Petersen et al., Gerontechnology (2023) 22:1
https://doi.org/10.4017/gt.2023.22.1.824.06

Adapting natural language processing and sentiment analysis methods for intervention in older adults: Positive perceptions of health and technology

Curtis L. Petersen MPH PhD[a,b], Xingyi Li BS[c], Courtney J. Stevens PhD[d], Tyler L. Goodinge BA[e], Elizabeth A. Carpenter-Song PhD[f], John A. Batsis MD[g,h,*]

[a]*The Dartmouth Institute for Health Policy, Dartmouth College, Hanover, New Hampshire, United States;* [b]*Quantitative Biomedical Sciences Program, Dartmouth College, Hanover, New Hampshire, United States;* [c]*Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, United States;* [d]*Department of Psychiatry, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire, United States;* [e]*Department of Medicine, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire, United States;* [f]*Dartmouth College, Hanover, New Hampshire, United States;* [g]*Division of Geriatric Medicine, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States;* [h]*Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States;*
**Corresponding author: john.batsis@gmail.com*

Abstract

Background: Older adults frequently participate in behavior change studies, yet it is not clear how to quantify a potential relationship between their perception of the intervention and its efficacy.
Objective: We assessed the relationship between participant sentiment toward the intervention from follow-up interviews with physical activity and questionnaires for the perception of health.
Methods: Sentiment was calculated using the transcripts of exit interviews through a bag of words approach defined as the sum of positive and negative words in 28 older adults with obesity (body mass index ≥30kg/m^2).
Results: Mean age was 73 years (82% female), and 54% lost ≥5% weight loss. Through linear regression we describe a significant association between positive sentiment about the intervention and weight loss; positive sentiment on technology and change in PROMIS-10 physical health and reduced physical activity time, while controlling for sex and age.
Conclusion: This analysis demonstrates that sentiment analysis and natural language processing in program review identified an association between perception and topics with clinical outcomes.

Keywords: older adults, mHealth, sentiment analysis, obesity, weight loss, natural language processing

Introduction

The United States has increased rates of obesity (a body mass index (BMI) ≥30 kg/m2) even in older adults that exceed 40% in 2016 (Hales, Carroll, Fryar, & Ogden, 2017). While obesity is associated with medical comorbidities, in older adults aged >65 years, it is associated with an increased risk of functional limitations (Jensen, 2005). Hence, weight management interventions consisting of combined dietary and exercise interventions are critical for overall wellness in improving physical function in this population (Batsis et al., 2017). Not all older adults with obesity are fully engaged in such programs. Higher levels of engagement are known to be associated with greater weight loss (Johnson & Wardle, 2011; LaRose, Fava, Lanoye, & Caccavale, 2019) despite barriers that need to be surmounted such as cost, confidence in technology use, transportation, and access to the internet (Liljas et al., 2017). A poorly explored barrier to engagement is the degree a participant actually likes or enjoys the intervention. While studies often provide qualitative evidence of a participant's emotions or perception of the intervention, it may be difficult to systematically assess repeatedly.

Mobile health (mHealth) is a rapidly developing field for engaging older adults in managing their health through electronic devices or monitoring

Table 1. Participant characteristics (n=28)

Age, years	72.93 (5.3)
Female sex	23 (82.1%)
Weight change, kg	-4.6 (3.2)
Reduction 5%	
Negligible weight loss or gain	13 (46.4%)
5% reduction in weight	15 (53.6%)
Mean daily physical activity	
Steps	5661.7 (2757.0)
Distance (miles)	2.9 (1.6)
Active (minutes)	205.0 (83.3)
Change in PROMIS	
Physical health t score	4.43 (4.95)
Mental health t score	5.00 (5.66)
Change in PAM	14.9 (13.1)

Values represented as means (standard deviations) or counts (%)
Patient activation measure (PAM), Patient-Reported Outcomes Measurement Information System (PROMIS)

systems (Changizi & Kaveh, 2017). While older adults are interested and willing to use mHealth modalities for health- promotion (Batsis et al., 2019), the amount of technical support they receive, such as live training or 1:1 support, may impact their success (Kampmeijer, Pavlova, Tambor, Golinowska, & Groot, 2016). Incorporating mHealth in weight loss programs can be beneficial; in studies of older adults, physical activity increased when self-management instructions were sent through smartphone devices (Batsis et al., 2020; Knight, Stuckey, & Petrella, 2014). To monitor physical activity, activity trackers such as Fitbit have become ubiquitous, demonstrating feasibility and acceptability when used with older adults participating in weight loss programs (Batsis et al., 2016).

