International Journal of Physical Activity and Health

Volume 1 | Issue 1 Article 5

4-2022

Trend Analysis of Physical Activity Measurement Research Using Text Mining in Big Data Analytics

Seungbak Lee
University of Mississippi, slee60@go.olemiss.edu

Minsoo Kang
University of Mississippi, kang@olemiss.edu

Recommended Citation

Lee, Seungbak and Kang, Minsoo (2022) "Trend Analysis of Physical Activity Measurement Research Using Text Mining in Big Data Analytics," *International Journal of Physical Activity and Health*: Vol. 1: Iss. 1, Article 5.
DOI: https://doi.org/10.18122/ijpah1.1.5.boisestate
Available at: https://scholarworks.boisestate.edu/ijpah/vol1/iss1/5

Trend Analysis of Physical Activity Measurement Research Using Text Mining in Big Data Analytics

Abstract

Measurements of physical activity taken in a valid and reliable way are essential in characterizing the relationship between physical activity and health outcomes. Given the steadily growing interest in the physical activity measurement and the lack of research to identify current trends, this study investigated the research trend of physical activity measurement by applying four text data mining techniques (i.e., future signal, keyword network analysis, keyword trend, and keyword association rule). A total of 54,670 publications from 1982 to 2021 were collected from PubMed. As a result, the current study 1) confirmed two weak signal topics (i.e., “validity of physical activity instrument” and “classification of physical activity patterns using machine learning algorithms”) that are likely to affect future research trends, 2) identified keywords (e.g., “youth,” “adult,” “woman,” “survey,” “questionnaire,” and “monitor”) from the perspective of populations and measurement tools, 3) examined that the relative importance of keyword, “senior” increased rapidly, and 4) indicated that new keywords (i.e., “smartphone,” “wearable device,” “GPS,” “tracker,” and “app”) appeared in the early 2000s. The findings of this study provided implications for the selection of research topics and the use of text mining techniques in physical activity measurement research.

International Journal of Physical Activity and Health, Vol 1 Issue 1

Trend Analysis of Physical Activity Measurement Research Using Text Mining in Big Data Analytics

Seungbak Lee[a] and Minsoo Kang[a]
[a]University of Mississippi

Abstract

Measurements of physical activity taken in a valid and reliable way are essential in characterizing the relationship between physical activity and health outcomes. Given the steadily growing interest in the physical activity measurement and the lack of research to identify current trends, this study investigated the research trend of physical activity measurement by applying four text data mining techniques (i.e., future signal, keyword network analysis, keyword trend, and keyword association rule). A total of 54,670 publications from 1982 to 2021 were collected from PubMed. As a result, the current study 1) confirmed two weak signal topics (i.e., "validity of physical activity instrument" and "classification of physical activity patterns using machine learning algorithms") that are likely to affect future research trends, 2) identified keywords (e.g., "youth," "adult," "woman," "survey," "questionnaire," and "monitor") from the perspective of populations and measurement tools, 3) examined that the relative importance of keyword, "senior" increased rapidly, and 4) indicated that new keywords (i.e., "smartphone," "wearable device," "GPS," "tracker," and "app") appeared in the early 2000s. The findings of this study provided implications for the selection of research topics and the use of text mining techniques in physical activity measurement research.

Key words: keyword association rule, keyword network analysis, data mining, future signal

Introduction

Physical activity is a key component in efforts to reduce mortality (Leitzmann et al., 2007), and it plays an important role in improving health (Piercy et al., 2018). Clinical studies are providing a growing body of evidence that moderate physical activity prevents the development of heart and chronic disease (Ignarro et al., 2007; Powell et al., 1987). Physical activity is associated with various positive health outcomes, such as reductions in obesity (Jakicic et al., 2018; Kim et al., 2016) and metabolic syndrome (Myers et al., 2019). As such, the demand for the promotion of physical activity has increased. For instance, in 2018, the Physical Activity Guidelines for Americans (U.S. Physical Activity Guidelines Committee, 2018) were published to provide better physical activity recommendations for people of all ages. Therefore, it is essential to measure physical activity in a valid and reliable way to promote healthy lifestyles and to characterize the relationship between physical activity and health outcomes (Kang & Rowe, 2015).

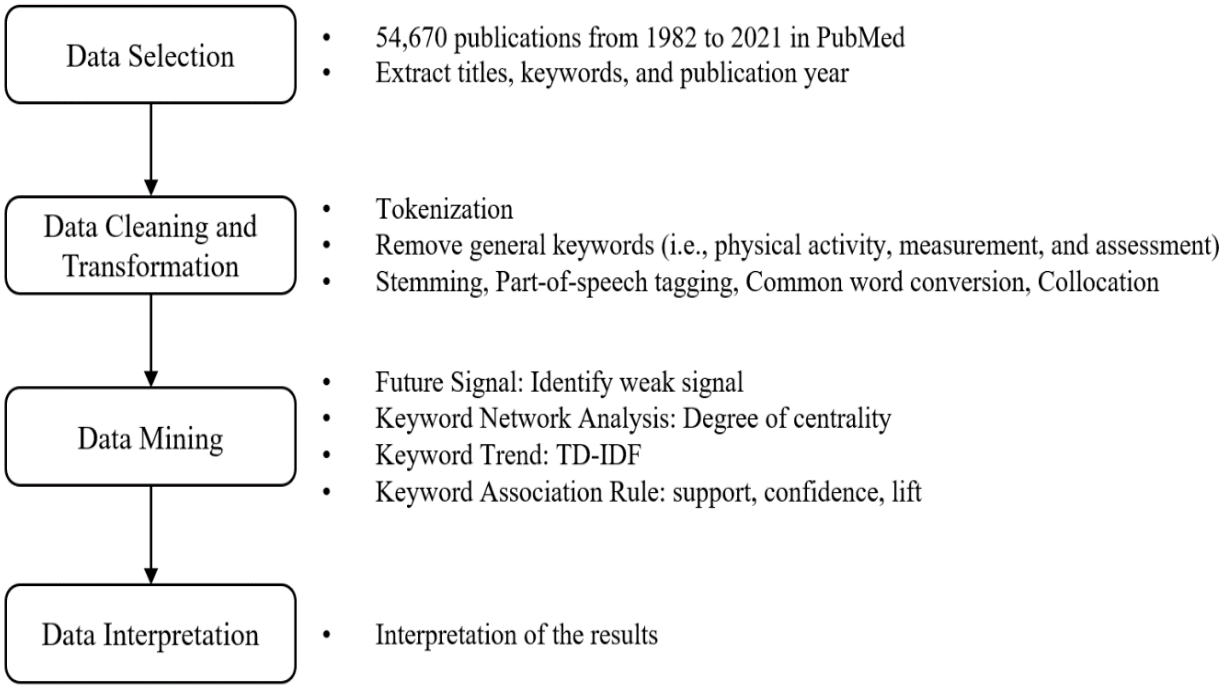
Methods to assess physical activity have changed over time; one popular choice is a questionnaire that asks respondents to describe the type of physical activity in which they engage as well as the frequency, intensity, and time spent (Craig et al., 2003). Despite the use of self-report measures in many studies, such data collection tools may miss most unstructured activities and could cause a considerable number of random errors (Adams et al., 2005). Additionally, self-report measures may lead to unreliable recall and bias (Shephard, 2003). To overcome the shortcomings of these subjective measures, objective physical activity monitoring devices (e.g., pedometers and accelerometers) have been used to gauge, test, and evaluate an individual's physical activity (Kang et al., 2009; Kang et al., 2012; Troiano et al., 2008). However, objective motion sensors also have limitations in their use in large-scale studies because of their high cost, uncertain reliability, and difficulties in data interpretation (Freedson & Miller, 2000). As

technology has advanced, internet-based tools (e.g., smartphones, smartwatches, and wearable devices) have become useful methods to measure physical activity (Gilson et al., 2013). Bort-Roig et al. (2014) identified the use of smart technologies (e.g., smartphones, smartwatches, and GPS) in physical activity measurement and promotion.

