Biased cancer-related tweets by congress members in the United States: implications for combating cancer disparities in the social media age

Jason T.-H. CHEUNG[1], Yabin YIN[2], Peter Keumseok KOH[3]

[1]Department of Public Health Sciences, The University of Chicago, Chicago, IL 60637, USA
[2]Division of Social Science, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR
[3]Department of Geography, The University of Hong Kong, Pok Fu Lam, Hong Kong Island, Hong Kong SAR

THE UNIVERSITY OF CHICAGO BIOLOGICAL SCIENCES
AT THE FOREFRONT OF MEDICINE®

INTRODUCTION & OBJECTIVES

Policymakers increasingly use social media platforms such as Twitter to release public statements to communicate representations and policy agendas. **We investigated to what extent partisanship influences attention to cancer-related topics on Twitter among congress members in the United States (US) and whether the attention is modifiable.**

DATA

DATA: Congress members' tweets about cancer from Jan 2010 to Mar 2022.

DATA SOURCE: i) The US politician tweets from 2008 to 2017 collected by a company named "Lobbyists4America" (Olson, 2020). **ii)** A project named "Tweets of Congress", which collects the daily tweets of congress members from 2017 to the present and is still updating daily. The data was provided on GitHub (Litel, 2022).

DATA PROCESSING: We combined the two datasets, and matched the tweets with the politicians' bio information such as gender, party, race, state, etc. based on Twitter screen names. The bio information was obtained via ProPublica API in the R package "*ppcong*". We limited the tweets to those containing keywords: "cancer", "oncology", "malignancy", "tumour". 17906 tweets were retrieved and analyzed.

METHODS

1. Keyword Approach

We used keywords such as "breast", "lung" and "stomach" to match tweets containing these words, producing tweets commenting on different types of cancers. Different types of cancer have very different incidence rates among races. We exploit such divergency to examine whether POC cancers (i.e., cancers that have higher incidence rates among people of color) are paid less attention on tweets by congress members. 5310 tweets have mentioned at least one keyword.

2. Structural Topic Model (STM)

However, the keyword approach only matched a small subsample of tweets (n=5310). We thus used Structural Topic Model to extract topics from cancer tweets and further examined how policymakers comment on cancer.

Pre-processing: First, we did traditional text cleaning, removed punctuations, digits, URLs, and stop words. Then we tokenized the text and kept unigrams and bigrams, and excluded tokens that appeared not more than 3 times.

Search K: To evaluate how many topics are the most ideal, we tried five models with the number of topics equal to 30, 35, 40, 45, 50, respectively. Then we evaluated the five models by simultaneously comparing semantic coherence (i.e., measures whether similar words are more likely to appear in the same topic) and exclusivity (i.e., measures whether topics are exclusive to each other). We found that the two variables were simultaneously optimal when K = 50.

STM: After identifying that the best K is 50, we ran a STM, setting K = 50, and added topic prevalence, party and year-quarter. We estimated the effect of party on topic prevalence. After manual coding and organizing, we finally produced 16 broad topics. Next, we assigned these 16 topics to each document. The rule is to assign the topic with the highest probability to a document.

MAIN TAKEAWAYS

Democrats published more cancer-related tweets than Republicans. Congress members tweeted more about cancers with higher incidence rates among Whites. A cancer type with higher incidence rates among Blacks was associated with significantly fewer tweets about that cancer. Democrats tweeted more about topics concerning cancer disparities (e.g., healthcare, environment, and minorities). Tweeting RMC topics was positively associated with tweets published after BLM, but inversely associated with tweets published during the COVID-19 pandemic, suggesting the differential attention is modifiable by external societal events. Understanding how to modify policymakers' attention to cancer may be helpful in addressing cancer disparities in the US.

