

STROBE Statement for

Sentiment analysis of internet posts on vaccination using ChatGPT and comparison with actual vaccination rates in South Korea

Sunyoung Park*

*Department of Psychiatry, National Health Insurance Service Ilsan Hospital, Goyang, Republic of Korea

Written on December 25, 2023.

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract The study's design is indicated in both the title page and the abstract. (b) Provide in the abstract an informative and balanced summary of what was done and what was found The abstract summarizes the research methods and key findings.
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported Described in the 1st~4th paragraph of the Introduction section.
Objectives	3	State specific objectives, including any prespecified hypotheses Described in the last paragraph of the Introduction section
Methods		
Study design	4	Present key elements of study design early in the paper Described in the first paragraph of the Method section
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection Presented in the "Web crawling and Data collection" paragraph of the Method section
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants Presented in the "Web crawling and Data collection" paragraph of the Method section
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable Described in the "Web crawling and Data collection" and "Sentiment analysis and Reasoning extraction using ChatGPT" paragraphs of the Method section
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group Explained in the "Sentiment analysis and Reasoning extraction using ChatGPT" paragraph of the Method section
Bias	9	Describe any efforts to address potential sources of bias Presented in the "Sentiment analysis and Reasoning extraction using ChatGPT" paragraph of the Method section and the 8th and 9th paragraphs of the Discussion section
Study size	10	Explain how the study size was arrived at Described in the "Web crawling and Data collection" paragraph of the Method section and the 1st and 2nd paragraphs of the Result section
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable,

		<p>describe which groupings were chosen and why</p> <p>Presented in the “Sentiment analysis and Reasoning extraction using ChatGPT” paragraph of the Method section and the 2nd paragraph of the Results section</p>
Statistical methods	12	<p>(a) Describe all statistical methods, including those used to control for confounding</p> <p>Explained in the “Statistical Analysis” paragraph of the Results section</p> <p>(b) Describe any methods used to examine subgroups and interactions</p> <p>In this study, analysis through subgroup and interaction was not conducted.</p> <p>(c) Explain how missing data were addressed</p> <p>This study, which conducted sentiment analysis on existing crawling data, did not have missing data in the analysis. However, for cases where sentiment analysis was uncertain in a small minority of cases, they were excluded from the analysis. This information is described in the 'Sentiment analysis and Reasoning extraction using ChatGPT' paragraph of the Method section.</p> <p>(d) If applicable, describe analytical methods taking account of sampling strategy</p> <p>Not relevant to the study</p> <p>(e) Describe any sensitivity analyses</p> <p>Not relevant to the study</p>
Results		
Participants	13*	<p>(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed</p> <p>In this study, no human participants were included. The research involved the analysis of data obtained through web crawling, specifically 550,559 posts, and extracted 851 posts relevant to the research topic. Among these, 3 posts that were uncertain in classifying emotional evaluations were excluded from the analysis. This process is described in the "Web crawling and Data collection" and "Sentiment analysis and Reasoning extraction using ChatGPT" paragraphs of the Method section.</p> <p>(b) Give reasons for non-participation at each stage</p> <p>Some data was excluded in cases where it was not relevant to the topic or when there was uncertainty in the results of sentiment analysis.</p> <p>(c) Consider use of a flow diagram</p>
Descriptive data	14*	<p>(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders</p> <p>This study utilized data obtained through web crawling, and it was not related to the aforementioned items.</p> <p>(b) Indicate number of participants with missing data for each variable of interest</p> <p>In this study, no human subjects were included. The research involved the analysis of data obtained through web crawling, specifically 550,559 posts, and extracted 851 posts relevant to the research topic. Among these, 3 posts that were uncertain in classifying emotional evaluations were excluded from the analysis.</p>
Outcome data	15*	<p>Report numbers of outcome events or summary measures</p> <p>Presented in Table 1.</p>
Main results	16	<p>(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included</p> <p>Not applicable</p> <p>(b) Report category boundaries when continuous variables were categorized</p>

		Not relevant to this study (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period Not relevant to this study
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses Additionally, an association analysis between variables was conducted. These findings are described in the 'Extraction of attitudinal reasons toward vaccination and association analysis between reasons' section of the Results.
Discussion		
Key results	18	Summarise key results with reference to study objectives Presented in the 1 st and 3 rd paragraphs in the Discussion section.
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias Presented in the 8 th and 9 th paragraphs in the Discussion section.
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence Presented in the Discussion section, specifically in the 3 rd , 4 th , and 5 th paragraphs.
Generalisability	21	Discuss the generalisability (external validity) of the study results Presented in the 10 th paragraph in the Discussion section
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based There was no funding source in this study.

*Give information separately for exposed and unexposed groups.