

## Original Paper

# Exploring Social Media Posts on Lifestyle Behaviors: Sentiment and Content Analysis

Yan Yee Yip<sup>1,2</sup>, MPharm, MClInPharm; Mohd Ridzwan Yaakub<sup>3</sup>, BSc, MSc, PhD; Mohd Makmor-Bakry<sup>1</sup>, BPharm, MClInPharm, PhD; Muhammad Iqbal Abu Latiffi<sup>3</sup>, BSc, MSc; Wei Wen Chong<sup>1</sup>, BPharm, PhD

<sup>1</sup>Centre for Quality Management of Medicines, Faculty of Pharmacy, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

<sup>2</sup>Centre for Clinical Epidemiology, Institute for Clinical Research, National Institutes of Health, Ministry of Health Malaysia, Setia Alam, Shah Alam, Malaysia

<sup>3</sup>Center for Artificial Intelligence Technology, Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor Darul Ehsan, Malaysia

**Corresponding Author:**

Wei Wen Chong, BPharm, PhD  
Centre for Quality Management of Medicines  
Faculty of Pharmacy, Universiti Kebangsaan Malaysia  
Jalan Raja Muda Abdul Aziz  
Kuala Lumpur, 50300  
Malaysia  
Phone: 60 392897265  
Email: [weiwen@ukm.edu.my](mailto:weiwen@ukm.edu.my)

## Abstract

**Background:** There has been an increase in the prevalence of noncommunicable diseases in Malaysia. This can be prevented and managed through the adoption of healthy lifestyle behaviors, including not smoking, avoiding alcohol consumption, maintaining a balanced diet, and being physically active. The growing importance of using social media to deliver information on healthy behaviors has led health care professionals (HCPs) to lead these efforts. To ensure effective delivery of information on healthy lifestyle behaviors, HCPs should begin by understanding users' current opinions about these behaviors and whether the users are receptive to recommended health practices. Nevertheless, there has been limited research conducted in Malaysia that aims to identify the sentiments and content of posts, as well as how well users' perceptions align with recommended health practices.

**Objective:** This study aims to examine social media posts related to various lifestyle behaviors, by using a combination of sentiment analysis to analyze users' sentiments and manual content analysis to explore the content of the posts and how well users' perceptions align with recommended health practices.

**Methods:** Using keywords based on lifestyle behaviors, posts originating from X (formerly known as Twitter) and published in Malaysia between November and December 2022 were scraped for sentiment analysis. Posts with positive and negative sentiments were randomly selected for content analysis. A codebook was developed to code the selected posts according to content and alignment of users' perceptions with recommended health practices.

**Results:** A total of 3320 posts were selected for sentiment analysis. Significant associations were observed between sentiment class and lifestyle behaviors ( $\chi^2_6=67.64$ ;  $P<.001$ ), with positive sentiments higher than negative sentiments for all lifestyle behaviors. Findings from content analysis of 1328 posts revealed that most of the posts were about users' narratives (492/1328), general statements (203/1328), and planned actions toward the conduct of their behavior (196/1328). More than half of tobacco-, diet-, and activity-related posts were aligned with recommended health practices, whereas most of the alcohol-related posts were not aligned with recommended health practices (63/112).

**Conclusions:** As most of the alcohol-related posts did not align with recommended health practices, the findings reflect a need for HCPs to increase their delivery of health information on alcohol consumption. It is also important to ensure the ongoing health promotion of the other 3 lifestyle behaviors on social media, while continuing to monitor the discussions made by social media users.

*JMIR Infodemiology* 2025;5:e65835; doi: [10.2196/65835](https://doi.org/10.2196/65835)

**Keywords:** consumer health information; content analysis; chronic illness; health promotion; healthy lifestyle; lifestyle; lifestyle risk reduction; internet; primary prevention; sentiment analysis; social media

## Introduction

Noncommunicable diseases (NCDs) have become a significant global health challenge, accounting for 74% of all deaths worldwide [1]. NCDs have also become a growing public health concern in middle-income countries within the Southeast Asia region. Malaysia is one of the countries in the region that has been significantly impacted by NCDs, with over 67% premature NCD-related mortality and over 70% disease burden [2]. Approximately, 2.5% of Malaysian adults, which accounts for over half a million people were affected by all 4 major NCDs, which are diabetes, hypertension, hypercholesterolemia, and obesity [3].

The World Health Organization (WHO) has identified 4 key modifiable risk factors associated with an elevated risk of NCDs, which are tobacco smoking, harmful use of alcohol, unhealthy diet, and physical inactivity [1,4]. The adoption of healthy lifestyle behaviors, including not smoking, avoiding alcohol consumption, maintaining a balanced diet, and being physically active can reduce modifiable risk factors, effectively preventing and managing NCDs. However, national surveys have indicated that the actual adoption of healthy behaviors among Malaysians remains low. For example, over 84% of Malaysian adults were inactive in sports, with half of the population leading a sedentary lifestyle, spending more than 2 hours sitting while awake [3].

In this regard, it is important to deliver information on healthy lifestyle behaviors, with health care professionals (HCPs) being ideally positioned to lead these efforts. Various technologies, such as mHealth applications, wearable devices [5], and social media platforms, can support the delivery of health information. Social media platforms have been widely used for health information delivery as these platforms are accessible to larger populations at a lower cost [6,7]. To effectively promote healthy lifestyle behaviors on social media, HCPs could begin by understanding users' current opinions on lifestyle behaviors and whether they are receptive to recommended health practices [8]. This could be achieved by examining social media posts discussing on lifestyle behaviors. X (formerly known as Twitter) is one of the microblog-based social media platform that allows users to freely express their opinions through posts, previously referred to as tweets. As of January 2024, approximately 5.71 million social media users in Malaysia were on X, accounting for 16.5% of the country's population [9], which highlights the growing popularity of X among Malaysians.

When users express their opinions in writing on social media, a range of emotions may be conveyed. Sentiment analysis is the process of classifying this textual data based on the emotions conveyed within the text as positive, negative, or neutral sentiments [10]. It can be conducted through manual annotation of posts or computational approaches [11]. Computational approaches in sentiment analysis are preferred as they are more cost-efficient, and

can leverage large amounts of publicly accessible and concise real-time data across different regions and demographics [12]. Methodologies of computational approaches include lexicon-based sentiment analysis, which uses pre-existing dictionaries containing words with pre-assigned sentiment scores of positive, negative, or neutral. Lexicon-based sentiment analysis is effective when limited labelled training data is available with a strong association of sentiments with specific words. The usage of this approach has been documented in numerous studies that analyze sentiments regarding lifestyle behaviors such as the examination of policies on electronic cigarettes [13], vegan-related posts [14], and organic food posts [15].

Lexicon-based sentiment analysis, however, may have limited coverage in terms of vocabulary and often misses sarcasm or irony [16]. Positive sentiments may not necessarily translate to good health practices and vice versa. For example, the sentence "I love tobacco" showed positive sentiments, but the actual context is related to the user's preference towards unhealthy lifestyle behaviors. In order to further understand the topics communicated on social media and how users' perceptions are aligned with recommended health practices, lexicon-based sentiment analysis can be supported with manual content analysis. A codebook can be used to manually assign labels to each post, which will provide a more in-depth analysis of the posts [17].

The examination of social media posts across multiple lifestyle behaviors of tobacco smoking, alcohol consumption, diet, and physical activity could facilitate effective comparisons of findings across these different behaviors. The use of such findings would enable HCPs to use social media to deliver information on healthy behaviors by targeting areas where the lifestyle behaviors are not aligned with recommended health practices. Analyses that are focused within a geographic location would provide opportunities for HCPs to prioritize region-targeted health information delivery on social media. Such health information could also potentially be replicated in other countries with similar digital and health ecosystems.

Nevertheless, there have been limited studies that used the combined approaches of lexicon-based sentiment and content analysis to examine social media posts on lifestyle behaviors. The majority of the available studies were focused on other health-related issues such as the examination of users' perceptions on diabetes [18] and marijuana usage [19]. A study by Kasson et al [20] have used both sentiment analysis and manual content analysis to examine users' vaping behaviors. However, the study was confined to vaping behaviors during the e-cigarette or vaping use-associated lung injury outbreak and did not address other lifestyle behaviors. In addition, there is a lack of studies examining the opinions of social media users in Malaysia on lifestyle behaviors, despite the increasing burden of NCDs and the rising prevalence of unhealthy lifestyle behaviors in the country.

Therefore, this study used a combination of lexicon-based sentiment analysis and manual content analysis to understand the discussions on lifestyle behaviors among social media users in Malaysia. This study had three objectives: (1) to determine the sentiments of social media users in Malaysia regarding lifestyle behaviors, (2) to identify the content of posts and ascertain if users' perceptions were aligned with recommended health practices, and (3) to explore the associations between the alignment of users' perceptions with recommended health practices and sentiment class.