Trend analysis has often been used to examine the research trend in physical activity fields. Trend analysis explores the intricate process of how a specific situation changes or develops in a certain direction (Lee et al., 2021). One of the most widely used trend analysis methods is content analysis, which calculates descriptive statistics about the tendency of specific concepts to appear in unstructured data based on text and synthesizes the opinions of experts (Hsieh & Shannon, 2005). Content analysis is useful for discovering trends and patterns in text data (Stemler, 2000), providing basic insights into how text is used (Babbie, 1992). Haegele and Lee (2015) examined research trends in Adapted Physical Activity Quarterly using content analysis. Additionally, Kebede et al. (2018) used content analysis to identify whether evidence-informed physical activity applications are providing evidence for the promotion of physical activity. Despite the widespread use of the content analysis, there are some limitations. First, content analysis is subjective (Guthrie & Abeysekera, 2006), and since researchers subjectively classify text data, it is difficult to generalize the results. Next, the findings from the content analysis may not provide more insights beyond the given data (Morgan, 1993).

Figure 1

A process of knowledge discovery in database in text mining



Another widely used method for the trend analysis is text mining. Hearst (1999) offers a definition of text mining as "the process of extracting previously unknown, understandable, potential, and practical patterns or knowledge from the collection of massive and unstructured text data or corpus." Recently, with the spread of big data in the form of text and the development of computing technology, efforts to automate reviews of trends are continuing (Park & Cho 2017). Text mining includes various analysis processes such as pattern matching, topic tracking, association rules, and text network visualization (Fan & Li, 2006). Using these processes, various

https://scholarworks.boisestate.edu/ijpah/vol1/iss1/5
DOI: https://doi.org/10.18122/ijpah1.1.5.boisestate

analysis methods (e.g., topic modeling, clustering, association rule analysis, and network analysis) have been created and developed.

Despite the increasing use of text mining in various fields, many studies have analyzed research trends of physical activity using content analysis instead of text mining. A few studies have applied text mining to analyze trends of physical activity research; however, by applying only one process of text mining analyses, these studies are unable to acquire data that could be extracted from other processes. Given the growing interest in measuring physical activity and the lack of research to identify the trends, it is essential to examine how the physical activity measurement tools have changed over time. Therefore, this study investigated the research trend of physical activity measurement by applying various text mining analyses. First, keywords that are likely to become issues in the future related to the measurement of physical activity were identified through future signal analysis. Second, the relationships between the keywords were visualized through the network analysis. Next, trends of keywords from the perspective of populations and physical activity measures were examined. Finally, keyword association rules were derived by applying association rule analysis.

Methods

Text mining is known as knowledge discovery from textual databases (Feldman & Dagna, 1995). It refers to the overall process of discovering meaningful and nontrivial knowledge from unstructured data based on text. Fayyad et al. (1996) suggested the knowledge discovery in databases (KDD) process, which involves five steps: 1) data selection, 2) data cleaning, 3) data transformation, 4) data mining, and 5) interpretation. This study followed the knowledge discovery in databases process to explore research trends of physical activity measurement using text mining (see Figure 1).

Data Selection

Data selection is defined as creating and selecting a target data set, focusing on a subset of data samples (Fayyad et al., 1996). PubMed was used to obtain research information on title, keyword, and publication year. PubMed has received attention from researchers who are interested in using text mining techniques because it provides various text data (e.g., abstract, keyword, published year, title) for more than 12 million papers (Feldman & Sanger, 2007). First, 16,732 studies that included "physical activity measurement" or "physical activity assessment" in their titles and abstracts were selected for the frequency analysis of keywords. Among the 50 most frequent keywords, 19 terms related to the physical activity measurement were selected for the data collection (i.e., physical activity, assessment, measurement, questionnaire, validity, accelerometer, reliability, energy expenditure, evaluation, monitoring, pedometer, report, validation, objective, wearable, actigraph, monitor, device, and sensor). Lastly, publications that frequently used these terms were searched. As a result, a total of 54,670 publications from 1982 to 2021 were collected for this study.

Data Cleaning and Transformation

Data cleaning and transformation include tasks such as removing noise or handling data for further analysis (Fayyad et al., 1996). Since text is unstructured data, preprocessing is essential to reduce the time required for the analysis and to increase accuracy. In this study, text preprocessing was performed in six steps. The first step is tokenization; this is the process of breaking up text data such as words, numbers, and punctuation marks, which can be considered tokens. The second step is stopwords, which are common words such as "a," "on," "is," and "all." These words are deleted because they are not essential. Since "physical activity," "measurement,"

and "assessment" appeared prominently in the studies, they were also removed to allow for examination of other meaningful keywords. The third step is stemming, which is the process of changing various forms of a word into their stem form (e.g., "students" to "student," "bodies" to "body").

Table 1

Example of text reprocessing

Raw data	Association of living physical activity and energy expenditure in women
Tokenization	['association', 'of', 'living', 'physical', 'activity', 'and', 'energy', 'expenditure', 'in', 'women']
Stopwords	['association', 'living', 'energy', 'expenditure', 'women']
Stemming	['association', 'living', 'energy', 'expenditure', 'woman']
Extraction Noun	['association', 'energy', 'expenditure', 'woman']
Common Word	female, woman → woman
Collocation	energy expenditure, sport participation, wearable device

The fourth step is part-of-speech tagging, which aims to assign a part of speech (noun, verb, adjective, etc.) to each word. In this study, only nouns were extracted and used for the analysis. The fifth step is a common word conversion, which changes words with the same meaning but different expressions into a single word (e.g., "female" and "woman" to "woman"). The final step is a collocation extraction. Collocations are continuous sequences of words occurring together more often than would be expected by chance. Examples of collocation are "energy expenditure," "sport participation," and "wearable device" (see Table 1).

Data Mining

This study uses four popular text data mining techniques—future signal, keyword network analysis, keyword trend, and keyword association rule—to find meaningful patterns from the study data.

Future Signal.

According to Hiltunen (2008), future signal indicates "current oddities and strange issues that are thought to be key in anticipating future changes in different environments." The degree of visibility (DoV) is the statistical representation of the signal level of a future sign that measures the degree of a keyword in a data set based on its occurrence frequency (McAbee et al., 2017). The degree of visibility is calculated by the ratio of the number of term occurrences to the total number of documents. Using a time weight, the DoV of keyword i in period j is defined as:

$$DoV_{ij} = \left(\frac{TF_{ij}}{NN_j}\right) \times \{1 - tw \times (n - j)\} \qquad (1)$$

where TF_{ij} is the total occurrence frequency of a keyword i in the period j, NN_j is the total number of documents in the period j, tw is the time-weight, and n is the number of periods. In this study, the time-weight was set to .05 so as to be uniformly assigned to all periods (Yoon, 2012).

