.e., the Straits Times [1,587,083 followers], TODAYonline [1,006,628 followers], and the New Paper [697,856 followers]). Last, we included pages of two community groups (i.e., Mothership.SG [697,856 followers] and Singapore Kindness Movement [104,855 followers]). A total of 240,593 posts from these pages for the study period of January to November 2020 form our Facebook data. We cleaned and preprocessed the text in the posts, following which we computed the sentiment scores from the text using an NLP tool, in order to obtain our main dependent variable (i.e., daily average public sentiment).

To account for extraneous factors that could influence public sentiment, we extracted data for our control variables from multiple sources (1, 21, 22). The first set of control variables consisted of policy indicators related to healthcare, containment and closure, economic, and miscellaneous policies (1). The second set comprised pandemic indicators (i.e., daily new case counts, daily COVID-19-related death counts, and daily positive rate of infection) (21). Third, we included economic indicators of retail sales index, consumer price index (CPI), unemployment rate, and inflation rate (22). The remaining control variables were user-defined and included lagged values of sentiment. Our independent variable for regression analyses was days elapsed. The analyses were performed by phase, where phases were defined by containment policy changes. The phase indicator had four values in the study period. The first phase was prelockdown (PtLD; January 15–April 6, 2020), when there were no containment policies. This was followed by the lockdown (LD) phase (April 7–June 1, 2020) called circuit breaker in Singapore, when offices, schools, shops, and leisure venues were closed, and only essential services were permitted. The postlockdown period was divided into two phases i.e., a period of initial reopening (PtLD1; June 2–August 3, 2020), and a period of further relaxation (PtLD2; August 4–November 30, 2020) when larger social gatherings were permitted. For causal analysis of the relationship between sentiment scores and independent/control variables we further aggregated these data at a daily average level. To obtain our final dataset (SI Appendix, section 1A), we computed correlations among independent and control variables and removed variables that could lead to collinearity for subsequent analysis (SI Appendix, section 1B).

We then regressed our dependent variable with the independent and remaining control variables. For each phase we used multiple linear regression (MLR) to identify the policy subtypes that are related to public sentiment. Subsequently, we used a regression discontinuity design (RDD) with time (days elapsed) as running variable, to determine the causal impact of change in containment policies (at the phase boundaries) on average daily sentiment levels. Last, we employed a bidirectional encoder representation from transformers (BERT)-based NLP model to identify key concerns of the public for the various phases, and used RDD to identify significant changes in the concerns at the phase boundaries. The results highlight the key policy measures that influenced sentiment, and the salient public concerns. Details of the steps are reported in the Materials and Methods. These techniques together form our approach (Fig. 1) that can aid policymakers in performing robust policy impact studies.

Estimating the Relationship between Policy Measures and Public Sentiment.

We used MLR to predict the daily sentiment values for each of four phases (PtLD, LD, PtLD1, and PtLD2) based on the independent variable, the policy indicators mentioned earlier, and the other control variables. The regression models for the four phases (shown in Table 1) are statistically significant, explaining between 31.6–41% of the variation in sentiment for the four phases. This indicates that our regression models work adequately in explaining the relationships between the policy measures and public sentiment for all phases. The sentiment plot for all phases exhibits a quadratic, inverted U relationship with days elapsed (Fig. 2A). However, the quadratic term for days elapsed was only statistically significant in phase PtLD1 (Table 1). The inverted U relationship in this phase indicates that the sentiment first increased with days elapsed, then started to decline after peaking.

International travel controls (including quarantine on return) were in place since March 21, 2020.* Such controls posed inconvenience to travelers and can help explain the negative relationship between this containment policy measure and COVID-19-related trends in sentiment.

