evolving interests, then the changes should be similar in places with similar culture, geography, and language. For example, Google Trends shows dieting spikes at the beginning of each year[16].

Hence, we compare the changes in interest from one sample to another between United States and United Kingdom (two Western countries sharing a language), and between US and India and Brazil. To do this, we calculated the correlation between the audience variation (the deltas) of one country to another country for each interest. For US vs. UK comparison, out of 29 interests, we found 17 to be significantly similar (having Pearson $r > 0$ at $p < 0.05$) and only one interest to be significantly negatively related between the two countries, implying a similarity in the *direction of change* over time. For US vs. India, however, only 5 interests shared a direction (1 negatively), and US vs. Brazil 9 (3 negatively). This suggests that comparisons between more similar countries may be more stable in the same time span.

Thus, upon a closer inspection we find more stability in the data across demographics *without including interests*. This may be due to the definition of such features to be stable in time, and possibly to Facebook's inference algorithms for age and gender being relying on more explicit user-specified information. However once we include interest information, the stability of the audience estimates is much worse, with only 45% of the interest-specific age groups having the same ordering (Spearman's $\rho > 0.7$).

7 CONCLUDING DISCUSSION

This paper explores the use of Facebook ad audience estimates for global lifestyle health concern surveillance. Particularly, we assess the quality of Facebook data by introducing placebo baselines, normalization alternatives, and performing temporal analysis. Among our findings, we show that within-country statistics are more statistically separable than statistics across countries, which is an important observation as previous efforts have showed very promising results, but they focused only on a single country analysis [4, 7]. More important, the high volatility observed between two data snapshots warrants extra caution in the use of Facebook Marketing API as a source of social interest.

We can only speculate about the causes of such instability. It is possible that Facebook is constantly updating and changing its NLP pipeline, unevenly improving its performance over different topics. Further, if the time span implied in the interest estimate is short-term, audience numbers may swing wildly as Facebook usage changes, say, during holiday seasons. Another source of variability could come from the redefinition of what content is included in an interest. For instance, if new brands are included in plus-size clothing interest, its volume will increase. Finally, the advertising market demand may drive Facebook to focus on developing one interest over another.

Still, it may be too early to give up on this potentially rich data source. Our effort is restricted to a specific kind of disease surveillance. Thus, we acknowledge that it is impossible to generalize the limitations we have found without future studies. For example, much as the search volume revealed by Google Trends [3], the sways in Facebook's interests may be seasonal or local. Unfortunately currently neither Facebook's Advertising interface or Marketing API provide historical data, so more work needs to be done to monitor the data – i.e. for over a year if one is to capture seasonal variation. Furthermore, to assess the accuracy of interest assignment, one could use the advertising platform to run a survey assessing the interests of the reached audience directly.

Thus, we hope that this work will encourage future efforts to use our methodology to gather user interest from the Facebook Ads for other applications and scenarios. We also hope that results presented here will encourage future researchers to test the reliability of this potentially rich data source, and inspire more rigor in the future data-driven studies that aim at correlating social media data with offline world data.

[16] https://trends.google.com/trends/explore?q=diet

REFERENCES

[1] Jisun An and Ingmar Weber. 2015. Whom should we sense in "social sensing"-analyzing which users work best for social media now-casting. *EPJ Data Science* 4, 1 (2015), 1.

[2] Matheus Araujo, Yelena Mejova, Michael Aupetit, and Ingmar Weber. 2017. Visualizing Health Awareness in the Middle East. In *ICWSM*. to appear.

[3] Yan Carrière-Swallow and Felipe Labbé. 2013. Nowcasting with Google Trends in an emerging market. *Journal of Forecasting* 32, 4 (2013), 289–298.

[4] Rumi Chunara, Lindsay Bouton, John W. Ayers, and John S. Brownstein. 2013. Assessing the Online Social Environment for Surveillance of Obesity Prevalence. *PLoS ONE* 8, 4 (04 2013), 1–8.

[5] Elia Gabarron, Luis Fernandez-Luque, and others. 2017. Impact of Facebook Ads for Sexual Health Promotion Via an Educational Web App: A Case Study. *IJEHMC* 8, 2 (2017), 18–32.

[6] Jeremy Ginsberg and others. 2009. Detecting influenza epidemics using search engine query data. *Nature* 457, 7232 (2 2009), 1012–4.

[7] Steven Gittelman, Victor Lange, A. Carol Gotway Crawford, A. Catherine Okoro, Eugene Lieb, S. Satvinder Dhingra, and Elaine Trimarchi. 2015. A New Source of Data for Public Health Surveillance: Facebook Likes. *JMIR* 17, 4 (2015), e98.

[8] J. Gomide, A. Veloso, W. Meira Jr., V. Almeida, F. Benevenuto, F. Ferraz, and M. Teixeira. 2011. Dengue surveillance based on a computational model of spatio-temporal locality of Twitter. In *WebSci*.

[9] L. Brianna Haskins and others. 2017. Health Evaluation and Referral Assistant: A Randomized Controlled Trial of a Web-Based Screening, Brief Intervention, and Referral to Treatment System to Reduce Risky Alcohol Use Among Emergency Department Patients. *JMIR* 19, 5 (May 2017), e119.

[10] Taha A. Kass-Hout and Hend Alhinnawi. 2013. Social media in public health. *British Medical Bulletin* 108, 1 (2013), 5–24.

[11] Yoonsang Kim, Jidong Huang, and Sherry Emery. 2016. Garbage in, Garbage Out: Data Collection, Quality Assessment and Reporting Standards for Social Media Data Use in Health Research, Infodemiology and Digital Disease Detection. *JMIR* 18, 2 (26 Feb 2016), e41.

[12] David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. The parable of Google flu: traps in big data analysis. *Science* 343, 6176 (2014).

[13] Yabing Liu, Chloe Kliman-Silver, Robert Bell, Balachander Krishnamurthy, and Alan Mislove. 2014. Measurement and analysis of osn ad auctions. In *COSN*. 139–150.

[14] Christopher J.L. Murray and Alan D. Lopez. 1997. Alternative projections of mortality and disability by cause 1990–2020: Global Burden of Disease Study. *Lancet* 349 (1997), 1498–1504. Issue 9064.

[15] Tobias Preis and Helen Susannah Moat. 2014. Adaptive nowcasting of influenza outbreaks using Google searches. *Royal Society Open Science* 1, 2 (2014).

[16] Ben Y Reis and John S Brownstein. 2010. Measuring the impact of health policies using Internet search patterns: the case of abortion. *BMC public health* 10 (2010), 514.

[17] Diego Saez-Trumper, Yabing Liu, Ricardo Baeza-Yates, Balachander Krishnamurthy, and Alan Mislove. 2014. Beyond cpm and cpc: Determining the value of users on osns. In *COSN*. 161–168.

[18] Koustuv Saha, Ingmar Weber, Michael L. Birnbaum, and Munmun De Choudhury. 2017. Characterizing Awareness of Schizophrenia Among Facebook Users by Leveraging Facebook Advertisement Estimates. *JMIR* 19, 5 (2017), e156.

[19] Elad Yom-Tov and Luis Fernandez-Luque. 2014. Information is in the eye of the beholder: Seeking information on the MMR vaccine through an Internet search engine. *AMIA Symposium* 2014 (2014), 1238–47.