To our knowledge, there are no studies that have quantified the features of exit interviews. Two methods that can quantify participants' perceptions and attitudes toward study interventions are natural language processing (NLP) and sentiment analysis, both of which use probabilistic models and word dictionaries to identify patterns within the written text. NLP and sentiment analysis can examine patient perceptions of treatments through social media posts (Gohil, Vuik, & Darzi, 2018) and investor perceptions of earning calls (Amenity Analytics, 2019). Through developed analytic pipelines and workflows NLP and sentiment analysis may potentially assist researchers and health professionals assess program implementation. The purpose of this study was to use NLP and sentiment analysis retrospectively for hypothesis generation as to how older adults perceive a mHealth-based weight loss intervention, and whether their perception was associated with a change in weight, physical activity, and questionnaires for the perception of health and use of the intervention.

Methods

Design and setting

Twenty-eight participants completed a 12-week behavioral weight-loss intervention consisting of regular dietary counseling and exercise classes. Eligibility criteria included age ≥65 years and a BMI ≥30kg/m2. Physical activity was monitored using a commercial, wearable fitness device (Fitbit Flex 2). Participants were recruited from a rural geriatric clinic at an academic medical center in rural New England. Details on the recruitment methods and study protocol have been previously reported (Batsis et al., 2020). All participants provided written informed consent to participate. The study was approved by the Committee for the Protection of Human Subjects at Dartmouth College and the Dartmouth-Hitchcock Institutional Review Board and was registered on clinicaltrials.gov (NCT03104192).

Study variables

Demographic and comorbidities were abstracted from the medical record by a member of the study team. All participants completed questionnaires using RedCAP (Harris et al., 2009) and objective measures at baseline and at the conclusion of the study. The Patient Activation Measure (PAM) was administered, comprised of 22 items, resulting in a 0-100 numerical measure of a patient's perceived ability to self-manage their care with 0 being no ability and 100 being full ability (Hibbard, Stockard, Mahoney, & Tusler, 2004). Participants also completed the 10-item Patient-Reported Outcomes Measurement Information System (PROMIS) Global Scale Short Form. Mental and physical health subscale scores were assessed (Cella et al., 2007) and standardized to a T-score with a mean of 50 with a 10-point standard deviation. Higher scores indicate greater health or function. Weight (kg) was assessed using an A&D digital scale, and height (m) was measured with a stadiometer by a trained research assistant. All study variables were stored in RedCap (Harris et al., 2009).

Wearable device

Participants wore a Fitbit Flex 2 tracker which recorded two physical activity metrics: (1) the number of daily steps; and (2) daily minutes spent "Lightly Active", "Fairly Active", and "Very Active". These activity levels are determined and defined by proprietary algorithms from Fitbit using metabolic equivalents (METs) based on heart rate. Total active time was calculated as the sum time of all three Fitbit-defined activity levels. Daily total steps were recorded on the Fitbit device and exported weekly by study staff. Daily non-wear time was defined as having fewer than 100 steps per

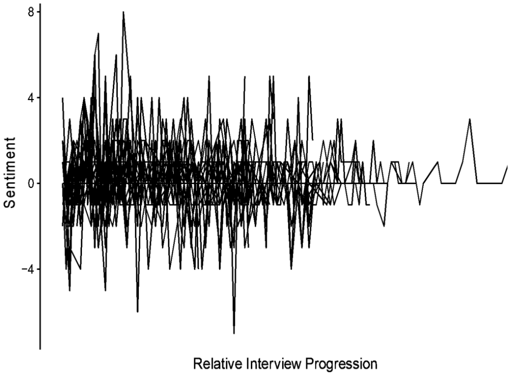

Figure 1. Sentiment of each statement over the course of each interview (Each line represents an individual participant interview with negative values corresponding to negative sentiment and positive values positive sentiment. Horizontal value is number of sentences within an interview. Here we demonstrate that we are able to measure the changing sentiment over the course of an interview.)

day. Compliance was defined as the percentage of days worn during their enrolled timeframe. At the study's conclusion, we calculated the mean active time and mean steps per day for all days that the device was worn for each participant.

