Original Research

Perceptions and Key Attributes of Exercise Among Korean Men and Women During the COVID-19 Pandemic: A Comparative Study Using Big Data Analysis

Sung-Un Park[1,2], Hye-Ri Song[2], Yoon-Kwon Yang[2,3,*]

[1]Department of Sport & Leisure Studies, Shingyeong University, 18274 Hwaseong City, Republic of Korea
[2]Department of Physical Education, Graduate School, Sungshin Women's University, 02844 Seoul, Republic of Korea
[3]Department of Sport & Leisure Studies, Sungshin Women's University, 02844 Seoul, Republic of Korea
*Correspondence: yangyk@sungshin.ac.kr (Yoon-Kwon Yang)

Abstract

Background: The COVID-19 pandemic has increased risks to people's health due to restrictions on movement and access to exercise facilities. This study compares and analyzes perceptions and key attributes of exercise among Korean men and women in the context of the COVID-19 pandemic by using data collected from social media. **Methods**: Data were collected from January 2020 to June 2021. The search keywords were classified into "COVID-19, men, and exercise" and "COVID-19, women, and exercise", and data were collected through Textom (The Imc Inc., Daegu, Republic of Korea), a big data collection and analysis platform. Frequency analysis, TF-IDF analysis, and degree centrality analysis were performed using Textom. In addition, a CONCOR (CONvergent and CORrelation) analysis was performed using UCINET 6 (Analytic Technologies Corp., Lexington, KY, USA) to visualize related words and analyze clusters. **Results**: In total, 10,703 data points of 5.935 MB were collected and analyzed. The top 50 words related to men's and women's exercises of COVID-19 were derived through frequency analysis, TF-IDF analysis, and degree centrality analysis. We also derived four common clusters (exercise method, COVID-19 situation, friends, and sports). In addition, three clusters (women's fitness, children and hashtags) for men's exercise and two clusters (weight loss and Olympics) for women's exercise were derived separately. **Conclusions**: The study results show that it is possible to provide suggestions and health improvement plans for exercise based on gender. In addition, our findings can serve as an academic basis for developing health promotion programs related to exercise during the COVID-19 pandemic, which can contribute to reducing the economic costs and health risks associated with the pandemic.

Keywords: COVID-19; men; women; exercise; Korea; social media

1. Introduction

The COVID-19 pandemic poses a worldwide threat to health and life [1]. As of December 13, 2021, 270,164,158 people have been infected by the virus and 5,306,162 people have died [2]. COVID-19 has impacted all levels of the economy, society, culture, education, and people's daily lives [3,4]. The social distancing protocols and national lockdowns implemented in many countries during the pandemic are of concern due to a potential decrease in overall physical activity (PA) levels, as well as an increase in sedentary behaviors [5].

As of February 15, 2022, Korea reported 57,117 active COVID-19 cases, 1,462,421 infections, and 7163 deaths (0.490%). Korea's metropolitan areas are densely populated, and COVID-19 outbreak rates are approaching 60% [6]. A fourth major wave began with the increasing number of confirmed COVID-19 cases [7]. As COVID-19 outbreaks continue to increase, public and private facilities, including health centers, gyms, and parks have been closed in many countries, including Korea [5].

Regular PA (e.g., exercise) has many benefits [8–12]. It can also reduce negative emotions—such as anxiety and depression—caused by social distancing [13]. The World Health Organisation (WHO), recommends that adults achieve a minimum of 150 min of moderate or 75 min of vigorous activity or equivalent PA per week [14]. Thus, regular exercise is one of the most important ways to increase PA [15,16]. Current studies indicate that the lack of PA has negative health effects and contributes to risk factors related to COVID-19 [17–19]. The results of these studies highlight the need to maintain an adequate level of PA [20]. New findings suggest that exercise may reduce risk factors such as acute respiratory distress symptoms, one of the leading causes of death in COVID-19 patients [21,22].

Therefore, there are concerns about the overall decrease in PA due to social distancing and restrictions to movement implemented in many countries during the COVID-19 pandemic due to their negative impact on people's physical, psychological, and social health [23]. Healthy individuals are forced to stay at home due to the COVID-19 pandemic; isolation and restrictions presents new challenges to achieve PA recommendations. [24]. Also, men and women report different exercise habits and reasons for exercising [25]. Moreover it is difficult for South Koreans to participate in exercise due to social dis-

tancing and COVID-19. Therefore, a lack of exercise exposes many Koreans to obesity and geriatric diseases.

Big data refers to very large and complex data sets that transcend the capabilities of existing data management systems [26]. Furthermore, big data can store and process data economically and efficiently. The use and analysis of big data can bring significant financial value to various industries and fields, including medical, healthcare, and life science research [26]. Therefore, modern data are typically generated faster than in the past, meaning data now have near real-time velocity and greater variety. Consequently, analysis techniques using big data are becoming increasingly diversified [27–29]. In addition, big data analysis has research potential in various fields, such as discourse analyses on social phenomena, cognitive analyses, securing product competitiveness through analyses of social trend changes, and creating value through the examination of new convergences. Therefore, big data has the potential to revolutionize many domains, including public health [30].

The purpose of this study is to compare and analyze perceptions and key attributes of exercise of Korean men and women in social media in the context of the COVID-19 pandemic. Through this big data analysis, perceptions and key attributes of the exercise and PA of men and women in Korea during the COVID-19 pandemic can be derived, proposing a strategic focus for exercise and PA for men and women.

2. Materials and Methods

2.1 Data Collection

The data were collected from January 1, 2010 to December 31, 2019; the search language was Korean. Data collection keywords were collected separately for "COVID-19, men, and exercise" (men's exercise) and "COVID-19, women, and exercise" (women's exercise). The data collection channel targeted unstructured text data content provided by Naver and Google. Naver (Naver Corp., Seongnam, Republic of Korea) and Google (Alphabet Inc., Mountain View, USA) as they are the most commonly used search engines in Korea [31]. The study used the data collection tool, TEXTOM (The Imc Inc., Daegu, Republic of Korea), a web crawling and analyzing program based on the Korean language. The information on the data collection is summarized in Table 1.

Table 1. Data collection.