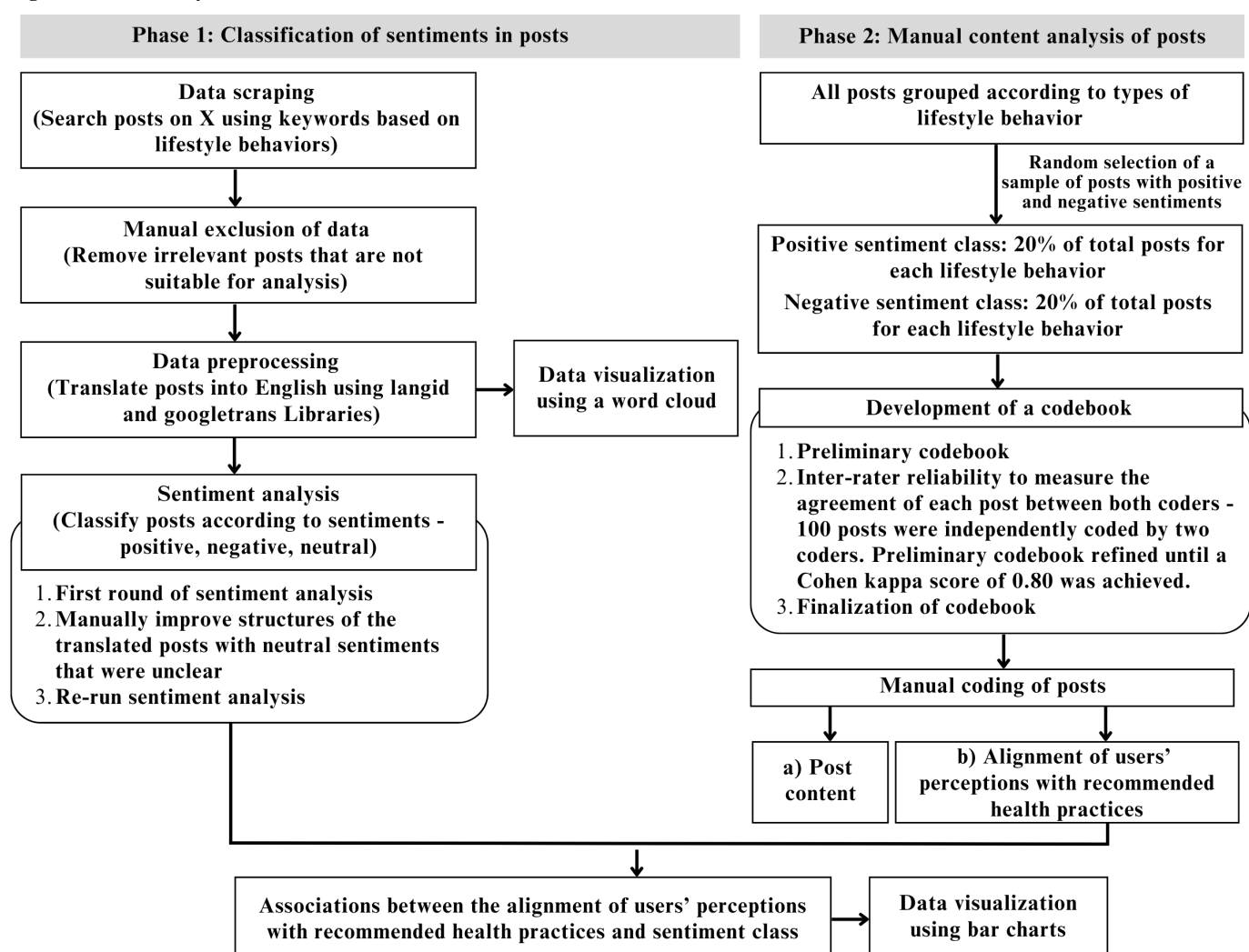
Following the manual exclusion of irrelevant posts, the data was cleaned, preprocessed and analyzed for sentiments. Data visualization was subsequently conducted. In the manual content analysis of posts, a random selection of posts with positive and negative sentiments for each lifestyle behavior was manually coded to identify the content of posts and the alignment of users' perceptions with recommended health practices. Associations between the alignment of users' perceptions and sentiment class were subsequently explored.

## Methods

### Overview

Figure 1 shows the overall study methods. In the classification of sentiments in posts, data was scraped from X.

Figure 1. Overall study methods.



### Classification of Sentiments in Posts

#### Data Scraping

The automated process of extracting large amounts of data from X is known as data scraping. All posts with keywords related to the 4 lifestyle behaviors aimed at reducing the 4

key modifiable risk factors for NCDs were scraped. These keywords are related to tobacco and its derivative products, alcohol, dietary, and physical activity. The 4 sets of keywords are provided in [Multimedia Appendix 1](#). The keywords were derived from published systematic reviews on the management of lifestyle behaviors using social media

[21,22]. Additional keywords commonly used locally were added upon discussion with all researchers.

Posts spanning 2 consecutive months were selected. This time frame was deemed to be appropriate, as similar studies analyzing health-related sentiments on the X platform have also utilized data across a 2-month period [23,24]. The selection of the 2 consecutive months was conducted by initially scraping all posts from January to December 2022 according to each month. The 2 months with the highest number of posts were from November to December 2022. Post based in Malaysia were determined using longitude and latitude metadata (4.2105°N, 101.9758°E). In terms of language, posts in Malay and English were scraped. Malay is the national language of Malaysia, whereas English is widely spoken and understood by Malaysians. They also use a mix of both languages, resulting in multilingual data.

Data scraping was conducted separately for each set of keywords using the SNScrape library on Python. In addition to posts, other X metadata such as timestamp, X username, number of reposts, language, and location were also collected.

## Manual Exclusion of Data

All posts were manually screened by 2 researchers to exclude those not suitable for analysis, with discrepancies resolved among the research team. The exclusion steps were as follows:

First, the exclusion of posts not from Malaysia—During data scraping, longitude and latitude data retained posts located within the coordinates but posted outside of Malaysia, such as parts of Singapore and Thailand. The “location” metadata was therefore used to manually exclude these posts.

Second, the exclusion of irrelevant posts—Irrelevant posts include posts with different definitions (eg, “exercising” your rights), posts not related to health care (eg, religious restrictions on alcohol), indecipherable posts and posts made by bots. Bots were verified by manually checking the user’s profile for any unusual activity patterns that exhibited automated behavior (eg, high frequency of posts without breaks).

## Data Preprocessing

Before conducting sentiment analysis, data preprocessing was carried out to ensure that the text data was cleaned, transformed, and prepared for analysis.

The model selected for sentiment analysis was the Valence Aware Dictionary and Sentiment Reasoner (VADER) on Python. It is a lexicon-based sentiment analysis tool [25] that is suitable for analyzing social media posts, generating results with high classification success [26]. As VADER is trained for sentiment analysis in English, all posts in Malay and mixed languages were preprocessed and translated into English using the langid and googletrans libraries on Python.

Other data preprocessing steps (tokenization, lower casing of texts, removal of stop words, html links, numbers, punctuations, emojis, and acronyms) were not executed. This is attributed to the unique advantages of VADER, in which

assessment scores would account for capitalism, punctuations, emojis, English acronyms (eg, “LOL”), and colloquialisms (eg, “meh”) [27-29].

Following data preprocessing, a word cloud used to provide a visual representation of the words in the overall X dataset.

## Sentiment Analysis

Computational, lexicon-based sentiment analysis was conducted to determine users’ sentiments. This approach was selected over manual annotations and other types of computational methods as it is cost-effective and does not require training for large datasets. As it relies on a predefined lexicon, it does not require significant computational resources. For general sentiment analysis, lexicon-based approaches often performs well enough to capture the overall sentiment trends in social media data [25].

The posts that have undergone data preprocessing were then analyzed for polarity using VADER. VADER classifies posts into positive, neutral, and negative sentiments. Positive sentiments have a compound score of  $\geq 0.05$ , neutral sentiments have a score between  $>-0.05$  and  $<0.05$  and negative sentiments have a score of  $\leq -0.05$  [25].

Computer-assisted translation tools may limit the extent of translation in posts that contain local dialects and slang. The translated posts may become indecipherable, causing them to be classified as having “neutral” sentiments. For example, “x” in Malay, which means “no” in English may not have undergone translation, resulting in sentiments not being classified accurately. Following the first round of sentiment analysis, the structures of the translated posts that were unclear and yielded neutral sentiments were manually improved. The sentiment analysis was then re-run to enhance the robustness of sentiment classification and to reduce the inaccurate labeling of posts.

## Manual Content Analysis of Posts With Positive and Negative Sentiments

Sentiment analysis classifies texts according to the emotions conveyed [10]. However, there was no further elaboration on the content and whether users’ perceptions were aligned with recommended health practices.

Therefore, a sample of posts with positive and negative sentiments were randomly selected for manual content analysis. Stratified sampling was conducted by dividing the posts according to the type of lifestyle behavior and sentiment class (positive or negative). For each type of lifestyle behavior and sentiment class, 20% of the total posts for the particular lifestyle behavior were subjected to random selection. This would allow the posts for each lifestyle behavior to have an equal number of positive and negative sentiments. The random sample of posts was generated by using the random number equation in Microsoft Excel relative to the ID number attached to each post. This approach was adopted from a previous content analysis study on X, which also manually coded a random sample of 20% of total posts [30].



A preliminary codebook with 2 categories was developed through discussions among the research team to classify the content of posts and the alignment of users’ perceptions with recommended health practices. This codebook is partially adapted from the codes used by Miller et al [31]. The recommended health practices are based on WHO’s health recommendations [32]. In brief, WHO advocates healthy practices, including abstaining from smoking and alcohol consumption, maintaining a balanced diet, and engaging in regular physical activities. The codes were mutually exclusive. Using the preliminary codebook, 100 posts (25

posts for each type of lifestyle behavior) were independently coded by 2 coders. Interrater reliability was conducted to measure the agreement of each post between both coders. The preliminary codebook was refined until a Cohen kappa score of 0.80 was achieved. The remaining posts were then coded independently using the finalized codebook by both coders.

Table 1 provides a brief description of the codes for the finalized codebook, with a more comprehensive codebook provided in Multimedia Appendix 2.

Table 1. Brief description of codes.

Category	Definition
Post content (Topical content dealing with the lifestyle behavior mentioned in the post)	
Self-narrative of current lifestyle behaviors	Narration of self’s current lifestyle behaviors.
Narrative of others’ current lifestyle behaviors	Talked about other people’s current lifestyle behaviors.
Planned action related to lifestyle behaviors	A planned action that will be conducted by the person who wrote the post.
Recommendations related to lifestyle behaviors	A recommendation by the person who wrote the post, providing instruction, advice, or suggestion to others.
Direct question	Direct question used in a post.
General statement	General statement that is not under any of the other categories above.
Alignment of users’ perceptions with recommended health practices (Whether users’ perceptions in posts are aligned with WHO’s health recommendations)	
Aligned with recommended health practices	Users agreed with the conduct of recommended health practices, that included not smoking, avoiding alcohol consumption, maintaining a balanced diet and being physically active.
Not aligned with recommended health practices	Users were not agreeable with the conduct of recommended health practices (eg, consumed oily food, refused to exercise).
Users’ perceptions cannot be defined	The perceptions of the user could not be defined or linked with health practices.