*https://www.moh.gov.sg/news-highlights/details/gradual-re-opening-of-travel-and-changes-to-border-measures

https://doi.org/10.1073/pnas.2117292119

pnas.org
sentiment during phase PrLD (Table 1). Cancellation of public events during the PtLD2 phase showed a negative relationship with sentiment (Table 1), which could be because the policy relaxation in this phase only permitted marriages and religious activities in controlled environments, while most other events still remained virtual (23). The positive relationships of income support economic policy with sentiment during the LD phase, and for facial covering healthcare policy with sentiment during both PrLD and LD phases (Table 1), suggest positive responses of the public to the government’s economic stimulus and mask policy for health safety, respectively. These results clearly show the opposing relationships of various economic, health, and containment policy subtypes with sentiment.

**Estimating Change in Sentiment Due to Containment Policy Interventions.** While the MLR analysis above showed us relationships between various policy measures and sentiment during phases, this technique does not indicate causal effects. Thus, we performed RDD analysis to identify the causal effects of phase changes on sentiment, when the lockdown was introduced (PrLD to LD), and when it was partially lifted in two phases (LD to PtLD1 and PtLD1 to PtLD2). We observed that the median sentiment changed with these containment policy measures/phases (Fig. 2B). At phase boundaries, we computed the discontinuity values (i.e., D1 for moving from PrLD to LD, D2 for moving from LD to PtLD1, and D3 for moving from PtLD1 to PtLD2) using the RD design (Fig. 2C).

We found that the implementation of lockdown on April 7, 2020 increased the average daily sentiment by 0.1 units (SI Appendix, section 2A), which indicates that people responded positively to the policy. Later, upon partial lifting of lockdown on June 2, 2020 there was another significant rise in average sentiment value by 0.2 units. Post August 4, we saw a significant decline of sentiment value by −0.15 units, despite further easing of restrictions, which seems counterintuitive. However, this could be due to the concerns regarding unemployment, on-going fear of infection, and restrictions on dining out at the time. We are able to make such inferences as we identified the major concerns during each phase using the topic modeling NLP approach and related them to the policy changes, as explained in the next two sections.

**Identifying Key Public Concerns and Their Changes across Containment Policy Phases.** Using BERT, we identified the prominent topics for each phase from the ∼240,000 posts. These topics reflect the most frequent concerns expressed on the highly popular Facebook pages (Fig. 3A). A mapping of the key topics to associated words found in posts can be seen in the SI Appendix, section 2B. The two major concerns present in all phases were virus infection, and wearing masks. Further, at the last phase boundary (PtLD1 to PtLD2), both the concerns rose by nearly 2 units (Fig. 3B and C). Such a large increase in these concerns could be a valuable indicator for administrators about the fears felt during the phase change.

**Table 1.** MLR results for relationships between days elapsed and policy variables on sentiment for the four phases: PrLD, LD, PtLD1, and PtLD2

| Days elapsed and policy variables/phase | PrLD               | LD                | PtLD1             | PtLD2              |
|----------------------------------------|--------------------|-------------------|-------------------|--------------------|
| Days elapsed                           | −0.001 (0.001)     | −0.007*** (0.001) | 0.077*** (0.025)  | −0.002*** (0.001)  |
| C8_International travel controls       | −0.128*** (0.025)  |                   | 0.101** (0.048)   |                   |
| E1_Income support                      | 0.177*** (0.063)   | 0.039** (0.016)   |                   | −0.0002*** (0.0001)|
| H6_Facial coverings                    |                    |                   |                   | −0.150*** (0.038)  |
| Days²                                  |                    |                   |                   | −0.048*** (0.016)  |
| C3_Cancel public events                | 0.749*** (0.072)   | 1.357*** (0.165)  | −9.448*** (3.022) | 1.090*** (0.247)  |
| Constant                               | 83                 | 56                | 63                | 118                |

Only significant results are shown. Standard Errors in brackets. **P < 0.05; ***P < 0.01; C8, C3, E1, H6 are types of containment policies, economic policies, and healthcare policies, respectively as per ref. (1). LD, lockdown; PrLD, prelockdown; PtLD1, postlockdown1; PtLD2, postlockdown2.
Analyzing by phase, at the start of the pandemic (PrLD) people were most concerned about two topics: 1) face masks and sanitizers, and 2) schools, parents, and students. This could be linked to the shortage of masks and sanitizers in February 2020 and the closure of schools (moving to full home-based learning) in the first week of April 2020 (24). We saw additional concerns surfacing during the LD phase (i.e., salaries and jobs, fines and enforcements, stay-at-home, virus infection, frontliners [including doctors and nurses], and foreign worker dormitories). The concern about salaries and jobs first surfaced during this phase, then declined by 1.7 units at the start of PtLD1, but rose again by 2.4 units in the PtLD2 phase (Fig. 3D). This concern captures people’s fear of losing jobs and the inability to find new jobs in a pandemic-stricken economy.