Collection category	Content
Period	January 1, 2020 to June 30, 2021
Language	Korean
Keywords	COVID-19, Men, Exercise
	COVID-19, Women, Exercise
Channel	Naver, Google
Tool	TEXTOM (http://textom.co.kr)

2.2 Research Process

First, data were collected and purified using Textom 4.5's big data collection and analysis solution. Second, the top 50 words were derived through a text mining analysis that included both frequency and term frequency-inverse document frequency (TF-IDF) analyses of the refined data. Third, the selected data were converted into a matrix format to perform the degree centrality analysis, with a convergence of iterated correlations (CONCOR) analysis performed using Ucinet 6 (Analytic Technologies Corp., Lexington, KY, USA). Fourth, the derived data were visualized using tables and figures. The research process is described in Fig. 1.
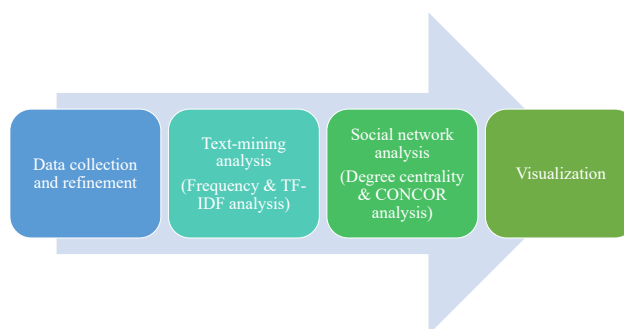

Fig. 1. Research process.

2.3 Data Analysis

In this study, (1) text mining and a (2) social network analysis (SNA) were performed to analyze the perception and key attributes of the men and women's exercises of COVID-19.

(1) Text mining is a field that uses algorithms to read and analyze unstructured text data. Specifically, it is defined as the discovery and extraction of useful knowledge from unstructured texts [32]. Generally, unstructured data refer to data generated in the form of text, pictures, and images from various sources. In modern society, with the popularization of smartphones and the development of the internet, unstructured data generated from platforms such as electronic newspaper articles, social media, and YouTube are increasing exponentially. Texts covered in news articles contain social perspectives on specific issues [33], and comments on social media and YouTube contain the thoughts and experiences of the general public [34].

Text mining can collect varied opinions across numerous categories based on freely written text, revealing psychological aspects that are difficult to capture in quantitative research [35]. Text mining analysis is the process of discovering new knowledge from a large volume and variety of a selected set of digital text [36–39] by extracting relevant terms, calculating their frequency of use, and finding their meaning [40]. Consequently, we used text mining

analysis using both frequency and TF-IDF analyses: Frequency analysis calculates the number of times a term appears in a given document, and the TF-IDF approach is used to weigh each term according to its uniqueness in a text document [41,42]. Thus, the TF-IDF value increases as the frequency of a word in a document increases, while the number of documents including the specific word decreases [41].

(2) SNA is a strategy for investigating social structures [43]; it is a method of quantitatively analyzing the characteristics of objects that constitute a social network model using nodes and links [44]. The SNA analysis in this study was performed using degree centrality and CONCOR analyses. Degree centrality refers to the sum of nodes connected to a central node in a network. This makes it possible to determine the number of nodes connected to a central node. Furthermore, CONCOR analysis is a method used to identify blocks of nodes according to the correlation of a matrix with words appearing simultaneously; it establishes a relationship between keywords and forms a cluster by identifying the relationships between blocks [45]. Therefore, CONCOR analysis was performed to analyze the relationships among the latent sub-clusters and determine the relationships and patterns between words. The greater the similarity between the relationships and patterns, the greater the degree of structural equivalence of other words [46]. In this study, clusters were derived from data collected using CONCOR analysis.

TEXTOM 4.5's big data analysis solution and the Netdraw visualization tool of the UCINET 6 social network analysis software (Analytic Technologies Corp., Lexington, KY, USA) [27] were used to perform the text mining and SNA.

3. Results

3.1 Results of the Data Collection

The data related to COVID-19 exercise for men and women are shown in Table 2. The number of men's data points was 5228 and the volume was 2.972 MB. The number of women's data points was 5475, and the volume was 2.963 MB. The number of datasets was slightly higher for women; however, the volume was slightly higher for men. In total, 10,703 data points and 5.935 MB were collected using TEXTOM.

Table 2. Men's and women's COVID-19 exercise-related datasets.

	Number of data points	Volume
Men's exercise	5228	2.972 MB
Women's exercise	5475	2.963 MB
Total	10,703	5.935 MB

3.2 Text Mining Analysis

Table 3 summarizes the results of the frequency analysis of the top 50 terms related to exercise for men and women in Korea during the COVID-19 pandemic.

Table 3. Results of the frequency analysis.

Rank	Men's exercise		Women's exercise	
	Term	Freq.	Term	Freq.
1	COVID-19	5070	COVID-19	4865
2	Exercise	4732	Exercise	4629
3	Men	2865	Women	2816
4	Home	665	Weight loss	938
5	Weight loss	644	Home	742
6	Gym	476	Fat	452
7	Muscle	455	Nowadays	436
8	Nowadays	454	Gym	414
9	Fat	442	Muscle	391
10	Women	408	Recommendation	381
11	Time	371	Men	355
12	Recommendation	362	Height	321
13	Height	355	Diet	300
14	Degree	343	Health	295
15	Health	340	Time	294
16	Start	312	Degree	284
17	Boyfriend	305	Start	280
18	Diet	290	Women who exercise	278
19	Boys	285	Body	246
20	Person	283	Person	244
21	Body	269	Home training	241
22	Thought	265	Thought	233
23	Weight training	251	Athlete	227
24	Home training	239	Method	221
25	Athlete	222	Weight	221
26	Day	221	Back	216
27	Weight	210	Today	186
28	Child	207	Day	185
29	Question	202	Mask	183
30	Back	198	Girlfriend	181
31	Today	195	Present	171
32	Above	192	Real	171
33	Now	190	Help	170
34	Mask	184	New virus	168
35	Men who exercise	183	Stamina	168
36	Cardio	181	Effect	160
37	Method	171	Truth	160
38	Situation	163	Grade	156
39	Help	161	This year	156
40	Women who exercise	158	Times	154
41	Potential	156	Weight training	153
42	Worry	155	Above	152
43	Routine	154	Now	152
44	New virus	153	Child	151
45	Belly fat	151	China	144
46	Friend	149	Diffusion	144
47	Times	147	Cardio	143
48	Football	147	Potential	143
49	Competition	146	Pilates	141
50	This year	146	Olympics	139

The keywords that occurred frequently in men's exercise are related to the TF-IDF analysis, as reflected in Table 4.