Data Analysis

Data analysis was performed using descriptive statistics with all variables expressed in frequencies and percentages. The Pearson chi-square test was used to compare the associations between the categorical variables, with *P* values <.05 considered to be statistically significant. IBM SPSS Statistics version 26.0 was used for data analysis.

Findings were also visualized using a word cloud and bar charts. In addition, examples of posts were provided to describe the study findings.

Ethical Considerations

This study was approved by the Medical Research and Ethics Committee, Ministry of Health Malaysia (NMRR ID-23-00293-CIM [IIR]) on March 23, 2023, and the Research Ethics Committee, Universiti Kebangsaan Malaysia (UKM PPI/111/8/JEP-2023-174) on April 13, 2023.

As this study relied solely on publicly available social media data on X and did not involve direct interaction with individuals, informed consent was not applicable. No compensation was offered or provided, as the study did not involve direct participation of human participants.

No identifiable private user information was collected or analyzed. All data used in the analysis were publicly available and did not contain personally identifiable information.

Results

Overview of X Dataset

Figure 2 shows the flowchart of the selection of the X dataset.

A total of 9581 posts were scraped from November to December 2022. Following the exclusion of 3047 posts that were not in Malaysia and 3214 irrelevant posts, 3320 posts across 4 types of lifestyle behaviors were retained for sentiment analysis. Almost half of the posts were dietary-related (1530/3320, 46.1%), followed by activity-related (810/3320, 24.4%) and tobacco-related (700/3320, 21.1%) posts. Alcohol-related posts were present in only one-tenth of the posts (280/3320, 8.4%).

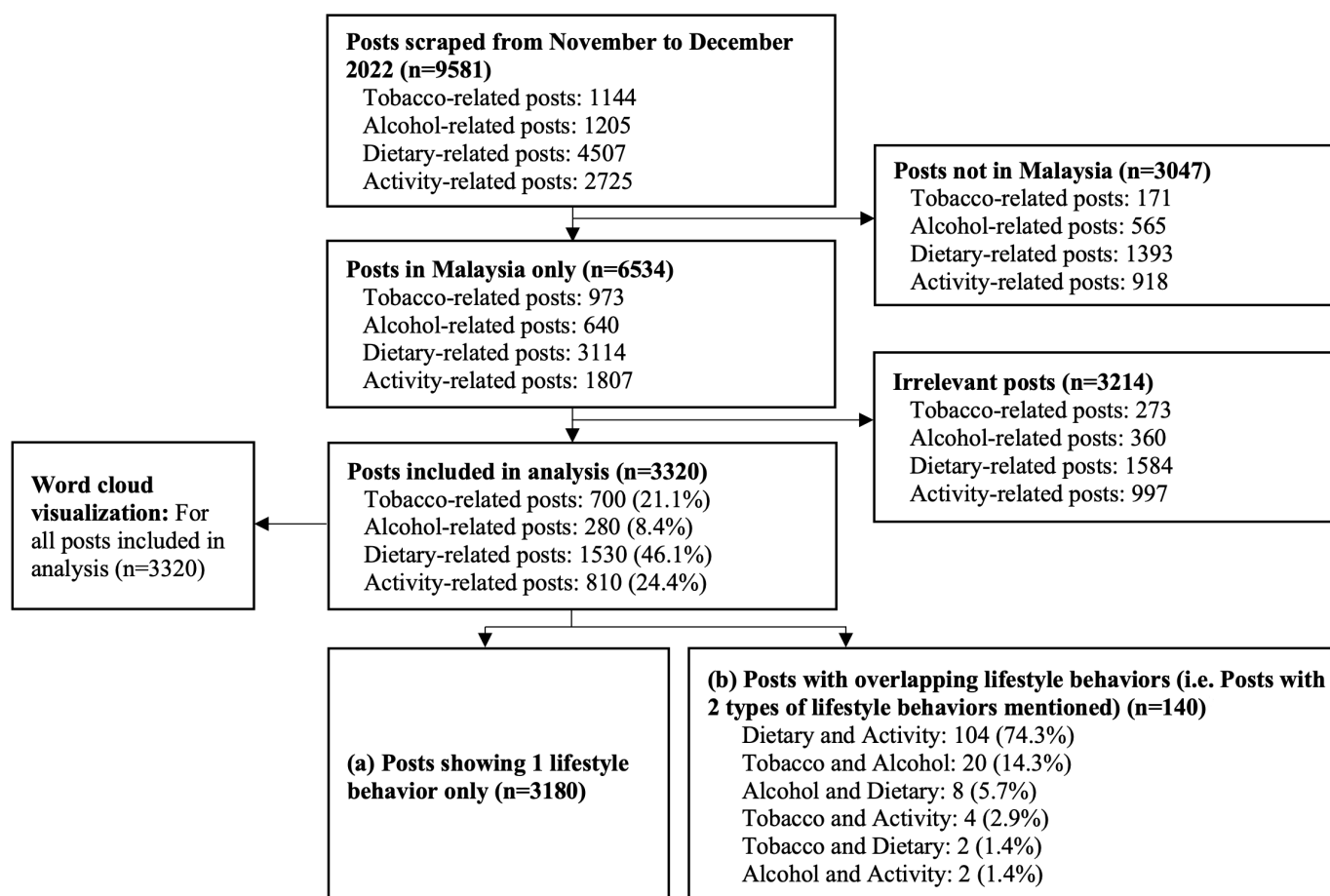
As data scraping was conducted separately for each lifestyle behavior, a post may appear more than two times across different behaviors. Out of the 3320 posts, 3180 (95.8%) posts showed 1 lifestyle behavior only. There were 140 posts with two types of lifestyle of behaviors mentioned,

with three-quarters (104/140, 74.3%) of posts mentioning both dietary- and activity-related behaviors.

most commonly mentioned by users were “diet,” “rice,” “eat,” “sugar,” and “smoke.”

A word cloud was used to visualize the overall X dataset (n=3320; see [Multimedia Appendix 3](#)). Overall, the 5 terms

**Figure 2.** Flowchart of selection of X dataset. Examples of irrelevant posts include (1) tobacco-related posts: gaming-type of posts (eg, smoke mentioned in a game), music band names (eg, Cigarettes After Sex), tweet mentioning terms as a location (eg, Hookah Island, Vape Shop); (2) alcohol-related posts: banning alcohol due to religion with no links to health (eg, at Qatar for the World Cup), sarcasm-based (eg, “you must be drunk,” which translates to “you must be kidding me”); (3) dietary-related posts: nephew in Malay (eg, anak buah), nonhealth posts (eg, fruit on the trees or plants, sugar daddy); (4) activity-related posts: posts mentioning terms as a location (eg, Sports Direct), nonhealth posts (eg, who will be Sports Minister?, “exercising” your rights). With regard to posts overlapping lifestyle behaviors, as data scraping was conducted separately for each lifestyle behavior, a post may appear more than 2 times across different behaviors (eg, the post talks about both smoking and dietary habits).



## Findings From Classification of Sentiments in Posts

**Table 2** presents the frequency distribution of sentiment analysis, with examples for each lifestyle behavior provided in [Multimedia Appendix 4](#).

The overall percentage of positive sentiments almost doubled that of negative sentiments (1874/3320, 56.5%) vs (1027/3320, 30.9%). Results have shown a significant association between sentiment class and lifestyle behaviors

( $\chi^2_6=67.64$ ;  $P<.001$ ), with positive sentiments being higher than negative sentiments for all lifestyle behaviors. The trends for dietary- and activity-related posts were similar, with both showing approximate percentages of 60% for positive sentiments and 27% for negative sentiments. This was followed by alcohol-related posts with positive sentiments of 54.7% (153/280). Less than half of tobacco-related posts (314/700, 44.9%) had positive sentiments, with the percentage differences between sentiment classes slightly below 2%.

**Table 2.** Frequency distribution of sentiment analysis before and after structure improvement of posts using Valence Aware Dictionary and Sentiment Reasoner (VADER; n=3320).

Lifestyle behaviors	Sentiment count before structure improvement of posts, n (%)			Sentiment count after structure improvement of posts, n (%) <sup>a</sup>			Total, n
	Positive	Neutral	Negative	Positive	Neutral	Negative	
Tobacco-related posts	254 (36.3)	197 (28.1)	249 (35.6)	314 (44.9)	84 (12)	302 (43.1)	700
Alcohol-related posts	120 (42.9)	95 (33.9)	65 (23.2)	153 (54.7)	41 (14.6)	86 (30.7)	280
Dietary-related posts	724 (47.3)	455 (29.7)	351 (23)	916 (59.9)	197 (12.9)	417 (27.2)	1530
Activity-related posts	378 (46.7)	246 (30.4)	186 (22.9)	491 (60.6)	97 (12)	222 (27.4)	810
Total	— <sup>b</sup>	—	—	1874 (56.5)	419 (12.6)	1027 (30.9)	3320

<sup>a</sup>Sentiment count after structure improvement of posts were used for analysis. A Pearson chi-square test was conducted to test the associations between sentiment class and lifestyle behaviors ( $\chi^2_6=67.64, P<.001$ ).  
<sup>b</sup>Not applicable.

Findings From Manual Content Analysis of Posts

A total of 1328 posts with an equal number of positive and negative sentiments for each lifestyle behavior were selected for manual content analysis. They comprised of 280 tobacco-related posts (140 posts for each sentiment class), 112 alcohol-related posts (56 posts for each sentiment class), 612 dietary-related posts (306 posts for each sentiment class), and 324 activity-related posts (162 posts for each sentiment class).

Prior to the manual content analysis of all 1328 posts, 100 posts were first subjected to interrater reliability testing. The Cohen kappa scores for both categories of post content and the alignment of users’ perceptions with recommended health practices were 0.807 and 0.801, respectively.