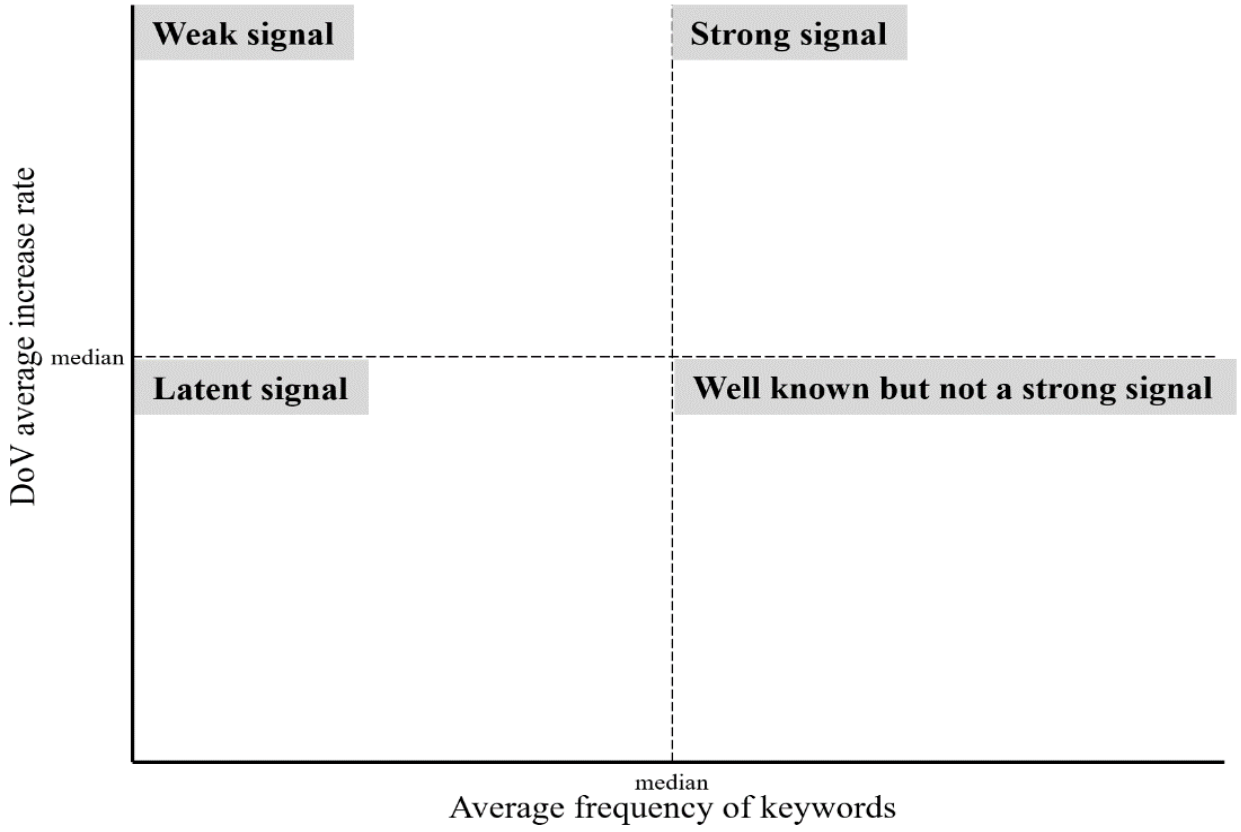
The visualization of the DoV can be accomplished by keyword emergency mapping (KEM). In KEM, the x-axis represents the average term frequency, and the y-axis indicates the average growth rate of DoV (Park & Cho, 2020). The quadrants of KEM are divided by median values. The high-right quadrant indicates strong signals (i.e., high frequency and growth rate of DoV), having the potential to become a trend since the topic pattern is relatively more stable and further exposed. The high-left quadrant means weak signals (i.e., low frequency, but high growth rate of DoV), suggesting that their relevance may increase in the future (Hiltunen, 2008). The low-

https://scholarworks.boisestate.edu/ijpah/vol1/iss1/5
DOI: https://doi.org/10.18122/ijpah1.1.5.boisestate

left quadrant refers to latent signals (i.e., low frequency and growth rate of DoV), which are not yet significantly noticeable. Lastly, the low-right quadrant can be identified as well-known but not a strong signal (i.e., high frequency, but low growth rate of DoV) because they are already familiar to people but their growth rate of DoV is not high.

Figure 2

Keyword emergency mapping (KEM) example.



Keyword Network Analysis.

Keyword network analysis, which is one of the social network analysis techniques, is the process of examining co-occurrence relationships between keywords (Su & Lee, 2010). In this study, keyword network analysis was performed using centrality, which identifies the nodes (keywords) that occupy important positions in a network (Freeman et al., 1979; Wasserman & Faust, 1994). The various centrality measures include degree of centrality, between centrality, closeness centrality, and eigenvector centrality (Freeman, 1978). Among these measures, the degree of centrality is used in this study; it represents the number of links a node (keyword) has (Freeman et al., 1979) and can be used to visualize a network map.

Keyword Trend.

The study data were divided into four periods (Phase 1: 1982–1991, Phase 2: 1992–2001, Phase 3: 2002–2011, and Phase 4: 2012–2021) to identify the trend of keywords. Word frequency analysis is often used, yet it has a limitation that cannot consider the total number of documents per period. To solve this problem, we used the term frequency-inverse document frequency (TF-IDF) measure, which indicates the relative importance of words (Yahav et al., 2018). Term frequency (TF) means the number of times that word t_i is in document d_j divided by the number of appearances of all words in document d_j. The equation of TF is given as:

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \qquad (2)$$

where inverse document frequency (IDF) indicates the total number of documents divided by the number of documents in which a specific word appears. IDF is calculated as the following:

$$IDF = log \frac{|D|}{|d_j|t_j \in d_j} \qquad (3)$$

where |D| is the total number of documents, and $|d_j|t_j \in d_j|$ is the number of documents in which word t_i appears. IDF assigns lower weight to frequent words and greater weight to infrequent words. Therefore, TF-IDF is the multiplication of term frequency and inverse document frequency.

Keyword Association Rule.

The association rule, widely known as market basket analysis, is a data mining technique to identify hidden patterns or rules in large data sets (Berry & Linoff, 1997). Agrawal et al. (1993) first introduced the association rule in mining large transaction databases using the Apriori algorithm. The rule generation process comprises two steps. The first step identifies all item sets whose support is greater than the predefined minimum support. The next step generates the association rules that satisfy a user-predefined minimum confidence.

Table 2

Degree of visibility, increasing rate, and average term frequency (top 15 keywords)

Keyword	Average Frequency	Degree of Visibility (DoV) (Phase 1 – Phase 4)				DoV increase rate average
youth	1387	0.064	0.068	0.099	0.103	0.182
health	1123	0.036	0.067	0.063	0.088	0.406
patient	1023	0.095	0.054	0.057	0.080	0.010
adult	1022	0.031	0.036	0.062	0.081	0.396
lifestyle	938	0.038	0.051	0.064	0.070	0.234
risk	894	0.058	0.095	0.077	0.058	0.068
woman	679	0.040	0.079	0.064	0.041	0.147
obesity	630	0.024	0.048	0.054	0.042	0.293
exercise	562	0.051	0.043	0.033	0.042	-0.035
behavior	510	0.018	0.018	0.030	0.041	0.344
exercise intensity	527	0.033	0.030	0.033	0.040	0.072
validation	508	0.055	0.034	0.036	0.036	-0.107
adolescent	463	0.004	0.014	0.034	0.035	1.210
senior	410	0.001	0.012	0.021	0.034	4.009
health program	406	0.024	0.017	0.026	0.031	0.138
disease	362	0.016	0.052	0.024	0.025	0.605

Support indicates how frequently a combination of antecedent and consequent of a rule appears together in the database (Kotsiantis & Kanellopoulos, 2006). If any combinations in the data occur more frequently than the minimum support level, they become candidates to be considered for a rule. Confidence indicates the strength of the rule by estimating the probability P(A|B), which is the portion of cases wherein the consequent appears given that the antecedent has appeared (Pande & Abdel-Aty, 2009). Any combinations that have a lower percentage than a predetermined confidence level is considered no association; therefore, they are dropped from the analysis. Lastly, there is a lift value to measure the quality of a rule. Lift indicates the probability of confidence increasing if A occurs. When lift (A → B) = 1, this indicates that the relationship

https://scholarworks.boisestate.edu/ijpah/vol1/iss1/5
DOI: https://doi.org/10.18122/ijpah1.1.5.boisestate

between antecedent and consequent is independent with no correlation. When lift (A → B) > 1, then A and B are dependent on one another. Therefore, the rules are potentially useful for predicting consequences in future databases. The support, confidence, and lift are calculated by the equations:

$$support\ (A \rightarrow B) = P(A \cap B) = \frac{P(A \cup B)}{|D|} \qquad (4)$$

$$confidence\ (A \rightarrow B) = \frac{support\ (A \rightarrow B)}{support\ (A)} = \frac{P\ (A \cap B)}{P\ (A)} \qquad (5)$$

$$lift\ (A \rightarrow B) = \frac{confidence\ (A \rightarrow B)}{support\ (B)} \qquad (6)$$

In this study, according to Chen et al. (2006), the minimum threshold for support and confidence is set as α = .01, and β = .70. In other words, only the rules above the 1% threshold for α and above the 70% threshold for β will be observed.