3.3 Social Network Analysis

In this study, an SNA was conducted using degree centrality and CONCOR analysis. A high degree centrality value would suggest a significant number of links among the terms, resulting in a significant impact on the overall network [27]. To test how connected the derived terms were to men and women's exercise, a degree centrality analysis was performed. The results are summarized in Table 5.

A CONCOR analysis was performed to analyze the structures of the relationships among the latent sub clusters in the network. The exercise results for the men are summarized in Table 6. Based on these results, homogenous groups were identified according to their shared relationships and correlations, resulting in seven clusters: exercise method, COVID-19 situation, friends, sports, women's fitness, children, and hashtags. The first cluster comprised the following terms: home, weight loss, gym, fat, time, recommendation, height, degree, diet, body, thought, weight training, day, weight, question, now, cardio, method, help, and belly fat; the terms were categorized as "exercise method". The second cluster comprised the following terms: COVID-19, exercise, men, muscle, nowadays, health, start, person, back, today, above, potential, football, and this year; the terms were categorized as "COVID-19 situation". The third cluster comprised the following terms: boyfriend, mask, situation, worry, and friend; the terms

were categorized as "friends". The fourth cluster comprised the following terms: athlete, new virus, times, and competition; the terms were categorized as "sports". The fifth cluster comprised the following terms: women, home training, and routine; the terms were categorized as "women's fitness". The sixth cluster comprised the following terms: boys and child; the terms were categorized as "children". The seventh cluster comprised the following terms: "men who exercise" and "women who exercise" and were categorized as "hashtag". The derieved clusters are show in Fig. 2.

The women's exercise results are summarized in Table 7. Based on these results, homogenous groups were identified according to their shared relationships and correlations, resulting in six clusters: exercise method, COVID-19 situation, loss of weight, friends, sports, and Olympics. The terms "new virus" and "women who exercise" did not form a cluster. The first cluster comprised the following terms: COVID-19, exercise, women, gym, start, home training, thought, back, stamina, grade, now, child, and Pilates; the terms were categorized as "exercise method". The second cluster comprised the following terms: home, nowadays, muscle, recommendation, height, health, time, body, present, real, and weight training; the terms were categorized as "COVID-19 situation". The third cluster comprised the following terms: weight loss, fat, diet, degree, method, weight, day, help, effect, truth, and cardio; the terms were categorized as "weight loss". The fourth cluster comprised the following terms: men, person, today, girlfriend, this year, above, and potential; the terms were categorized as
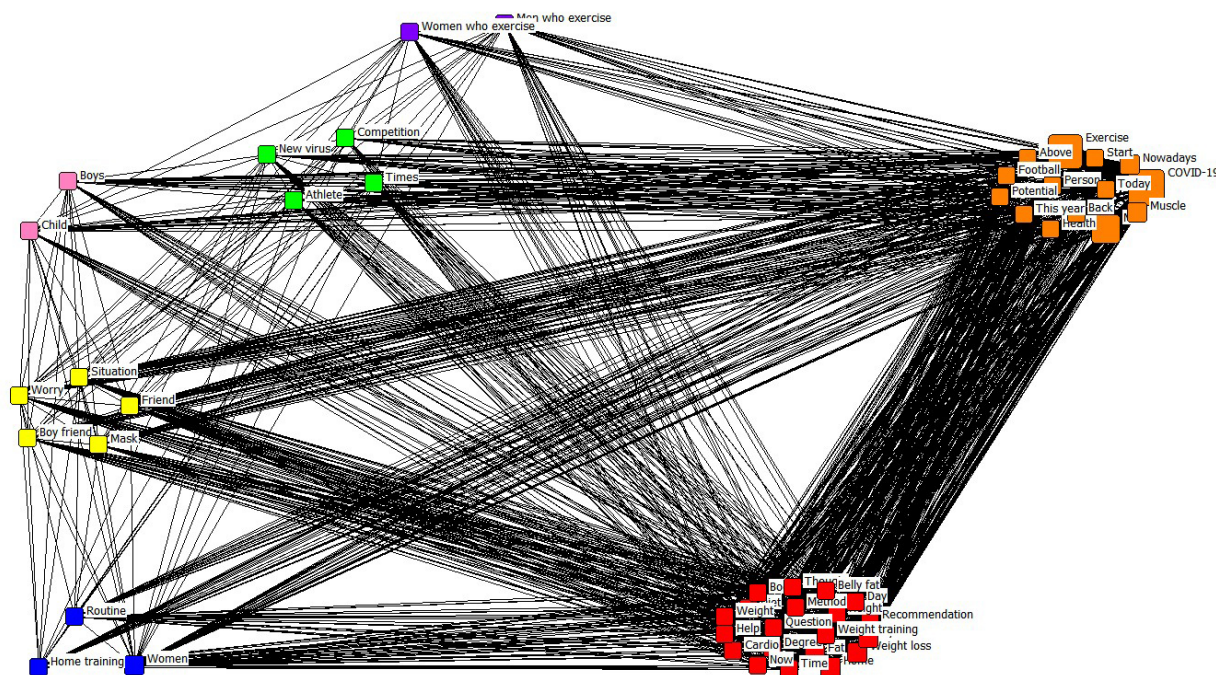

Fig. 2. Men's exercise CONCOR analysis results. Red cluster = exercise method; orange cluster = COVID-19 situation; yellow cluster = friends; green cluster = sports; blue cluster = women's fitness; pink cluster = children; violet cluster = hashtag.

Table 4. TF-IDF results.

