Public Health Nutrition: 24(6), 1193–1209 doi:10.1017/S1368980020005248

Review Article

Social media analytics in nutrition research: a rapid review of current usage in investigation of dietary behaviours

Emma Stirling[1,2,*], Jane Willcox[1], Kok-Leong Ong[3] and Adrienne Forsyth[1]
[1]School of Allied Health, La Trobe University, Melbourne 3086, Australia: [2]School of Behavioural and Health Sciences, Australian Catholic University, Melbourne 3065, Australia: [3]La Trobe Business School, La Trobe University, Melbourne, Australia

Submitted 7 August 2020: Final revision received 16 December 2020: Accepted 17 December 2020: First published online 23 December 2020

Abstract

Objective: Social media analytics (SMA) has a track record in business research. The utilisation in nutrition research is unknown, despite social media being populated with real-time eating behaviours. This rapid review aimed to explore the use of SMA in nutrition research with the investigation of dietary behaviours.
Design: The review was conducted according to rapid review guidelines by WHO and the National Collaborating Centre for Methods and Tools. Five databases of peer-reviewed, English language studies were searched using the keywords 'social media' in combination with 'data analytics' and 'food' or 'nutrition' and screened for those with general population health using SMA on public domain, social media data between 2014 and 2020.
Results: The review identified 34 studies involving SMA in the investigation of dietary behaviours. Nutrition topics included population nutrition health investigations, alcohol consumption, dieting and eating out of the home behaviours. All studies involved content analysis with evidence of surveillance and engagement. Twitter was predominant with data sets in tens of millions. SMA tools were observed in data discovery, collection and preparation, but less so in data analysis. Approximately, a third of the studies involved interdisciplinary collaborations with health representation and only two studies involved nutrition disciplines. Less than a quarter of studies obtained formal human ethics approval.
Conclusions: SMA in nutrition research with the investigation of dietary behaviours is emerging, nevertheless, if consideration is taken with technological capabilities and ethical integrity, the future shows promise at a broad population census level and as a scoping tool or complementary, triangulation instrument.

Keywords
Social media
Social media analytics
Dietary behaviour
Nutrition
Surveillance

The exponential rise of social media, where users create online communities to share information for social networking and microblogging including Twitter, Facebook, YouTube, Instagram and Reddit[1], presents an unprecedented opportunity for nutrition research. As of January 2020, there were 3·88 billion worldwide social media users, growing more than 9 % since 2019[2], with sharing food and eating behaviours one of the most popular online community activities[3,4]. Social listening is recognised as an emerging type of communication monitoring as a means of attaining interpersonal information and social intelligence[5], and when applied to social media, it can be further analysed using social media analytics (SMA) to uncover how people behave and talk in relation to food, nutrition and health, thus, providing a future solution or complement to traditional nutrition research methods augmenting the investigation or tracking of trends in changing dietary behaviours.

SMA can provide economies of scale, and most importantly, in real time, surveillance models[6] compared with traditional research methods such as surveys. SMA is an interdisciplinary research field that aims to extend methods of analysis of social media data[6]. SMA has grown exponentially in recent years and is well established in the

*Corresponding author: Email emma.stirling@acu.edu.au

research domains of information systems, business and consumer marketing, political science, crisis identification and communication[6]. The popularity of SMA came about because of social media big data, that is characterised by the four Vs of volume, velocity, variety and veracity. The rich data that can be collected at scale and quickly have enabled new SMA use cases. Common applications of SMA are in opinion mining (automatic systems to determine human opinion from text written in natural language), sentiment analysis (using natural language processing (NLP), computational linguistics and text analytics to identify subjective information) and content analytics (often of text and the study of word frequency, distributions, pattern recognition, and visualisation)[7]. Predictive analytics is often described as the pinnacle of SMA and uses data mining of historical and current social media data, machine learning plus statistical techniques, to predict future or other unknown events or behaviours[8].

As SMA has evolved, so have developed frameworks including defined steps across data discovery, collection, preparation and analysis[6]. SMA often uses established tools across these steps which include in-platform analytic tools (e.g. Application Programming Interface (API)), open source tools (e.g. SentiStrength, NodeXL) which generally require little programming experience, commercial tools (e.g. Radian 6, SAS, IBM Watson Analytics[9]) obtained through subscription and often requiring an experienced data analyst or team, and the design of customised machine learning systems and dashboarding requiring far more sophisticated programming capabilities[10].

In the field of nutrition research, there is evidence of the use of social media platforms; however, there does not appear to be an established track record of SMA related to the investigation or tracking of dietary behaviours with general population health. Nutrition and health researchers have used social media platforms for recruitment of study participants[11–13] and as part of intervention studies[14,15]. There is also evidence that nutrition researchers have utilised social media platforms for gathering primary research data; however, in some studies, the researchers have used manual data collection, preparation and analysis methods (such as this example with Instagram posts[16]) in comparison with the use of sophisticated machine learning, artificial intelligence and customised programming that define the steps of SMA. In other cases, nutrition and health researchers have relied on traditional qualitative methods such as focus groups (as in this study examining the use of Instagram for healthy eating[17]), interviews or surveys[18] to understand social media user content via self-reporting, as opposed to direct data extraction and analysis with SMA. One example of the use of SMA tools is NLP techniques explored with Reddit posts from an online eating disorder community, and the top techniques, compared with coding by clinical psychologists, were found to have an error rate of only 4 % when assessing if a person required immediate professional help[19]. A second example using predictive SMA is the customised dashboard nEmesis machine learning system. This system was designed to automatically detect food outlets that pose a food safety risk, by following any adverse reactions self-reported via Twitter of geotagged diners, and had a 64 % greater effectiveness than traditional methods[20].

In contrast, in the field of health research, there is evidence from systematic reviews of the use of both SMA applications along with evidence of SMA in investigating general population health behaviours. In a recent review of the popular microblogging site Twitter in health research, SMA was commonly used in cross-sectional content analysis (56 %) and also in longitudinal surveillance (26 %), and to a lesser extent with participant recruitment (7 %) or intervention (7 %)[21]. In this review, the research fields represented by studies were public health (23 %), infectious diseases (20 %) behavioural medicine (18 %) and psychiatry (11 %), with the most common research topics being influenza (8 %), smoking (7 %), cancer (5 %) and Ebola (4 %). A second, recent scoping review on the use of Twitter for data collection with health care consumers concluded that the platform is utilised to search and mine primary research data[22]. In this review, a wide range of health topics and research questions were explored including health challenges such as pain, migraines and cancer; social discourse of conditions like perceptions or portrayal of seizures; and cyberspace in comparison with real-world phenomena, with data obtained via posts (such as keywords or phrases) or profiles (such as geolocation). Established SMA applications in health behaviours research include as a complementary data source for pharmacovigilance[23,24], in the monitoring of changing health habits such as smoking[25–30], in public health surveillance such as predicting flu outbreak[31,32], with sentiment and content analysis such as in distinct, online communities[33,34] and as a predictor of morbidity and psychosocial health such as in suicide risk[35–39]. These findings in health research provide promising support for the application of SMA in nutrition research, with the investigation or tracking trends in dietary behaviours in general population health.

Social media analytics potential with investigating dietary behaviours

To our knowledge, there has not been a literature review which has explored the potential use of SMA in nutrition research with the investigation or trend tracking of dietary behaviours. SMA can provide economies of scale, and most importantly, in real time, surveillance models and is well developed in other fields that indicate promise for nutrition research[6]. Conversely, SMA also has known limitations which must be further investigated for applications to the field of nutrition research. For example, sentiment analysis can fail to accurately identify semantics and pragmatics (such as irony and slang) which are commonly used in

https://doi.org/10.1017/S1368980020005248 Published online by Cambridge University Press

personal conversations to describe eating behaviours[40]. Finally, as SMA is a fast moving, new area of research, it also presents challenges for health and nutrition researchers which warrant further investigation, such as different skill sets required to plan a study, to derive and analyse data and to conduct a study under existing ethical standards for medical research with human subjects[41].

Given the established track record of SMA with health research and the limited use in nutrition research, the aim of this rapid review is to understand how SMA is currently being used with the investigation or the trend tracking of dietary behaviours in general population health. It behoves us to investigate the studies, platforms, applications, nutrition topics, research disciplines and ethical considerations to map current opportunities and inform future research.

Due to the broad area of investigation, likely heterogenous nature of the study methods, focus on current research due to the rapidly changing landscape of SMA and the aim to draw together timely evidence, a rapid review was the appropriate approach for this knowledge synthesis[42]. The aim of this paper was to investigate: How are researchers currently using SMA for investigating dietary behaviours related to general population health?

The key considerations of the rapid review were to:

1. Extract and collate studies that involve dietary behaviours in general population health using SMA on public domain, social media data.
2. Rank the social media platforms by data set size and by number of included studies to assess the scope of use and common research platforms.
3. Collate the included studies by nutrition topics to assess common categories of relevance to public health nutrition.
4. Map each study design for an overview of the scope of SMA applications across the recognised steps of data discovery, collection, preparation and analysis.
5. Explore the types of research disciplines involved and the presence of collaborative research with particular focus on nutrition or health discipline involvement.
6. Record the presence of formal human ethics review or considerations.

Methods

Protocol

The protocol was drafted using guidance in the WHO Rapid Reviews to Strengthen Health Policy and Systems: A Practical Guide[43] and the National Collaborating Centre for Methods and Tools Rapid Review Guidebook: Steps for Conducting a Rapid Review[44]. An overview of the rapid review approach was recorded with the Open Science Framework on 27 October 2019 and updated on 9 March 2020 and 9 December 2020[45] with further information on the protocol available via the authors.