Results

Future Signal

Table 2 indicates the average growth rate in the DoV and the average frequency of each keyword involved in the search between 1982 and 2021. This study was conducted in four phases divided into 10-year increments: Phase 1 (1982–1991), Phase 2 (1992–2001), Phase 3 (2002–2011), and Phase 4 (2012–2021). Table 2 reveals that the DoV average growth rates for most keywords increase across the four phases, demonstrating a positive trend. In particular, the keywords indicating relatively more drastic trend changes are "health," "adult," "senior," and "disease." Unlike the overall increasing trend, the keywords "validation" and "exercise" exhibit a decreasing trend. The average increased rate of exercise declines rapidly from Phase 2 (1992–2001) to Phase 3 (2002–2011) and validation decreases between Phase 1 (1982–1991) and Phase 2 (1992–2001).

Table 3 presents the summary of future signals, which were classified into four quadrants using the median of the DoV average increase rate and the average frequency of each keyword. The noticeable keywords in the strong signal quadrant are related to population (e.g., "youth," "adult," and "patient") and to health (e.g., "fitness," "obesity," and "health program"), and these words represent the current physical activity measurement research trend. On the contrary, the keywords "testing," "instrument," "algorithm," "prediction," "accuracy," "motor," "walking," "medicine," and "hormone" are considered to have weak signals, with some keywords such as "algorithm," "prediction," and "testing" exhibiting the potential to become strong signals in the future. The main keywords in the latent signals are "sensors," "treadmill," "heart rate," "personality," "endurance," "nursing," and "smoking," with these keywords all falling below the median of both the DoV average increase rate and the average frequency of keywords. These keywords may be obscured until they become appropriate for future signals, or they may have already been studied enough to be of little interest. Keywords with well-known but not strong signals are "man," "questionnaire," "monitor," "pattern," "evaluation," "validation," "exercise," "weight," "body," and "intake." These words have been exposed to people and used in physical activity measurement research, but the increase rate is stagnant.

Table 3
Summary of future signals

Strong	Weak	Latent	Well known, but not strong
youth	employee	clinic	exercise
health	testing	endurance	validation
patient	instrument	sensor	evaluation
adult	mother	chemotherapy	man
lifestyle	accuracy	nursing	weight
risk	motor	restriction	cancer
woman	prediction	treadmill	pattern
obesity	anxiety	chronic	questionnaire
behavior	safety	personality	monitor
exercise intensity	walking	heartrate	mortality
adolescent	medicine	smoking	hypertension
senior	water	physician	injury
health program	algorithm	protein	sport
disease	hormone	style	body
fitness		power	blood pressure
protocol			bone
diabetes			stress
quality			intake
student			density
nutrition			work
diet			consumption

Table 4
Degree of centrality related to populations

No	Keyword	Degree Centrality
1	youth	2.916
2	adult	2.217
3	woman	1.818
4	man	1.012
5	senior	0.764
6	student	0.512
7	family	0.272
8	parent	0.251
9	player	0.183
10	children	0.107

Keyword Network Analysis

Keyword network analysis was performed with studies containing keywords related to populations (i.e., “youth,” “adult,” “woman,” “man,” “adolescent,” “senior,” “student,” “sport participation,” “family,” “parent,” “player,” and “children”) and measurement tools (i.e., “accelerometer,” “pedometer,” “smartphone,” “questionnaire,” “monitor,” “survey,” “treadmill,”

https://scholarworks.boisestate.edu/ijpah/vol1/iss1/5
DOI: https://doi.org/10.18122/ijpah1.1.5.boisestate

"wearable device," "GPS," "sensor," "self-report," "tracker," "app," "scale," and "device"), respectively. Table 4 and Figure 3 present the results of the centrality analysis and visualization using studies of physical activity measurement keywords related to populations. Table 4 presents the top 10 keywords based on the degree of centrality. As per result, "youth" (2.916) exhibited the highest degree of centrality, followed by "adult" (2.217), "woman" (1.818), "man" (1.012), "senior" (0.764), and "student" (0.512). These population-related keywords have been widely used in physical activity measurement research.

Figure 3 illustrates the keyword network visualization based on the degree of centrality. Keywords such as "adult," "woman," "man," "senior," and "student" are situated around the center of "youth" and are mainly used in physical activity measurement studies related to populations. Likewise, "adult" forms a cluster with keywords such as "health," "obesity," "exercise intensity," and "fitness." "Woman" is associated with "postmenopausal," "persistence," "anxiety disorder," and "incontinence."