The frequency of posts is tabulated in Table 3. Overall, the content with the 3 highest number of posts were self-narratives of current lifestyle behaviors (492/1328, 37%), general statements (203/1328, 15.3%) and planned actions (196/1328, 14.8%). Self-narratives were the most popular content for all types of lifestyle behaviors except for tobacco-related posts, in which the majority were narratives of others’ current behaviors (96/280, 34.3%). Question-based posts were the least popular content for tobacco-, alcohol-, and activity-related posts, with less than 10% present in all types. Users’ perceptions in more than half of the posts were aligned with recommended health practices (769/1328, 57.9%). Similar proportions were observed for all types of lifestyle behaviors except for alcohol-related posts, in which posts not aligned with recommended health practices were double those aligned with recommended health practices (63/112, 56.3% vs 33/112, 29.4%).

Figure 3 shows the frequency of posts that demonstrated the alignment of users’ perceptions with recommended health practices according to the sentiment classification. A total of 3 main findings were observed. First, in dietary- and activity-related posts, significant associations between sentiment class and alignment of users’ perceptions with recommended health practices were observed ( $\chi^2_2=30.98, P<.001$  and  $\chi^2_2=24.16,$

$P<.001$ ; respectively). In both positive and negative sentiment classes, the percentages of posts aligned with recommended health practices were significantly higher than those not aligned with recommended health practices and those with undefined user perceptions, with percentages ranging from 49.3% to 80.2%. Posts with positive sentiments that aligned with recommended healthy practices showcased users’ optimism to stay healthy (eg, “I’m ready to cut sugar. Let’s go” [D-919-positive]). Meanwhile, negative sentiments that aligned with recommended healthy practices highlighted users’ worries to stay healthy (eg, “Feel the weight.. rise suddenly. Sad. Have to fix it” [P-522-negative]).

Second, in tobacco-related posts, there was no significant association between sentiment class and alignment of users’ perceptions with recommended health practices ( $\chi^2_2=5.76; P=.06$ ). Among posts with positive sentiments, the percentage of posts that aligned with recommended health practices was similar to those not aligned with recommended health practices, with a percentage difference of 7.1%. When users posted about tobacco with positive emotions, the likelihood of their perceptions aligning with the recommended health practices of not smoking (eg, “Please pray that I can stop smoking...” [T-563-positive]) or aligning with hazardous smoking practices (eg, “My kind of chill with cigar” [T-412-positive]) were similar. Despite the lack of significant association, the percentage of posts with negative sentiments that aligned with recommended health practices was noticeably higher than those that did not align with recommended health practices (86/140, 61.4% vs 41/140, 29.3%).

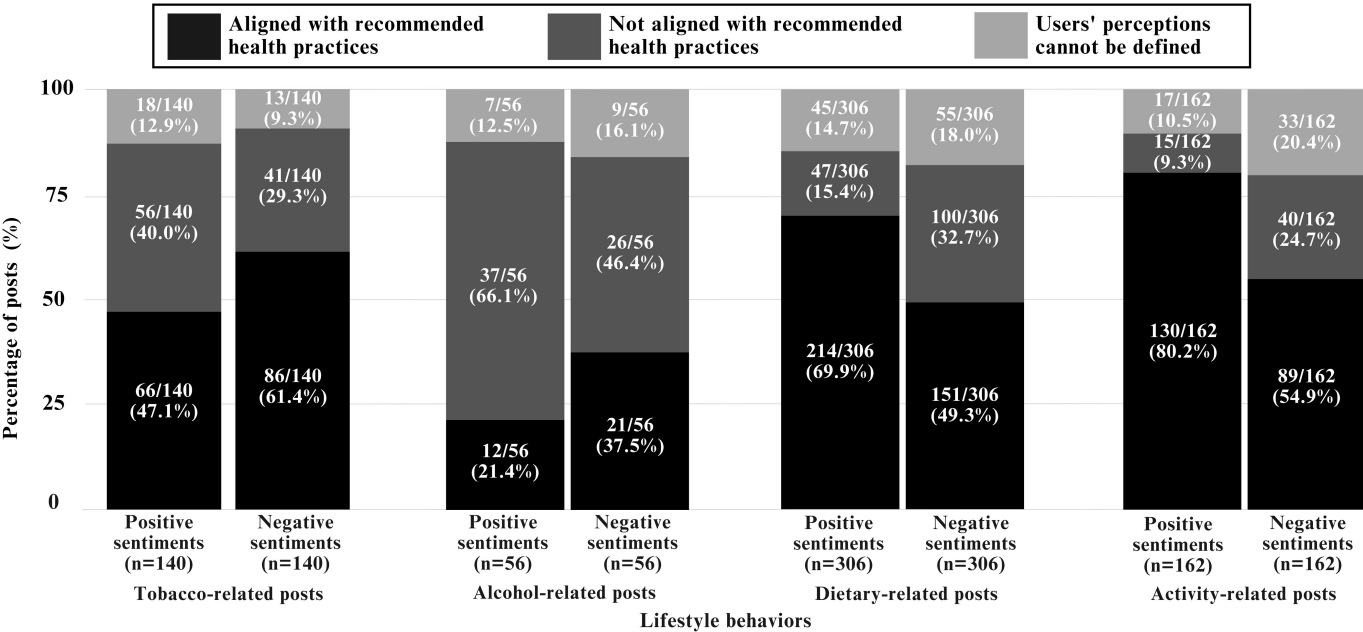
Third, a lack of significant association between sentiment class and alignment of users’ perceptions with recommended health practices was also observed in alcohol-related posts ( $\chi^2_2=4.62; P=.10$ ). Although the findings were not statistically significant, the percentage of posts not aligned with recommended health practices was higher than those aligned with recommended health practices for both positive (37/56, 66.1% vs 12/56, 21.4%) and negative (26/56, 46.4% vs 21/56, 37.5%) sentiment classes. Positive sentiments leading to alcohol consumption included celebratory posts (eg, “I’m

gonna have so much wine this weekend” [A-55-positive]), whereas negative sentiments involved users coping with worries (“I am going to drown my sorrows in alcohol and pick things back up tomorrow” [A-198-negative]). Examples of posts selected for manual content analysis are provided in [Multimedia Appendix 5](#).

**Table 3.** Frequency of posts according to post content and alignment of users’ perceptions with recommended health practices (n=1328).

Category	Lifestyle behaviors				Total posts (N=1328)
	Tobacco-related (n=280)	Alcohol-related (n=112)	Dietary-related (n=612)	Activity-related (n=324)	
Post content, n (%)					
Self-narrative of current lifestyle behaviors	70 (25)	36 (32.1)	243 (39.7)	143 (44.1)	492 (37)
Narrative of others’ current lifestyle	96 (34.3)	17 (15.2)	40 (6.5)	22 (6.8)	175 (13.2)
Planned action related to lifestyle behaviors	26 (9.3)	15 (13.4)	97 (15.8)	58 (17.9)	196 (14.8)
Recommendations related to lifestyle behaviors	35 (12.5)	13 (11.6)	83 (13.6)	46 (14.2)	177 (13.3)
Direct question	17 (6.1)	6 (5.4)	45 (7.4)	17 (5.3)	85 (6.4)
General statement	36 (12.8)	25 (22.3)	104 (17)	38 (11.7)	203 (15.3)
Alignment of users’ perceptions with recommended health practices, n (%)					
Aligned with recommended health practices	152 (54.3)	33 (29.4)	365 (59.7)	219 (67.6)	769 (57.9)
Not aligned with recommended health practices	97 (34.6)	63 (56.3)	147 (24)	55 (17)	362 (27.3)
Users’ perceptions cannot be defined	31 (11.1)	16 (14.3)	100 (16.3)	50 (15.4)	197 (14.8)

**Figure 3.** Stacked bar charts of alignment of users’ perceptions with recommended health practices stratified according to sentiment classification. A Pearson chi-square test was conducted for each lifestyle behavior to test the associations between sentiment class and alignment of users’ perceptions with recommended health practices (tobacco:  $\chi^2_2=5.76$ ,  $P=.06$ ; alcohol:  $\chi^2_2=4.62$ ,  $P=.10$ ; dietary:  $\chi^2_2=30.98$ ,  $P<.001$ ; activity:  $\chi^2_2=24.16$ ,  $P<.001$ ).



# Discussion

## Overview

To the best of our knowledge, this is the first study in the region that examined discussions on X across multiple lifestyle behaviors. This study is also the first of its kind that used dual approaches of lexicon-based sentiment analysis and manual content analysis of posts to examine users’ sentiments, post content and the alignment of users’ perceptions with recommended health practices. Positive sentiments were significantly higher than negative sentiments for all 4

lifestyle behaviors. In dietary- and activity-related posts, users exhibited twice as many positive sentiments as negative ones. The majority of the sampled posts were self-narratives of current lifestyle behaviors. More than half of the sampled tobacco-, dietary-, and activity-related posts were aligned with WHO’s recommended health practices, with contrasting results in alcohol-related posts.

## Principal Findings

Data scraping has shown that dietary-related topics were the lifestyle behaviors most frequently discussed. The usage of a more extensive set of search terms in scraping dietary-related



posts covered a variety of nutrition-based topics involving individuals across all age groups. This resulted in dietary-related discourses among young children and adolescents (eg, formula milk, vegetables, and fruits consumption). Users frequently mentioned “rice,” which is attributed to Malaysians’ staple diet, with the average Malaysian adult consuming 82.3 kilograms of rice annually [33]. In contrast, alcohol-related topics were the least discussed lifestyle behaviors among users. Alcohol-related discussions in Malaysia were largely anchored on themes related to cultural and religious beliefs. Alcohol consumption among Malaysians is generally lower as behaviors are influenced by compartmentalization among the three main races in Malaysia [34]. Malays who are Muslims are not allowed to consume alcohol as it is forbidden in Islam [35], whereas no restrictions were imposed on the Chinese and Indian communities [34]. The prohibition of alcohol consumption in certain communities was hypothesized as one of the reasons for the lower frequency of alcohol-related posts, compared to other lifestyle behaviors.