Rank	Men's exercise		Women's exercise	
	Term	Freq.	Term	Freq.
1	Exercise	2089.726	Exercise	2156.583
2	Men	1836.046	Weight loss	1951.933
3	Weight loss	1566.305	Women	1808.240
4	Home	1351.450	Home	1425.425
5	Fat	1128.454	Fat	1151.576
6	Gym	1128.224	Gym	1032.729
7	Muscle	1105.366	Muscle	1008.943
8	COVID-19	1058.254	Nowadays	1003.269
9	Nowadays	1037.466	COVID-19	996.4361
10	Women	1004.978	Recommendation	992.1044
11	Height	1002.943	Height	926.0372
12	Recommendation	979.2557	Men	920.7977
13	Time	972.4507	Diet	869.5370
14	Boyfriend	953.2157	Health	839.3077
15	Degree	924.1232	Time	814.4022
16	Health	913.5945	Degree	810.4810
17	Diet	875.7817	Women who exercise	800.7391
18	Start	847.4343	Start	766.7667
19	Boys	804.0095	Home training	744.4566
20	Person	798.3674	Person	725.7347
21	Weight training	781.6762	Body	715.2772
22	Body	771.2811	Athlete	714.1113
23	Home training	755.0486	Method	692.6663
24	Thought	751.9681	Thought	689.5990
25	Athlete	725.5844	Weight	688.8668
26	Day	689.4664	Girlfriend	655.9602
27	Child	666.2063	Back	651.1382
28	Weight	656.3125	Mask	610.9946
29	Question	637.0006	Day	596.6127
30	Above	636.3471	Today	591.7744
31	Today	617.1677	Stamina	579.8094
32	Mask	616.1131	Real	566.1459
33	Back	613.3839	China	562.3923
34	Cardio	612.4643	Help	553.7054
35	Present	604.6666	Present	553.6420
36	Men who exercise	596.9476	Year	552.6450
37	Belly fat	583.4500	Effect	543.5487
38	Routine	581.3438	Times	542.9046
39	Method	579.8611	New virus	541.7888
40	Help	541.3508	Above	541.1414
41	Competition	540.6575	Child	533.6247
42	Worry	538.4804	Olympics	531.2052
43	Situation	536.9829	Truth	530.8344
44	Women who exercise	535.7781	This year	518.6507
45	Friend	531.1350	Weigh training	518.6224
46	Vaccine	528.4607	Cardio	507.8402
47	Football	526.5622	Now	505.3519
48	Potential	522.3568	Pilates	504.4977
49	Times	520.2522	Korea	495.1830
50	Taekwondo	516.3295	Potential	492.3978

TF-IDF = Term frequency-inverse document frequency analysis.

Table 5. Results of the degree centrality analysis.

Rank	Men's exercise		Women's exercise	
	Term	Freq.	Term	Freq.
1	Exercise	0.136805	COVID-19	0.135545
2	COVID-19	0.133798	Exercise	0.132800
3	Men	0.089062	Women	0.085664
4	Home	0.023826	Weight loss	0.029547
5	Muscle	0.021457	Home	0.025127
6	Weight loss	0.021047	Muscle	0.019915
7	Nowadays	0.018951	Gym	0.017821
8	Gym	0.018496	Nowadays	0.017682
9	Fat	0.017494	Recommendation	0.016612
10	Recommendation	0.016172	Fat	0.016379
11	Time	0.015626	Time	0.014425
12	Start	0.014305	Men	0.014332
13	Body	0.014213	Start	0.013494
14	Person	0.014168	Person	0.012936
15	Degree	0.014031	Body	0.012470
16	Women	0.013758	Thought	0.012424
17	Thought	0.013667	Height	0.012424
18	Boyfriend	0.013393	Back	0.011819
19	Height	0.012938	Athlete	0.011726
20	Health	0.012756	Degree	0.011679
21	Boys	0.011571	Health	0.011400
22	Athlete	0.011571	Today	0.010981
23	Today	0.011070	Women who exercise	0.010609
24	Diet	0.010751	Home training	0.010051
25	Back	0.010660	Girlfriend	0.010051
26	Child	0.010296	Diet	0.009865
27	Day	0.009703	Real	0.009539
28	Weight training	0.009658	Method	0.009167
29	Present	0.009293	Help	0.008655
30	Home training	0.008929	Day	0.008608
31	Mask	0.008337	Stamina	0.008608
32	Potential	0.008063	Truth	0.008376
33	Help	0.007881	Child	0.008329
34	Situation	0.007881	Weight training	0.008189
35	Friend	0.007745	Now	0.008189
36	Taekwondo	0.007699	Mask	0.007864
37	Weight	0.007517	Pilates	0.007864
38	Method	0.007426	Weight	0.007817
39	Year	0.007380	Present	0.007631
40	Effect	0.007289	Year	0.007585
41	Question	0.007198	Worry	0.007492
42	Above	0.007061	Potential	0.007492
43	Progress	0.006970	Situation	0.007352
44	Resistance exercise	0.006925	School	0.007259
45	Worry	0.006788	Effect	0.007212
46	Football	0.006788	Concern	0.006980
47	Concern	0.006742	Picture	0.006700
48	Now	0.006742	Student	0.006654
49	All	0.006697	Above	0.006654
50	Men who exercise	0.006423	Stress	0.006561

Table 6. Results of the men's exercise CONCOR analysis.