Figure 3

Keyword network related to populations

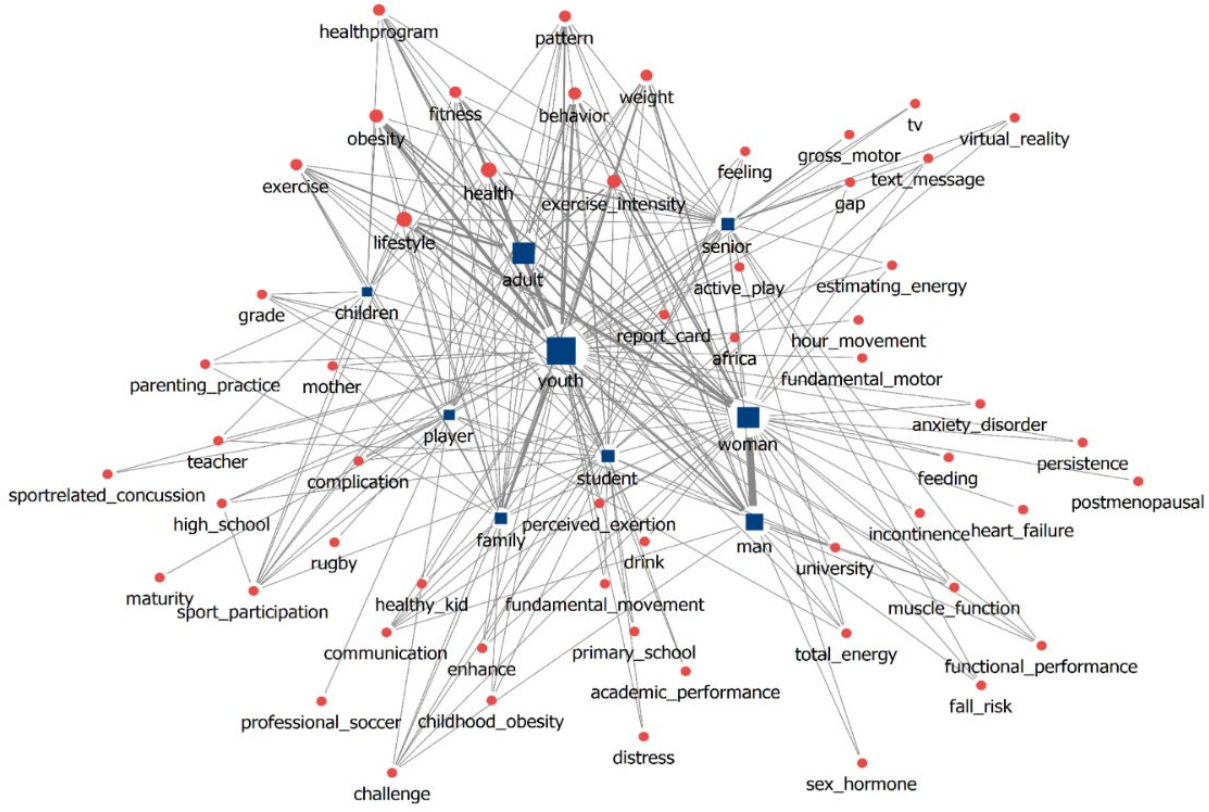


Table 5 and Figure 4 present the results of the centrality analysis and visualization using studies of physical activity measurement keywords related to measurement tools. The results of the analysis revealed that "survey" (1.383) has the highest degree of centrality, followed by "questionnaire" (1.007), "monitor" (0.775), "accelerometer" (0.447), "scale" (0.395), and "device" (0.314).

Table 5
Degree of centrality related to tools

No	Keywords	Degree Centrality
1	survey	1.383
2	questionnaire	1.007
3	monitor	0.775
4	accelerometer	0.447
5	scale	0.395
6	device	0.314
7	app	0.255
8	sensor	0.194
9	pedometer	0.181
10	self-report	0.144

Figure 4
Keyword network related tools

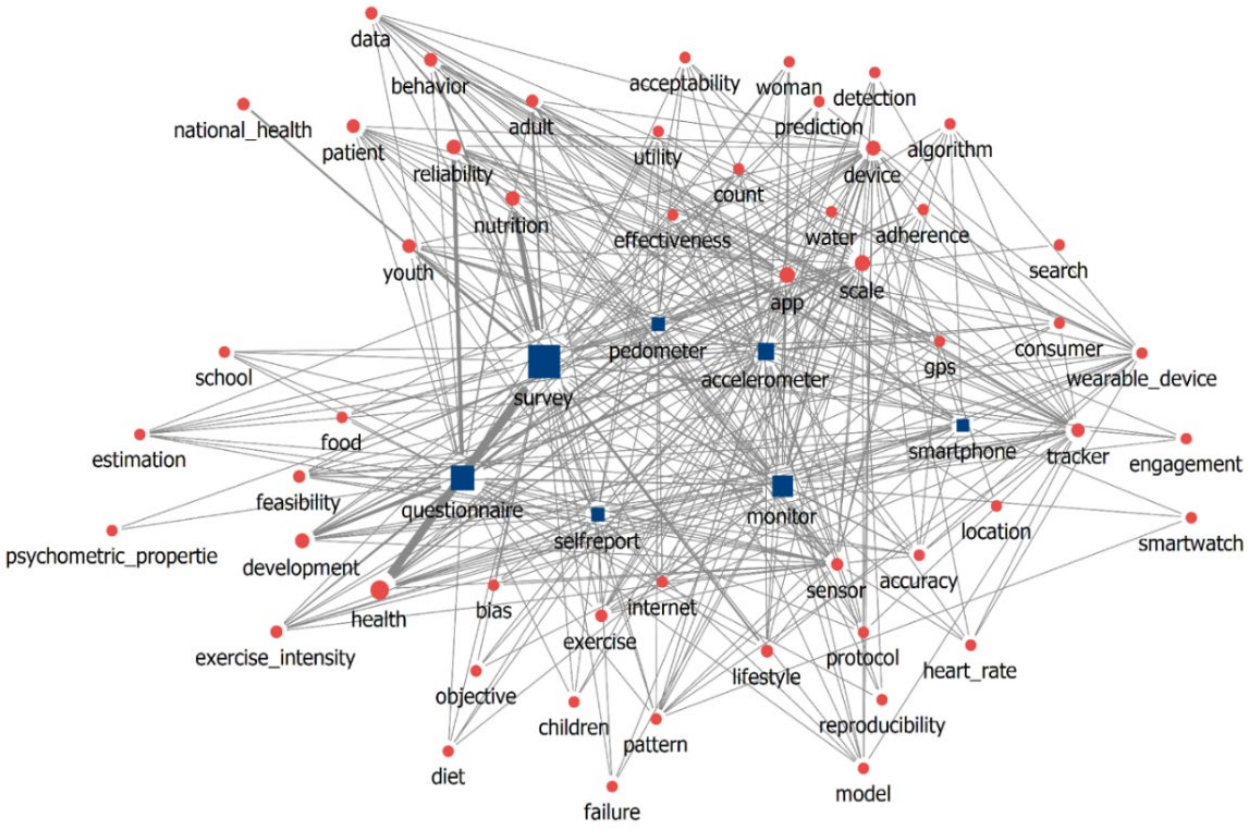


Figure 4 visualizes the results of the keyword network based on the degree of centrality. "Survey," "questionnaire," "monitor," "accelerometer," "pedometer," and "smartphone" are main keywords related to measurement tools and they are widely used in physical activity measurement studies. Specifically, keywords such as "survey," "self-report," and "questionnaire," representing subjective tools, are related to exercise (e.g., exercise intensity, exercise), health (e.g., health, national health), nutrition (e.g., nutrition, food, and diet), and scale development (e.g., reliability,

https://scholarworks.boisestate.edu/ijpah/vol1/iss1/5
DOI: https://doi.org/10.18122/ijpah1.1.5.boisestate

development, and estimation). "Accelerometer," "monitor," "smartphone," and "pedometer," representing objective tools, are related with protocol and algorithm development (e.g., protocol, algorithm, and effectiveness), validation (e.g., accuracy, acceptability, and prediction), and advanced instruments (e.g., app, wearable device, tracker, and smartwatch).

Table 6

Keywords trend of populations

Keywords	Phase 1	Phase 2	Phase 3	Phase 4
youth	0.020	0.020	0.024	0.023
adult	0.012	0.013	0.018	0.020
woman	0.015	0.022	0.019	0.013
man	0.013	0.016	0.011	0.007
adolescent	0.003	0.007	0.012	0.012
senior	0.000	0.006	0.009	0.011
student	0.007	0.004	0.007	0.008
sport participation	0.000	0.000	0.001	0.001
family	0.004	0.003	0.005	0.004
parent	0.002	0.002	0.003	0.003
player	0.006	0.004	0.003	0.002
children	0.002	0.001	0.003	0.003

Table 7

Keywords trend of tools

Keywords	Phase 1	Phase 2	Phase 3	Phase 4
accelerometer	0.003	0.003	0.004	0.004
pedometer	0.002	0.001	0.005	0.002
smartphone	0.000	0.000	0.000	0.001
questionnaire	0.009	0.009	0.008	0.006
monitor	0.015	0.008	0.006	0.006
survey	0.007	0.009	0.007	0.005
treadmill	0.002	0.001	0.001	0.001
wearable device	0.000	0.000	0.000	0.001
GPS	0.000	0.000	0.001	0.000
sensor	0.003	0.002	0.001	0.002
self-report	0.006	0.003	0.002	0.002
tracker	0.000	0.000	0.000	0.002
app	0.000	0.000	0.000	0.003
scale	0.006	0.003	0.003	0.003
device	0.003	0.001	0.002	0.003

TF-IDF was calculated to examine the trend of keywords related to populations and measurement tools.

Keyword Trend

Table 6 presents the trends of keywords related to populations. "Youth," "adult," "adolescent," and "senior" demonstrate an upward trend over time, while studies involving "man" decline over time. Research on "senior" begins to appear in Phase 2.