Sentiment analysis and natural language processing

All interviews were conducted by the same member of the research team, digitally recorded, and transcribed. De-identified interview transcripts were separated by question and response type (general feedback, dislikes of the program, likes of the program, study personnel, technology, and usefulness). The text was separated by utterance per speaker after removing the interviewer's component. All stop words were removed to develop the final document corpus. Using the Bing sentiment lexicon (Hu & Liu, 2004; Liu, 2016), we calculated positive and negative sentiment through the bag-of-words method that matches words to a dictionary. The total sentiment was defined as the sum of positive and negative sentiments with positive sentiments having a +1 value and negative sentiments having a -1 value. Thus, a higher numerical value of sentiment can be interpreted as positive sentiment while a lower numerical value can be interpreted as more negative.

Statistical analysis

Descriptive statistics are used to mean and standard deviation for continuous values and frequency of categorical data. Differences in pre/post-PAM and PROMIS-10 scores and weight were calculated and used for analyses. Univariate and multivariate linear regression models were used to test the relationship between the sentiment of each question group (independent outcomes), while independent variables of interest were mean daily steps, mean daily time active, and change in weight (kg), PAM, and PROMIS-10. Multivariate linear models controlled for age and sex. All analyses were conducted using R version 3.6.0 (R Core Team, 2019); significance was defined as $\alpha<0.05$.

Results

Study population characteristics

The study cohort consisted of 28 individuals (82% female) with a mean age of 73±3 years; all were white. Mean compliance with wearing the Fitbit per day was 94%. At baseline, the mean BMI was 36.6±5.7 kg/m2 and weight was 97.2±17.2 kg with 15 (54%) experiencing a ≥5% reduction in weight over the course of the study period (*Table 1*). A full description of cohort demographics and change in body metrics have been previously published (Batsis et al., 2020). A total of 28 interviews were conducted with an average length of 30.5 min.

Interviews

Participants noted an average of 1,817±809 words; a mean of 79.1% represented the most commonly used words ("the", "a", or "and"), which are considered non-informative and thus removed. Within the interviews, an average of 67±20 statements were made – statements are continuous comments or opinions made without the interviewer making a statement. Per the interview, the average sentiment was more positive (32.0±11.6) than negative (22.7±12.3). This relationship was also true at the statement level with each statement being more positive (0.7±1.0) than negative (0.5±0.9) with the majority of the statement being neutral (6.3±6.3). The sentiment was variable over the course of the interview and generally became more neutral (*Figures 1&2*).

Models

In unadjusted models, non-significant associations were observed between the sentiment of feedback on technology, and the pre-post change in PROMIS physical health, average daily steps, and average daily activity. After adjusting for age and sex, statistically significant associations between the sentiment of feedback on technology and the pre-post change in PROMIS physical health (β = -0.22, 95%CI = [-0.42, -0.01], p = 0.039) were found. There was no association between sentiment and change in PAM score or average daily steps. In independent models, there was a statistically significant association between the sentiment of feedback on technology and average daily activity (β = -0.013, 95%CI = [-0.02, -0.002], p = 0.04) and the sentiment of feedback regarding positive aspects of the intervention and weight change (β = -0.31, 95%CI = [-0.60, -0.01], p = 0.04), when

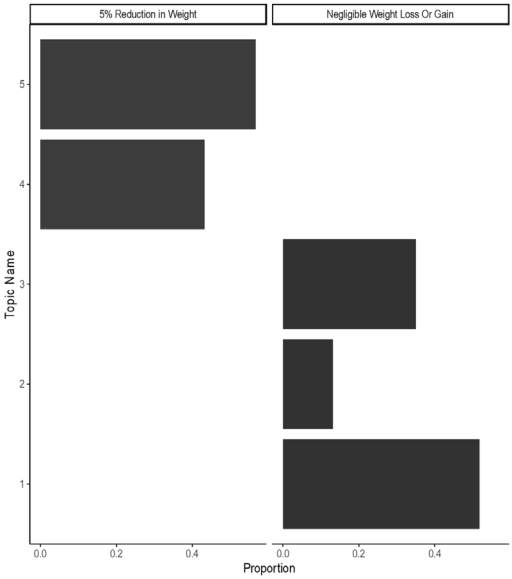

Figure 2. Proportion of each interview comprised by each topic by those achieving or not achieving at least a 5% reduction in weight

adjusting for age and sex (*Table 2*).