Findings indicate that positive sentiments significantly outweighed negative sentiments for all lifestyle behaviors. In dietary- and activity-related posts, positive sentiments were found to be twice as many as negative sentiments, which is consistent with the sentiment analysis findings of Shaw et al [36] who analyzed over 1.5 million posts on dietary and exercise topics. In our study, more than half of the sampled posts were either self-narratives or planned actions for self-implementation. It could be postulated that the posts with positive sentiments were driven by self-determination theory (SDT), a comprehensive theory of human motivation and personality that focuses on individuals’ intrinsic tendencies for growth. SDT assumes the importance of autonomous motivation, which is a type of self-emanating motivation that is consistent with users’ innate values to engage in behaviors or pursue a goal [37,38]. The field of autonomous motivation has been extensively studied in the context of dietary and exercise lifestyle behaviors [39-41]. Individuals who are autonomously motivated have a sense of self-control over their actions (eg, choose to exercise regularly), leading to an increase in positive sentiments, personal fulfilment and enjoyment in the actions pursued [40].

In tobacco-related posts, positive sentiments were found to be higher than negative sentiments, albeit with a small percentage difference. This suggests that in tobacco-related discourses, users tend to either feel positive emotions (eg, satisfaction, happiness, and trust) or negative emotions (eg, dissatisfaction, unhappiness, and worry). Mixed sentiments were found to be prevalent in discussions related to vaping among both the scientific community [42], and the general public [43]. In Malaysia, the Health Ministry has proposed the Generational End Game plan, which would ban tobacco sales for those born after 2005. The bill was first tabled at the country’s parliamentary discussions in July 2022 and has yet to be finalized at the time when the social media posts were scraped from X [44]. Such uncertainties towards health policy changes have generated both positive and negative reactions, with the issue being debated constantly throughout

the year. Users either praised the government’s efforts to mitigate smoking behaviors or expressed concerns about such “untested” plans [45]. The negative reactions may also stemmed from users’ awareness of the adverse effects of smoking, with over 90% of male lung cancer patients in Malaysia having a significant history of smoking [46]. In addition, 3500 out of 10,000 annual deaths were linked to smoking [47].

As discussed, the predominance of self-narratives in posts related to diet, physical activity and alcohol consumption is likely due to users’ autonomous motivation and self-awareness to perform a behavior. Conversely, most posts related to tobacco were found to be linked to narratives of other users. As tobacco smoking has been associated with social stigmatization due to its negative health impact on others, users may have been more reluctant to post from a first-person perspective. Instead, users opt to openly discuss the smoking habits of others. In addition, sharing experiences in a third-person perspective may be preferred by users to maintain anonymity. During the 20th century, smokers were often viewed as “mysterious” or “cool,” but this social status has slowly diminished over the past two decades [48,49]. In Malaysia, this was propelled by smoking reduction strategies, such as the ban on smoking in public eateries implemented in 2019, that socially impacted users’ impression toward smoking [50].

In recent years, there has been a growing trend of health influencers using online platforms to actively share their dietary and fitness regimens. Previous studies have shown that social media users who were exposed to this information delivered by health influencers as well as content from other social media users, were more likely to be receptive to adopting healthy practices, such as maintaining a balanced diet and being physically active [51-53]. The results were consistent with the findings of this study, which showed a significantly higher percentage of dietary- and activity-related posts by social media users that were aligned with recommended health practices. Nevertheless, HCPs must remain active in advocating positive lifestyle behaviors on social media. Although almost one-fifth of posts for these two lifestyle behaviors were on planned actions, this may not always translate into actions by the population. This is a caveat of much research that relies on social media or self-reported data on social media. It is often unclear whether individuals actually follow through on what they post about, highlighting the intention-behavior gap [54]. Findings from the Malaysian National Health and Morbidity Survey (NHMS) survey conducted in 2023 have shown that the actual adoption of healthy practices was still lacking among the Malaysian public. Almost 95.1% Malaysian adults did not meet the recommended daily intake of fruits and vegetables, consuming only two servings of fruit or vegetables daily instead of the recommended five servings daily [3]. The prevalence of physical inactivity among Malaysian adults was at 29.9% [3], which was also considerably higher than other Asian countries, including China and India [55,56].

There was a lack of significant association between sentiment class and alignment with recommended health

practices in both tobacco- and alcohol-related posts. Despite the smaller number of sampled alcohol-related posts, it is interesting to note that users' perceptions with recommended health practices had contrasting outcomes compared to the other three lifestyle behaviors. In more than half of the alcohol-related posts with positive and negative sentiments, users' perceptions were not aligned with recommended health practices. Most users perceived alcohol consumption as a casual and an affordable social activity and did not acknowledge the potential health risks involved. A survey conducted in Thailand, a country of similar income setting, had previously mentioned the popularity of alcohol being a social activity among urban communities [57]. In Malaysia, alcoholic beverages were available for purchase at neighborhood convenience stores, which allowed for easy purchases of takeaway alcohol [35]. This further downplayed users' awareness of the negative consequences of alcohol consumption [58].

Assessment of posts made by social media users on X allows HCPs to identify priority areas for social media-based health information delivery on this platform. As most alcohol-related posts do not align with health recommendations, it is postulated that greater emphasis should be placed on strategies to limit alcohol consumption among users in Malaysia. The WHO has proposed collaborative efforts with HCPs and journalists to improve targeted public health messaging to the public. A guide was recently developed for journalists to facilitate media reporting to communities on the harms of alcohol consumption [58]. While the other 3 lifestyle behaviors were mostly aligned with recommended health practices, it remains essential for HCPs to continuously deliver information advocating healthy behaviors. Online approaches allow HCPs to deliver information beyond geographic barriers, reaching a wider audience in diverse community settings. Therefore, health information can be adopted by users in countries with similar cultural beliefs, including countries within the Southeast Asia region.

## Strengths and Limitations

The strengths of this study included its comprehensive coverage of 4 lifestyle behaviors aimed at reducing the 4 key modifiable risk factors under WHO's health priority [4]. This allowed for the simultaneous analysis of posts across different lifestyle behaviors. Unlike most studies that focus on global contexts, this research uniquely focused on the Malaysian context. It provides insights into the cultural and social dynamics that influence discussions around lifestyle behaviors in this specific region. Notably, the inclusion of alcohol-related posts in analysis shed light on culturally and socially nuanced discussions within the region. This is particularly valuable in regions like Malaysia, where religious and cultural factors strongly influence alcohol consumption. In addition, the retrospective examination of social media posts utilized approaches of lexicon-based sentiment analysis and manual content analysis. The dual approach provided real-time and spontaneous insights into users' opinions on lifestyle behaviors while addressing limitations of single-method studies. Findings from this study could assist HCPs in

prioritizing the delivery of region-specific health information through social media.

Nevertheless, a few limitations should be considered. First, potential bias may exist during the selection of social media dataset. In self-selection bias, users who choose to share their opinions on social media may not represent the broader population. The study may be subjected to data selection bias as it included only social media posts in Malay and English, excluding other spoken languages in Malaysia, such as Chinese and Tamil. Nevertheless, sentiment analysis studies on X in Malaysia have largely concentrated on data scraped in Malay and English [59,60]. Demographic bias may be present due to the overrepresentation or underrepresentation of certain groups on X. For instance, the majority of users in Malaysia who post on social media are aged between 25 and 34 years old [9]. The limitations in the availability of metadata on X also prevented the collection of demographic data such as age, gender and race, as most users did not disclose this information in their profiles. Population bias may occur in geotagged posts utilizing longitude and latitude metadata. Previously literature has indicated that only 1% of users would geotag their location in posts [61]. Nevertheless, this is the most effective method to scrape posts that are published within a specific location.

Second, there are limitations in the study design and methods used for sentiment analysis and manual content analysis. The study is cross-sectional in nature and provides a snapshot of discussions at a specific time. Therefore, temporal bias may exist, making it challenging to track changes in sentiments or behaviors over time. In addition, posts that were collected for sentiment analysis and manual content analysis over two consecutive months may not accurately reflect year-round sentiments and discussions, as findings may vary due to the presence of health-related events occurring at certain times of the year. The events may include a change in legislations, prominent public health campaigns or disease outbreaks. The quality of the dataset was ensured by verifying that there were no notable health-related occurrences between November and December 2022. In addition, previous studies analyzing users' sentiments and content have similarly explored health data over two consecutive months [23,24]. The manual content analysis of social media posts can be time-consuming due to the involvement of large datasets, therefore, only 20% of the total posts were randomly selected using stratified sampling. This percentage was previously utilized in a content analysis study by Mathieson et al [30]. While analyzing the full dataset would provide more comprehensive findings, the randomized sample offers a reliable snapshot for identifying the thematic content without the substantial time and resource demands of manual analysis for the entire dataset. Furthermore, prior to sentiment analysis, the computer-assisted translation of posts from Malay to English may have led to inaccuracies due to the usage of local dialects, sarcasm or slangs. To enhance sentiment labeling, the structures of translated posts with neutral sentiments that were unclear were manually refined, and sentiment analysis was repeated.

Third, the study results should be interpreted with caution, regarding posts on alcohol consumption due to the smaller sample size of 112 posts. A power analysis indicated that this sample size is adequate for detecting effects, with a power of 0.82. Furthermore, we acknowledge the presence of potential interactions in posts with overlapping lifestyle behaviors (eg, a post that talks about diet and physical activity). In sentiment analysis, the conduct of Pearson chi-square tests also did not account for potential confounding factors or interactions in posts with overlapping lifestyle behaviors. To account for this limitation, we compared the proportions between sentiment count for posts showing 1 lifestyle behavior only ( $n=3180$ ), and sentiment count for posts across 4 types of lifestyle behaviors ( $n=3320$ ). The proportions of sentiment counts for both were similar to each other. In addition, while many of the Pearson chi-square associations were significant, these may not imply causality and thus may not inform categorically that the observed sentiments result in practicing different lifestyle behaviors or the direction of the relationship.

### ***Implications and Further Research***

The findings from this study could help HCPs to prioritize the delivery of health information on lifestyle behaviors using social media tailored to the targeted region, which is Malaysia. Given the low number of alcohol-related posts by social media users in Malaysia, HCPs could focus on initiating positive discussions around this topic to raise awareness about the harmful effects of alcohol consumption. In addition, most of the alcohol-related posts made by social media users were not aligned with recommended health practices. There is an increased need for HCPs to emphasize on limiting and stopping alcohol consumption, while also acknowledging that the users' attitudes towards alcohol consumption may still vary among different religions in Malaysia. Health advocacy for positive lifestyle behaviors on social media should continue for the other three lifestyle behaviors.