Table 7 indicates the trends of keywords related to measurement tools. Changes in the relative importance of each period are modest except for "monitor" and "scale." Notably, new tools measuring physical activity emerge over time. Specifically, the keywords such as "smartphone," "wearable device," "GPS," "tracker," and "app" begin to appear in Phases 3 and 4.

Keyword Association Rule

Association rule mining was performed with studies containing keywords related to populations and measurement tools, respectively. Additionally, keyword association rules were derived for each period, and support and confidence thresholds were set at .01 and .70, respectively. However, in the analysis using population-related keywords, no rule was derived above the confidence .70 threshold. Thus, rules with more than .50 were suggested.

Table 8

Association rules analysis results (Populations)

Phase	antecedents	consequents	support	confidence	lift
Phase 1	blood pressure	youth	0.03	1.00	2.88
	history	family	0.03	1.00	23.75
	mortality	man	0.03	1.00	4.52
	epidemiology, school	female	0.03	1.00	23.75
	epidemiology, female	student	0.03	1.00	10.56
Phase 2	cardia	adult	0.01	1.00	6.48
	development, cardia	adult	0.01	1.00	6.48
	risk, cardia	adult	0.01	1.00	6.48
	development, risk, cardia	adult	0.01	1.00	6.48
	development, risk	adult	0.01	0.92	5.94
Phase 3	density	woman	0.01	0.61	2.74
	parent	youth	0.01	0.55	1.75
	school	youth	0.01	0.54	1.74
	risk, obesity	youth	0.01	0.50	1.60
	bone	youth	0.01	0.48	1.55
Phase 4	parent	youth	0.02	0.66	2.03
	screen	youth	0.01	0.64	1.97
	school	youth	0.01	0.60	1.84
	diabetes	adult	0.01	0.59	2.43
	fitness	youth	0.01	0.51	1.58

https://scholarworks.boisestate.edu/ijpah/vol1/iss1/5
DOI: https://doi.org/10.18122/ijpah1.1.5.boisestate

Table 8 indicates the top five association rules related to populations for each period. All of the association rules generated in Phase 1 have a 100% confidence rate. The lift value of 2.88 implies that the probability of "blood pressure" and "youth" appearing together is 2.88 times higher than when it is not. The remainder of the rules are interpreted similarly. In Phase 2, most of the rules consist of "adult," "cardia," and "risk." In Phase 3, the rules are related to "woman" and "youth." For example, "density" is associated with "woman" at a 61% confidence rate. Moreover, the probability of "risk" and "obesity" appearing with "youth" is about 1.6 times higher than when it is not. The keyword rules indicated in Phase 4 are related mostly to "youth" and "adult." Specifically, "parent," "screen," "school," and "fitness" formed rules with "youth," and "diabetes" formed rules with "adult."

Table 9

Association rules analysis results (Tools)

Phase	antecedents	consequents	support	confidence	lift
Phase 1	adult	sensor	0.02	1.00	17.00
	alcohol	survey	0.02	1.00	7.29
	algorithm	treadmill	0.02	1.00	51.00
	Alzheimer	scale	0.02	1.00	10.20
	computer	treadmill	0.02	1.00	51.00
Phase 2	consumption	survey	0.01	1.00	3.79
	fitness	survey	0.03	1.00	3.79
	heartrate	monitor	0.02	1.00	4.27
	motion	sensor	0.02	1.00	23.50
	nutrition	survey	0.03	1.00	3.79
Phase 3	continuous glucose	monitor	0.01	1.00	5.95
	national health	survey	0.01	1.00	4.42
	reproducibility	questionnaire	0.01	1.00	3.91
	health, adult	survey	0.01	1.00	4.42
	nutrition, adult	survey	0.01	1.00	4.42
Phase 4	national health	survey	0.02	1.00	6.37
	adult, nutrition	survey	0.01	1.00	6.37
	health, nutrition	survey	0.01	1.00	6.37
	national health, nutrition	survey	0.01	1.00	6.37
	reliability, youth	questionnaire	0.01	0.78	4.21

Table 9 indicates the top five keyword association rules-related measurement tools for each period. In Phase 1, various tools (e.g., "sensor," "survey," "treadmill," and "scale") appear as consequents. All association rules have a 100% confidence rate in Phase 1. "Algorithm" and "computer" appear with "treadmill" with a 100% confidence rate. Additionally, the probability that "Alzheimer" and "scale" coexist is about 10 times higher than when it is not. Notably, in terms

of measurement tools, most keywords from Phases 2, 3, and 4 form the rules with "survey." In Phase 2, "consumption," "fitness," and "nutrition" are related to "survey." In Phases 3 and 4, keywords such as "national health," "nutrition," "youth," and "adult" form the rules with "survey."

Discussion

To our knowledge, this study is among the first to examine the trend of physical activity measurement using various text mining techniques. The findings of this study 1) confirmed keywords that are likely to affect future research trends through future signal analysis, 2) identified keywords and contents from the perspective of populations and measurement tools through keyword network analysis, 3) examined the relative importance of keywords by period using TF-IDF, and 4) formed keyword association rules in terms of populations and measurement tools based on the association rule analysis. From the results of the future signal analysis, the keywords (i.e., "testing," "instrument," "accuracy, "motor," "prediction," "walking," "algorithm," "employee," "mother," "anxiety," "safety," "medicine," "water," and "hormone") represent weak signals, indicating their potential to be trends in the physical activity measurement field. Two weak signal topics are identified from those keywords. The first is "validity of physical activity instrument." Along with technology development, the studies about establishing validity and reliability evidence of devices measuring physical activity will continue (e.g., Brodie et al., 2018; Degroote et al., 2018; Holbrook et al., 2009; Holbrook et al., 2011; Kim & Kang, 2019; Kim et al., 2013; Lamont et al., 2018).

The second weak signal topic is "classification of physical activity patterns using machine learning algorithms" and includes keywords such as "prediction," "walking," "algorithm," "testing," and "accuracy." For example, a machine learning model was used to classify walking types (Hu et al., 2018) and to develop algorithms to measure and recommend physical activity (Liao et al., 2020; Mohammadi et al., 2020). With the rapid increase in the generation of data and the development of computational science, data mining and machine learning techniques will continue to identify more informed interpretations of physical activity behavior.

The keyword network analysis results revealed the top three keywords (i.e., "youth," "adult," and "woman") with a high degree of centrality using physical activity measurement studies related to populations. Previous studies have examined physical activity measurement issues in children. For example, Kang et al. (2016) described the background and issues that often arise when measuring physical activity in youth. Loprinzi and Cardinal (2011) reviewed various physical activity and sedentary behavior measurement tools (e.g., self-report surveys, self-report diaries, parental reporting, and heart rate monitoring) for children. Furthermore, several researchers have conducted studies targeting adults and, specifically, women. For example, previous studies have identified the validity and responsivity of the adult physical activity measurement questionnaire (Deng et al., 2008; Hallal & Victora, 2004; Tomioka et al., 2011); examined the relationship between physical activity, health, and nutrition (Bassett et al., 2010; Heller et al., 2011); and confirmed the association between physical activity and disease in postmenopausal women (LaMonte et al., 2018; Segev et al., 2018).