RESULTS – KEYWORD APPROACH

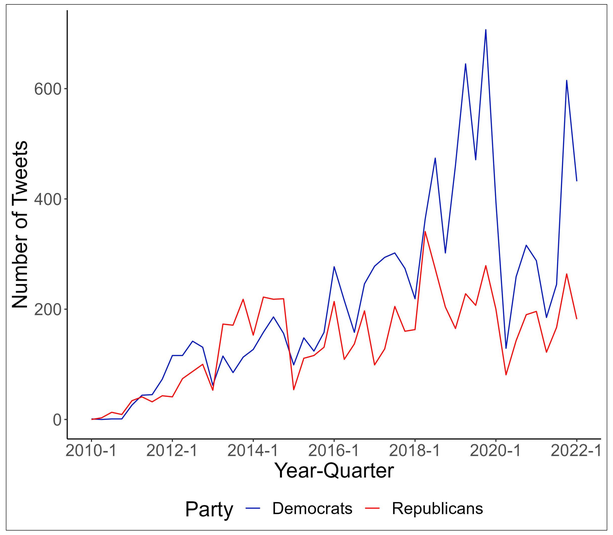

Figure 1. Trend of policymakers' tweets by party.

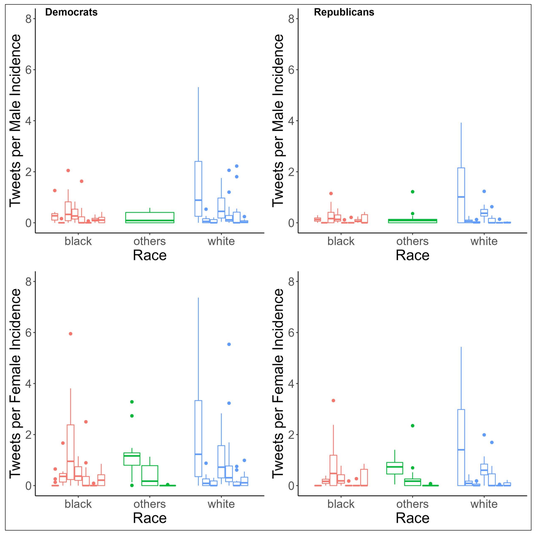

Figure 2. Boxplots of number of tweets by cancer type.

RESULTS – KEYWORD APPROACH (REGRESSION)

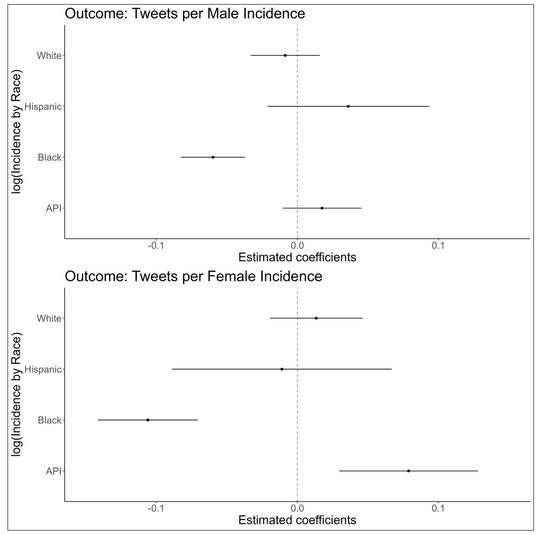

Figure 3. Associations between cancer incidence rates and number of tweets.

RESULTS – STRUCTURAL TOPIC MODEL & ATTENTION MODIFIABILITY

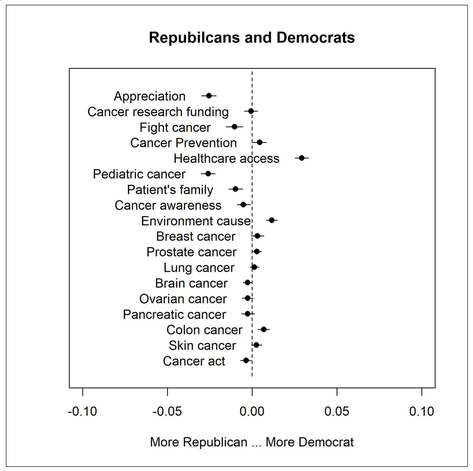

Figure 4. The effect of party on topic prevalence.