Further research could be proposed to explore the opinions of social media users toward lifestyle behaviors in Malaysia. First, despite the statistical significance observed in the associations between sentiment classification and lifestyle behaviors, the percentage difference between both sentiment classes in tobacco-related posts was small. Therefore, it would be interesting to investigate whether tobacco sentiments would vary over time. We may want to further track sentiments by time series analysis to explore changes in users' emotions towards tobacco across a time period. The tracking of real-time sentiments across a time period was previously conducted in a review examining public health data on X that included posts on alcohol consumption [62]. In addition, since posts are scraped based on location meta-data, future studies could leverage on this data to explore the relationship between the prevalence of specific lifestyle behaviors in certain locations (eg, urban areas in Malaysia) and the intensity of lifestyle behavior-related discussions on social media. A similar study has previously been conducted

in the United States; therefore, conducting such studies in the Malaysian context would be beneficial [63].

Second, the majority of posts involved content related to self-narratives of lifestyle behaviors. These self-narratives outlined X's roles as a microblog for users to freely express the behaviors they practice from a first-person perspective. As self-narratives encompass a broad and generalized category, it may be beneficial to conduct a more detailed examination of posts that only described users' self-narratives. This in-depth analysis would provide insights into the specific themes commonly discussed by users from a first-person perspective. In addition, the examination of posts could be extended to other lifestyle behaviors such as sleep patterns, which is particularly relevant as active social media users are mainly adolescents and young adults who are commonly affected by sleep-related issues [64].

Third, this study was conducted on the microblogging platform X. It is also important to examine social media posts made by users on other platforms, such as Facebook. Future research is proposed to analyze the sentiments and content of posts on these platforms. Audience demographics can vary across these platforms. For instance, younger millennials may be more active on X, whereas Facebook often attracts a slightly older audience [65,66]. Comparing our study findings with those obtained from Facebook could help HCPs to deliver health messages that suit the audiences of different social media platforms.

Fourth, our study emphasizes accessibility and simplicity in data visualization and reporting to effectively communicate findings to a diverse audience, including non-technical stakeholders such as HCPs, public health practitioners and policymakers. To achieve this, we employed techniques like word clouds, which provide a visually appealing representation of frequently mentioned terms in the dataset, and lexicon-based sentiment analysis, which is straightforward to implement as it does not require additional labeled data or extensive training. We recognize the potential value of more advanced methods and suggest exploring these techniques in future studies related to the conduct of in-depth text analysis. These may include approaches like topic modeling or keyword co-occurrence analysis to summarize text data through word groups, as well as training machine learning models such as support vector machines or Naïve Bayes to classify sentiments. Furthermore, hybrid methods of sentiment analysis could be explored by integrating machine learning models with lexicon-based approaches. These combined models can then be assessed for accuracy and robustness through comparative analysis. Similar studies have been conducted previously in both health and non-health posts [67,68].

### ***Conclusion***

In conclusion, the incorporation of lexicon-based sentiment analysis holds significance as it enabled the use of large amounts of data to capture users' emotions whilst posting on lifestyle behaviors. Positive sentiments were significantly expressed in posts for all lifestyle behaviors. Nevertheless, there was a small percentage difference observed in

tobacco-related posts, indicating a more varied sentiment among users. Most of the posts showed users' own narratives and planned actions towards the conduct of a behavior. As the majority of alcohol-related discussions were not aligned with recommended health practices, this reflects the need for individual HCPs and health organizations to increase their

delivery of health information pertaining to alcohol consumption on social media platforms. It is also equally important for HCPs to continue providing health information on other lifestyle behaviors to social media users, while monitoring ongoing discussions by users on social media.

---

## Acknowledgments

The authors would like to thank the director-general of Health Malaysia for his permission to publish this article. The work was supported by the Ministry of Higher Education of Malaysia's Fundamental Research Grant Scheme under grant FRGS/1/2020/SS0/UKM/02/11. The funders played no role in study design, collection, analysis, interpretation of data, or writing of the report.

---

## Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

---

## Authors' Contributions

YYY, MRY, and WWC contributed to the conception or design of the study. YYY and MIAL contributed to data collection of the study. YYY, MIAL, and WWC contributed to data analysis of the study. All authors (YYY, MRY, MM-B, MIAL, and WWC) contributed to data interpretation and provided scientific inputs and technical improvement. YYY drafted the manuscript while MRY and WWC guided the revisions. All authors (YYY, MRY, MM-B, MIAL, and WWC) read and approved the final version for publication.

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

List of keywords for data scraping of posts.

[\[DOCX File \(Microsoft Word File\), 19 KB-Multimedia Appendix 1\]](#)

---

## Multimedia Appendix 2

Codebook.

[\[DOCX File \(Microsoft Word File\), 28 KB-Multimedia Appendix 2\]](#)

---

## Multimedia Appendix 3

Word cloud representation of overall X dataset (n=3320).

[\[PNG File \(Portable Network Graphics File\), 2513 KB-Multimedia Appendix 3\]](#)

---

## Multimedia Appendix 4

Examples of posts with positive and negative sentiments according to each lifestyle behavior.

[\[DOCX File \(Microsoft Word File\), 21 KB-Multimedia Appendix 4\]](#)

---

## Multimedia Appendix 5

Example of posts selected for manual content analysis.

[\[DOCX File \(Microsoft Word File\), 23 KB-Multimedia Appendix 5\]](#)

---

## References

1. Noncommunicable diseases. World Health Organization. 2022. URL: <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases> [Accessed 2022-12-06]
2. Malaysian Burden of Disease and Injury Study 2009-2014. Institute for Public Health; 2017. URL: <https://iku.moh.gov.my/images/IKU/Document/REPORT/BOD/BOD2009-2014.pdf> [Accessed 2022-12-06]
3. National Health and Morbidity Survey (NHMS) 2023: Non-Communicable Diseases and Healthcare Demand. Institute for Public Health; 2024. URL: <https://iku.nih.gov.my/images/nhms2023/key-findings-nhms-2023.pdf> [Accessed 2024-05-24]
4. Reducing modifiable risk factors for noncommunicable diseases. World Health Organization. 2022. URL: <https://www.who.int/westernpacific/activities/reducing-modifiable-risk-factors-for-noncommunicable-diseases> [Accessed 2022-12-06]



5. Miyamoto SW, Henderson S, Young HM, Pande A, Han JJ. Tracking health data is not enough: a qualitative exploration of the role of healthcare partnerships and mhealth technology to promote physical activity and to sustain behavior change. *JMIR Mhealth Uhealth*. Jan 20, 2016;4(1):e5. [doi: [10.2196/mhealth.4814](https://doi.org/10.2196/mhealth.4814)] [Medline: [26792225](https://pubmed.ncbi.nlm.nih.gov/26792225/)]
6. Islam SMS, Tabassum R, Liu Y, et al. The role of social media in preventing and managing non-communicable diseases in low-and-middle income countries: Hope or hype? *Health Policy Technol*. Mar 2019;8(1):96-101. [doi: [10.1016/j.hlpt.2019.01.001](https://doi.org/10.1016/j.hlpt.2019.01.001)]
7. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. *P T*. Jul 2014;39(7):491-520. [Medline: [25083128](https://pubmed.ncbi.nlm.nih.gov/25083128/)]
8. Balicer RD, Luengo-Oroz M, Cohen-Stavi C, et al. Using big data for non-communicable disease surveillance. *Lancet Diabetes Endocrinol*. Aug 2018;6(8):595-598. [doi: [10.1016/S2213-8587\(17\)30372-8](https://doi.org/10.1016/S2213-8587(17)30372-8)] [Medline: [29146206](https://pubmed.ncbi.nlm.nih.gov/29146206/)]
9. Digital 2024: Malaysia. Data Reportal. 2024. URL: <https://datareportal.com/reports/digital-2024-malaysia> [Accessed 2024-04-24]
10. Tonkin EL. Tonkin EL, Tourte GJL, editors. *A Day at Work (with Text): A Brief Introduction*. 1st ed. Chandos Publishing; 2016:23-60. ISBN: 978-1-84334-749-1
11. van Atteveldt W, van der Velden M, Boukes M. The validity of sentiment analysis: comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. *Commun Methods Meas*. Apr 3, 2021;15(2):121-140. [doi: [10.1080/19312458.2020.1869198](https://doi.org/10.1080/19312458.2020.1869198)]
12. Abualigah L, Alfar HE, Shehab M, Hussein AMA. Sentiment analysis in healthcare: a brief review. In: Abd Elaziz M, Al-qaness MAA, Ewees AA, Dahou A, editors. *Recent Advances in NLP: The Case of Arabic Language*. Springer International Publishing; 2020:29-141. ISBN: 978-3-030-34614-0
13. Lu X, Sun L, Xie Z, Li D. Perception of the food and drug administration electronic cigarette flavor enforcement policy on twitter: observational study. *JMIR Public Health Surveill*. Mar 29, 2022;8(3):e25697. [doi: [10.2196/25697](https://doi.org/10.2196/25697)] [Medline: [35348461](https://pubmed.ncbi.nlm.nih.gov/35348461/)]
14. Shamo E, Turdybay A, Shamo P, Akhmetov I, Jaxylykova A, Pak A. Sentiment analysis of vegan related tweets using mutual information for feature selection. *PeerJ Comput Sci*. 2022;8:e1149. [doi: [10.7717/peerj-cs.1149](https://doi.org/10.7717/peerj-cs.1149)] [Medline: [36532810](https://pubmed.ncbi.nlm.nih.gov/36532810/)]
15. Rintyarna B. Mapping acceptance of indonesian organic food consumption under COVID-19 pandemic using sentiment analysis of Twitter dataset. *J Theor Appl Inf Technol*. 2021;99:1009-1019. URL: <https://www.jatit.org/volumes/Vol99No5/1Vol99No5.pdf> [Accessed 2022-12-12]
16. Kaity M, Balakrishnan V. An integrated semi-automated framework for domain-based polarity words extraction from an unannotated non-English corpus. *J Supercomput*. Dec 2020;76(12):9772-9799. [doi: [10.1007/s11227-020-03222-0](https://doi.org/10.1007/s11227-020-03222-0)]
17. Mogaji E, Balakrishnan J, Kieu TA. Examining consumer behaviour in the UK Energy sector through the sentimental and thematic analysis of tweets. *J of Consumer Behaviour*. Mar 2021;20(2):218-230. [doi: [10.1002/cb.1820](https://doi.org/10.1002/cb.1820)]
18. Karmegam D, Mappillairaju B. Social media analytics and reachability evaluation - #Diabetes. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*. Jan 2022;16(1):102359. [doi: [10.1016/j.dsx.2021.102359](https://doi.org/10.1016/j.dsx.2021.102359)]
19. Najafizada M, Rahman A, Donnan J, Dong Z, Bishop L. Analyzing sentiments and themes on cannabis in Canada using 2018 to 2020 Twitter data. *J Cannabis Res*. Apr 13, 2022;4(1):22. [doi: [10.1186/s42238-022-00132-1](https://doi.org/10.1186/s42238-022-00132-1)] [Medline: [35418169](https://pubmed.ncbi.nlm.nih.gov/35418169/)]
20. Kasson E, Singh AK, Huang M, Wu D, Cavazos-Rehg P. Using a mixed methods approach to identify public perception of vaping risks and overall health outcomes on Twitter during the 2019 EVALI outbreak. *Int J Med Inform*. Nov 2021;155:104574. [doi: [10.1016/j.ijmedinf.2021.104574](https://doi.org/10.1016/j.ijmedinf.2021.104574)] [Medline: [34592539](https://pubmed.ncbi.nlm.nih.gov/34592539/)]
21. Chatterjee A, Prinz A, Gerdes M, Martinez S. Digital Interventions on healthy lifestyle management: systematic review. *J Med Internet Res*. Nov 17, 2021;23(11):e26931. [doi: [10.2196/26931](https://doi.org/10.2196/26931)] [Medline: [34787575](https://pubmed.ncbi.nlm.nih.gov/34787575/)]
22. Kohl LFM, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. *J Med Internet Res*. Jul 16, 2013;15(7):e146. [doi: [10.2196/jmir.2665](https://doi.org/10.2196/jmir.2665)] [Medline: [23859884](https://pubmed.ncbi.nlm.nih.gov/23859884/)]
23. Ong SQ, Pauzi MBM, Gan KH. Text mining and determinants of sentiments towards the COVID-19 vaccine booster of Twitter users in Malaysia. *Healthcare (Basel)*. May 27, 2022;10(6):994. [doi: [10.3390/healthcare10060994](https://doi.org/10.3390/healthcare10060994)] [Medline: [35742045](https://pubmed.ncbi.nlm.nih.gov/35742045/)]
24. Kent EE, Prestin A, Gaysynsky A, et al. "Obesity is the New Major Cause of Cancer": connections between obesity and cancer on Facebook and Twitter. *J Cancer Educ*. Sep 2016;31(3):453-459. [doi: [10.1007/s13187-015-0824-1](https://doi.org/10.1007/s13187-015-0824-1)] [Medline: [25865399](https://pubmed.ncbi.nlm.nih.gov/25865399/)]
25. Hutto C, Gilbert E, editors. *VADER: a parsimonious rule-based model for sentiment analysis of social media text*. Presented at: Proceedings of the International AAAI Conference on Web and Social Media; Jun 1-4, 2014; Ann Arbor, Michigan. [doi: [10.1609/icwsm.v8i1.14550](https://doi.org/10.1609/icwsm.v8i1.14550)]