During the next (PtLD1) phase, people continued to have concerns about health of frontline workers, people not wearing masks, and salary and jobs. Additionally, new concerns surfaced about suicide and depression, and opening borders with Malaysia. Malaysia is Singapore’s only land neighbor—with significant cross-border goods and manpower movement. The limited opening of the border in August 2020 raised concerns about the spread of the virus across the two countries, as well as the
rule for Malaysians working in Singapore to stay for a minimum 3-mo period, instead of commuting back and forth (25). Somewhat surprisingly, the topic of suicide and depression emerged after more relaxations were declared in PtLD1, possibly indicating that people were anxious and depressed about the severity/duration of the pandemic, and what will happen next. The concern about depression rose further in the final (PtLD2) phase by 1.91 units (Fig. 3E). Such an increase indicates that people may need counseling, information, and reassurance from administrators on what to expect next. The salient concerns persisted during the PtLD2 phase (i.e., suicide and depression, masks, and salaries and jobs), with a new concern emerging about food and restaurants (Fig. 3A). The topic of food and restaurants indicates that people were looking forward to eating out again after months of restaurant closures.

It is noteworthy that throughout the four phases, the topic of leadership, pride, and admiration was salient. This indicates that overall, the public viewed government leadership positively and admired their leaders’ handling of the pandemic. This could have been further reinforced by the global recognition that Singapore received for its effective pandemic management policies and low fatalities (19).

**Relating Topics/Concerns Uncovered to Change in Public Sentiment during Policy/Phase Changes.** To understand the reasons for the changes in public sentiment when moving from one phase to the next, we analyzed the relationships between the frequencies of major topics/concerns discussed in the social media posts and the sentiment changes. At the start of the pandemic, the PrLD phase saw an initial fall in sentiment, followed by a rise around March to April (Fig. 2A) once the lockdown was announced for April 3, 2020. Masks were the dominant topic in this phase and rose in frequency with sentiment values, as people felt safer with the mask mandate. The second major
concern in this phase, virus infection was negatively associated with sentiment. After this, the LD phase saw a consistent decline in daily sentiment (Fig. 2A). The dominant concern around salaries and jobs rose as the sentiment values declined. Initially the topics of frontliners, stay-at-home, and dormitories were very common but gradually saw a decline in frequency. The PtLD phase saw a rise in sentiment initially (Fig. 2A). As the sentiment rose, the discussions on masks and virus infection went down in frequency. People were seen to praise government leadership (the dominant topic in this phase), which rose with sentiment. The PtLD2 phase saw a decline in sentiment (Fig. 2A). The topic of virus infection spread was negatively associated with sentiment, suggesting an on-going fear of catching the virus. The topic of depression, which was dominant in this phase, declined gradually. Also, the topic of food and restaurants declined with sentiment. Thus, we were able to utilize the topics from our NLP analysis to understand the changes in sentiment when moving from one phase to the next.

Discussion

In this study, we present a near real-time, multistep approach, comprising of statistical regression and NLP machine learning methods to study pandemic containment policy impacts on public sentiment, and the concerns expressed on social media associated with the policy changes. Using the approach, we were able to determine and quantify how various containment policy measures impacted public sentiment. Further, we identified the salient public concerns in phases demarcated by policy changes, and computed significant changes in the concerns across phase boundaries. The results provide rich information to policymakers on the impacts of their containment policies in the presence of multiple covariates, which facilitate policy revisions toward improving public sentiment, and implementation of similar policies in future.