	Cluster	Terms
1	Exercise method	Home, weight loss, gym, fat, time, recommendation, height, degree, diet, body, thought, weight training, day, weight, question, now, cardio, method, help, belly fat
2	COVID-19 situation	COVID-19, exercise, men, muscle, nowadays, health, start, person, back, today, above, potential, football, this year
3	Friends	Boyfriend, mask, situation, worry, friend
4	Sports	Athlete, new virus, times, competition
5	Women's fitness	Women, home training, routine
6	Children	Boys, child
7	Hashtag	Men who exercise, women who exercise
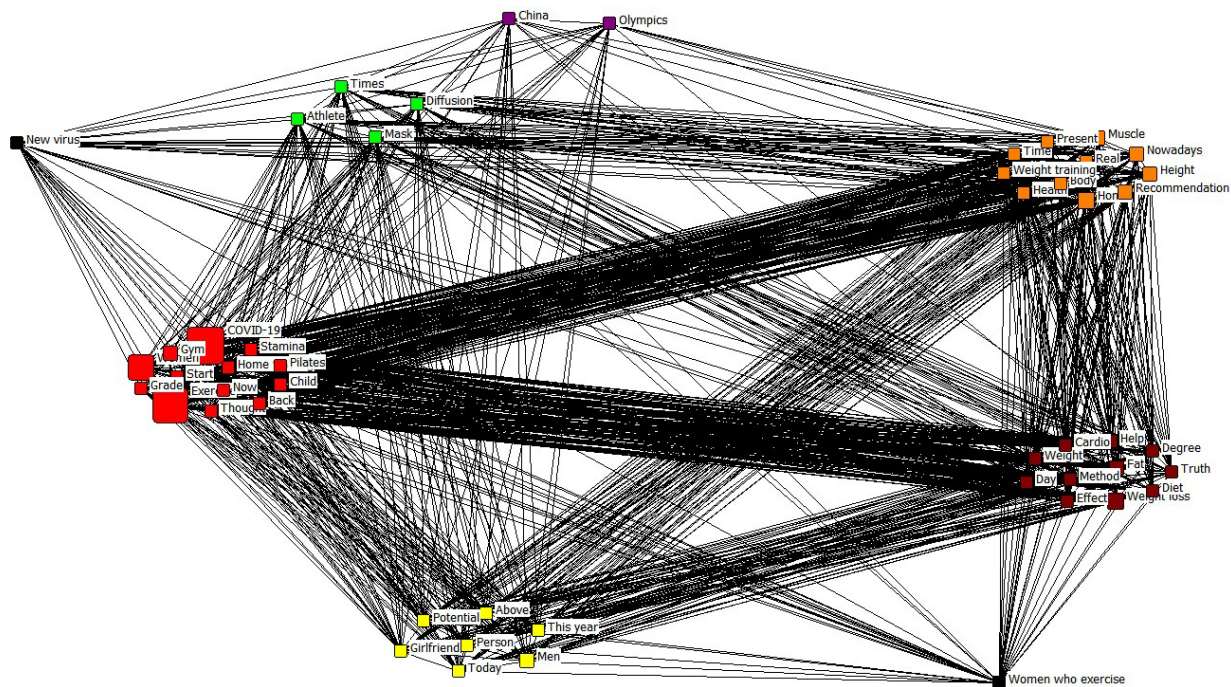
Fig. 3. Women's exercise CONCOR analysis results. Red cluster = exercise method; orange cluster = COVID-19 situation; brown cluster = weight loss; yellow cluster = friends; green cluster = sports; violet cluster = Olympics; black = could not form a cluster.

Table 7. Results of the women's exercise CONCOR analysis.

	Cluster	Terms
1	Exercise method	COVID-19, exercise, women, gym, start, home training, thought, back, stamina, grade, now, child, Pilates
2	COVID-19 situation	Home, nowadays, muscle, recommendation, height, health, time, body, present, real, weight training
3	Weight loss	Weight loss, fat, diet, degree, method, weight, day, help, effect, truth, cardio
4	Friends	Men, person, today, girlfriend, this year, above, potential
5	Sports	Athlete, mask, times, diffusion
6	Olympics	China, Olympics

"friends". The fifth cluster comprised the following terms: athlete, mask, times, and diffusion; the terms were categorized as "sports". The sixth cluster comprised the following terms: China and Olympics; the terms were categorized as "Olympics". The derieved clusters are show in Fig. 3.

4. Discussion

The aim of this study was to comparitively analyze perceptions and key attributes of exercise among Korean men and women, in social media, during the COVID-19 pandemic. Big data analysis was performed to achieve the study's aim.

The frequency analysis of keywords related to men's and women's exercise in the COVID-19 pandemic shows that the most frequently used words were home, weight loss, gym, muscle, fat, and nowadays. The data on men and women's exercise were derived from exercise-related words at home and in the gym to reduce weight and fat and increase muscle mass. Most public health organizations recommend exercise for weight management [47–52]. The results show that "home" was considered a safe space to exercise during the COVID-19 pandemic because the virus spreads when an infected person exhales droplets and extremely small particles containing the virus [53]. Exercising in nature, like walking, boosts immunity and improves chronic disease conditions [54]. Moreover, staying home is one of the best ways to protect individuals from COVID-19 [55]. The terms "weight loss" and "muscle" appeared more frequently in the men's data, and the words "weight loss" and "fat" in the women's data. Weight and muscularity are crucial factors in men's body image; however, body dissatisfaction, particularly regarding weight and shape, is common among women in all societies [56–58]. Our results support previous studies which indicated that physical attractiveness affects women more than men [59–62]. Recent Japanese studies (which has similar socio-cultural characteristics to Korea), in contrast to previous studies, also found women were more physically active than men [63]. During the COVID-19 pandemic, this study confirmed that Korean men and women have different goals and perspectives on exercise, and that physical attractiveness was accentuated by women.

As a result of the TF-IDF analysis, both data sets (men's and women's exercise) show a similar tendency in terms of superordinate words. In the men's data, words for "women" were ranked relatively high, and in the women's data, the words "weight loss" and "recommendation" were ranked relatively high. Generally, the dimension of strength had a greater influence on exercise participation for men; however for women, it was physical attractiveness. According to Silberstein *et al.* [64] and Tiggemann and Williamson [65], women are more likely to participate in exercises than men for appearance-related reasons, whereas men are more likely to exercise to appear more attractive to women [66]. In other words, the reason the word "women" appears in the men's data and "weight loss" in the women's data shows that appearance is closely related to motivation to exercise. The ideal of slender beauty is constantly being glorified and that women who achieve this ideal will be positively evaluated by others [66]. In addition, women often experience objectification through continuous evaluation, entertainment, and advertising media [66]. Therefore, exercise is used as a means to achieve the ideal through the term "recommendation," which is positively expressed in social media.

The results of the degree centrality analysis are similar to those derived from the frequency analysis and TF-IDF

analysis. In the degree centrality analysis, words related to exercising at home and the home as a safe place to exercise had high rankings due to the COVID-19 pandemic. In addition, there was a relative difference in terms of "muscle" in the men's data and "weight loss" in the women's data. The ranking of "recommendation" in both data sets also increased. The home makes exercise more accessible and safer during the COVID-19 pandemic but may require supervision by an exercise professional [67]. Therefore, the demand for safe and efficient exercise methods during the COVID-19 pandemic manifests through "recommendation".