26. Elbagir S, Yang J, editors. Twitter sentiment analysis using natural language toolkit and VADER sentiment. Presented at: Proceedings of the International Multiconference of Engineers and Computer Scientists; Mar 13-15, 2019; Hong Kong. [doi: [10.1142/9789811215094\\_0005](https://doi.org/10.1142/9789811215094_0005)]
27. Wan Min WNS, Zulkarnain NZ. Comparative evaluation of lexicons in performing sentiment analysis. *J Adv Res Comput Tech Software Appl*. Jun 18, 2020;2(1):14-20. URL: <https://jacta.utem.edu.my/jacta/article/view/5207/3684> [Accessed 2022-12-12]
28. Agarwal A, Xie B, Vovsha I, Rambow O. Sentiment analysis of Twitter data. Presented at: Proceedings of the Workshop on Languages in Social Media; Jun 23, 2011; Portland, Oregon.
29. Emblem H. Are you scared, VADER? Understanding how NLP pre-processing impacts VADER scoring. Medium. 2021. URL: <https://medium.com/data-science/are-you-scared-vader-understanding-how-nlp-pre-processing-impacts-vader-scoring-4f4edadb91d> [Accessed 2023-07-19]
30. Mathieson S, O'Keeffe M, Traeger AC, Ferreira GE, Abdel Shaheed C. Content and sentiment analysis of gabapentinoid-related tweets: an infodemiology study. *Drug Alcohol Rev*. Jan 2024;43(1):45-55. [doi: [10.1111/dar.13590](https://doi.org/10.1111/dar.13590)] [Medline: [36539307](https://pubmed.ncbi.nlm.nih.gov/36539307/)]
31. Miller CA, Jung Kim S, Schwartz-Bloom RD, Bloom PN, Murphy SK, Fuemmeler BF. Informing women about the risks of exposing babies to tobacco smoke: outreach and education efforts using Facebook "boost posts". *Transl Behav Med*. May 26, 2022;12(5):714-720. [doi: [10.1093/tbm/ibab158](https://doi.org/10.1093/tbm/ibab158)] [Medline: [35532323](https://pubmed.ncbi.nlm.nih.gov/35532323/)]
32. Healthy living: what is a healthy lifestyle? WHO Regional Office for Europe. 1999. URL: [https://iris.who.int/bitstream/handle/10665/108180/EUR\\_ICP\\_?jsessionid=A62C6B339873362BC61C7C8BC6088C38?sequence=1](https://iris.who.int/bitstream/handle/10665/108180/EUR_ICP_?jsessionid=A62C6B339873362BC61C7C8BC6088C38?sequence=1) [Accessed 2022-12-06]
33. Norimah AK, Safiah M, Jamal K, Haslinda S, Zuhaida H, Rohida S, et al. Food consumption patterns: Findings from the Malaysian Adult Nutrition Survey (MANS). *Malays J Med Sci*. Mar 2008;14(1). URL: <https://nutriweb.org.my/mjn/publication/14-1/b.pdf> [Accessed 2022-12-12]
34. Kortteinen T. Alcohol in Malaysia: the impact of social transformation. *Contemp Drug Probl*. Sep 1999;26(3):391-411. [doi: [10.1177/009145099902600303](https://doi.org/10.1177/009145099902600303)]
35. Robert Lourdes TG, Abd Hamid HA, Riyadzi MR, et al. Findings from a nationwide study on alcohol consumption patterns in an upper middle-income country. *Int J Environ Res Public Health*. Jul 21, 2022;19(14):35886700. [doi: [10.3390/ijerph19148851](https://doi.org/10.3390/ijerph19148851)] [Medline: [35886700](https://pubmed.ncbi.nlm.nih.gov/35886700/)]
36. Shaw G, Zimmerman M, Vasquez-Huot L, Karami A. Deciphering latent health Information in social media using a mixed-methods design. *Healthcare (Basel)*. Nov 19, 2022;10(11):2320. [doi: [10.3390/healthcare10112320](https://doi.org/10.3390/healthcare10112320)] [Medline: [36421644](https://pubmed.ncbi.nlm.nih.gov/36421644/)]
37. Hagger MS, Hardcastle SJ, Chater A, Mallett C, Pal S, Chatzisarantis NLD. Autonomous and controlled motivational regulations for multiple health-related behaviors: between- and within-participants analyses. *Health Psychol Behav Med*. Jan 1, 2014;2(1):565-601. [doi: [10.1080/21642850.2014.912945](https://doi.org/10.1080/21642850.2014.912945)] [Medline: [25750803](https://pubmed.ncbi.nlm.nih.gov/25750803/)]
38. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*. 2000;55(1):68-78. [doi: [10.1037/0003-066X.55.1.68](https://doi.org/10.1037/0003-066X.55.1.68)]
39. Rutten GM, Meis JJM, Hendriks MRC, Hamers FJM, Veenhof C, Kremers SPJ. The contribution of lifestyle coaching of overweight patients in primary care to more autonomous motivation for physical activity and healthy dietary behaviour: results of a longitudinal study. *Int J Behav Nutr Phys Act*. Dec 2014;11(1):86. [doi: [10.1186/s12966-014-0086-z](https://doi.org/10.1186/s12966-014-0086-z)]
40. Teixeira PJ, Patrick H, Mata J. Why we eat what we eat: the role of autonomous motivation in eating behaviour regulation. *Nutr Bull*. Mar 2011;36(1):102-107. [doi: [10.1111/j.1467-3010.2010.01876.x](https://doi.org/10.1111/j.1467-3010.2010.01876.x)]
41. Silva MN, Markland D, Carraça EV, et al. Exercise autonomous motivation predicts 3-yr weight loss in women. *Med Sci Sports Exerc*. Apr 2011;43(4):728-737. [doi: [10.1249/MSS.0b013e3181f3818f](https://doi.org/10.1249/MSS.0b013e3181f3818f)] [Medline: [20689448](https://pubmed.ncbi.nlm.nih.gov/20689448/)]
42. Hartmann-Boyce J. Why can't scientists agree on e-cigarettes. *The Guardian*. 2016. URL: <https://www.theguardian.com/science/sifting-the-evidence/2016/sep/14/why-cant-scientists-agree-on-e-cigarettes-vaping> [Accessed 2023-05-06]
43. Kwon M, Park E. Perceptions and sentiments about electronic cigarettes on social media platforms: systematic review. *JMIR Public Health Surveill*. Jan 15, 2020;6(1):e13673. [doi: [10.2196/13673](https://doi.org/10.2196/13673)] [Medline: [31939747](https://pubmed.ncbi.nlm.nih.gov/31939747/)]
44. Health minster: MOH to review tobacco generational endgame policy. *Malay Mail*. 2022. URL: <https://www.malaymail.com/news/malaysia/2022/12/08/health-minster-moh-to-review-tobacco-generational-endgame-policy/44238> [Accessed 2023-05-06]
45. Mixed sentiments towards smoking 'end game'. *New Straits Times*. 2022. URL: <https://www.nst.com.my/news/nation/2022/07/811829/mixed-sentiments-towards-smoking-end-game> [Accessed 2023-05-06]
46. National strategic plan for cancer control programe, 2021-2025. International Cancer Control Partnership. 2021. URL: [https://www.iccp-portal.org/system/files/plans/National\\_Strategic\\_Plan\\_for\\_Cancer\\_Control\\_Programme\\_2021-2025.pdf](https://www.iccp-portal.org/system/files/plans/National_Strategic_Plan_for_Cancer_Control_Programme_2021-2025.pdf) [Accessed 2024-08-20]