Eligibility criteria

Original, full-text research studies published in peer-reviewed journals from January 2014 to February 2020 in the English language were included for review. Due to rapid changes in social media platforms and functions, along with equally rapidly changing technology driving SMA, the authors limited this review to studies published in the prior 5–6 year period in order to obtain data most reflective of current usage. Table 1 outlines the full inclusion and exclusion criteria. Studies were included if they involved SMA investigating individual dietary behaviours related to general population health and not with acute or chronic diseases/conditions or with advertising, campaigns or policy. Nutrition was not required to be the primary focus of studies, and broad fields of research were considered. For the purposes of this review, social media was defined as third-party sites such as Twitter, Facebook, Instagram, Reddit and YouTube and not health-targeted apps, Internet of Things (e.g. wearables) or health-targeted blogs and websites as they often require subscription services with user-generated content not in the public domain or are subject to copyright or privacy policy of the site ownership. Unlike a website or blog hosted and managed by an individual, company or organisation, social media platforms enable users to create online communities to share information, ideas, personal messages and other content (such as videos) typically in the public domain and often available for extraction and

Table 1 Eligibility criteria for the present rapid review of social media analytics in nutrition research, current usage in investigation of dietary behaviours

Included studies	Excluded studies
1. Original full access research published from 1 January 2014 to 29 February 2020.	1. Abstracts, reviews, conference proceedings unless published as full papers, commentaries, conference posters and unpublished thesis.
2. English language studies.	2. Non-English language studies.
3. Social media data on individual dietary behaviours in general population health.	3. Social media data with acute or chronic diseases or with campaigns, policy or advertising.
4. Public domain, user-generated data from third-party social media sites.	4. Data from mhealth, IOT, platforms specifically health-targeted/hosted/moderated including apps, blogs and websites or obtained by recruiting participants.
5. Studies that used social media data as above and included at least one description of SMA steps.	5. Studies that used social media data as above without any description of any of the SMA steps.

IOT, Internet of Things.

https://doi.org/10.1017/S1368980020005248 Published online by Cambridge University Press

SMA often via an API[1]. The final eligibility criteria were independently assessed by two authors (ES and JW) who are nutrition scientists and accredited dietitians.

Information sources and search strategy

One researcher (ES) conducted the systematic search using the keywords 'social media' in combination with 'data analytics' and 'food' or 'nutrition' (1 January 2014 until 29 February 2020). A full list of search term keyword synonyms, along with the search strategy across all databases, is displayed in the Supplementary Materials. Social media and data analytics search terms were verified with a research librarian through formative searching and confirmed with data analytics expert author (KL) and also cross-referenced with keywords identified by Taylor *et al.*[46]. Specific keywords were included for the major platforms such as Facebook, Twitter and YouTube as initial scoping had shown evidence of use in SMA and additional platforms were identified through generic social media keywords. Nutrition-related search terms were purposely kept broad in order to capture studies from a diverse range of research fields that may contain pertinent information on dietary behaviours of interest to public health nutrition. The first subset of nutrition keywords included synonyms for nutrition, diet and health. This subset also contained a targeted list of non-communicable diseases such as 'diabetes' and 'cancer' to assess if studies on populations with acute or chronic conditions, although not the focus here, contained relevant information on general population health dietary behaviours. The second subset of nutrition keywords were related to general food or eating behaviours such as 'snack', 'takeaway' and 'supermarket' in order to capture studies outside the fields of health, such as hospitality, travel or marketing, that also contained relevant information to this investigation. Searched bibliographic databases included Medline, CINAHL, Scopus, ACM Digital Library and Engineering Village, and all articles were exported into EndNote. The reference lists of all included articles were also hand-searched to capture related texts.

Selection of sources of evidence

The primary author (ES) undertook a first abstract and title assessment to exclude duplicate and irrelevant articles. Full-text articles were then retrieved and screened against the inclusion criteria by two authors who are nutrition scientists and accredited dietitians (ES and JW).

Data charting, data items and synthesis

Data were extracted from eligible studies by ES using data-charting forms jointly developed by three authors (ES, AF and JW), and key study characteristics were discussed by the authors. Data analytics expert author (KL) provided advice on SMA methods as required.

Extracted data items included year, country of origin, social media platform(s), data set size(s), purpose of the study and where available, information on the SMA steps across data discovery, preparation, collection and analysis. The aim of this information was to explore the scope of, and most popular, social media platforms utilised in research by extracting information on the number of studies per platform and the data set size. Secondly, extracting the purpose of the study allowed for collation of the range of nutrition topics under current investigation and alignment with public health nutrition. Thirdly, information on the SMA steps would reveal common applications such as if there was a heavy reliance on SMA for data discovery (such as searching social media data), but less so in advanced analytic applications (such as sophisticated machine learning models and not human coding). Due to the heterogenous nature of the research investigations, diverse research fields and study designs, in-depth comparison of SMA techniques across studies was not warranted, rather the focus was on identifying any common trends across the SMA steps.

Sinnenberg *et al.* have provided a recent taxonomy to describe the roles of Twitter in health research in order to assign each study to recognised categories in the field of SMA of either content analysis, surveillance (monitoring trends in a particular topic or metric over time), engagement (user interactions with content produced by other users) or network analysis (the connections between users or influencers)[21]. This taxonomy was applied to the included studies in this review, across all social media platforms, given it provided a timely and up-to-date framework that covered appropriate categories for the current field of SMA as confirmed with data analytics expert author (KL). The two additional Sinnenberg categories of recruitment and intervention were outside the scope of this review[21].

Additional data items were collected on the affiliations or disciplines listed of all study researchers with the aim to explore if current research is driven by particular disciplines or collaborative research. Finally, the presence of institutional ethics board review or ethical mentions was recorded for all included studies to assess ethical considerations, and with the aim to explore any discipline differences in ethical approaches. The final versions of the data-charting forms are displayed in the Supplementary Materials.

Included sources of evidence were not critically appraised in this rapid review as the aim was to capture a broad range of heterogeneous studies to inform a summary of current topics, platforms and other information and not evaluate specific effects on a particular area of public health nutrition.

Results

Selection of sources of evidence

On completion, the original database searches yielded 5220 results (Fig. 1). Duplicate records were removed

https://doi.org/10.1017/S1368980020005248 Published online by Cambridge University Press

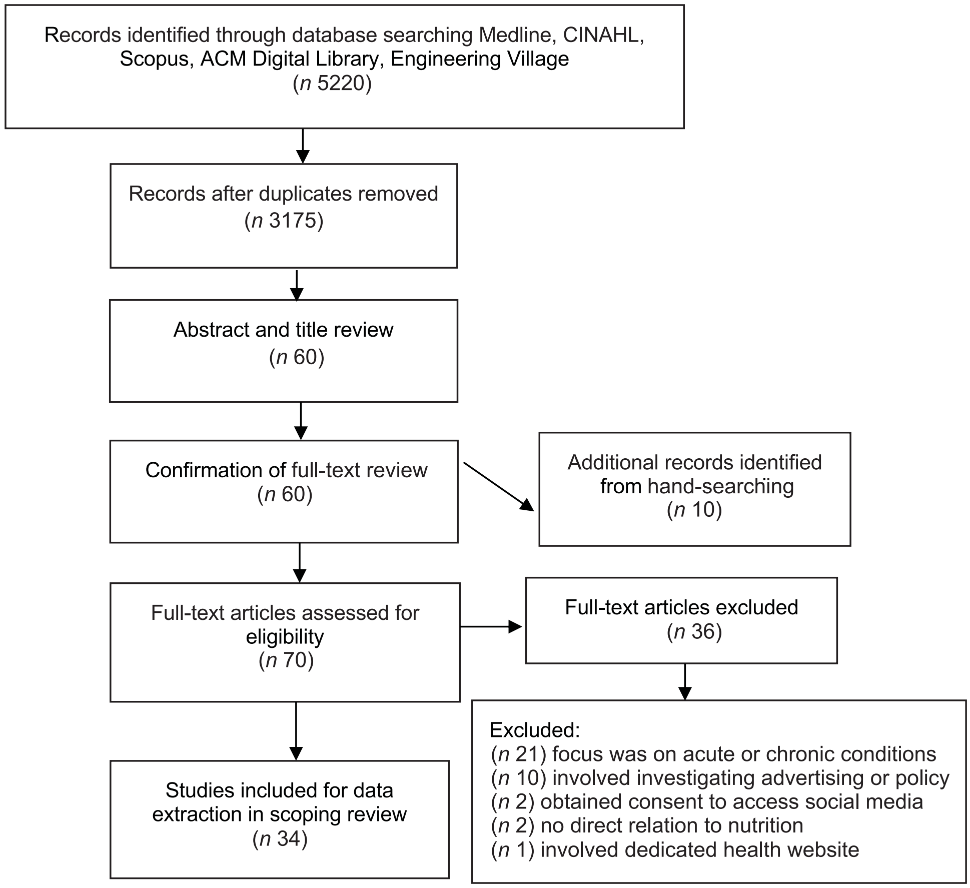

Fig. 1 Flowchart showing the search for studies included in the rapid review on Social Media Analytics in Nutrition Research: a rapid review of current usage in investigation of dietary behaviours

resulting in 3175 studies. Abstract and title searching was conducted to remove irrelevant articles against eligibility criteria and resulted in 60 retained for full-text review. An additional 10 records were identified through hand-searching reference lists. A total of 70 full-text articles were screened for inclusion in the final review and subsequently 34 of these met the inclusion criteria. Four of the included studies involved extracting posts of user reviews from platforms with major e-commerce or product and service info sharing features (Amazon, Yelp, Weibo and Koubei), which were deemed to meet the criteria of user-generated social media in publicly available data on the Internet. Excluded studies were due to the following reasons: focus was on individuals with acute or chronic conditions and there was a lack of relevant information to general population health dietary behaviours (*n* 21), studies involved tracking or investigating advertising, campaigns or policy in relation to health and nutrition and not individual dietary behaviours (*n* 10), in two studies researchers obtained consent and access to the social media accounts of individual participants rather than extracting public domain data, on further review two additional studies were found to have no direct relation to nutrition and one study involved a dedicated health website that was not part of the definition of social media sites in this review.