The CONCOR analysis produced seven clusters for the male data and six clusters for the female data. First, both data sets produced a cluster for "exercise methods" and "COVID-19". Regular moderate-intensity training improves immune capacity and reduces the incidence of respiratory diseases [68]. In addition, regular exercise helps to improve immunity, control weight, improve mood, sleep, and relieve anxiety [69]. Therefore, people in both data sets were looking for effective exercise methods to improve physical, mental, and mental health. In addition, both data sets produced clusters for "friends" and "sports". Social distancing is the most effective measure to reduce the spread of airborne viruses [70]; therefore, people searched for information related to face-to-face sports, such as exercising with friends and participating in sports, due to social distancing in Korea. The results show that governments, the media, and healthcare professionals must accurately communicate public health guidance in the context of COVID-19 [71]. It is also necessary to develop a safe and effective exercise manual according to the social distancing stage and systematically disseminate it through social media.

Second, the CONCOR analysis of men's data showed a cluster for women's fitness and children. According to Park, Lee, Kim, Park, and Jang [12], traditional Asian cultural differences such as sexual conservatism exist in Korean female college students' participation in sports. In Korea, men still play a patriarchal role in providing information on women's fitness and exercise information for children. In other words, the men's data show that men were looking for relevant information because of a belief that men need to know more about exercise than women. In addition, the data showed a hashtag cluster. On social media such as Instagram and Twitter, hashtags connect people with similar interests, allowing users to share their photos and writings [72]. The hashtag function can serve as a topic management tool or help users to categorize photos and express emotions [73]. Those who use the hashtag function to search "men and women who exercise" may be curious about the ideal body or the effects of exercise.

Third, the women's data showed that the "weight loss" and "Olympics" clusters differed from the men's data. As discussed in the TF-IDF, the "weight loss" cluster was once again important in the women's data. The "Olympics" clus-

ter in the women's data was highly related to the collection period during which the Tokyo Olympics were postponed due to COVID-19 and the Olympic qualifiers competed fiercely against one another. Matches between Korea and China's women's national teams were frequently held and posted on social media. Therefore, words related to the most symbolic female athletes (women's national teams) appeared in the women's data.

Fourth, "new virus" and "women who exercise" did not form a cluster; despite this, the women's data revealed concerns about the new virus and as indirect evidence for evaluating the image of a woman who exercises. Therefore, it was still necessary to examine the two words that appeared in the network.

5. Conclusions

This study conducted big data analysis to determine the perception and key attributes related to the exercise of Korean men and women due to COVID-19, based on social media posts. The 50 words derived through frequency, TF-IDF, and degree centrality analyses provide pivotal information on the relationship between the keywords and a set period. In addition, four common clusters (exercise method, COVID-19 situation, friends, and sports) were derived: In men's exercise, the clusters were around women's fitness, children, and hashtags; in women's exercise, around weight loss and the Olympics. In other words, during COVID-19, the issues related to exercise for men and women were exercise methods, friends, and sports in common. The gender difference in men's and women's exercise is that the latter focused on data and knowledge that they could teach or refer to. Despite the COVID-19 pandemic, weight loss and physical attractiveness were important factors was an important factor in women's exercise. Women also tended to be sensitive to current trends.

However, the use of big data analysis applied in this study has the following limitations. First, there is a limit to exploring the motivations for exercise participation such as attitudes and self-concepts in relation to men's and women's exercise. Second, since the data was collected mainly from Korean-based social media, it should be cautiously interpreted as it cannot be generalised to other countries and cultures. Third, the results derived from the vast amount of big data results can be summarized differently, depending on the researcher's viewpoint.

Based on the results of this study, it is possible to provide social suggestions and improvement plans for exercise according to gender. In addition, it can serve as an academic basis for developing health-promotion programs related to exercise during the COVID-19 pandemic, which can contribute to reducing economic benefits and costs. Therefore, in a follow-up study, a gender-specific exercise program for Koreans during the COVID-19 pandemic can be developed based on the results and clusters derived from this study.

Author Contributions

S-UP designed the research study. S-UP, H-RS, and Y-KY performed the research. S-UP analyzed the data. S-UP, H-RS, and Y-KY wrote the manuscript. All authors read and approved the final manuscript.

References

[1] Chang CL, McAleer M. Alternative global health security indexes for risk analysis of COVID-19. International Journal of Environmental Research and Public Health. 2020; 17: 3161.

[2] Johns Hopkins University & Medicine. COVID-19 dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). 2021. Available at: https://coronavirus.jhu.edu/map.html (Accessed: 13 December 2021).

[3] United Nations. Statement on the coronavirus disease (COVID-19) pandemic and economic, social and cultural rights. 2021. Available at: https://digitallibrary.un.org/record/3856957 (Accessed: 1 December 2021).

[4] Lorente LML, Arrabal AA, Pulido-Montes C. The right to education and ict during covid-19: an international perspective. Sustainability. 2020; 12: 9091.

[5] Lim S, Lim H, Després J. Collateral Damage of the COVID-19 Pandemic on Nutritional Quality and Physical Activity: Perspective from South Korea. Obesity. 2020; 28: 1788–1790.

[6] Korean Ministry of Health and Welfare. 2022. Available at: http://ncov.mohw.go.kr/tcmBoardView.do?brdId=3&brdGubun=31&dataGubun=&ncvContSeq=6393&contSeq=6393&board_id=312&gubun=ALL (Accessed: 15 February 2022).

[7] Korean Ministry of Health and Welfare. 2021. Available at: http://ncov.mohw.go.kr/upload/140/202107/1625404288815_20210704221131.pdf (Accessed: 9 November 2021).

[8] Bassuk SS, Manson JE. Epidemiological evidence for the role of physical activity in reducing risk of type 2 diabetes and cardiovascular disease. Journal of Applied Physiology. 2005; 99: 1193–1204.

[9] Kohrt WM, Bloomfield SA, Little KD, Nelson ME, Yingling VR. American College of Sports Medicine Position Stand: phys-

ical activity and bone health. Medicine and Science in Sports and Exercise. 2004; 36: 1985–1996.

[10] Hale JW, Pacheco JA, Lewis CS, Swimmer L, Daley SM, Nazir N, *et al*. Everyday discrimination for American Indian tribal college students enrolled in the Internet All Nations Breath of Life program. Journal of American College Health. 2021. (in press)

[11] Sigal RJ, Kenny GP, Wasserman DH, Castaneda-Sceppa C, White RD. Physical activity/exercise and type 2 diabetes: a consensus statement from the American Diabetes Association. Diabetes Care. 2006; 29: 1433–1438.

