Media Coverage and Reality; How News Frequency of Violence Against Women Correlates with Official Statistics

News media is one significant source for people to keep track of what's going on in the world. However, with all the events happening every moment, news media has to focus on reporting certain types of stories that are deemed newsworthy. Rarer or unexpected events have a greater chance to be reported. In general, the number of negative news exceeds positive news. Study has provided evidence for the negativity bias that makes people physiologically more activated by negative news (Soroka et al., 2019).

Violence is one of the negative news types that is reported out of the proportion of real statistics (Naylor, 2001). Violence against women (VAW) has increased since the COVID-19 pandemic. According to a 2021 UN report with data collected from 13 countries, nearly 7 in 10 women (68%) think the incidence of physical or verbal abuse by a spouse/partner violence has increased during the pandemic in the area where they live. I'm interested in how news media reports reflect (or ignore) the increase of VAW. To investigate this, I chose one dataset from an India news source *Times of India* (TOI), and one dataset from the Australia news source *Australian Broadcasting Corporation* (ABC). My rationale for these two datasets is that I expected the geographical and cultural differences between these two news agencies would lead to variations in reporting frequency and style. This allows me to compare and analyze any potential patterns. The 13 countries mentioned in the 2021 UN report didn't include India or Australia. According to the National Crime Records Bureau (NCRB) of India, the number of VAW decreased slightly in 2020 (405326) from 2019 (371503) but increased in 2021 (428278). In Australia, the rate of women suffered physical violence declined to 2.9% in 2020-21 from 4.7% in 2005.

With the official statistics at hand, I can compare whether media coverage frequency correlates with the data. One possible trend is that news reports on VAW increased with the crime rate against women. However, new reports don't necessarily reflect reality and the more frequently a certain type of event happens, the less likely it will attract readers' attention. Therefore, it's likely that higher rate of VAW will be reported less frequently. Obviously, different news agencies have different standards and criteria for evaluating events, and I'm hoping to be able to detect such differences between ABCNews and TOI.

Research questions

1. How news headlines reporting violence against women differ in India and Australia news agencies?

2. Does media coverage frequency correlate with the change of number in violence against women during the pandemic?

Datasets

The dataset Times of India News Headlines includes news headlines from *Times of India* between 2001 – q1-2022 in one single csv file. The file also includes location data (i.e., city name) of the event reported.

The dataset A Million News Headlines includes news headlines published on the Australian Broadcasting Corporation website between early-2001 – end-2021 in one single csv file.

Data cleaning and processing

After downloading the csv files, I first used Python to extract news headlines from the years 2019, 2020, and 2021 from the India news dataset; the years of 2005, 2020, and 2021 from the Australian news dataset, and output them into separate csv files. The reason for my different choice of years was to match the year with official statistics before and after the pandemic from Indian and Australian governments. I then set a couple of keywords to search for headlines of violence against women. After the grouping, I ended up with six csv files, three of them for ABCNews, and three for TOI.

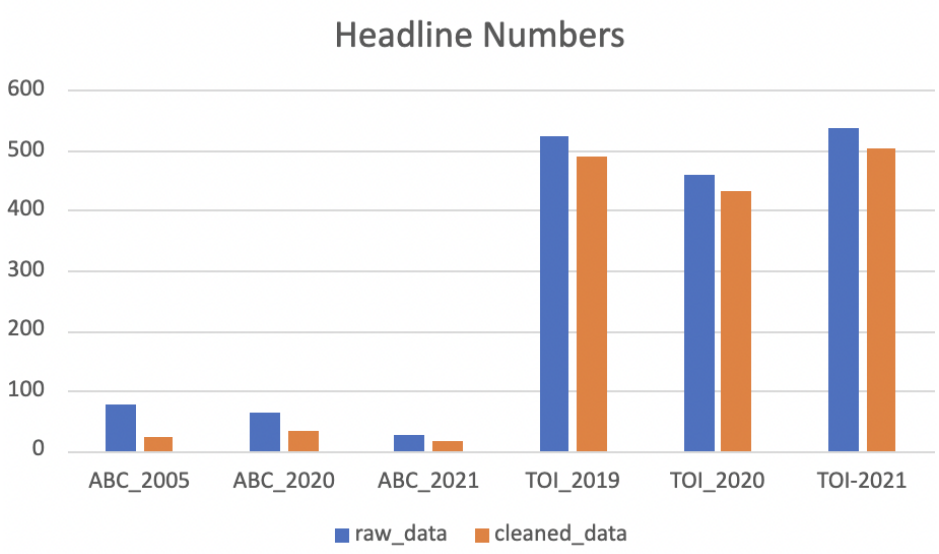
Keywords combination:

['woman', 'robbed'], ['woman', 'beaten'], ['woman', 'murder'], ['woman', 'killed'], ['woman', 'raped'], ['woman', 'violent'], ['woman', 'violence'],

['women', 'violence'], ['women', 'murder'],

['female', 'burned'], ['female', 'robbed'], ['female', 'murder'], ['female', 'beaten'], ['female', 'raped'], ['female', 'violence'], ['female', 'violent']

While examining the output files, I noticed that the keywords “female” and “killed” also showed up in headlines involving accidents. Therefore, I manually removed headlines indicating traffic accidents, animal attacks, and natural disasters. Such as “Elderly woman killed in a road accident”, “Speeding car hits bike; pregnant woman killed”, “Quepem woman killed by 'lightning”, “Woman killed in leopard attack in Nainital”. I also deleted headlines clearly stating the violence happening abroad (i.e., not in Australia for the ABCNews dataset). For instance, “japanese woman killed in bali attacks”. As well as headlines indicated woman as perpetrator, not victim (e.g., “woman charged with sydney mans murder”). Manually examining and removing these headlines allowed me to accurately identify headlines unrelated to VAW, however, this gave rise to human error and there might be irrelevant VAW headlines unspotted. The change of headline numbers before and after the manual cleaning is shown below.



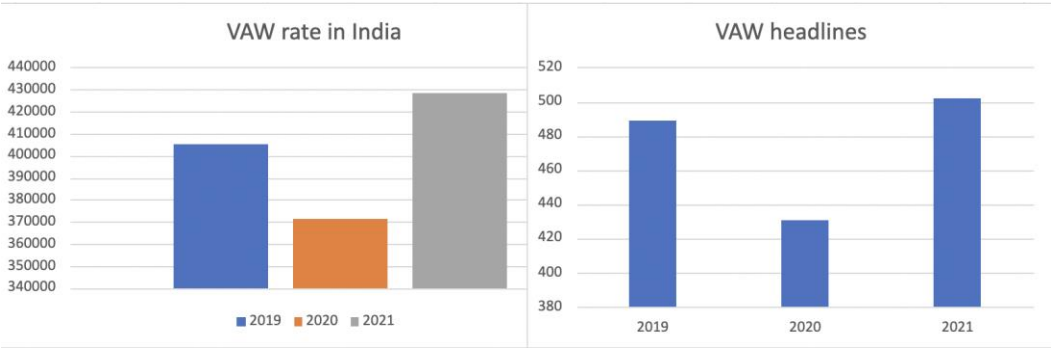
Data analysis and visualization

While I was manually cleaning the data, I had the opportunity to study the phrasing and wording of headlines from both datasets. To my disappointment, I didn’t notice any pattern that distinguished the headlines between two news sources. No obvious media bias was observed. I could easily tell the source of most headlines, but it was only due to the content of headlines (e.g., honor kill or other culturally specific phenomenon), not by reporting style (e.g., one source uses more emotional words for VAW). The same keywords combination yielded significantly

more headlines from the TOI dataset than the ABCNews dataset which was expected as there were much more VAW in India than in Australia.