Synthesis of results

Data characteristics of social media analytics studies involving dietary behaviours

The data characteristics of 34 research articles are described in Table 2, including the social media platform(s), data set size(s), study purpose, Sinnenberg's Taxonomy and SMA steps. Each study was mapped for the usage of SMA steps across data discovery, collection, preparation and analysis and common tools or processes captured if clearly determinable. The full data-charting and additional details are available in the Supplementary Materials.

Social media platforms and data set sizes

Table 3 displays the ranking of social media platforms accessed in included studies with a description of each platform.

Of the 34 studies, the greatest number utilised Twitter (62 %, *n* 21), followed by Instagram (21 %, *n* 7). Facebook (9 %, *n* 3) and Foursquare (9 %, *n* 3) had equal next ranking. Amazon.com user product reviews, YouTube, Yelp reviews and Tumblr were used in single studies at times in combination with another platform(s) under investigation. Two studies utilised social media platforms in China, Weibo and Koubei. The majority of studies

https://doi.org/10.1017/S1368980020005248 Published online by Cambridge University Press

Table 2 Characteristics of included studies ranked by original data set size per platform in present rapid review

First author, year and country	SoMe platforms	Data set size SMA summary	Purpose of study	Taxonomy	SMA steps* Data discovery	Data collection	Data preparation	Data analysis
Nguyen *et al.* 2016, USA[49]	Twitter	80M geotagged tweets	Validated machine learning for constructing indicators of happiness, food and physical activity via location (neighbourhoods) compared with census data	Content analysis	✓ Twitter API and geolocator	✓ Twitter API and geolocator	✓ Stanford Tokenizer	✓ MALLET
Alajajian *et al.* 2017, USA[50]	Twitter	~50M tweets	Developed a machine learning device called 'Lexicocalorimeter' with the attempt to measure energy/energetic content of foods mapped across the US population.	Content analysis	✓ Customised machine learning model	✓ Customised machine learning model	✓ Customised machine learning model	✓ Customised machine learning model
Huang *et al.* 2017, USA[69]	Twitter	50M	Applied natural language processing and machine learning methods to examine alcohol (and tobacco) tweets for temporal patterns by age group (including underage).	Content analysis	✓	✓ Twitter API	✓ Python	x/✓ Python Amazon Mechanical Turk
Alhabash *et al.* 2018, USA[73]	Twitter	47·5M tweets	Collected a large sample of tweets and performed content analysis to examine how the prevalence of tweeting about alcohol, along with tweet features, can predict tweet virality.	Content analysis, engagement	✓	✓ Twitter API	✓	✓ LIWC
Sun *et al.* 2018, China and UK using US data[51]	Twitter	41M tweets	Collected and analysed tweets to show relation with government obesity data and Twitter derived dietary habits of users.	Content analysis, (view to surveillance)	✓	✓	✓	✓
Kershaw *et al.* 2014, UK[70]	Twitter	31·6M	Designed exploratory machine learning model with the aim to assess alcohol-related tweets for consumption patterns and compared to an existing health data set.	Content analysis	✓	✓ Twitter API	✓	✓
Vydiswaran *et al.* 2020, USA[48]	Twitter	21·19M tweets, multiple subsets	Demonstrated that Twitter can be used to characterise neighbourhood-level food-related behaviours and attitudes/sentiment in food-related tweets.	Content analysis	✓ Twitter API	✓	✓/x	✓/x
Cavazos-Rehg *et al.* 2015, USA[65]	Twitter	11,966,381 tweets Subset 5000	Assessed the content of tweets with alcohol or drinking-related keywords to explore binge-drinking behaviours.	Content analysis	✓ Purchased data	✓ Purchased data	✓	✓/x Manual coding of subset by Crowd Flower
Shah *et al.* 2020, USA[47]	Twitter	~10M tweets 50 000 food subset	Designed a high-performance machine learning model to classify food tweets across Canada and demonstrated correlation between the location of high-energetic food tweets and government consumption data.	Content analysis	✓ Twitter API	✓ Customised machine learning model	✓ Customised machine learning model	✓ Customised machine learning model
Karami *et al.* 2018, USA[55]	Twitter	4·5M tweets	Analysed using machine learning text in tweets to characterise health opinions in diet and exercise (and obesity and diabetes).	Content analysis	✓	✓ Twitter API	✓	✓ MALLET LIWC
Nguyen *et al.* 2017, USA[64]	Twitter Yelp	4 041 521 tweets 505 554 reviews	Collected food data from Twitter and Yelp to characterise food environments of different locations to create a state-level food database in the USA and compare with government obesity data.	Content analysis	✓	✓ Twitter API Yelp API OpenStreetMap	✓	✓ MALLET Python

Table 2 *Continued*

First author, year and country	SoMe platforms	Data set size SMA summary	Purpose of study	Taxonomy	SMA steps* Data discovery	Data collection	Data preparation	Data analysis
Fried *et al.* 2014, USA[58]	Twitter	3M tweets	Analysed food-related tweets by hashtags (such as #dinner and #breakfast) and developed machine learning system to predict population characteristics such as overweight.	Content analysis	✓	✓	✓	✓ Standford Core NLP LDA
Abbar *et al.* 2015, Qatar on USA data[57]	Twitter	892 000 tweets	Collected and analysed tweets from a sample of users and aimed to correlate food-related messages with state-wide health data (obesity rates).	Content analysis	✓	✓ Twitter API	✓	✓ LIWC Genderise API
Nguyen *et al.* 2017, USA[52]	Twitter	422 094 tweets	Collected food and physical activity tweets sent from one US state (Utah) and compared sentiment to health data on overweight and obesity to demonstrate validity of Twitter-derived neighbourhood characteristics	Content analysis	✓ Twitter API and geolocator	✓ Twitter API and geolocator	✓ Stanford Tokenizer	✓ MALLET
Widener *et al.* 2014, USA[53]	Twitter	148 533 tweets	Collected geolocated Twitter data in the USA and applied advanced data-mining framework to explore if prevalence of healthy and unhealthy food mentions could be mapped in line with USDA data.	Content analysis	✓ Twitter API	✓	✓	✓ Alchemy API
Chen *et al.* 2014, USA[63]	Twitter	81 543 tweets Subset ~350	Collected tweets with mentions of grocery stores and fast food outlets to map relationship between food choices and exposure to local food environment.	Content analysis	✓ Open Source Python library and keyword search	✓ Open Source Python library	✓/x Esri ArcMap 2.0	✓/x ANOVA
Vidal *et al.* 2015, Uruguay & NZ[54]	Twitter	48 746 tweets Subset 12 000	Explored main topics in tweets on breakfast, lunch, snack and dinner to evaluate Twitter's potential as a research tool.	Content analysis	✓	✓ TwitterR Package	✓ TwitterR Package	x
Wombacher *et al.* 2017, USA[66]	Twitter	851 tweets	Analysed Twitter data using hashtag #NeckNominate to identify normative forces at play in a popular, dare-based alcoholic drinking game	Content analysis	✓ Topsy	✓ Topsy	?	x
Turner-McGrievy *et al.* 2015, USA[75]	Twitter	?	Collected tweets over 1 year with #fitness, #diet, #health and 'weight' to explore temporal trends (influence of holiday period on frequency).	Content analysis, surveillance	✓ Hashtagify.me	✓ peoplebrowsr	?	?
ElTayeby *et al.* 2018, USA[67]	Facebook group	4266 posts	Investigated the feasibility of mining Facebook posts within 'I'm Shmacked' Facebook group and used machine learning to identify between alcohol drinking and non-alcohol posts with a view to exploring drinking behaviours in US college students	Content analysis	✓	✓ Python SDK and Facebook API	✓ Human annotation to train machine learning	✓ SVM or LLDA AlexNet
Blackstone *et al.* 2018, USA[76]	Facebook	617 posts and comments	Analysed posts from two public extreme fitness and nutrition Facebook groups for harmful health messages	Content analysis	✓	x	?	x
Ofli *et al.* 2017, USA and Qatar[60]	Instagram	1·9M images	Collected Instagram food images and hashtags and used image recognition technology to study differences in how a human describes the food compared with machine learning.	Content analysis	✓	✓ Python using Shapely	✓	x/✓ Manual classification using Amazon Mechanical Turk
Sharma *et al.* 2015, USA[62]	Instagram	1·8M	Collected posts with food hashtags and designed a machine learning model to link posts to USDA National Nutrient Database and estimate energy information.	Content analysis	✓ Instagram API	✓	✓	x/✓
Phan *et al.* 2019, Switzerland[72]	Instagram	1·6M	Analysed Instagram alcohol-related posts and compared with an existing data set in order to develop machine learning systems that monitor alcohol consumption.	Content analysis	✓	✓	✓	✓