Specifically, we analyzed ~240,000 posts on the most-followed Facebook pages in Singapore spanning 11 mo from the start of the pandemic. Using multiple regression analysis, we uncovered key policy types that had significant negative (international travel controls and cancelling public events) and positive (income support and facial covering) associations with public sentiment during various phases of the study period. Our RDD analysis then determined the impact of the lockdown and subsequent gradual reopening phases on public sentiment. An increase in the average public sentiment immediately after the lockdown was related to people’s desire for containment measures in order to prevent the virus spread. Another rise in average public sentiment upon the partial lifting of lockdown indicated that people, on average, were positive about the relaxation of some restrictions. Subsequently, the average sentiment decreased with the further easing of restrictions in August, which appears counter intuitive at first but can be understood by observing the on-going concerns about jobs, masking, dining out, and depression. We identified the salient public concerns during each phase of the study using BERT NLP model. These concerns and the significant changes in their levels due policy changes (computed using RDD) could be an important input for policymakers.

However, the study findings need to be considered in light of its limitations. First, the model variables were selected via literature reviews, but there could be additional factors influencing public sentiment. To address this issue, we performed RDD in time (RDIT) analysis at the policy intervention boundaries with multiple control variables, while simultaneously controlling for unobservable factors. By incorporating a
time trend, we accounted for time variant confounding factors, such as seasonality or the pandemic evolution, as long as these factors are modeled by the RDD function. Second, the sentiments reflected on social media may not be fully representative of the population. To address this issue, we sampled data from the highest-followed Facebook groups over a time period of nearly a year. This issue is also alleviated by the high degree of internet penetration (88.5%) and the widespread use of social media like Facebook (82%) in Singapore. A future extension of this study can consider using data from multiple social media platforms and offline sources. Further, misinformation campaigns may deliberately target highest-followed Facebook groups that we have focused on. Future research can include protocols for identifying and eliminating misinformation campaigns during the data preprocessing phase of the analysis. The third limitation arises from the absence of a completely randomized experiment. To address this issue, we used RDD, which allows observations to derive from different subjects before and after the policy intervention, and can uncover causal relationships (26, 27). In conclusion, this study provides policymakers with a tool to study pandemic containment policy impacts on public sentiment in social media in near real-time, and to understand the public concerns related to the changes in policies and sentiment. Overall, it proposes a people-centric, data-driven, and evidence-based approach for fine-tuning existing policies and implementing similar policies in future.

Our study has significant strengths in that we present a methodology to study real-time, high-frequency changes in public sentiment and concerns, in light of changing containment policies—by combining multiple methods such as MLR, RDD, and state-of-art NLP technique, BERT. Further, our method is scalable and general enough to be adopted in other similarly managed geographies with internet savvy populations (e.g., Hong Kong, Taiwan, and Brunei). The approach can be incorporated into the policy formulation process to design similar policies in future. A direct practical advantage of this methodology is that government bodies can plan appropriate interventions when they start to see a rise in levels of negative public sentiment or concerns, under conditions where restrictive containment and closure policies are unavoidable.

Materials and Methods

Once we collected the Facebook posts from the seven highly popular Facebook pages, we preprocessed the text by segmentation of sentences, tokenization, part-of-speech tagging, lemmatization, and stop-word removal. We then computed the daily public sentiment scores from the text using the VADER NLP library in Python, to obtain our main dependent variable (DV), daily average public sentiment. Subsequently, we applied the MLR, RDD, and topic modeling NLP techniques on the data as detailed below. More information about our processed data and analysis code can be found in the Data Availability section.

MLR. After preprocessing the posts and computing the daily sentiment values, we used MLR to predict the average value of daily sentiment with respect to the independent variable (IV) days elapsed, and the control variables (CVs) (SI Appendix, section 1A). Daily mean values of all variables were used for MLR analysis, and a separate model was fitted to each phase. Using MLR we were able to identify the containment policy subtypes, which significantly predicted sentiment values during each of the 4 phases. The functional relationship explored is shown below:

\[
\text{Sentiment} = f(\text{policy, pandemic, economic, user defined variables})
\]

RDD in Time. We used a RDD to assess whether the policy interventions/changes caused a statistically significant change in levels of average sentiment.