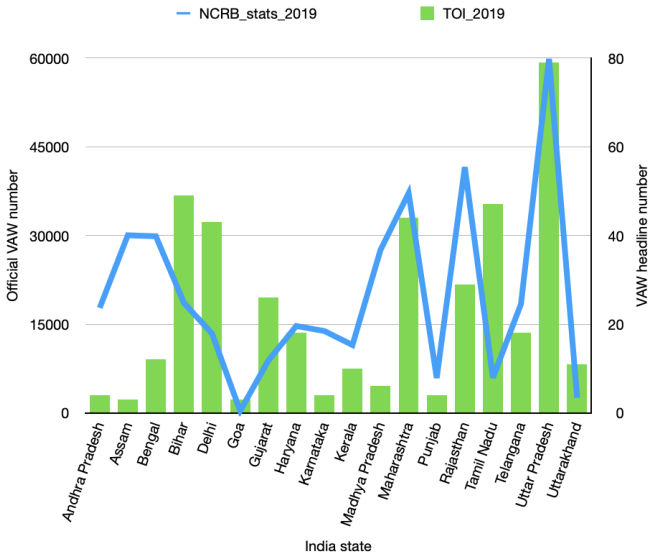
One difference between ABCNews and TOI headlines was that ABCNews had relatively more international violence coverage while TOI coverage was more domestically focused. This was likely the result of different target audiences.

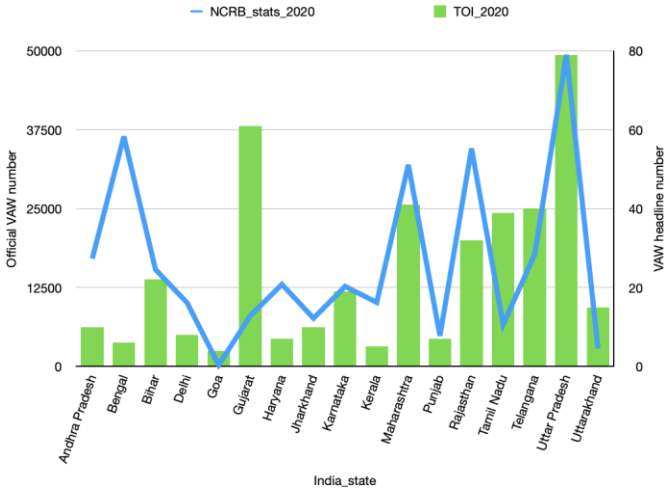
Surprisingly, the number of VAW headlines from the TOI dataset actually reflected the trend of VAW rate by year. As can be seen from the graph below, the year of 2021 had the highest number of VAW in India within the three years, and the number of headlines covering this type of violence was also the highest in 2021. The cleaned ABCNews dataset, on the other hand, didn't yield any meaningful results as the number of headlines were too negligible to be useful.