47. Jusoh S, Naing NN, Wan-Arfah N, et al. Prevalence and factors influencing smoking behavior among female inmates in Malaysia. *Healthcare (Basel)*. Jan 9, 2023;11(2):203. [doi: [10.3390/healthcare11020203](https://doi.org/10.3390/healthcare11020203)] [Medline: [36673571](https://pubmed.ncbi.nlm.nih.gov/36673571/)]
48. Evans-Polce RJ, Castaldelli-Maia JM, Schomerus G, Evans-Lacko SE. The downside of tobacco control? Smoking and self-stigma: a systematic review. *Soc Sci Med*. Nov 2015;145:26-34. [doi: [10.1016/j.socscimed.2015.09.026](https://doi.org/10.1016/j.socscimed.2015.09.026)] [Medline: [26439764](https://pubmed.ncbi.nlm.nih.gov/26439764/)]
49. Castaldelli-Maia JM, Ventriglio A, Bhugra D. Tobacco smoking: From “glamour” to “stigma”. A comprehensive review. *Psychiatry Clin Neurosci*. Jan 2016;70(1):24-33. [doi: [10.1111/pcn.12365](https://doi.org/10.1111/pcn.12365)] [Medline: [26449875](https://pubmed.ncbi.nlm.nih.gov/26449875/)]
50. Flashback #star50: when smoking was banned at eateries. *Star*. 2021. URL: <https://www.thestar.com.my/news/nation/2021/10/20/flashback-star50-when-smoking-was-banned-at-eateries#:~:text=Malaysia%20enforced%20a%20no%2Dsmoking,rooms%20and%20ashtrays%20were%20prohibited> [Accessed 2023-05-06]
51. Hawkins LK, Farrow C, Thomas JM. Do perceived norms of social media users’ eating habits and preferences predict our own food consumption and BMI? *Appetite*. Jun 1, 2020;149:104611. [doi: [10.1016/j.appet.2020.104611](https://doi.org/10.1016/j.appet.2020.104611)] [Medline: [31958481](https://pubmed.ncbi.nlm.nih.gov/31958481/)]
52. Durau J, Diehl S, Terlutter R. Motivate me to exercise with you: the effects of social media fitness influencers on users’ intentions to engage in physical activity and the role of user gender. *Digit Health*. 2022;8:20552076221102769. [doi: [10.1177/20552076221102769](https://doi.org/10.1177/20552076221102769)] [Medline: [35615268](https://pubmed.ncbi.nlm.nih.gov/35615268/)]
53. Johnston C, Davis WE. Motivating exercise through social media: is a picture always worth a thousand words? *Psychol Sport Exerc*. Mar 2019;41:119-126. [doi: [10.1016/j.psychsport.2018.12.012](https://doi.org/10.1016/j.psychsport.2018.12.012)]
54. Mohamad Saleh MS, Mehellou A, Huang M, Briandana R. Social media impact on sustainable intention and behaviour: a comparative study between university students in Malaysia and Indonesia. *JARHE*. [doi: [10.1108/JARHE-10-2023-0479](https://doi.org/10.1108/JARHE-10-2023-0479)]
55. Bauman A, Bull F, Chey T, et al. The International Prevalence Study on Physical Activity: results from 20 countries. *Int J Behav Nutr Phys Act*. Mar 31, 2009;6(1):21. [doi: [10.1186/1479-5868-6-21](https://doi.org/10.1186/1479-5868-6-21)] [Medline: [19335883](https://pubmed.ncbi.nlm.nih.gov/19335883/)]
56. Nik-Nasir NM, Md-Yasin M, Ariffin F, et al. Physical activity in Malaysia: are we doing enough? Findings from the REDISCOVER Study. *Int J Environ Res Public Health*. Dec 15, 2022;19(24):16888. [doi: [10.3390/ijerph192416888](https://doi.org/10.3390/ijerph192416888)] [Medline: [36554769](https://pubmed.ncbi.nlm.nih.gov/36554769/)]
57. Assanangkornchai S, Sam-Angsri N, Rerngpongpan S, Lertnakorn A. Patterns of alcohol consumption in the Thai population: results of the National Household Survey of 2007. *Alcohol Alcohol*. 2010;45(3):278-285. [doi: [10.1093/alcalc/agg018](https://doi.org/10.1093/alcalc/agg018)] [Medline: [20348437](https://pubmed.ncbi.nlm.nih.gov/20348437/)]
58. Reporting about alcohol: a guide for journalists. World Health Organization. 2023. URL: <https://iris.who.int/bitstream/handle/10665/366715/9789240071490-eng.pdf?sequence=1> [Accessed 2023-09-06]
59. Abu Samah KAF, Nor Azharludin NM, Riza LS, Hasrol Jono MNH, Moketar NA. Classification and visualization: Twitter sentiment analysis of Malaysia’s private hospitals. *IJ-AI*. Mar 10, 2023;12(4):1793. [doi: [10.11591/ijai.v12.i4.pp1793-1802](https://doi.org/10.11591/ijai.v12.i4.pp1793-1802)]
60. Mohd Yuswardi P, Ahmad NA. Sentiment analysis of Malaysians citizen’s emotion towards cyberbullying in Twitter. *IJARBS*. Apr 2023;13(4):769-780. [doi: [10.6007/IJARBS/v13-i4/16777](https://doi.org/10.6007/IJARBS/v13-i4/16777)]
61. Malik M, Lamba H, Nakos C. Population bias in geotagged tweets. Presented at: Proceedings of the International AAAI Conference on Web and Social Media; May 26-29, 2015; University of Oxford, Oxford, UK. [doi: [10.1609/icwsm.v9i4.14688](https://doi.org/10.1609/icwsm.v9i4.14688)]
62. Lane JM, Habib D, Curtis B. Linguistic methodologies to surveil the leading causes of mortality: scoping review of twitter for public health data. *J Med Internet Res*. Jun 12, 2023;25:e39484. [doi: [10.2196/39484](https://doi.org/10.2196/39484)] [Medline: [37307062](https://pubmed.ncbi.nlm.nih.gov/37307062/)]
63. Gore RJ, Diallo S, Padilla J. You are what you tweet: connecting the geographic variation in America’s obesity rate to Twitter content. *PLoS ONE*. 2015;10(9):e0133505. [doi: [10.1371/journal.pone.0133505](https://doi.org/10.1371/journal.pone.0133505)] [Medline: [26332588](https://pubmed.ncbi.nlm.nih.gov/26332588/)]
64. Owens J, Adolescent Sleep Working Group, Committee on Adolescence. Insufficient sleep in adolescents and young adults: an update on causes and consequences. *Pediatrics*. Sep 2014;134(3):e921-32. [doi: [10.1542/peds.2014-1696](https://doi.org/10.1542/peds.2014-1696)] [Medline: [25157012](https://pubmed.ncbi.nlm.nih.gov/25157012/)]
65. How to determine the best social media platforms for your business. *Emphatic*. 2024. URL: <https://emphatic.co/how-to-determine-the-best-social-media-platforms-for-your-business> [Accessed 2024-01-08]
66. Riserbato R. What a social media target audience is and how to find it. *Hubspot*. 2024. URL: <https://blog.hubspot.com/marketing/social-media-target-audience> [Accessed 2024-01-08]
67. HuangM, Rasool A, Jiang Q, Qu Q, Kamyab M, editors. HSMC: hybrid sentiment method for correlation to analyze COVID-19 tweets. In: *Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery*. Cham: Springer International Publishing; 2021. [doi: [10.1007/978-3-030-89698-0\\_101](https://doi.org/10.1007/978-3-030-89698-0_101)]
68. Rasool A, Tao R, Marjan K, Naveed T. Twitter sentiment analysis: a case study for apparel brands. *J Phys: Conf Ser*. Mar 1, 2019;1176(2):022015. [doi: [10.1088/1742-6596/1176/2/022015](https://doi.org/10.1088/1742-6596/1176/2/022015)]

---

**Abbreviations**

**HCP:** health care professional  
**NCD:** noncommunicable disease  
**NHMS:** National Health and Morbidity Survey  
**SDT:** self-determination theory  
**VADER:** Valence Aware Dictionary and Sentiment Reasoner  
**WHO:** World Health Organization

---

*Edited by Michael Haupt; peer-reviewed by Abdur Rasool, Ross Gore; submitted 24.09.2024; final revised version received 26.01.2025; accepted 13.02.2025; published 25.06.2025*

*Please cite as:*

Yip YY, Yaakub MR, Makmor-Bakry M, Abu Latiffi MI, Chong WW  
*Exploring Social Media Posts on Lifestyle Behaviors: Sentiment and Content Analysis*  
JMIR Infodemiology 2025;5:e65835  
URL: <https://infodemiology.jmir.org/2025/1/e65835>  
doi: [10.2196/65835](https://doi.org/10.2196/65835)

© Yan Yee Yip, Mohd Ridzwan Yaakub, Mohd Makmor-Bakry, Muhammad Iqbal Abu Latiffi, Wei Wen Chong. Originally published in JMIR Infodemiology (<https://infodemiology.jmir.org>), 25.06.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on <https://infodemiology.jmir.org/>, as well as this copyright and license information must be included.