The keyword network analysis using physical activity measurement studies related to measurement tools identified "survey," "questionnaire," and "monitor" with the high degree of centrality. Surveys and questionnaires seem to play an essential role despite the development of new instruments measuring physical activity. The US National Health and Nutrition Examination Survey and the International Physical Activity Questionnaire are widely used to confirm the relationship between physical activity and health outcomes.

https://scholarworks.boisestate.edu/ijpah/vol1/iss1/5
DOI: https://doi.org/10.18122/ijpah1.1.5.boisestate

The results of TF-IDF indicated the relative importance of keywords according to the period. First, among keywords related to population, the keywords with increasing importance were "youth," "adult," "adolescent," and "senior," and the keyword exhibiting the most rapid growth was "senior." According to a 2019 United Nations report, people over 65 years old will account for 16% of the world population by 2025. As the elderly population increases, measurement of physical activity grows in emphasis. The finding is consistent with previous research (Koch, 2010), demonstrating that measurement tools to evaluate physical activity and health outcomes are essential to preserving older people's overall health. Next, the relative importance of keywords related to measurement tools did not change much across the various phases. However, new keywords (i.e., "smartphone," "wearable device," "GPS," "tracker," and "app") appeared between Phase 3 and Phase 4. These keywords indicate that physical activity measurement tools were developed with technological advances. The result follows the previous findings that pointed out the use of smart devices to measure physical activity (Stella et al., 2021; Zheng et al., 2014).

Keyword association rule analysis was used to find meaningful associations or correlations among a large set of keywords. The results of the keyword association rule analysis support the findings derived from keyword network analysis. Specifically, the association rules related to population consisted mainly of "adult" and "youth." However, some earlier studies included other consequents associated with population (e.g., "family," "student") in Phase 1. The association rules related to measurement tools generally comprised surveys and questionnaires. The findings also indicate similar results with keyword network analysis. Self-reported measures, including questionnaires and surveys, have been widely used in physical activity measurement studies, especially in population-based observational research.

Although these findings provide significant implications for research trends of physical activity measurement, this study has several limitations. First, text preprocessing is a controversial step. Text data can have different meanings depending on the context and can result in multiple errors due to various interpretations. Second, there is a limit to generalizing the research trend results of physical activity measurement in this study because the data were collected only from PubMed. Despite these limitations, this study can provide researchers with information on research trends about physical activity measurement. Moreover, there is the implication of using text mining techniques in physical activity measurement. Traditionally, studies have been conducted to confirm and compare the validity and reliability evidences of subjective and objective tools for measuring physical activity. However, text mining technology can be utilized to evaluate the tools in new ways by synthesizing the written or spoken information (e.g., expert's opinion) in varied textual form (e.g., transcripts, speeches, articles) on the physical activity measurement tools. Future researchers are advised to use various, credible databases to conduct text-mining research on physical activity measurement and other areas associated with physical activity.

References

Adams, S. A., Matthews, C. E., Ebbeling, C. B., Moore, C. G., Cunningham, J. E., Fulton, J., & Hebert, J. R. (2005). The effect of social desirability and social approval on self-reports of physical activity. *American journal of epidemiology*, *161*(4), 389-398.

Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data* (pp. 207-216).

Babbie, E. (1992). The practice of social research. Belmont, CA: Wadsworth. *Journal of MultiDisciplinary Evaluation*, *4*(7), 112-314.

Bassett Jr, D. R., Wyatt, H. R., Thompson, H., Peters, J. C., & Hill, J. O. (2010). Pedometer-measured physical activity and health behaviors in United States adults. *Medicine and science in sports and exercise*, *42*(10), 1819.

Berry, M. J., & Linoff, G. (Eds.) (1997). Market basket analysis. In *Data Mining Techniques: For Marketing, Sales, and Customer Support* (pp. 124-156). Indianapolis, IN: John Wiley & Sons, Inc.

Bort-Roig, J., Gilson, N. D., Puig-Ribera, A., Contreras, R. S., & Trost, S. G. (2014). Measuring and influencing physical activity with smartphone technology: a systematic review. *Sports medicine*, *44*(5), 671-686.

Brodie, M. A., Pliner, E. M., Ho, A., Li, K., Chen, Z., Gandevia, S. C., & Lord, S. R. (2018). Big data vs accurate data in health research: large-scale physical activity monitoring, smartphones, wearable devices and risk of unconscious bias. *Medical hypotheses*, *119*, 32-36.

Chen, G., Liu, H., Yu, L., Wei, Q., & Zhang, X. (2006). A new approach to classification based on association rule mining. *Decision Support Systems*, *42*(2), 674-689.

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., ... & Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine & science in sports & exercise*, *35*(8), 1381-1395.

Degroote, L., De Bourdeaudhuij, I., Verloigne, M., Poppe, L., & Crombez, G. (2018). The accuracy of smart devices for measuring physical activity in daily life: validation study. *JMIR mHealth and uHealth*, *6*(12), e10972.

Deng, H. B., Macfarlane, D. J., Thomas, G. N., Lao, X. Q., Jiang, C. Q., Cheng, K. K., & Lam, T. H. (2008). Reliability and validity of the IPAQ-Chinese: the Guangzhou Biobank Cohort study. *Medicine and science in sports and exercise*, *40*(2), 303-307.

Fan, J., & Li, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. *arXiv preprint math/0602133*.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, *39*(11), 27-34.

Feldman, R., & Dagan, I. (1995, August). Knowledge Discovery in Textual Databases (KDT). In *KDD* (Vol. 95, pp. 112-117).

Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press.

Freedson, P. S., & Miller, K. (2000). Objective monitoring of physical activity using motion sensors and heart rate. *Research quarterly for exercise and sport*, *71*(sup2), 21-29.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, *1*(3), 215-239.

Freeman, L. C., Roeder, D., & Mulholland, R. R. (1979). Centrality in social networks: II. Experimental results. *Social networks*, *2*(2), 119-141.

Gilson, N. D., Faulkner, G., Murphy, M. H., Meyer, M. R. U., Washington, T., Ryde, G. C., ... & Dillon, K. A. (2013). Walk@ Work: An automated intervention to increase walking in university employees not achieving 10,000 daily steps. *Preventive medicine*, *56*(5), 283-287.

Guthrie, J., & Abeysekera, I. (2006). Content analysis of social, environmental reporting: what is new?. *Journal of Human Resource Costing & Accounting*.

https://scholarworks.boisestate.edu/ijpah/vol1/iss1/5
DOI: https://doi.org/10.18122/ijpah1.1.5.boisestate

Haegele, J. A., Lee, J., & Porretta, D. L. (2015). Research trends in adapted physical activity quarterly from 2004 to 2013. *Adapted Physical Activity Quarterly*, *32*(3), 187-2016.

Hallal, P. C., & Victora, C. G. (2004). Reliability and validity of the international physical activity questionnaire (IPAQ). *Med Sci Sports Exerc*, *36*(3), 556.

Hearst, M. A. (1999, June). Untangling text data mining. In *Proceedings of the 37th Annual meeting of the Association for Computational Linguistics* (pp. 3-10).

Heller, T., McCubbin, J. A., Drum, C., & Peterson, J. (2011). Physical activity and nutrition health promotion interventions: what is working for people with intellectual disabilities?. *Intellectual and developmental disabilities*, *49*(1), 26-36.

Hiltunen, E. (2008). The future sign and its three dimensions. *Futures*, *40*(3), 247-260.

Holbrook, E. A., Stevens, S. L., Kang, M., & Morgan, D. (2011). Validation of a talking pedometer for adults with visual impairment. *Medicine and science in sports and exercise*, *43*(6), 1094-1099.

Holbrook, E., Barreira, T., & Kang, M. (2009). Validity and reliability of Omron pedometers for prescribed and self-paced walking. *Medicine+ Science in Sports+ Exercise*, *41*(3), 670.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, *15*(9), 1277-1288.

Hu, B., Dixon, P. C., Jacobs, J. V., Dennerlein, J. T., & Schiffman, J. M. (2018). Machine learning algorithms based on signals from a single wearable inertial sensor can detect surface-and age-related differences in walking. *Journal of biomechanics*, *71*, 37-42.