Table 2 *Continued*

					SMA steps*			
First author, year and country	SoMe platforms	Data set size SMA summary	Purpose of study	Taxonomy	Data discovery	Data collection	Data preparation	Data analysis
Rich *et al.* 2016, UK[61]	Instagram	800 000 images	Analysed Instagram food images with the aim to improve machine learning food image recognition.	Content analysis	✓ Instagram API	✓ Instagram API	✓	✓
Pang *et al.* 2015, USA[71]	Instagram	195 000 images	Collected Instagram images with alcohol hashtags and used facial recognition technology to determine patterns of drinking in underage consumers.	Content analysis	✓/ x	✓	✓	✓ Face processing toolkit Face++
Silva *et al.* 2017, Brazil and UK[56]	Foursquare	~5M	Collected data from location-based network (Foursquare) check-ins on eating and drinking habits and proposed a methodology to identify boundaries and similarities across populations.	Content analysis	✓	✓ Foursquare/ Twitter API	✓	✓
Mejova *et al.* 2015, UK and Qatar using USA data[59]	Instagram Foursquare	20 K 194 K	Designed machine learning model to analyse posts for food consumption patterns with the aim to correlate with obesity rates and location of fast-food restaurants.	Content analysis	✓	✓	✓	✓/x Datascience toolkit API Instagram Endpoints API CrowdFlower
Rahman *et al.* 2016, Bangladesh[78]	Foursquare and Twitter	731 Twitter users	Collected tweets on Foursquare restaurant check-ins and designed predictive machine learning model with aim to assess if tweet words can predict eating out preferences.	Content analysis	✓ Greptweet.com	✓	✓	✓ LIWC WEKA
Carrotte *et al.* 2017, Australia[77]	Instagram Facebook Twitter Tumblr	360 12 4 39	Collected posts with hashtag #fitspo to characterise food, body image and dieting messages by gender.	Content analysis	✓ Tagboard.com	x	x	x
Primack *et al.* 2015, USA[68]	YouTube	200 videos retrieved and 70 subset	Collected and manually coded YouTube videos for sentiment related to alcohol intoxication to characterise the popular content	Content analysis	✓	x	x	x
Sullivan *et al.* 2016, USA[81]	Amazon.com	400 000 reviews of 2708 products	Collected Amazon.com user-generated reviews of nutritional supplements and used unsupervised natural language processing techniques to capture adverse reactions and design system to score products on adverse reaction potential	Content analysis, surveillance	✓	✓	✓ LDA	✓ LDA
Yan *et al.* 2015, China[79]	Koubei	10 136 reviews	Analysed online restaurant reviews in Harbin, China, to assess revisit intent and obtain evaluation indicators of which healthy was important in food quality.	Content analysis	?	?	?	✓
Zhou *et al.* 2018, China[80]	Weibo	3 975 800 microblogs	Used popular Chinese social media site to detect different preferences in types of meals using sentiment analysis and trends over time.	Content analysis, surveillance	✓	✓	✓	✓

SMA, social media analytics; API, Application Programming Interface; MALLET, Machine Learning for LanguagE Toolkit; LIWC, Linguistic Inquiry and Word Count; NLP, natural language processing; LDA, linear discriminant analysis; LLDA, local linear discriminant analysis; SVM, support vector machine; WEKA, waikato environment for knowledge analysis.
*SMA key:
✓ SMA was described for the relevant step.
X Manual extraction, coding or analysis by humans was described for the relevant step.
✓/x A combination of SMA and manual processes were described for the relevant step.
? The details could not be clearly identified for the relevant step from the paper. Authors were not contacted for additional information.

Table 3 Social media platforms in included studies in present rapid review

Platform	Description of platform	Articles No.	%	Range in data set size
Twitter www.twitter.com	An American online microblogging and social networking service launched in 2006 on which users post and interact with short messages known as 'tweets' containing text, video, images and links to the Internet. Registered users can post, like and retweet tweets.	21	62	4–80M
Instagram www.instagram.com	An American online social networking photo and video sharing platform launched in 2010 and now owned by Facebook in which users share and interact with 'followers'.	7	21	360–1·9M
Facebook www.facebook.com	An American online social networking service launched in 2004 in which users register and create a profile revealing information about themselves via text, photos and multimedia which is shared with any other users that have agreed to be their 'friend', or, with a different privacy setting, with any reader.	3	9	12–4250
Foursquare www.foursquare.com	An American online social networking location platform launched in 2009 in which users register and create a profile to 'check-in' to places such as restaurants or attractions and share real-time locations with followers.	3	9	194 K-5M
YouTube www.youtube.com	An American online video sharing social media platform launched in 2005 in which registered users upload, view, rate, share, add to playlists, report, comment on videos and subscribe to other users.	1	3	200
Tumblr www.tumblr.com	An American online social networking photo and video sharing platform launched in 2007 in which registered users interact via blogs.	1	3	39
Yelp www.yelp.com	An American company launched in 2004 running the yelp.com online platform in which users publish reviews in the public domain about businesses and services as well as make online reservations.	1	3	500 000
Amazon.com www.amazon.com	An American company running a major e-commerce site (and subsequent IOT platforms) launched in 1994 in which users purchase goods and provide user-generated ratings and reviews in the public domain.	1	3	40 000
Weibo www.weibo.com	A major Chinese online social networking platform launched in 2009 with similar features to Twitter.	1	3	4M
Koubei www.koubei.com	A major Chinese social media platform launched in 2004 with location features for shopping, entertainment and restaurants.	1	3	10 000

IOT, Internet of Things.

(88 %, *n* 30) utilised only one type of social media platform in the research investigation with only four studies (12 %, *n* 4) using two or more platforms.

The majority of the studies originated in the USA (62 %, *n* 21) with the remainder across a range of countries and collaborations including UK, Australia, China, Brazil, Uruguay, Bangladesh, Qatar and New Zealand.

Data set sizes ranged from 4 to 80M individual data points of social media content. Twitter studies had the largest data set sizes with 9 studies (26 %, *n* 9 of total included studies) having data sets of at least 10M tweets. Data sets in the millions were also seen in the Instagram, Weibo and FourSquare studies.

Nutrition topics of relevance to public health nutrition

Included studies were categorised by nutrition topics of relevance to public health nutrition and displayed in Table 4.

The majority of studies (53 %, *n* 18)[47–64] were cross-sectional in design and were categorised as investigations into general population health level, dietary behaviours as part of using social media with the aim to define demographic characteristics or sentiment. Studies in this category ranged from being exploratory in nature to test innovative, machine learning models, through to national population health-level dietary behaviour investigations validated against existing health data or other independent data such as location of fast-food outlets. The most promising studies were able to demonstrate a correlation between the location of high-energetic food mentions on Twitter and independent, government tracked, obesity consumption data[47] or validated machine learning as a tool for constructing indicators of food intakes compared with government census data[52]. Over two-thirds of studies in this category used Twitter (38 %, *n* 13) as the social media platform and source of data via tweets, and Instagram (12 %, *n* 4) and Foursquare (6 %, *n* 2) also utilised in more than one study.

The next largest category of studies involved insights into dietary behaviours with alcohol consumption (24 %, *n* 8) related to intoxication, binge-drinking and social norms[65–72]. An additional study in this category examined how to predict virality of content promoting drinking[73]. A post that 'goes viral' means one that becomes very popular on a social media platform in a short time period (hours or days), due to it being shared by a significant number of users, often in the millions, which increases the reach and amplitude[74].

The next category identified was dieting behaviours (13 %, *n* 3) (including food, body image, weight loss and dieting messages) to explore a range of investigations relevant to public health nutrition. These included detecting temporal trends such as patterns of weight loss dieting messages related to periods when weight gain is common, during or after Christmas or in particular seasons such as winter[75]. Blackstone *et al.* explored posts in two Facebook groups that were dedicated food and exercise

https://doi.org/10.1017/S1368980020005248 Published online by Cambridge University Press

Table 4 Categorisation of included studies by nutrition topics of relevance to public health nutrition in present rapid review

Nutrition topic	Articles, No.	%	Research focus
General population health dietary behaviours	18	53	Characterised food behaviour social media data across a national or local population level with some studies comparing or validating results with government census data, existing health or nutrition data, or with the local food environment[47–64].
Alcohol behaviours	9	26	Characterised social media data on alcohol to explore drinking-related behaviours related to intoxication, binge-drinking and social norms[65–72] and examined how to predict which content on alcohol could go viral (be commonly shared by social media users)[73].
Dieting behaviours	3	9	Characterised social media data on dieting behaviours (including food, body image and dieting messages) to explore temporal trends related to holiday periods[75], harmful messages[76] and gender differences[77].
Eating out of the home behaviours	3	9	Characterised social media data on restaurant check-ins to assess if content analysis can predict eating out preferences[78], and on restaurant reviews to assess customer preferences including healthy menu options[79]and trends in types of meals[80].
Dietary supplement behaviours	1	3	Collected Amazon.com user-generated reviews of nutritional supplements and designed machine learning system to score products on adverse reaction potential[81].

programmes led by the fitness industry advertised to 'jump start' a healthy lifestyle and found 88·6 % of content promoted harmful messages about dieting restraint, body image and losing body fat/weight[76]. Finally, a study focused on gender differences and stereotypes in social media using the hashtag #fitspo (fit inspiration) also revealed that 19·6 % of posts were thematically linked to food or dietary behaviours[77].