Discussion

We used a novel method, sentiment analysis, to evaluate participant perceptions of an older adult, technology-based weight-loss trial. In post-study interviews, we observed the sentiment of technology use increased with decreasing physical activity and decreasing change in physical PROMIS score. Further, we observed that the sentiment of the positive aspects of the intervention increased with a higher degree of weight loss.

Our findings demonstrate that NLP and sentiment analyses may have utility for evaluating mHealth interventions. Traditionally, qualitative analyses are important and critical parts of evaluating the feasibility, acceptability, and participant satisfaction of interventions (Ayala & Elder, 2011). However, there are unique challenges and limitations of performing qualitative analysis: (1) the interpretation of data could be influenced by preconceptions of the researchers (Malterud, 2001); (2) the generalizability of results to the general population may be limited (Malterud, 2001); (3) these methods cannot be deployed on a large scale without specialized training and oversight; and (4) a desire to please the investigator that may appear as an overall bias towards a positive sentiment. By using a quantitative method of sentiment analysis to evaluate questions typically answered through qualitative methods, we are able to overcome these limitations.

Though not specifically examined, the quantification of interviews may benefit from augmentation with qualitative methods. The use of quantitative NLP methods can assist in hypothesis generation or identification of perceptions that may not have been found otherwise. For instance, using traditional qualitative approaches, we would not have demonstrated an association between weight loss and increasing sentiment when asked about elements participants liked about the program. Such inferences make intuitive sense – individuals who lose more weight may more positively reflect on things they appreciated and enjoyed about a study intervention. Such quantification enables researchers to assess the magnitude and direction of the association with other behaviors or factors. Quantitative analysis allows for large sets of interviews to be examined with relative ease compared to qualitative analysis.

We decided to use an open-source library for our sentiment analysis; this increases reproducibility by allowing others to use the same methodology (Liu, 2016). This lexicon measured the sentiment of an interview and allowed us to compute summary statistics for different components of the intervention and study. The more positive sentiment toward technology was associated with both decreasing PROMIS-10 score and decreasing active time. Though the effects are small this may indicate that those that positively viewed the technology were less active and perceived their physical health as being poorer. This is separate from the association between the positive sentiment of the program more broadly and an increase in weight loss. This indicates that enthusiasm for technology may be needed for weight loss in older adults but also indicates a transfer of responsibility for the outcome away from themselves and to the technology – that is if they do not lose weight, it would be the fault of the technology.

Given the association between weight loss and positive sentiment, our findings highlight that patient perception differs with the degree of weight loss. Additional research should be conducted on using methods for the prediction of outcomes of interest. Specifically examining the relationship between quantified features of pre-intervention interviews and outcomes like physical activity or blood pressure could help predict the efficacy of the mHealth or behavior-based intervention. This would be an important and useful tool to identify those who may have the greatest impact on their health.

This study has notable limitations. First, the study cohort was relatively homogeneous in a rural part of the country and may not be representative of the population. Second, we attempted to mitigate the potential bias in self-report ques-

Table 2. Unadjusted and adjusted associations between the positive-negative sentiment of each interview component and each intervention measure

Outcome	Measure	Unadjusted		Adjusted*	
		B (95%CI)	p-value	B (95%CI)	p-value
General Feedback	Active	-0.02 (-0.05-0.02)	0.37	-0.01 (-0.05-0.03)	0.58
	Weight change	0.50 (-0.43-1.43)	0.28	0.58 (-0.37-1.53)	0.22
Dislikes of the program	Active	0.03 (0-0.05)	0.03	0.02 (0-0.05)	0.06
	Weight change	-0.14 (-0.81-0.53)	0.68	-0.09 (-0.76-0.59)	0.79
Likes of the program	Active	0.01 (-0.01-0.02)	0.42	0.01 (0-0.03)	0.16
	Weight change	-0.34 (-0.63--0.04)	0.03	-0.31 (-0.6--0.01)	**0.04**
Study Personnel	Active	-0.01 (-0.02-0.01)	0.53	0 (-0.02-0.02)	0.86
	Weight change	-0.11 (-0.57-0.35)	0.64	-0.07 (-0.53-0.38)	0.75
Technology	Active	-0.01 (-0.02-0)	0.16	-0.01 (-0.02-0)	**0.04**
	Weight change	0.01 (-0.3-0.32)	0.95	-0.01 (-0.33-0.32)	0.97
Usefulness	Active	-0.01 (-0.02-0.01)	0.40	0 (-0.02-0.01)	0.66
	Weight change	0.13 (-0.19-0.45)	0.41	0.13 (-0.21-0.46)	0.44

*Adjusted for sex and age

tionnaires by using previously validated questionnaire tools. Third, given that the Fitbit Flex 2 tracker is not a research-grade device, our physical activity measures may not be accurate and leading to increased variation within the models (Feehan et al., 2018). Fourth, interview bias was minimized using a single trained interviewer following a script. Fifth, we recognize the potential of social desirability bias yet due to the quantified range of negative and positive statements in each interview, such risk is likely minimal. Finally, given the small sample size of this pilot study, additional adequately powered studies are needed to validate that the approach that we have demonstrated is feasible. We caution our readers that while we present trends and statistical testing, we are likely underpowered, and estimates, while potentially of use to power future trials, should be interpreted with caution.

Conclusions

These results demonstrate that the use of NLP and sentiment analysis can be used in assessing the implementation of an intervention. These methods demonstrate an association between weight loss and positive sentiment of the program and those who lost a clinically significant amount of weight discussed different topics than those that did not lose weight.

Conflict of interest statement

Drs. Batsis and Petersen have equity in SynchroHealth LLC.

Abbreviations

NLP: Natural Language Processing
MET: Metabolic Equivalent of Task
PAM: Patient Activation Measure
PROMIS: Patient-Reported Outcomes Measurement Information System

Funding statement

This work was supported by the Burroughs-Wellcome Fund: Big Data in the Life Sciences at Dartmouth (Mr. Petersen); National Institutes of Health Training Geriatric Mental Health Services Researchers: T32 MH073553 (Dr. Stevens); the National Institute on Aging and the Office of Dietary Supplements of the National Institutes of Health under Award Number K23AG051681 (Dr. Batsis).

Acknowledgements

We would like to acknowledge the work of the Aging Resource Center and all patient participants.

References

Amenity Analytics (2019). Earnings Call Sentiment Analysis & Text NLP Tool | Amenity. Retrieved from https://www.amenityanalytics.com/amenity-viewer

Ayala, G. X., & Elder, J. P. (2011). Qualitative methods to ensure acceptability of behavioral and social interventions to the target population. J Public Health Dent, 71(0 1), S69-79.

Batsis, J. A., Gill, L., Masutani, R., Adachi-Mejia, A., Blunt, H., Bagley, P., . . . Bartels, S. (2017). Weight Loss Interventions in Older Adults With Obesity: A Systematic Review of Randomized Controlled Trials Since 2005. Journal of the American Geriatrics Society, 65(2). https://doi.org/10.1111/jgs.14514

Batsis, J. A., Naslund, J. A., Gill, L. E., Masutani, R. K., Agarwal, N., & Bartels, S. J. (2016). Use of a Wearable Activity Device in Rural Older Obese Adults: A Pilot Study. Gerontol Geriatr Med, 2. https://doi.org/10.1177/2333721416678076

Batsis, J. A., Naslund, J. A., Zagaria, A. B., Kotz, D., Dokko, R., Bartels, S. J., & Carpenter-Song, E. (2019). Technology for Behavioral Change in Rural Older Adults with Obesity. J Nutr Gerontol Geriatr, 38(2), 130-148. https://doi.org/10.1080/21551197.2019.1600097

Batsis, J. A., Petersen, C. L., Clark, M. M., Cook, S. B., Lopez-Jimenez, F., Jensen, M. D., . . . Bartels, S. J. (2020). A Weight-Loss Intervention Augmented by a Wearable Device in Rural Older Adults With Obesity: A Feasibility Study. The journals of gerontology. Series A, Biological sciences and medical sciences. https://doi.org/10.1093/gerona/glaa115

Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., . . . Group, o. b. o. t. P. C. (2007). The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap Cooperative Group During its First Two Years. Medical Care, 45(5), S3-S11. https://doi.org/10.1097/01.mlr.0000258615.42478.55

Changizi, M., & Kaveh, M. H. (2017). Effectiveness of the mHealth technology in improvement of healthy behaviors in an elderly population—a systematic review. Mhealth, 3. https://doi.org/10.21037/mhealth.2017.08.06

Feehan, L. M., Geldman, J., Sayre, E. C., Park, C., Ezzat, A. M., Yoo, J. Y., . . . Li, L. C. (2018). Accuracy of Fitbit Devices: Systematic Review and

Narrative Syntheses of Quantitative Data. JMIR Mhealth Uhealth 2018;6(8):e10527. https://doi.org/10.2196/10527. Retrieved from https://mhealth.jmir.org/2018/8/e10527/.

Gohil, S., Vuik, S., & Darzi, A. (2018). Sentiment Analysis of Health Care Tweets: Review of the Methods Used. JMIR Public Health Surveill, 4(2), e43. https://doi.org/10.2196/publichealth.5789

Hales, C., Carroll, M., Fryar, C., & Ogden, C. (2017). Prevalence of Obesity Among Adults and Youth: United States, 2015–2016. Retrieved from https://www.cdc.gov/nchs/products/databriefs/db288.htm

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of biomedical informatics, 42(2), 377-381.

Hibbard, J. H., Stockard, J., Mahoney, E. R., & Tusler, M. (2004). Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res, 39(4 Pt 1), 1005-1026. https://doi.org/10.1111/j.1475-6773.2004.00269.x

Hu, M., & Liu, B. (2004, 08/22/2004). Mining and summarizing customer reviews. Paper presented at the Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining.

Jensen, G. L. (2005). Obesity and Functional Decline: Epidemiology and Geriatric Consequences. Clinics in geriatric medicine, 21(4). https://doi.org/10.1016/j.cger.2005.06.007

Johnson, F., & Wardle, J. (2011). The Association Between Weight Loss and Engagement With a Web-Based Food and Exercise Diary in a Commercial Weight Loss Programme: A Retrospective Analysis. The international journal of behavioral nutrition and physical activity, 8. https://doi.org/10.1186/1479-5868-8-83

Kampmeijer, R., Pavlova, M., Tambor, M., Golinowska, S., & Groot, W. (2016). The Use of E-Health and M-Health Tools in Health Promotion and Primary Prevention Among Older Adults: A Systematic Literature Review. BMC health services research, 16 (Suppl 5). https://doi.org/10.1186/s12913-016-1522-3

Knight, E., Stuckey, M., & Petrella, R. (2014). Health Promotion Through Primary Care: Enhancing Self-Management With Activity Prescription and mHealth. The Physician and sportsmedicine, 42(3). https://doi.org/10.3810/psm.2014.09.2080

LaRose, J. G., Fava, J. L., Lanoye, A., & Caccavale, L. (2019). Early Engagement is Associated with Better Weight Loss in Emerging Adults. Am J Health Behav, 43(4), 795-801. https://doi.org/10.5993/ajhb.43.4.12

Liljas, A. E. M., Walters, K., Jovicic, A., Iliffe, S., Manthorpe, J., Goodman, C., & Kharicha, K. (2017). Strategies to improve engagement of 'hard to reach' older people in research on health promotion: a systematic review. BMC Public Health, 17. https://doi.org/10.1186/s12889-017-4241-8

Liu, B. (2016). Sentiment analysis and opinion mining: Synthesis Lectures on Human Language Technologies.[si]: Morgan & Claypool Publishers, 2012.

Malterud, K. (2001). Qualitative Research: Standards, Challenges, and Guidelines. Lancet (London, England), 358(9280). https://doi.org/10.1016/S0140-6736(01)05627-6

R Core Team (2019). R: A Language and Environment for Statistical Computing (Version 3.6.0). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/