Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240405

Using Natural Language Processing on Expert Panel Discussions to Gain Insights for Recruitment, Retention and Intervention Adherence for Online Social Support Interventions on a Stage II-III Clinical Trial Among Hispanic and African American Dementia Caregivers

Michelle ODLUM[a], Soyoung MOON[b], Peter BROADWELL[c], Niya HUANG[d], Frederick SUN[e], Dante TIPANI[f], Nicole DAVIS[g], Milea KIM[h] and Sunmoo YOON[b,1]
[a]*School of Nursing, Columbia University Medical Center, New York, NY, USA*
[b]*General Medicine, Department of Medicine, Columbia University, New York, USA*
[c]*Center for Interdisciplinary Digital Research, Stanford University, Stanford, CA, USA*
[d]*University of California in San Francisco, San Francisco, CA, USA*
[e]*Department of Rehabilitation Medicine, Columbia University, NewYork, NY, USA*
[f]*CaringKind, New York, NY, USA*
[g]*School of Nursing, Clemson University, Clemson, SC, USA*
[h]*NewYork-Presbyterian Hospital, New York, NY, USA*
ORCID iD: Michelle Oldum 0000-0001-6241-6667, Soyoung Moon 0009-0002-3294-3372, Peter Broadwell 0000-0003-4371-9472, Sunmoo Yoon 0000-0001-9413-8980

Abstract. We applied natural language processing (NLP) to a corpus extracted from 4 hours of expert panel discussion transcripts to determine the sustainability of a Stage II-III clinical trial of online social support interventions for Hispanic and African American dementia caregivers. Prominent topics included Technology/hard to reach populations, Training younger populations, Building trust, Privacy and security issues, Simplification of screening questions and recruitment procedures, Understanding participants' needs, Planning strategies and logistics, Potential recruitment places, Adjusting intervention size downwards to engage elderly participants, Targeting different generations, Internet-based interventions by age range, and Providing step-by-step instructions and an overview of the entire research process during recruitment. The application of NLP to qualitative data on a dementia caregiving clinical trial provides useful insights for recruitment, retention, and adherence to guidelines for such interventions serving Hispanic and African American dementia caregivers.

Keywords. Twitter, dementia caregiving, health disparity, natural language processing

[1] Corresponding Author: Sunmoo Yoon, PhD, MS, 630 W 168th Street, PH105, New York, NY, 10032, USA; E-mail: sy2102@cumc.columbia.edu.

1. Introduction

Hispanic and African Americans have higher prevalences of dementia in the US [1]. Further, caregiving for a person with dementia involves greater demands than other chronic conditions [1]. Because of socio-economic disadvantages, racially and ethnically underrecognized caregivers struggle from poor health status and poor quality of life [1]. Although online interventions to provide caregiving support on digital platforms became critically important during the COVID-19 pandemic, what makes such an intervention sustainable, relevant and its mechanism of efficacy remains largely unknown [2]. The expansion of social media use among underserved populations, particularly the X platform (formerly Twitter), can potentially provide social support including informational, emotional, instrumental and appraisal support for Hispanic and African American dementia caregivers [2]. However, analyzing Tweet data and network structures remains a rather novel form of study, and Twitter has rarely been used as an intervention delivery tool in health [2].

Participatory design is a user-centered approach for developing digital platform-based interventions for health consumers [3]. Needs assessment using qualitative methods (e.g., expert panels) has been commonly applied during the participatory design process [3]. Little is known about how best to involve older ethnic and racial minorities in the development design process for digital platform-based interventions, to support continually engagement [3]. Therefore, the purpose of this study was to apply a participatory qualitative methods approach to gain insights to conduct Stage II-III clinical trials of social media-based interventions to enhance the social support (based upon 4 constructs: informational, instrumental, emotional, appraisal) among Hispanic and African American dementia caregivers [4].

2. Methods

The study was approved by the Institutional Review Board (IRB). We applied natural language processing (NLP) and word clustering algorithms to qualitative data collected from seven expert panel members regarding a Stage II-III clinical trial to enhance social support for Hispanics and African American dementia caregivers using Twitter (clinical trial NCT03865498). A total of seven experts—including four dementia caregivers (two African Americans, one of whom was male and the other female, and two Hispanics, both female) and three health professionals participated together in the expert panel discussions. The expert panel was conducted in June 2023 for approximately 4.0 hours of synchronous discussions via Zoom after enrolling 966 Hispanic and African American family members of a person with memory issues and dementia in the trial [5].

First, we extracted a text corpus from audio recorded during the expert panel meeting, via the auto-transcription function in Microsoft Word. Next, we applied natural language processing (NLP) to this text corpus using the AutoMap text mining tool and Perl regular expressions via the following steps: 1) cleaning (e.g., fixing common typos, expanding common contractions and abbreviations, removing extra spaces, converting British to American spelling); 2) preprocessing (e.g., removing single letters, n-gram conversion, removing pronouns and prepositions, dates and possessive forms, converting hyphenated words to n-grams, reconciling full names); and 3) refining (removing punctuation, lower case conversion, concept tuning and merging by a domain expert) [5].We next applied the Newman clustering algorithm to the corpus, and visualized the detected term clusters

via a network diagram (Newman modularity: 0.876) [6]. Last, domain experts in dementia caregiving contributed interpretations of the detected clusters as semantic topics. Resources and analytic code are available on GitHub and OSF.io (https://osf.io/qruf3).

3. Results

Twelve salient semantic topics were extracted from the qualitative text data from the expert panel corpus via the Newman clustering algorithm (Figure 1; Table 1): 1) Technology and hard to reach populations, e.g., "Technology can save us."; 2) Training younger populations: "Now, bringing all of this to the community, teaching somebody that you trust in the community, some like younger people and then they kind of talk to their friends of families."; 3) Building trust: "I understand social media is a big thing, but I am more of a people person, so I want to get out in the streets, and be face to face with people because it is one thing to go online and do something."; 4) Privacy/security issues: "People ashamed about dementia and they whisper it to you, … they realize that it is not just happening to them. I am saying sometimes you see the relief in their eyes too.";

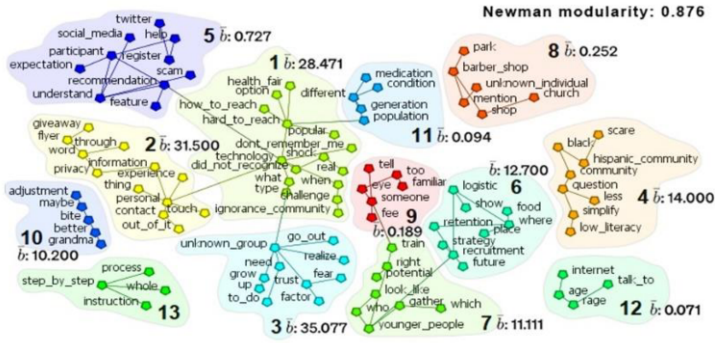

Figure 1. Term-cluster topics extracted from text corpora of expert panel discussion for recruitment, retention, and intervention adherence for online social support intervention for a Stage II-III clinical trial for Hispanic and African American Dementia Caregivers (Newman $\bar{b}$ unit: 10^{-3})

5) Simplify screening questions and recruitment procedures to engage the low-literacy population among Hispanic and African American communities; 6) Understanding the participants' needs and providing features in a social media-based clinical trial is important for intervention adherence: "It was kind of daunting for them. Even if you are computer literate, that is a lot of questions to answer."; 7) Planning strategies and logistics for recruitment and retention (e.g., food, locations); 8) Potential recruitment places including church, barber shop, park, or shops to engage new candidates: "If you talk to people, they would be willing to learn."; 9) Adjusting intervention to a smaller size to engage elderly participants; 10) Targeting different generations: "That people may not listen to me because they may not know me, or I might be from a different generation.";11) Internet-based interventions by age ranges: "We could do it like an outreach program with the schools."; 12) Providing step-by-step instructions and an overview of the whole research process during recruitment: "Register help them with the whole step by step process."

Table 1. n-gram centrality rankings from expert panel discussions for a dementia caregiving clinical trial

n-grams (topic)	katz	PageRank	n-grams (topic)	katz	PageRank
Unknown_group (3)	1	0.031	Contact (4)	0.381	0.01

Technology (1)	0.982	0.03	Logistic (7)	0.376	0.011
Personal (4)	0.893	0.03	Participant (6)	0.364	0.014
Hard_to_reach (1)	0.723	0.023	Understand (6)	0.357	0.014
Younger_people (2)	0.503	0.014	Word (4)	0.344	0.013
Grow (3)	0.502	0.012	Where (7)	0.335	0.011
Right (2)	0.5	0.014	Train (2)	0.335	0.009
Recruitment (7)	0.451	0.014	Privacy (4)	0.301	0.01
Trust (3)	0.409	0.012	Age (11)	0.272	0.007
Information (4)	0.401	0.01	Community (5)	0.25	0.011

4. Discussion and Conclusions

We applied NLP to qualitative research transcripts for a Stage II-III clinical online social support intervention trial for Hispanic and African American dementia caregivers where we prioritized community engagement, cultural competence, and user-centered design. Our detected topics from NLP thoroughly cover all "5W" essential points of clinical trial recruitment and retention strategies, including the who, what, when, where, why and how of underrepresentation in clinical trial participation. Besides the benefits of objectively detecting topics and avoiding human bias via NLP, the credibility, confirmability, meaningfulness, and direct transferability to practice of our findings may be explained by the time of the data collection: the data was collected soon after having the potential participants form long lines at recruitment events, eager to learn how to use Twitter/X and to participate in the intervention to support their caregiving needs [5].

The topics identified via network clustering analysis of term co-occurrences exhibited a notable pattern, with terms regarding the target population (who: hard to reach, younger, older, different generations) as the core component, surrounded by subsidiary information such as the motivation of the selected population (why: technology, literacy), instrumental intervention vehicles (what: internet, technology), methods of connection (how: peers, community outreach), time of recruitment (when: after establishing trust), and place of recruitment (where: schools, barber shops).

Five of the topics involve defining the target population by age, covering young, old or all generations. Although it may be intuitive to assume that caregiving support through technology platforms such as Twitter/X requires participation from the younger population, as noted in topic 1, three topics include terms referencing the older population (e.g., grandma) or various age ranges as a target of such clinical trials. Approximately 80% of dementia care is provided by family members, and a relatively older population participates in the caregiver community, while the younger population participates in the professional workforce. Conversely, two of the topics concern training the younger population to become involved in care as a family member in the community, and call for the older population to take charge in learning technology and social media.

The encouragement of "grandma" and "grandpa" to use technology and social media to utilize caregiving resources in topic 9 is aligned with the effort to fight against invisible ageism. Although the elderly are often viewed as having difficulties learning and adopting new technologies, thus missing out on health education using new digital media, our social media-based clinical trial for dementia caregiving found that this is not the case. As topics 5, 9, 11 and 12 indicate, step-by-step instruction to ease reluctance and

fear among the elderly, school-based training of younger generations to explain technologies to their family members, and using simplified language while screening to enhance understanding of the study, may contribute to the fight against invisible ageism in social media and technology-based clinical trials in general [5]. Digital marginalization disproportionately affects older individuals, those with less education, lower income, ethnic minorities, and residents of remote areas. Adjusting the intervention to a "bite-size" format also aids in making content easily digestible for older adults. This could involve breaking down information into smaller, more manageable units and using simple, clear language. Intuitive and easy-to-navigate systems provide support and guidance for users new to digital health technologies and ensure that the content is relevant and culturally sensitive. Moreover, incorporating interactive elements and visual aids can further elevate understanding and retention [7]. Engagement strategies that resonate with older and younger generations should be implemented, such as storytelling, group activities, or intergenerational programming.

Optimizing the logistics of culturally welcomed and appropriate foods and venues also can significantly impact participation in a clinical trial. Food can be a compelling incentive for participation throughout the trial, from recruitment to follow-up. Similarly, selecting accessible and comfortable venues can encourage attendance and adherence. These strategies underscore the need for meticulous planning in the initial stages of intervention design to ensure inclusivity and high participation rates [7].

In conclusion, the application of NLP techniques to qualitative research data provides credible, meaningful, and transferable insights for recruitment, retention, and adherence to guidelines for technology-based interventions serving Hispanic and African American dementia caregivers. Our findings emphasize a holistic approach to designing and implementing social media-based social support interventions that tailor to the needs of all age groups including elderly and socially disadvantaged groups in dementia caregiving. By focusing on logistical planning, making design adjustments for accessibility, and addressing the broader issue of digital marginalization, it is possible to create more inclusive and effective technological interventions. This amplifies the user experience for these groups to overcome the digital divide and reduce health disparities.

Acknowledgements: This research was supported by US federal grant TweetS2 R01AG060929 (PI: Yoon).

References

[1] Alzheimer's Association. Race, ethnicity and Alzheimer's in America, 2021.
[2] Yoon S, et al. Analyzing Topics and Sentiments from Twitter to Gain Insights to Refine Interventions for Family Caregivers of Persons with Alzheimer's Disease and Related Dementias During COVID-19 Pandemic. Informatics and Technology in Clinical Care and Public Health; 2022. Jan;IOS Press; p. 170-3
[3] Vandekerckhove P, et al. Generative participatory design methodology to develop electronic health interventions: a systematic literature review. J of med Inter research. 2020 Apr 27;22(4):e13780.
[4] Langford CP, Bowsher J, Maloney JP, Lillis PP. Social support: a conceptual analysis. Journal of advanced nursing. 1997 Jan;25(1):95–100.
[5] Yoon S, et al. Comparing Emotional Valence Scores of Twitter Messages from Human Coding and Machine Learning Algorithms among Hispanic and African American Family Caregivers of Persons with Dementia. Studies in Health Technology and Informatics. 2023;305:440. doi:10.3233/SHTI210104.
[6] Newman ME, Girvan M. Finding and evaluating community structure in networks. Physical review E. 2004 Feb 26;69(2):026113.
[7] Cheng C, et al. Applying the Electronic Health Literacy Lens: Systematic Review of Electronic Health Interventions Targeted at Socially Disadvantaged Groups. J Med Internet Res. 2020;22(8):e18476.