A further category characterised social media data on eating out of the home behaviours via studies looking at a range of investigations with restaurant check-ins or reviews. Rahman *et al.* correlated the analysis of the psycholinguistic content of tweets, the common words written by the user, with Foursquare restaurant check-ins to predict eating out preferences[78]. In this study, it was revealed that people who frequently use the words 'health' are more likely to visit mid-priced restaurants (as opposed to expensive, e.g. indulgent, fine dining or cheaper, e.g. fast food). The two studies on Chinese social media platforms Koubei and Weibo used online restaurant reviews to provide insights into dietary behaviours when eating out, whether healthy options were important to consumers and changing cuisine trends, respectively[79,80].

The single study on Amazon.com designed a surveillance system to monitor dietary behaviours with and flag adverse reactions to dietary supplements via user reviews, as an innovative alternative to traditional methods of therapeutic goods reporting[81].

Taxonomy

As displayed in Table 2, applying Sinneberg's Taxonomy[21] all of the studies included in this review performed some type of content analysis (100 %, *n* 34), followed by a smaller category also performing or planning for surveillance (12 %, *n* 4)[51,75,80,81]. One study also performed analysis of social media engagement[73]. There were no included studies that performed network analysis.

Social media analytics methods

Studies involving SMA methods clearly observed across all steps of data discovery, collection, preparation and analysis were prominent in Twitter (38 %, *n* 13)[47,49–53,55,57,58,64,69,70,73] with the largest data set at 80M tweets. Two large-scale studies designed sophisticated, customised machine learning systems with the 'Lexicocalorimeter' attempting to measure energy content of food tweets[50], and Shah *et al.* designed a machine learning classifier to assess the content of food-related tweets across the whole of Canada with demonstrated correlation with government consumption data[47]. In contrast, a number of studies (approximately 18 %, *n* 6)[48,54,63,66,68,77] across several platforms collected data sets using SMA for data discovery, collection and (often) preparation but relied on manual processes across the steps of data analysis. In these studies, data analyses typically involved human, manual coding, often using a code book designed for the study, and resulted in a much smaller subset of social media data making up the investigation.

The common use of in-platform search tools, along with third-party search tools (such as tagboard.com, Topsy and hashtagify.me), was observed in the data discovery phase to find data with relevant topics, hashtags or key words. In the data collection step, the use of a platform API to mine data such as the Twitter API, Foursquare API, Instagram API and Yelp API was common. In the data preparation step, the use of Stanford Tokenizer was named in two studies, a software that divides text into a sequence of tokens, which roughly correspond to words[82]. Across the data analysis step, four studies mentioned the use of MALLET (Machine Learning for LanguagE Toolkit) which is a Java-based package for statistical NLP, clustering, topic modelling, information extraction and other machine learning applications to text[83]. In addition, four studies included the use of LIWC (Linguistic Inquiry and Word

https://doi.org/10.1017/S1368980020005248 Published online by Cambridge University Press

Count) in data analysis which is a commercial software program used in NLP[84].

Interdisciplinary collaborations

The types of disciplines by the author listed affiliations or disciplines on the publication were grouped into broad categories by health/medicine-related disciplines and non-health-related disciplines including business/computer science and are presented in Table 5.

Approximately, a third of the studies with available affiliations (32 %, *n* 11) involved interdisciplinary collaborations between health/medicine or related disciplines, and business/computer science or related disciplines. Over an additional third (38 %, *n* 13) involved business/computer science or related disciplines only. Under a quarter of studies involved health/medicine disciplines (21 %, *n* 7) only.

Specific nutrition or dietetics discipline involvement was identified in only two of the studies, with one involving only health/medicine disciplines across health promotion, exercise science and nutrition[75] and the other a broader interdisciplinary collaboration between public health, nutrition science, urban planning and information disciplines[48].

Ethics status

Table 6 presents information extracted on the ethics status of included studies. Nearly, three-quarters of the studies (74 %, *n* 25) did not mention ethical considerations or approvals. Less than a quarter of studies (15 %, *n* 5) obtained formal human ethics approval[52,64,76,85] or exemption[65] by a relevant institutional review board. In the absence of ethical review process, only four studies (12 %) mentioned ethical considerations[65–67,75] including

Table 5 Extracted information on disciplinary scope as listed on publication* of included studies in present rapid review

Study	Platform	Health, Medicine, other related	Business, Informatics, Computer Science, other related
Interdisciplinary collaborations			
Alhabash *et al.*[73]	Twitter	Health	Computer Science, Advertising, Data Mining
Alajajian *et al.*[50]	Twitter	Culinary Arts and Food Science	Mathematical and Computer Science
Huang *et al.*[69]	Twitter	Public Health	Statistics, Engineering
Karami *et al.*[55]	Twitter	Public Health	Information Science
Nguyen *et al.*[49]	Twitter	Health, Global Health	Computing, Geography, Sociology
Nguyen *et al.*[52]	Twitter	Epidemiology, Biostatistics, Public Health	Computing, Geography, Sociology
Nguyen *et al.*[64]	Twitter Yelp	Epidemiology, Health	Computing, Geography, Sustainability
Shah *et al.*[47]	Twitter	Medicine	Mathematics and Computer Science
Sullivan *et al.*[81]	Amazon	Biomedical Informatics	Biomedical Informatics
Vydiswaran *et al.*[48]	Twitter	Public Health, Nutritional Science†	Information, Urban Planning
Wombacher *et al.*[66]	Twitter	Health	Communication
Health-related disciplines only			
Blackstone *et al.*[76]	Facebook	Public Health, Health Science	x
Carrotte *et al.*[77]	Instagram Facebook Twitter, Tumblr	Health, Medicine	x
Cavazos-Rehg *et al.*[65]	Twitter	Psychiatry, Medicine	*Assisted by external social media analytics companies*
ElTayeby *et al.*[67]	Facebook group	Health Science	x
Primack *et al.*[68]	YouTube	Medicine, Public Health	x
Turner-McGrievy *et al.*[75]	Twitter	Health Promotion, Exercise Science, RD†	x
Vidal *et al.*[54]	Twitter	Psychology, Food	x
Non-health-related disciplines only including business, computing and informatics			
Abbar *et al.*[57]	Twitter	x	Computing
Chen *et al.*[63]	Twitter	x	Geography
Kershaw *et al.*[70]	Twitter	x	Computing, Management Science
Ofli *et al.*[60]	Instagram	x	Computing
Phan *et al.*[72]	Instagram	x	Artificial Intelligence
Rahman *et al.*[78]	Twitter Foursquare	x	Engineering and Technology
Rich *et al.*[61]	Instagram	x	Engineering, Computer Science
Sharma *et al.*[62]	Instagram	x	Technology
Silva *et al.*[56]	Foursquare	x	Computer Science, Informatics, Geography
Sun *et al.*[51]	Twitter	x	Computer Science, Network Computing and Security
Widener *et al.*[53]	Twitter	x	Geography, Geospatial Analysis and Computation
Yan *et al.*[79]	Koubei	x	Management Science and Engineering
Zhou *et al.*[80]	Weibo	x	Information Management

*Studies where author affiliation not clearly identifiable: Fried *et al.*[58], Mejova *et al.*[59], Pang *et al.*[71]
†Presence of nutrition or dietetic discipline involvement. RD is the US credential for Registered Dietitian.

Table 6 Extracted information on the ethics status of included studies in present rapid review

Study	SoMe platform(s)	Ethics approval or mention if present
Formal ethics approval		
Blackstone *et al.*[76]	Facebook	Approved by the IRB at the researchers' universities.
Nguyen *et al.*[52]	Twitter	The University of Utah Institutional Review Board
Nguyen *et al.*[64]	Twitter Yelp	The University of Utah Institutional Review Board
Vyidiswaran *et al.*[85]	Twitter	University of Michigan Institutional Review Board
Formal ethics review with exemption granted		
Cavazos-Rehg *et al.*[65]	Twitter	Institutional review board exemption.
Mention of ethics consideration in the absence of formal review		
ElTayeby *et al.*[67]	Facebook group	Deidentified posts.
Karami *et al.*[55]	Twitter	Publicly available tweets.
Turner-McGrievy *et al.*[75]	Twitter	Conducted in accordance with ethical standards
Wombacher *et al.*[66]	Twitter	Messages on Twitter are publicly available
No ethical approval or mention		
Abbar *et al.* Twitter[57]; Alhabash *et al.* Twitter[73]; Alajajian *et al.* Twitter[50]; Carrotte *et al.* Twitter, Facebook, Tumblr, Instagram[77]; Chen *et al.* Twitter[63]; Fried *et al.* Twitter[58]; Huang *et al.* Twitter[69], Kershaw *et al.* Twitter[70], Mejova *et al.* Instagram, Foursquare[59], Ofli *et al.* Instagram[60], Pang *et al.* Instagram[71], Phan *et al.* Instagram[72], Primack *et al.* YouTube[68]; Rahman *et al.* Twitter, Foursquare[78], Rich *et al.* Instagram[61], Nguyen *et al.* Twitter[49]; Shah *et al.* Twitter[47]; Sharma *et al.* Instagram[62], Silva *et al.* Foursquare[56]; Sullivan *et al.* Amazon[81]; Sun, *et al.* Twitter[51]; Vidal *et al.* Twitter[54]; Widener *et al.* Twitter[53]; Yan *et al.* Koubei[79]; Zhou *et al.* Weibo[80].		