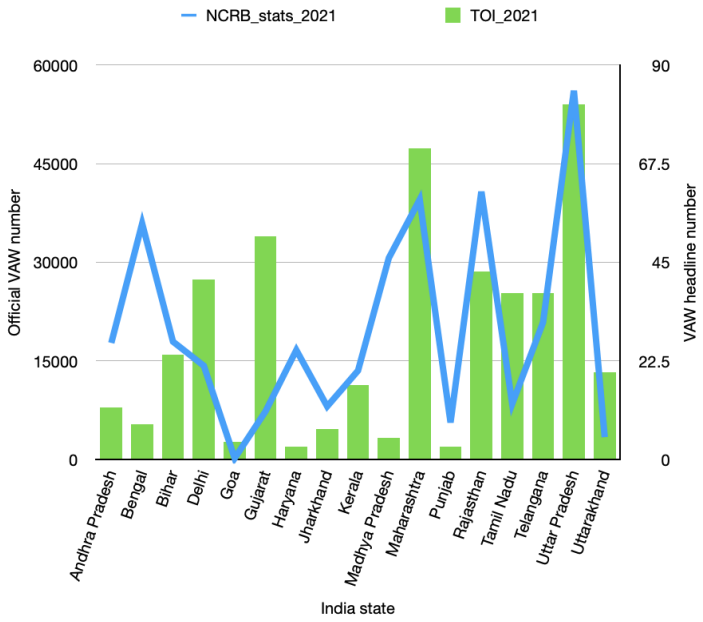


After seeing this graph, I decided to further compare the actual VAW numbers in each India state with the location data from the TOI dataset. The official statistics from the National Crime Records Bureau are divided into states and union territories, but the TOI data is categorized by city. I googled each city and recorded the state it is in, cities with less than three news headline entries were not included. I then entered the data from the NCRB statistics together with the headline data to generate graphs. I didn't expect to see much correlation between actual VAW numbers and headline data but there were a few interesting results. Uttar Pradesh had the highest VAW among all Indian states and union territories in 2019, 2020, and 2021, and the numbers of VAW headline in Uttar Pradesh were also the highest within the three years. Gujarat had the second highest and third highest VAW in 2020 and 2021, yet VAW coverage of this state was very limited. West Bengal had one of the lowest numbers of VAW between 2019–21 but there was significantly more VAW news coverage by TOI. I looked this

state up, but it didn’t appear to be economically or culturally predominant to justify this high frequency of news coverage.







The graphs demonstrate that the frequency of news coverage does not reflect the frequency of events in reality. However, there are still certain correlations between reality and news coverage frequency. It's not completely the case that the rarer an event, the more likely it gets reported. India as a country has much higher number of VAW than Australia, and this is also reflected by the higher frequency of VAW coverage. State-wise, although there were some anomalies, the general trend is that higher VAW is correlated with more coverage.

Potential bias and issues

The two datasets I used were produced by an individual data scientist, not from an official database. This can be a problem as I have no way to validate the datasets actually contained all the headlines within the stated periods. Human error is also one big concern of mine. I just mentioned that no reporting style difference was observed from manual examination between the two news sources. Manual inspection was not my first option as I didn't want to bring subjective factors into the process and made possible reproduction difficult. However, after experimenting with Voyant and Palladio in vain, I had to do it manually. I also had a difficult time navigating the Matlab. In the end, the only tools I made some progresses were Python, Excel, and Voyager (graphs not included in the final project). I imagine natural language processing models can be used to analyze linguistic style (which I don't know how to utilize) but I am unaware of any tool that can group cities into states automatically.

TOI is only one of the many Indian news agencies and its focus and coverage may not be generalized to all Indian media. Although my research question was about the relationship between media coverage and reality, the results I got were not enough to answer the question. At best, they offered some ideas about TOI's style and priority in reporting VAW. Another issue is that TOI is an English-language newspaper, but English is not the only official language in India. News agencies with other languages (e.g., Hindi) may have very different reporting styles or frequency on VAW. To be able to answer my initial research question, data with other popular languages in India need to be analyzed to understand the relationship between media coverage and reality.

References

Kulkarni, Rohit, 2018, "A Million News Headlines", https://doi.org/10.7910/DVN/SYBGZL, Harvard Dataverse, V6

Kulkarni, Rohit, 2020, "Times of India News Headlines" https://doi.org/10.7910/DVN/DPQMQH, Harvard Dataverse, V3

Naylor, B. (2001), Reporting Violence in the British Print Media: Gendered Stories. The Howard Journal of Criminal Justice, 40: 180-194. https://doi.org/10.1111/1468-2311.00200

Soroka, S., Fournier, P., & Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. Proceedings of the National Academy of Sciences of the United States of America, 116(38), 18888–18892. https://doi.org/10.1073/pnas.1908369116

https://ncrb.gov.in/uploads/nationalcrimerecordsbureau/custom/1696831798CII2021

National Crime Records Bureau, 2021 Report, Volume 1

The Shadow Pandemic: Violence against women during COVID-19, https://www.unwomen.org/en/news/in-focus/in-focus-gender-equality-in-covid-19-response/violence-against-women-during-covid-19