†https://www.statista.com/statistics/284466/singapore-social-network-penetration/
An RD design is appropriate here because: 1) treatment in a RD design is deterministic based on cutoff set on the running variable, which is time (days elapsed) in our case, and 2) it only assumes that the individuals before and after an intervention are similar (as in our case) to establish a causal relationship between the DV and running variable (26, 27). The model is given below:

$$Y_t - \beta_0 + \beta_1 D_t + h(t) + \delta + \epsilon_t.$$  

Where, Yt is the daily average sentiment values, Dt is the magnitude of discontinuity caused by variables other than the running variable. This adds additional flexibility in curve fitting, it also poses a risk of overfitting. We examined outputs with polynomials of orders 2/3/4, and stopped increasing the order when the model ceased to improve in terms of adjusted R-squared. Since the adjusted R-squared did not improve beyond order 2 for any of the RDit implementations (SI Appendix, section 2A), we stopped at order 2. As seen in SI Appendix, Table S3 (SI Appendix, section 2A), D1 was best fitted by a second order polynomial, while D2 and D3 were best fitted by first order polynomials (Fig. 2C). Second, we addressed the concern of the presence of auto regressive (AR) component in RDIT using auto.arima() method from the forecast package. Third, the sorted observations check of the running variable (McCray test for density) for conventional RD designs is not relevant here, since the values of our running variable, time/days elapsed, are inherently sorted.

**Topic Modeling using BERT.** We used BERT to identify the key concerns from Facebook posts, surfacing during various phases of the pandemic from January to November 2020. The BERT model by Google AI (31) was chosen because it has outperformed existing state-of-art models on multiple NLP tasks, such as question answering and general language understanding evaluation benchmark. By applying semantic similarity between word vectors, BERT was seen to perform better than existing popular techniques such as latent Dirichlet allocation and probabilistic latent semantic analysis in terms of informativeness and representativeness of the topics (32).

For all four pandemic phases described earlier, we selected the top 10 concerns by frequency, which are related to COVID-19. This was followed by an analysis of changes in frequencies of the key concerns at the phase boundaries. This allowed us to estimate the causal impact of policy interventions and associate them with the change in concerns.

**Ethics Statement.** The data from Facebook posts was publicly available and was collected using Facepager (33), which uses the Facebook Graph application programming interface (34). We used only publicly available data and did not collect data from users with privacy restrictions for our research. We abided by the terms, conditions, and privacy policies of Facebook. We did not seek ethical approval because all the data were preexisting.

**Data Availability.** Anonymized data have been deposited in GitHub (https://github.com/prakashshukwal/covid_policy_impacts) (35). The economic indicators were purchased from The Global Economy (https://www.theglobaleconomy.com/), and can be provided for verification. The Facebook posts were from public groups and were collected using Facepager (https://github.com/strohne/Facepager/), which uses the Facebook Graph application programming interface (https://developers.facebook.com/docs/graph-api/overview).
30. Y. Chen, A. Whalley. Green infrastructure: The effects of urban rail transit on air quality. *Am. Econ. J. Econ. Policy* **4**, 58–97 (2012).

31. J. Devlin, M. W. Chang, K. Lee, K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv Preprint* (2018). arXiv:1810.04805.

32. D. Angelov. Top2vec: Distributed representations of topics. *arXiv Preprint* (2020). arXiv:2008.09470.

33. J. Jünger, T. Jakob, Keyling. Facepager. An application for automated data retrieval on the web (2020) https://github.com/strohne/Facepager/. Accessed 28 April 2022.

34. Facebook. Using the graph API (Facebook, Menlo Park, CA, 2013) https://developers.facebook.com/docs/graph-api/overview. Accessed 28 April 2022.

35. P. C. Sukhwal. COVID19 policy impacts [Data set]. https://github.com/prakashsukhwal/covid_policy_impacts. Accessed 28 April 2022.