IRB, institutional review board.

anonymising data and that social media data were in the public domain.

The available data on the type of research disciplines (Table 5) were compared with that of available data in the presence of ethics approval or considerations (Table 6). All studies that obtained formal ethics approval[52,64,76,85] or exemption[65] involved a health-related discipline. Of the twenty-five studies with no mention of ethical approval or considerations, over half (54 %, *n* 13)[51,53,56,57,60–63,70,72,78–80] were conducted by business/computing disciplines only, with no health-related discipline involvement. The remaining studies without ethics approval or mentions showed mixed results with involvement of health disciplines only (13 %, *n* 3)[54,68,77] and interdisciplinary collaborations with the presence of health disciplines (25 %, *n* 6)[47,49,50,69,73,81].

Discussion

Principal results

This was the first ever rapid review, to the authors' knowledge, that involved SMA in nutrition research, particularly in the investigation of dietary behaviours. The review identified thirty-four studies involving general population health using SMA on public domain, social media data between 2014 and 2020. Nutrition topics were segmented into the main category of broad population nutrition health investigations with subcategories seen in alcohol consumption behaviours, along with minor categories in dieting behaviours and eating out of the home. It was identified that all studies involved content analysis with some evidence of surveillance and limited evidence of engagement when applying Sinnenberg's Taxonomy to describe the roles of social media in health research. Moreover, Twitter was found to be the predominant social media platform under investigation with large data set sizes into the tens of millions. Across all platforms, the use of SMA tools were observed in the steps of data discovery, collection and preparation, but to a lesser extent in data analysis. Approximately, a third of the studies involved interdisciplinary collaborations between health/medicine and business/computer science or related disciplines, and only two studies involved nutrition or dietetic discipline involvement. Less than a quarter of studies obtained formal human ethics approval or exemption by a relevant institutional review board and these all were lead or included health-related discipline research team members. These findings reveal existing SMA topics and platforms of relevance to nutrition researchers along with important considerations to inform future collaborative research.

Implications for future nutrition research

A key strength of this rapid review is being the first of a kind knowledge synthesis on SMA in nutrition research, particularly with the investigation of dietary behaviours. For nutrition researchers, it reveals existing social media platforms and categories of nutrition topics to inform further investigations and collaborative research, along with important considerations with ethical standards and technology.

Despite demonstrated capabilities to rapidly collect and analyse millions of individual social media posts and perform social listening or surveillance, this review supports further technological innovation before SMA will augment traditional research methods with investigating or tracking dietary behaviours. However, SMA shows promise for investigating dietary behaviours at a broad population census level and as a tool for mixed methods research.

https://doi.org/10.1017/S1368980020005248 Published online by Cambridge University Press

For example, in mixed methods research, SMA could be beneficial as a scoping or scanning tool to investigate dietary behaviours to inform surveys, focus groups or other participant investigations or interventions. SMA could also form an important triangulation instrument[86] in conjunction with multiple sources of data (such as interviews, focus groups or other sources) to increase the credibility and validity of research findings. One of the most promising studies, using new approaches to NLP, was able to demonstrate a correlation between the location of high-energetic food mentions on Twitter and independent, government-tracked, obesity consumption data[47]. According to Shah *et al*., their study showed that social media analysis on a large scale, with the use of NLP, can help identify food- and activity-related tweets and is readily available as a close representation of real time[47]. In addition, Nguyen *et al.* was also able to demonstrate machine learning as a validated tool for constructing indicators of food intakes across local government areas compared with census data[52]. However, in both studies, the approaches were unable to provide in-depth, quantitative data on food or nutrient intakes on par with that obtained from surveys. Technological innovation is required before SMA could be considered as a novel source to augment major public health nutrition surveillance tools like national nutrition surveys. Innovation will likely be needed across all steps of SMA such as being able to more accurately extract posts from individuals and filter those of businesses or brands in data collection, right through to more sophisticated machine learning models in the analysis steps, such as determining dietary behaviours from posts with slang, text abbreviations or colloquial language.

Collaborative and interdisciplinary research was explored in this review as it is often a requirement of research funders and vital in the modern era with the study of complex phenomena[87]. Complementing the capabilities of an expert data analytics team with experienced health or nutrition researchers could enable large data sets to be processed and produce promising insights and outcomes for public health nutrition. Due to the small number of included studies and top line assessment of disciplinary involvement (via assessment of author listed affiliation on each publication), it is not possible to draw conclusions from this rapid review; however, relevant examples were observed in studies with and without health discipline involvement that warrant further investigation. In the multidisciplinary team of Vydiswaran *et al*.[48], which included nutrition expertise, Twitter was demonstrated at being able to characterise neighbourhood-level food-related behaviours and attitudes/sentiment in food-related tweets. The researchers concluded that social media data can provide a reliable signal for dietary patterns and food-related attitudes at the census level, despite the noisy nature of user-generated text data, the limited fraction of geolocated tweets and access only to public discussions rather than actual dietary patterns. In contrast, in an example of computer science researchers working in isolation of nutrition-health researchers, there was an emphasis on demonstrating technology processes and at times lack of in-depth questioning or judgements on human dietary behaviours under investigation[53]. In this study by Widener *et al*., researchers from the fields of geography and computation compiled their own narrow list of 'healthy' and 'unhealthy foods' to inform coding rather than accessing existing, comprehensive, validated dietary assessment instruments[53]. In contrast, as another example, in Vidal *et al.* study, health researchers working in isolation of computer informatics expertise relied on manual content analysis which resulted in a much smaller subset being coded or collated as findings and not the larger data set extracted[54]. It is also warranted to point out that specific nutrition or dietetics discipline involvement was identified in only two of the studies[48,75] which could indicate an underutilised opportunity for nutrition researchers and collaborative research.

Future investigations into social media platforms capabilities by nutrition researchers is warranted, along with keeping abreast of new and emerging platforms that may provide more in-depth or reliable insights into dietary behaviours[88]. Even though this review found that Twitter was the prominent social media platform, it should not be concluded that this represents a preferred tool for investigating dietary behaviours with SMA. Social media tools and the digital landscape are rapidly changing, and the prominence of Twitter could be explained by the ease of extraction of data via the Twitter API[89] and the fact that Twitter, launched in 2006, has been established longer compared with relatively newcomer platforms such as Instagram launched in 2010[21].

Paramount to nutrition research is the conduction of ethical research according to the Declaration of Helsinki including human ethics review and formal participant consent[41]. However, there are significant ethical challenges facing nutrition and health researchers accessing social media data that require further exploration. The rapid rise in the popularity of social media has presented new dilemmas for tech companies and governments as privacy laws, data sharing policies and consumer protections have failed to keep pace[90]. In this review, all studies that did obtain formal ethics approval[52,64,76,85] or exemption[65] involved a health-related discipline. However, nearly three-quarters of the studies were found to make no mention of ethical considerations at all, with the majority here involving business/computing disciplines only; however, there were overall mixed results in the presence of health disciplines and ethical approval. These findings complement a recent review into social media data mining and health, highlighting that many researchers do not seek, or view the need to seek, formal human ethics approval[91].

Nutrition and health professionals also face an array of additional, ethical practice dilemmas when utilising

https://doi.org/10.1017/S1368980020005248 Published online by Cambridge University Press

social media for research more broadly, such as when performing surveillance on at-risk communities with no intention-to-treat or with unclear relational boundaries in researcher–participant interactions. This review demonstrated that there was a lack of consistency and frameworks for the ethical assessment of social media data in research. Assumptions were made by some that, given users have chosen to place content in the public domain, social media data are outside ethical jurisdiction. Fundamentally, mining data from public domain social media accounts does involve easily identifiable individuals, at times sharing intimate details of their habits, who do not provide informed consent. Simply complying with company-derived privacy policies, along with established API terms for third-party use, is inordinately insufficient to maintain ethical integrity when collecting and storing personal social media data for nutrition research. It will be vital that these framework gaps, particularly in confidentiality and anonymity, should be addressed, and several groups including those from the US Council for Big Data Ethics and Society, the UK Data Service as well as the UK Society for Data Miners' are currently working on ethical guidelines for researchers[91].

Limitations

The number of databases searched was limited to five to complete the rapid review within the time frame set. For this reason, there were also no further searches conducted following the execution of the planned search strategy. Limitations also included English-language-only studies which may have excluded studies from other markets such as China where technology-related research is well established. Assessment of the public health nutrition relevance of studies to the investigation of dietary behaviours was a subjective process by two Accredited Dietitians and may have different interpretations by other nutrition researchers, for example, studies involving user-generated data on cuisine preferences in restaurant reviews may be interpreted as relevant or irrelevant to public health nutrition. Analysis of the disciplinary involvement was performed by listed author affiliations, and this review did not delve deeper into credentials or relevant researcher experience and conclusions could not be drawn on the probable benefits of collaborative research. Charting of the SMA steps uncovered areas of undefined scope and therefore was incomplete and generalised towards top line assessment of the scope of usage across the key SMA steps. Study authors were not contacted to clarify reporting gaps due to the nature of this rapid review. When assessing SMA use in each step across data discovery, preparation, collection and analysis, there were studies where manual coding was charted and SMA was not observed. However, the reasons for manual coding were not assessed and may not represent a lack of SMA expertise, rather there may have been limitations in machine learning capabilities or reliability warranting manual coding, or human coding may have been used to inform machine learning, or a combination, and specific interpretation was outside the scope of this review.