Ignarro, L. J., Balestrieri, M. L., & Napoli, C. (2007). Nutrition, physical activity, and cardiovascular disease: an update. *Cardiovascular research*, *73*(2), 326-340.

Jakicic, J. M., Rogers, R. J., Davis, K. K., & Collins, K. A. (2018). Role of physical activity and exercise in treating patients with overweight and obesity. *Clinical chemistry*, *64*(1), 99-107.

Kang, M., & Rowe, D. A. (2015). Issues and challenges in sedentary behavior measurement. *Measurement in physical education and exercise science*, *19*(3), 105-115.

Kang, M., Bassett, D. R., Barreira, T. V., Tudor-Locke, C., & Ainsworth, B. E. (2012). Measurement effects of seasonal and monthly variability on pedometer-determined data. *Journal of Physical Activity and Health*, *9*(3), 336-343.

Kang, M., Bassett, D. R., Barreira, T. V., Tudor-Locke, C., Ainsworth, B., Reis, J. P., ... & Swartz, A. (2009). How many days are enough? A study of 365 days of pedometer monitoring. *Research quarterly for exercise and sport*, *80*(3), 445-453.

Kang, M., Mahar, M. T., & Morrow Jr, J. R. (2016). Issues in the assessment of physical activity in children. *Journal of Physical Education, Recreation & Dance*, *87*(6), 35-43.

Kebede, M., Steenbock, B., Helmer, S. M., Sill, J., Möllers, T., & Pischke, C. R. (2018). Identifying evidence-informed physical activity apps: content analysis. *JMIR mHealth and uHealth*, *6*(12), e10314.

Kim, H., & Kang, M. (2019). Validation of sedentary behavior record instrument as a measure of contextual information of sedentary behavior. *Journal of Physical Activity and Health*, *16*(8), 623-630.

Kim, Y., Barreira, T. V., & Kang, M. (2016). Concurrent associations of physical activity and screen-based sedentary behavior on obesity among US adolescents: a latent class analysis. *Journal of epidemiology*, *26*(3), 137-144.

Kim, Y., Park, I., & Kang, M. (2013). Convergent validity of the international physical activity questionnaire (IPAQ): meta-analysis. *Public health nutrition*, *16*(3), 440-452.

Koch, S. (2010). Healthy ageing supported by technology–a cross-disciplinary research challenge. *Informatics for Health and Social Care*, *35*(3-4), 81-91.

Kotsiantis, S., & Kanellopoulos, D. (2006). Association rules mining: A recent overview. *GESTS International Transactions on Computer Science and Engineering*, *32*(1), 71-82.

Lamont, R. M., Daniel, H. L., Payne, C. L., & Brauer, S. G. (2018). Accuracy of wearable physical activity trackers in people with Parkinson's disease. *Gait & posture*, *63*, 104-108.

LaMonte, M. J., Manson, J. E., Chomistek, A. K., Larson, J. C., Lewis, C. E., Bea, J. W., ... & Eaton, C. B. (2018). Physical activity and incidence of heart failure in postmenopausal women. *JACC: Heart Failure*, *6*(12), 983-995.

Lee, Y., Kim, M. L., & Hong, S. (2021). Big-data Analytics: Exploring the Well-being Trend in South Korea Through Inductive Reasoning. *KSII Transactions on Internet and Information Systems (TIIS)*, *15*(6), 1996-2011.

Leitzmann, M. F., Park, Y., Blair, A., Ballard-Barbash, R., Mouw, T., Hollenbeck, A. R., & Schatzkin, A. (2007). Physical activity recommendations and decreased risk of mortality. *Archives of internal medicine*, *167*(22), 2453-2460.

Liao, P., Greenewald, K., Klasnja, P., & Murphy, S. (2020). Personalized heartsteps: A reinforcement learning algorithm for optimizing physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *4*(1), 1-22.

Loprinzi, P. D., & Cardinal, B. J. (2011). Measuring children's physical activity and sedentary behaviors. *Journal of exercise science & fitness*, *9*(1), 15-23.

McAbee, S. T., Landis, R. S., & Burke, M. I. (2017). Inductive reasoning: The promise of big data. *Human Resource Management Review*, *27*(2), 277-290.

Mohammadi, R., Atif, M., Centi, A. J., Agboola, S., Jethwani, K., Kvedar, J., & Kamarthi, S. (2020). Neural Network–Based Algorithm for Adjusting Activity Targets to Sustain Exercise Engagement Among People Using Activity Trackers: Retrospective Observation and Algorithm Development Study. *JMIR mHealth and uHealth*, *8*(9), e18142.

Morgan, D. L. (1993). Qualitative content analysis: a guide to paths not taken. *Qualitative health research*, *3*(1), 112-121.

Myers, J., Kokkinos, P., & Nyelin, E. (2019). Physical activity, cardiorespiratory fitness, and the metabolic syndrome. *Nutrients*, *11*(7), 1652.

Pande, A., & Abdel-Aty, M. (2009). Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. *Safety science*, *47*(1), 145-154.

Park, C., & Cho, S. (2020). Analysis on trends and future signs of smart grids. *Int. J. Smart Grid Clean Energy*, *9*(3), 533-543.

Piercy, K. L., Troiano, R. P., Ballard, R. M., Carlson, S. A., Fulton, J. E., Galuska, D. A., ... & Olson, R. D. (2018). The physical activity guidelines for Americans. *Jama*, *320*(19), 2020-2028.

Powell, K. E., Thompson, P. D., Caspersen, C. J., & Kendrick, J. S. (1987). Physical activity and the incidence of coronary heart disease. *Annual review of public health*, *8*(1), 253-287.

Segev, D., Hellerstein, D., & Dunsky, A. (2018). Physical activity-does it really increase bone density in postmenopausal women? A Review of articles published between 2001-2016. *Current aging science*, *11*(1), 4-9.

Shephard, R. J. (2003). Limits to the measurement of habitual physical activity by questionnaires. *British journal of sports medicine*, *37*(3), 197-206.

https://scholarworks.boisestate.edu/ijpah/vol1/iss1/5
DOI: https://doi.org/10.18122/ijpah1.1.5.boisestate

Stella, A. B., AjČeviĆ, M., Furlanis, G., Cillotto, T., Menichelli, A., Accardo, A., & Manganotti, P. (2021). Smart technology for physical activity and health assessment during COVID-19 lockdown. *The Journal of sports medicine and physical fitness*, *61*(3), 452-460.

Stemler, S. (2000). An overview of content analysis. *Practical assessment, research, and evaluation*, *7*(1), 17.

Su, H. N., & Lee, P. C. (2010). Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in Technology Foresight. *Scientometrics*, *85*(1), 65-79.

Tomioka, K., Iwamoto, J., Saeki, K., & Okamoto, N. (2011). Reliability and validity of the International Physical Activity Questionnaire (IPAQ) in elderly adults: the Fujiwara-kyo Study. *Journal of epidemiology*, 1109210254-1109210254.

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine and science in sports and exercise*, *40*(1), 181.

United Nations Department of Economic and Social Affairs. *2019 Revision of World Population Prospects;* United Nations: New York, NY, USA, 2019.

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications.

Yahav, I., Shehory, O., & Schwartz, D. (2018). Comments mining with TF-IDF: the inherent bias and its removal. *IEEE Transactions on Knowledge and Data Engineering*, *31*(3), 437-450.

Yoon, J. (2012). Detecting weak signals for long-term business opportunities using text mining of Web news. *Expert Systems with Applications*, *39*(16), 12543-12550.

Zheng, Y. L., Ding, X. R., Poon, C. C. Y., Lo, B. P. L., Zhang, H., Zhou, X. L., ... & Zhang, Y. T. (2014). Unobtrusive sensing and wearable devices for health informatics. *IEEE Transactions on Biomedical Engineering*, *61*(5), 1538-1554.