[12] Park SU, Lee CG, Kim DK, Park JH, Jang DJ. A developmental model for predicting sport participation among female Korean college students. International Journal of Environmental Research and Public Health. 2020; 17: 5010.

[13] De Oliveira Neto L, Elsangedy HM, Tavares VDDO, Teixeira CVLS, Behm DG, Da Silva-Grigoletto ME. #TrainingInHome - Home-based training during COVID-19 (SARS-COV2) pandemic: physical exercise and behavior-based approach. Revista Brasileira de Prescrição e Fisiologia do Exercício. 2020; 19: S9–S19.

[14] WHO. 'Global Recommendations on Physical Activity for Health', Geneva World Heal Organ. WHO Press: Geneva. 2010.

[15] Pepe MV, Gandee RF. Ohio senior Olympics: creating the new adult image. Physical activity, aging and sports: practice, program and policy. The center for the study of aging. 1992: 75–82.

[16] Stevenson CL. Seeking Identities: towards an understanding of the athletic careers of masters swimmers. International Review for the Sociology of Sport. 2002; 37: 131–146.

[17] Kim Y, Lee E. The association between elderly people's sedentary behaviors and their health-related quality of life: focusing on comparing the young-old and the old-old. Health and Quality of Life Outcomes. 2019; 17: 131.

[18] Chen P, Mao L, Nassis GP, Harmer P, Ainsworth BE, Li F. Coronavirus disease (COVID-19): the need to maintain regular physical activity while taking precautions. Journal of Sport and Health Science. 2020; 9: 103–104.

[19] Halabchi F, Ahmadinejad Z, Selk-Ghaffari M. COVID-19 epidemic: exercise or not to exercise; that is the question! Asian Journal of Sports Medicine. 2020; 11: e102630.

[20] Cortez ACL, Pitanga FJG, Almeida-Santos MA, Nunes RAM, Botero-Rosas DA, Dantas EHM. Centers of physical activities and health promotion during the COVID-19 pandemic. Revista da Associacao Medica Brasileira. 2020; 66: 1328–1334.

[21] Nyenhuis SM, Greiwe J, Zeiger JS, Nanda A, Cooke A. Exercise and Fitness in the Age of Social Distancing during the COVID-19 Pandemic. The Journal of Allergy and Clinical Immunology: in Practice. 2020; 8: 2152–2155.

[22] University of Virginia Health System. COVID-19: exercise may help prevent deadly complication. 2020. Available at: https://newsroom.uvahealth.com/2020/04/15/covid-19-exercise-may-help-prevent-deadly-complication/ (Accessed: 9 August 2021).

[23] Ammar A, Trabelsi K, Brach M, Chtourou H, Boukhris O, Masmoudi L, *et al*. Effects of home confinement on mental health and lifestyle behaviours during the COVID-19 outbreak: insight from the "ECLB-COVID19" multi countries survey. MedRxiv. 2020. (in press)

[24] Castañeda-Babarro A, Arbillaga-Etxarri A, Gutiérrez-Santamaría B, Coca A. Physical Activity Change during COVID-19 Confinement. International Journal of Environmental Research and Public Health. 2020; 17: 6878.

[25] Craft BB, Carroll HA, Lustyk MKB. Gender differences in exercise habits and quality of life reports: assessing the moderating effects of reasons for exercise. International Journal of Liberal Arts and Social Science. 2014; 2: 65–76.

[26] Nambiar R, Bhardwaj R, Sethi A, Vargheese R. A look at chal-

lenges and opportunities of Big Data analytics in healthcare. 2013 IEEE International Conference on Big Data. 2013; 17–22.

[27] Carter P. Big data analytics: future architectures, skills and roadmaps for the CIO: September. White pa-per, IDC sponsored by SAS. 2011: 1–16.

[28] Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, *et al*. Big Data: the next frontier for innovation, competition, and productivity. McKinsey Global Institute: Seattle. 2011.

[29] Xing EP, Ho Q, Dai W, Kim JK, Wei J, Lee S, *et al*. Petuum: a New Platform for Distributed Machine Learning on Big Data. IEEE Transactions on Big Data. 2015; 1: 49–67.

[30] Sahay S. Big data and public health: challenges and opportunities for low and middle income countries. Communications of the Association for Information Systems. 2016; 39: 419–438.

[31] Park TH, Kim WI, Park S, Ahn J, Cho MK, Kim S. Public interest in acne on the internet: comparison of search information from Google trends and Naver. Journal of Medical Internet Research. 2020; 22: e19427.

[32] Kao A, Poteet SR, (eds.) Natural language processing and text mining. Springer science & business media. Springer: London. 2007.

[33] Iyengar S. Is anyone responsible? How television frames political issues. University of Chicago Press: Illinois. 1994.

[34] Rapp A, Beitelspacher LS, Grewal D, Hughes DE. Understanding social media effects across seller, retailer, and consumer interactions. Journal of the Academy of Marketing Science. 2013; 41: 547–566.

[35] Ganu G, Kakodkar Y, Marian A. Improving the quality of predictions using textual information in online user reviews. Information Systems. 2013; 38: 1–15.

[36] Feldman R, Sanger J. The text mining handbook: advanced approaches in analyzing unstructured data. Cambridge University Press: United Kingdom. 2007.

[37] Aggarwal CC, Zhai C, editors. Mining text data. Springer science & business media. Springer: Boston. 2012.

[38] Aggarwal CC. Machine learning for text. Springer: Switzerland. 2018.

[39] Sinoara RA, Camacho-Collados J, Rossi RG, Navigli R, Rezende SO. Knowledge-enhanced document embeddings for text classification. Knowledge-Based Systems. 2019; 163: 955–971.

[40] Voss KE, Spangenberg ER, Grohmann B. Measuring the Hedonic and Utilitarian Dimensions of Consumer Attitude. Journal of Marketing Research. 2003; 40: 310–320.

[41] Park SU, Ahn H, Kim DK, So WY. Big Data analysis of sports and physical activities among Korean adolescents. International Journal of Environmental Research and Public Health. 2020; 17: 5577.

[42] Zhang Y, Gong L, Wang Y. An improved TF-IDF approach for text classification. Journal of Zhejiang University Science. 2005; 6A: 49–55.