Conclusions

This rapid review was the first knowledge synthesis on SMA in nutrition research, particularly with the investigation of dietary behaviours in general population health using SMA on public domain, social media data and revealed that it is still emerging. The review revealed existing topics and platforms for nutrition and dietetics professionals to inform future collaborative research and to use SMA at a broad population census level, as a scoping tool or complementary, triangulation instrument to traditional research methods. However, careful consideration and planning need to be taken to investigate technological capabilities and maintain ethical integrity. Nevertheless, with the strong track record of innovation from rapidly advancing digital technology, an established track record of SMA in health and other research fields, along with the popularity of sharing eating behaviours on social media, the future of SMA in nutrition research with the investigation of dietary behaviours shows promise.

Acknowledgements

Acknowledgements: Craig Mack provided insights into social listening expertise in commercial organisations. Liz Harris, Senior Research Advisor La Trobe University Library, provided assistance with the search strategy. Emma Stirling is an Advanced Accrediting Practising Dietitian currently undertaking her PhD studies at La Trobe University and is also a Senior Lecturer in the Discipline of Nutrition and Dietetics, Australian Catholic University. *Financial support:* This research received no external funding. *Conflict of interest:* None. *Authorship:* ES, JW, K-L O and AF conceived and designed the study. ES, JW and AF contributed to data acquisition, analysis and interpretation. ES drafted the manuscript. JW and AF assisted with critically revising the manuscript. All authors approved the final version of this manuscript. *Ethics of human subject participation:* Not applicable.

Supplementary material

For supplementary material accompanying this paper, visit https://doi.org/10.1017/S1368980020005248

References

1. Webster M Dictionary (2019) Social media definition. https://www.merriam-webster.com/dictionary/social%20media (accessed October 2019).

2. Kemp S (2020) Digital 2020 Global Digital Overview. https://datareportal.com/reports/digital-2020-global-digital-overview (accessed October 2019).
3. Lewis T & Phillipov M (2018) Food/media: eating, cooking, and provisioning in a digital world. *Commun Res Pract* **4**, 207–211.
4. Lewis T (2018) Digital food: from paddock to platform. *Commun Res Pract* **4**, 212–228.
5. Stewart MC & Arnold CL (2018) Defining social listening: recognizing an emerging dimension of listening. *Int J Listening* **32**, 85–100.
6. Stieglitz S, Mirbabaie M, Ross B *et al.* (2018) Social media analytics – challenges in topic discovery, data collection, and data preparation. *Int J Inf Manage* **39**, 156–168.
7. Batrinca B & Treleaven PC (2015) Social media analytics: a survey of techniques, tools and platforms. *AI Soc* **30**, 89–116.
8. Hernandez I & Zhang Y (2017) Using predictive analytics and big data to optimize pharmaceutical outcomes. *Am J Health Syst Pharm* **74**, 1494–1500.
9. Hoyt RE, Snider D, Thompson C *et al.* (2016) IBM Watson analytics: automating visualization, descriptive, and predictive statistics. *JMIR Public Health Surveill* **2**, e157.
10. Gohil S, Vuik S & Darzi A (2018) Sentiment analysis of health care tweets: review of the methods used. *JMIR Public Health Surveill* **4**, e43.
11. Wasilewski MB, Stinson JN, Webster F *et al.* (2018) Using Twitter to recruit participants for health research: an example from a caregiving study. *Health Inf J* **25**, 1485–1497.
12. Volkova E, Michie J, Corrigan C *et al.* (2017) Effectiveness of recruitment to a smartphone-delivered nutrition intervention in New Zealand: analysis of a randomised controlled trial. *BMJ Open* **7**, e016198.
13. Arigo D, Pagoto S, Carter-Harris L *et al.* (2018) Using social media for health research: methodological and ethical considerations for recruitment and intervention delivery. *Digital Health* **4**, 2055207618771757.
14. Loh IH, Schwendler T, Trude ACB *et al.* (2018) Implementation of text-messaging and social media strategies in a multilevel childhood obesity prevention intervention: process evaluation results. *INQUIRY: J Health Care Org Provis Financing* **55**, 0046958018779189.
15. Jane M, Foster J, Hagger M *et al.* (2018) Psychological effects of belonging to a Facebook weight management group in overweight and obese adults: results of a randomised controlled trial. *Health Social Care Commun* **26**, 714–724.
16. Baker SA & Walsh MJ (2018) 'Good Morning Fitfam': top posts, hashtags and gender display on Instagram. *New Media Society* **20**, 4553–4570.
17. Chung CF, Agapie E, Schroeder J *et al.* (2017) When personal tracking becomes social: examining the use of Instagram for healthy eating. *Proc SIGCHI Conf Hum Factor Comput Syst* **2017**, 1674–1687.
18. Lyons AC, Goodwin I, McCreanor T *et al.* (2015) Social networking and young adults' drinking practices: innovative qualitative methods for health behavior research. *Health Psychol* **34**, 293.
19. Yan H, Fitzsimmons-Craft EE, Goodman M *et al.* (2019) Automatic detection of eating disorder-related social media posts that could benefit from a mental health intervention. *Int J Eat Disord* **52**, 1150–1156.
20. Sadilek A, Kautz H, Di Prete L *et al.* (2017) Deploying nemesis: preventing foodborne illness by data mining social media. *AI Mag* **38**, 37–48.
21. Sinnenberg L, Buttenheim AM, Padrez K *et al.* (2017) Twitter as a tool for health research: a systematic review. *Am J Public Health* **107**, e1–e8.
22. Zhang AJ, Albrecht L & Scott SD (2018) Using twitter for data collection with health-care consumers: a scoping review. *Int J Qual Methods* **17**, 1609406917750782.
23. Lardon J, Bellet F, Aboukhamis R *et al.* (2018) Evaluating Twitter as a complementary data source for pharmacovigilance. *Expert Opin Drug Saf* **17**, 763–774.
24. Tricco AC, Zarin W, Lillie E *et al.* (2018) Utility of social media and crowd-intelligence data for pharmacovigilance: a scoping review. *BMC Med Inf Decis Making* **18**, 38.
25. Krauss MJ, Sowles SJ, Moreno M *et al.* (2015) Hookah-related twitter chatter: a content analysis. *Prev Chronic Dis* **12**, E121.
26. Grant A & O'Mahoney H (2016) Portrayal of waterpipe (shisha, hookah, nargile) smoking on Twitter: a qualitative exploration. *Public Health* **140**, 128–135.
27. Allem JP, Chu KH, Cruz TB *et al.* (2017) Waterpipe promotion and use on Instagram: #Hookah. *Nicotine Tob Res* **19**, 1248–1252.
28. Allem JP, Escobedo P, Chu KH *et al.* (2017) Images of little cigars and cigarillos on Instagram identified by the hashtag #swisher: thematic analysis. *J Med Internet Res* **19**, e255.
29. Allem JP, Ramanujam J, Lerman K *et al.* (2017) Identifying sentiment of Hookah-related posts on Twitter. *JMIR Public Health Surveill* **3**, e74.
30. Ben Taleb Z, Laestadius LI, Asfar T *et al.* (2018) #Hookahlife: the rise of Waterpipe promotion on Instagram. *Health Educ Behav*. Published online: 28 June 2018. doi: 10.1177/1090198118779131.
31. Alkouz B, Al Aghbari Z & Abawajy JH (2019) Tweetluenza: predicting Flu trends from Twitter data. *Big Data Mining Anal* **2**, 273–287.
32. Zadeh AH, Zolbanin HM, Sharda R *et al.* (2019) Social media for nowcasting Flu activity: spatio-temporal big data analysis. *Inf Syst Front* **21**, 743–760.
33. Sutton J, Vos SC, Olson MK *et al.* (2018) Lung cancer messages on Twitter: content analysis and evaluation. *J Am Coll Radiol* **15**, 210–217.
34. Teoh D, Shaikh R, Vogel RI *et al.* (2018) A cross-sectional review of cervical cancer messages on twitter during cervical cancer awareness month. *J Low Genit Tract Dis* **22**, 8–12.
35. Seabrook EM, Kern ML, Fulcher BD *et al.* (2018) Predicting depression from language-based emotion dynamics: longitudinal analysis of Facebook and Twitter status updates. *J Med Internet Res* **20**, e168.
36. Ricard BJ, Marsch LA, Crosier B *et al.* (2018) Exploring the utility of community-generated social media content for detecting depression: an analytical study on Instagram. *J Med Internet Res* **20**, e11817.
37. Daiphule L, Reddy B, Savith A *et al.* (2019) Tracking suicidal tendency using twitter data and machine learning algorithms. *Int J Eng Adv Technol* **8**, 188–191.
38. Liu X, Sun J, Yu NX *et al.* (2019) Proactive suicide prevention online (PSPO): machine identification and crisis management for Chinese social media users with suicidal thoughts and behaviors. *J Med Internet Res* **21**, e11705.
39. Thorstad R & Wolff P (2019) Predicting future mental illness from social media: a big-data approach. *Behav Res Methods* **51**, 1586–1600.
40. Hussein DME-DM (2018) A survey on sentiment analysis challenges. *J King Saud Univ Eng Sci* **30**, 330–338.
41. National Library of M (2001) World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Bull World Health Organ* **79**, 373–374.
42. Pang T (2018) Rapid guidelines – timely and important guidance needed for setting standards and best practices. *Health Res Policy Syst* **16**, 56.
43. WHO (2017) *Rapid Reviews to Strengthen Health Policy and Systems: A Practical Guide*. Geneva: WHO
44. Dobbins M (2017) *Rapid Review Guidebook*. Hamilton, ON: National Collaborating Centre for Methods and Tools. https://www.nccmt.ca/tools/rapid-review-guidebook (accessed October 2020).