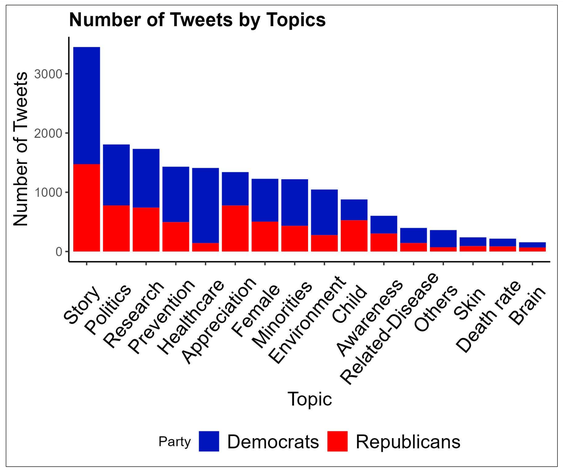

Figure 5. Number of tweets by topic by party.

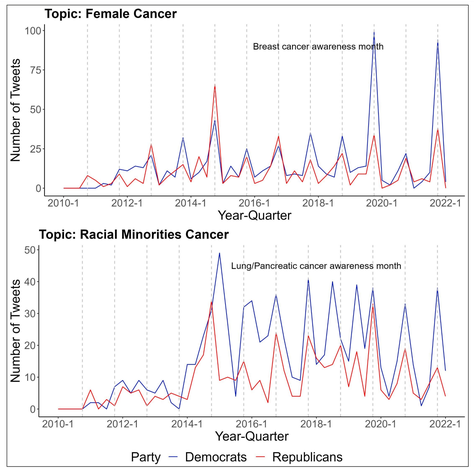

Figure 6. The validation of topic classification.

To determine whether our classification is reasonable, we validated it by plotting the trend of two topics: **"female cancer"** and **"racial minorities cancer" (RMC)** from 2010 to 2022 (i.e., the unit is the number of tweets of two topics by party by year-quarter). We annotated the time of breast cancer awareness month (every October) and that of lung/pancreatic (higher incidence rate among black people) cancer awareness month (every November) with grey dashed lines on the plots. We found the number of related tweets peaked at the time of the awareness months, which to some extent implies our model is valid.

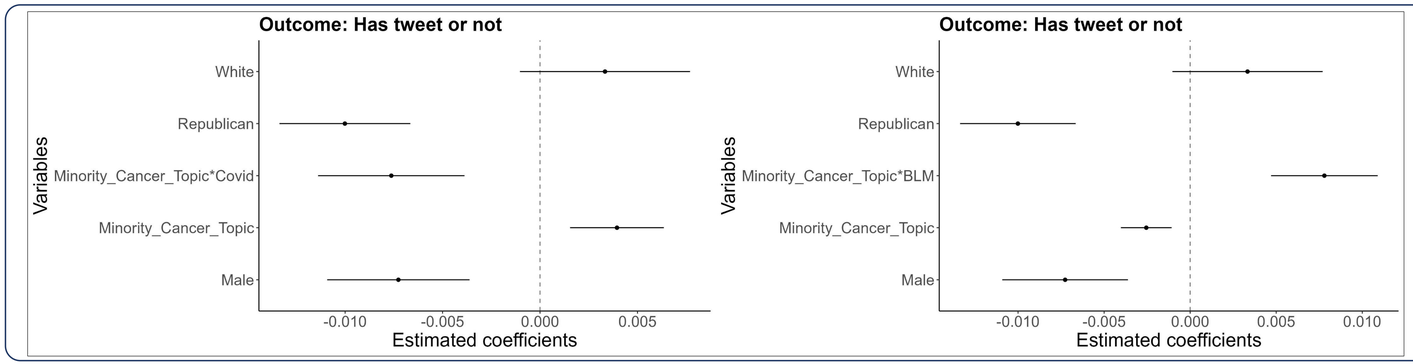

Figure 7. Associations between societal events (i.e., Black Lives Matter Movement (BLM) and COVID-19) and the number of tweets on "racial minorities cancer" (RMC) topic.

To test whether Twitter attention is modifiable, we investigated whether external societal events were associated with congress members' tweets on the RMC topic. We specifically estimated how **Black Lives Matter Movement (BLM)** and **the COVID-19 pandemic** are associated with the number of tweets with the RMC topic. We controlled the congress members' party, gender, and race. We also added state fixed effects and time fixed effects.

References are here: