9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Crowdsourcing the potential negative effects of the most popular health apps: Evidence from Twitter

Keywords: e-health, evaluation, topic modelling, crowdsourced data

Extended Abstract

The last 10 years have seen a dramatic growth in the number and popularity of mobile applications (henceforth "apps") created to address health issues. However, most of the publicly available apps have not undergone scientific evaluation for their safety or effectiveness, warranting worrying implications for the health of users (Byambasuren, Glasziou, Beller, & Sanders, 2018). Of the health apps found to be effective in randomised controlled trials, less than 25% were functioning and publicly available (Rogers, Lemmen, Kramer, Mann, & Chopra, 2017). A systematic review by McKay and colleagues (2018) highlighted a lack of best practices for evaluating mobile health apps, in a landscape of general lack of regulation in the field of health promotion. The risks involved are not trivial – many of the apps currently in the market could have significant adverse effects on certain user groups. For example, a fitness app designed to monitor calorie intake and aid in weight loss could be effective in encouraging healthy habits for the general population, but the same app could pose serious risks to users with eating disorders (Neal et al., 2022). Feedback on user experience is central to the successful development of apps and a core component of the evaluation of app-delivered health interventions. Negative user experiences lead to decreased uptake, engagement, retention, and compliance with guidance (Szinay, Jones, Chadborn, Brown, & Naughton, 2020).

In absence of empirical studies evaluating health app, social media might offer a valuable source of insight from the experience of thousands providing feedback on health apps, with the potential to uncover important usability, safety and efficacy concerns that developers might not have taken into account. Lagan and colleagues (2020) highlighted the need for a public, interactive and accessible approach to data collection when it comes to app evaluation, for which social media present a promising avenue. In this study, we explore the application of a computational social science approach to the evaluation of health apps. We investigate whether crowdsourcing feedback from Twitter and using a semi-automated method of qualitative analysis could lead to novel insights into the user experience, especially as related to negative effects of health apps. The final aim of this study was to generate actionable and low-cost recommendations for the improvement of health apps. We focus on evaluating top-ranked apps in three core health domains: fitness, or mental wellbeing, and fertility and menstruation tracking apps.

A sample of the 5 top-grossing apps from each of the three pre-defined domains was obtained through Statista on February 7, 2023. Both free and paid apps were included in the study, as recommended by research indicating an association between price and involvement of behaviour change techniques in health apps. A single exclusion criterion was specified; apps were excluded if they did not have stand-alone functionality (e.g. required the purchase of additional equipment). The final sample of apps included in each domain is detailed in Table 1. Tweets were extracted using Twitter's official API with an academic license. The tweets were identified using search queries including different variations of the app name and the words "app" or "application". We cleaned the dataset and filtered the tweets through a sentiment lexicon so that the final dataset included only tweets labelled as having negative sentiment.

To analyse the content of the tweets we applied machine-assisted topic analysis (MATA; Towler et al., 2022), a method integrating machine learning and human qualitative thematic analysis. MATA is based on structural topic modelling and used to assist qualitative researchers to analyse large amounts of textual data. Our analysis was conducted in 2 stages: (1) a separate model and analysis for each of the three domains, including all apps and (2) a more granular analysis consisting of a separate model for each app. The rational for a two-stage analysis is as follows; the first stage provides a broad overview of the user experience of apps in each health domain, and the second a more granular analysis of app-specific features and issues that has the potential to lead to more concrete and actionable recommendations. The analysis procedures starts with determining the optimal number of topics for each corpus and fitting a structural topic model using the stm package in R (Roberts et al., 2019). The output of that model is then analysed by human coders. The outputs examined consist of two main elements; the 10 most representative quotes for each topic and two lists of weighted words that constitute the topic. In order to analyse the model's output systematically, two coders interpret the output and agree upon narrative labels for the topics. Finally, the coders analyse the topics generated by the text analysis and created broader themes. Preliminary results reveal 10 topics for each domain. In the fitness domain, the main themes revolved around negative consequences of calorie tracking, technical issues, pay-walled features, and experiences of eating disorder onset triggered by the app. In the fertility and menstruation domains, themes included reported paranoia related to fear of pregnancy, reported paranoia related to menstruation symptoms, empowerment, and UI issues.

References

Byambasuren, O., Glasziou, P., Beller, E., & Sanders, S. (2018). Prescribable mhealth apps identified from an overview of systematic reviews. *npj Digital Medicine*, *1*.

Lagan, S., Aquino, P., Emerson, M., Fortuna, K., Walker, R., & Torous, J. (2020). Actionable health app evaluation: translating expert frameworks into objective metrics. *npj Digital Medicine*, *3*.

McKay, F. H., Cheng, C., Wright, A., Shill, J., Stephens, H., & Uccellini, M. (2018). Evaluating mobile phone applications for health behaviour change: A systematic review. *Journal of Telemedicine and Telecare*, *24*, 22 - 30.

Neal, D., Engelsma, T., Tan, J., Craven, M., Marcilly, R., Peute, L., … Dröes, R.-M. (2022, 01). Limitations of the new iso standard for health and wellness apps. *The Lancet*, *4*, E80-82.

Rogers, M. A., Lemmen, K., Kramer, R., Mann, J., & Chopra, V. (2017). Internet-delivered health interventions that work: Systematic review of meta-analyses and evaluation of website availability. *J Med Internet Res*, *19*(3), e90.

Szinay, D., Jones, A., Chadborn, T., Brown, J., & Naughton, F. (2020). Influences on the uptake of and engagement with health and well-being smartphone apps: Systematic review. *J Med Internet Res*, *22*(5), e17572.

Towler, L., Bondaronek, P., Papakonstantinou, T., Amlôt, R., Chadborn, T., Ainsworth, B., & Yardley, L. (2022). Applying machine-learning to rapidly analyse large qualitative text datasets to inform the covid-19 pandemic response: Comparing human and machine-assisted topic analysis techniques. *medRxiv*. doi: 10.1101/2022.05.12.22274993

Table 1: Apps included in the analysis and number of tweets per app after data cleaning.

Corpus	**App**	**N**
Fitness	MyFitnessPal: Calorie Counter	8464
Fitness	Strava: Run, Ride, Swim	2264
Fitness	WW (formerly Weight Watchers)	2902
Fitness	Workouts by Muscle Booster	11
Fitness	Fitness Coach & Diet: FitCoach	158
Mental well-being	Calm.com	574
Mental well-being	Headspace	9984
Mental well-being	Sleep Cycle	5527
Mental well-being	Ipnos Software	912
Mental well-being	Waking Up Course	2606
Fertility and menstruation	Flo	2266
Fertility and menstruation	LunaLuna App	545
Fertility and menstruation	Period Calendar	525
Fertility and menstruation	Glow	1270
Fertility and menstruation	Clue Period Tracker & Ovulation app	484