https://doi.org/10.1017/S1368980020005248 Published online by Cambridge University Press

45. Stirling E (2019) Social Media Analytics in Nutrition Related Research Scoping Review Protocol. https://osf.io/ed72t/ (accessed October 2019).
46. Taylor J & Pagliari C (2018) Comprehensive scoping review of health research using social media data. *BMJ Open* **8**, e022931.
47. Shah N, Srivastava G, Savage DW *et al.* (2020) Assessing Canadians health activity and nutritional habits through social media. *Front Public Health* **7**, 400.
48. Vydiswaran VGV, Romero DM, Zhao X *et al.* (2020) Uncovering the relationship between food-related discussion on Twitter and neighborhood characteristics. *J Am Med Inf Assoc* **27**, 254–264.
49. Nguyen QC, Li D, Meng HW *et al.* (2016) Building a national neighborhood dataset from geotagged twitter data for indicators of happiness, diet, and physical activity. *JMIR Public Health Surv* **2**, e158.
50. Alajajian SE, Williams JR, Reagan AJ *et al.* (2017) The Lexicocalorimeter: Gauging public health through caloric input and output on social media. *PLoS ONE* **12**, e0168893.
51. Sun Q, Wang N, Li S *et al.* (2018) Local spatial obesity analysis and estimation using online social network sensors. *J Biomed Inf* **83**, 54–62.
52. Nguyen QC, Brunisholz KD, Yu W *et al.* (2017) Twitter-derived neighborhood characteristics associated with obesity and diabetes. *Sci Rep* **7**, 1–10.
53. Widener MJ & Li W (2014) Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US. *Appl Geogr* **54**, 189–197.
54. Vidal L, Ares G, Machín L *et al.* (2015) Using Twitter data for food-related consumer research: a case study on "what people say when tweeting about different eating situations". *Food Qual Preference* **45**, 58–69.
55. Karami A, Dahl AA, Turner-McGrievy G *et al.* (2018) Characterizing diabetes, diet, exercise, and obesity comments on Twitter. *Int J Inf Manage* **38**, 1–6.
56. Silva TH, Melo POSVD & Almeida JM *et al.* (2017) A large-scale study of cultural differences using urban data about eating and drinking preferences. *Inf Syst* **72**, 95–116.
57. Abbar S, Mejova Y & Weber I (2015) You tweet what you eat: studying food consumption through twitter. https://doi.org/10.1145/2702123.2702153 (accessed April 2015).
58. Fried D, Surdeanu M, Kobourov S *et al.* (2014) Analyzing the language of food on social media. *IEEE Int Conf Big Data*. Published online: 8 January 2015. doi: 10.1109/BigData.2014.7004305.
59. Mejova Y, Haddadi H, Noulas A *et al.* (2015) #FoodPorn: Obesity Patterns in Culinary Interactions. Proceedings of the 5th International Conference on Digital Health 2015. Florence, Italy: ACM. https://doi.org/10.1145/2750511.2750524 (accessed May 2015).
60. Ofli F, Aytar Y, Weber I *et al.* (2017) Is Saki #delicious? The Food Perception Gap on Instagram and Its Relation to Health. Proceedings of the 26th International Conference on World Wide Web. Perth: International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/3038912.3052663 (accessed April 2017).
61. Rich J, Haddadi H & Hospedales TM (2016) Towards Bottom-Up Analysis of Social Food. Proceedings of the 6th International Conference on Digital Health Conference. Montreal, QC: ACM. https://doi.org/10.1145/2896338.2897734 (accessed April 2016).
62. Sharma S & De Choudhury M (2015) Measuring and Characterizing Nutritional Information of Food and Ingestion Content in Instagram. https://doi.org/10.1145/2740908.2742754 (accessed May 2015).
63. Chen X & Yang X (2014) Does food environment influence food choices? A geographical analysis through "tweets". *Appl Geogr* **51**, 82–89.
64. Nguyen QC, Meng H, Li D *et al.* (2017) Social media indicators of the food environment and state health outcomes. *Public Health* **148**, 120–128.
65. Cavazos-Rehg PA, Krauss MJ, Sowles SJ *et al.* (2015) "Hey Everyone, I'm Drunk." an evaluation of drinking-related Twitter chatter. *J Stud Alcohol Drugs* **76**, 635.
66. Wombacher K, Reno JE & Veil SR (2017) NekNominate: social norms, social media, and binge drinking. *Health Commun* **32**, 596–602.
67. ElTayeby O, Eaglin T, Abdullah M *et al.* (2018) A feasibility study on identifying drinking-related contents in Facebook through mining heterogeneous data. *Health Inf J*. Published online: 19 September 2018. doi: 10.1177/1460458218798084.
68. Primack BA, Colditz JB, Pang KC *et al.* (2015) Portrayal of alcohol intoxication on YouTube. *Alcohol: Clin Exp Res* **39**, 496–503.
69. Huang T, Elghafari A, Relia K *et al.* (2017) High-resolution temporal representations of alcohol and tobacco behaviors from social media data. Proceedings of the ACM on Human-Computer Interaction 1. https://doi.org/10.1145/3134689 (accessed December 2017).
70. Kershaw D, Rowe M & Stacey P (2014) Towards tracking and analysing regional alcohol consumption patterns in the UK through the use of social media. https://doi.org/10.1145/2615569.2615678 (accessed June 2014).
71. Pang R, Baretto A, Kautz H *et al.* (2015) Monitoring adolescent alcohol use via multimodal analysis in social multimedia. Proceedings of the 2015 IEEE International Conference on Big Data (Big Data): IEEE Computer Society. https://doi.org/10.1109/BigData.2015.7363914 (accessed December 2015).
72. Phan T-T, Muralidhar S & Gatica-Perez D (2019) Drinks & crowds: characterizing alcohol consumption through crowdsensing and social media. Proceedings of ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3, 1–30. https://doi.org/10.1145/3328930 (accessed June 2019).
73. Alhabash S, VanDam C, Tan PN *et al.* (2018) 140 characters of intoxication: exploring the prevalence of alcohol-related Tweets and predicting their virality. *SAGE Open* **8**, 2158244018803137.
74. Himelboim I & Golan GJ (2019) A social networks approach to viral advertising: the role of primary, contextual, and low influencers. Social Media + Society. Published online: 21 July 2019. doi: 10.1177/2056305119847516.
75. Turner-McGrievy GM & Beets MW (2015) Tweet for health: using an online social network to examine temporal trends in weight loss-related posts. *Transl Behav Med* **5**, 160–166.
76. Blackstone SR & Herrmann LK (2018) Extreme body messages: themes from Facebook posts in extreme fitness and nutrition online support groups. *Began 2015* **4**, 33.
77. Carrotte ER, Prichard I & Lim MSC (2017) "Fitspiration" on social media: a content analysis of gendered images. *J Med Internet Res* **19**, e95.
78. Rahman MM, Majumder MTH, Mukta MSH *et al.* (2016) Can we predict eat-out preference of a person from tweets? In Proceedings of the 8th ACM Conference on Web Science. Hannover, Germany: ACM. https://doi.org/10.1145/2908131.2908199 (accessed May 2016).
79. Yan X, Wang J & Chau M (2015) Customer revisit intention to restaurants: evidence from online reviews. *Inf Syst Front* **17**, 645–657.
80. Zhou Q & Zhang C (2018) Detecting users' dietary preferences and their evolutions via Chinese social media. *J Database Manag* **29**, 89–110.
81. Sullivan R, Sarker A, O'Connor K *et al.* (2016) Finding potentially unsafe nutritional supplements from user reviews with topic modeling. *Pac Symp Biocomput* **21**, 528–539.

Public Health Nutrition

82. Stanford Natural Processing Group (2020) Stanford Tokeniser. https://nlp.stanford.edu/software/tokenizer.shtml (accessed June 2020).
83. MALLET (2020) Machine Learning Language Toolkit. http://mallet.cs.umass.edu/ (accessed June 2020).
84. LIWC Linguistic Inquiry & Word Count (2015) DISCOVER LIWC. https://liwc.wpengine.com/ (accessed June 2020).
85. Zhang L, Hall M & Bastola D (2018) Utilizing Twitter data for analysis of chemotherapy. *Int J Med Inf* **120**, 92–100.
86. Noble H & Heale R (2019) Triangulation in research, with examples. *Evid Based Nurs* **22**, 67–68.
87. Nyström M, Karltun K, Keller C *et al.* (2018) Collaborative and partnership research for improvement of health and social services: researcher's experiences from 20 projects. *Health Res Policy Syst* **16**, 1–17.
88. Hermann J (2019) How TikTok is rewriting the world. https://www.nytimes.com/2019/03/10/style/what-is-tik-tok.html (accessed March 2019).
89. Moessner M, Feldhege J, Wolf M *et al.* (2018) Analyzing big data in social media: text and network analyses of an eating disorder forum. *Int J Eat Disord* **51**, 656–667.
90. Commission ACAC (2019) Digital Platforms Inquiry – Final Report. Canberra, Australia: Commonwealth of Australia.
91. Taylor J & Pagliari C (2017) Mining social media data: how are research sponsors and researchers addressing the ethical challenges? *Res Ethics* **14**, 1–39.