[43] Otte E, Rousseau R. Social network analysis: a powerful strategy, also for the information sciences. Journal of Information Science. 2002; 28: 441–453.

[44] Scott N, Baggio R, Cooper C. Network analysis and tourism from theory to practice. Cromwell Press: United Kingdom. 2008.

[45] Wasserman S, Faust K. Social network analysis: Methods and applications. Cambride University Press: Cambridge, United Kingdom. 1994.

[46] Ban HJ, Choi H, Choi EK, Lee S, Kim H. Investigating key attributes in experience and satisfaction of hotel customer using online review data. Sustainability. 2019; 11: 6570.

[47] Shephard RJ, Lankenau B, Pratt M, Neiman A, Puska P, Benaziza H, *et al*. Physical Activity Policy Development: a synopsis of the who/CDC Consultation, September 29 through October 2,

2002, Atlanta, Georgia. Public Health Reports. 2004; 119: 346–351.

[48] Pate RR, Pratt M, Blair SN, Haskell WL, Macera CA, Bouchard C, *et al*. Physical activity and public health. a recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. Journal of the American Medical Association. 1995; 273: 402–407.

[49] National Center for Health Statistics. Healthy people 2000 review 1997. 1997. Available at: https://www.cdc.gov/nchs/data/hp2000/hp2k97.pdf (Accessed: 23 December 2021).

[50] Centers for Disease Control and Prevention. Physical activity and health: A report the surgeon general. 1999. Available at: https://www.cdc.gov/nccdphp/sgr/index.htm (Accessed: 1 December 1 2021).

[51] US Department of Health and Human Services Public Health Service. Office of Disease Prevention and Health Promotion. Healthy people. 2010. Available at: http://www.healthypeople.gov/2010/ (Accessed: 1 December 2021).

[52] Donnelly JE, Honas JJ, Smith BK, Mayo MS, Gibson CA, Sullivan DK, *et al*. Aerobic exercise alone results in clinically significant weight loss for men and women: midwest exercise trial 2. Obesity. 2013; 21: E219–E228.

[53] Centers for Disease Control and Prevention. How COVID-19 spreads. 2021. Available at: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html (Accessed: 1 December 2021).

[54] Remick RA. A walk in nature: The superfood of physical activities. British Columbia Medical Journal. 2021; 63: 74.

[55] Centers for Disease Control and Prevention. Improving ventilation in your Home. 2021. Available at: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/Improving-Ventilation-Home.html (Accessed: 1 December 2021).

[56] Tiggemann M, Martins Y, Churchett L. Beyond Muscles: unexplored parts of men's body image. Journal of Health Psychology. 2008; 13: 1163–1172.

[57] Rodin J, Silberstein L, Striegel-Moore R. Women and weight: a normative discontent. In: Sondregger TB, (ed.) Psychology and gender (pp. 267–307). University of Nebraska Press: Nebraska. 1985.

[58] Thompson JK, Heinberg LJ, Altabe M, Tantleff-Dunn S. Exacting beauty: Theory, assessment, and treatment of body image disturbance. American Psychological Association: Washington. 1999.

[59] Moreno JA, Moreno R, Cervelló E. Relación del autoconcepto físico con las conductas de consumo de alcohol y tabaco en adolescentes. Adicciones. 2009; 21: 147–154.

[60] Estévez M, Muros JJ, Torres B, Pradas F, Zurita F, Cepero M. Influencia de la composición corporal y la aceptación por las clases de educación física sobre la autoestima de niños de 14-16 años de Alicante, España. Nutricion Hospitalaria. 2015; 31: 1519–1524.

[61] Franco E, Coterón J, Gómez V, Laura de Franza A. Relación entre motivación, actividad física realizada en el tiempo libre y la intención futura de práctica de actividad física. Estudio comparativo entre adolescentes argentinos y españoles. SPORT TK-Revista EuroAmericana de Ciencias del Deporte. 2017; 6: 25–34.

[62] Sáez I, Solabarrieta J, Rubio I. Physical Self-Concept, Gender, and Physical Condition of Bizkaia University Students. International Journal of Environmental Research and Public Health. 2020; 17: 5152.

[63] Amagasa S, Inoue S, Ukawa S, Sasaki S, Nakamura K, Yoshimura A, *et al*. Are Japanese Women less Physically Active than Men? Findings from the DOSANCO Health Study. Journal of Epidemiology. 2021; 31: 530–536.

[64] Silberstein LR, Striegel-Moore RH, Timko C, Rodin J. Behavioral and psychological implications of body dissatisfaction: do men and women differ? Sex Roles. 1988; 19: 219–232.

[65] Tiggemann M, Williamson S. The effect of exercise on body satisfaction and self-esteem as a function of gender and age. Sex Roles. 2000; 43: 119–127.

[66] Strelan P, Hargreaves D. Reasons for Exercise and Body Esteem: Men's Responses to Self-Objectification. Sex Roles. 2005; 53: 495–503.

[67] Federighi Baisi Chagas E, Biteli P, Moreira Candeloro B, Angelo Rodrigues M, Henrique Rodrigues P. Physical exercise and COVID-19: a summary of the recommendations. AIMS Bioengineering. 2020; 7: 236–241.

[68] Bernardi E, Pomidori L, Cassutti F, Cogo A. Home-Based, Moderate-Intensity Exercise Training Using a Metronome Improves the Breathing Pattern and Oxygen Saturation during Exercise in Patients with COPD. Journal of Cardiopulmonary Rehabilitation and Prevention. 2018; 38: E16–E18.

[69] Burnham JM. Exercise is medicine: health benefits of regular physical activity. The Journal of the Louisiana State Medical Society. 1998; 150: 319–323.

[70] Qian M, Jiang J. COVID-19 and social distancing. Journal of Public Health. 2020; 30: 259–261.

[71] Venkatesh A, Edirappuli S. Social distancing in covid-19: what are the mental health implications? British Medical Journal. 2020; 369: m1379.

[72] Small TA. What the HASHTAG? : A content analysis of Canadian politics on Twitter. Information, Communication & Society. 2011; 14: 872–895.

[73] Hu Y, Manikonda L, Kambhampati S. What we Instagram: a first analysis of Instagram photo content and user type (pp. 21–24). 8th International Conference on Weblogs and Social Media: Michigan. 2014.