Instagram use and mortality risk

Siyi Fan

University of Chicago

**Introduction**

The Instagram was launched in October 2010, characterized by posting pictures and videos by using different manipulation tools (e.g., filters), which has attracted 500 million daily active users (www.cnbc.com/2017/09/25/how-many-users-does-instagram-have-now-800-million).

Instagram as an image-based social media allows users to transform the appearance of the pictures and share instantly on multiple media channels (e.g., Facebook, Twitter, and Flickr) in addition to Instagram page. Besides, users could use the @symbol to mention other users in the picture before posting them. In this way, Instagram not only enables users to share content with networks they construct online, but also reflects how people interact with each other offline.

Previous research shows that online friendship could promote offline social activities (Rainie & Wellman, 2012), which suggests that online social media could be used as a way to help users engage in face-to-face social behaviors. Therefore, I expect there is a positive relationship between online social platform use and physically health.

The reason I choose Instagram instead of Facebook or Twitter is that people tend to trust in pictures over textual descriptions according to MAIN model, which suggests that our brains implicitly believe in visual modalities such as pictures and videos more than text, because we as social animals get used to communicating with an actual person rather than objects or words. Just image a photo of your friend making a silly face lying on the beach, particularly with audible voice. Then, you are more likely to feel your friend is really there. Therefore, this perceived authenticity may play an important role in providing social support and encouraging socially-motivated behaviors that may prevent illness.

**Purpose of the Present Study**

The purpose of this study is to determine the relationship between use of image-based platform (Instagram) and health problems, specifically for morality risk. I will measure six-months’ Instagram activity of 12 million people and evaluate whether those activities were associated with decreased mortality risk in the following two years. To be clear, there is a 6-month gap between these two measurement periods.

I will examine the following hypothesis and expect to see that people who receive more tags experience lower mortality rates than those who receive less.

* Research Question 1: Is Instagram use associated with health problems, specifically with morality risk?
* Hypothesis 1: Instagram use is associated with lower risk of morality.

**Method**

This study is a natural experiment, start with the always-on data source and looking for random events on Instagram. I don’t have to set a randomized controlled experiment where users are randomly assigned to different levels and then resulting morality risk is measured for everyone. Instead, big data sources can provide me the outcome data to compare the results for users with different tags received. Besides, we can possess more outcome data from automatically collected data sources, which enables us to measure the effects of changes created by exogenous variation.

The strategy with observational data in this study is forecasting - predicting morality risk from current Instagram use. I will match Instagram users with their health records from Department of Public Health, analyzing online activity, specifically for tags received over a six-month period and comparing that of those who are still living to those who have died after 2.5 years.

**Participants**

In order to obtain a random sample of Instagram users and their activities, I plan to analyze social network by using Instagram API. Users should have at least 90 days of experience with the Instagram before I start to measure Instagram use. Also, every user is required to provide a first name, nickname, or date of birth not shared by others, so that those individuals could be linked to their health records uniquely.

**Measures**

**Activity analysis.** I will mainly focus on how many tags received rather than sent tags for every Instagram by using “counting” strategy. First of all, “tag” is a tool, which allows users to label friends in the photos. In other words, they must have already met with each other in real life, otherwise they can’t take a picture together. Thus, tags in photos could be a predictor that two people have a face-to-face relationship and this social interaction maintains once having met by posting a picture online. Secondly, received tags instead of sending tags indicate that the recipient is in the photo and therefore has engaged in a real world social interactions. Therefore, I need to collect observational data through Instagram API and big data enables me to count how many tags received directly.

**Assessment of morality risk.** I will turn to Department of Public Health for records of users’ mortality status and causes of mortality. In this study, causes of death include: infections, sexually transmitted diseases, colorectal cancer, pancreatic cancer, lung cancer, blood lymphatic cancers, all other cancers, diabetes, mental illness dementia, ischemic heart disease, stroke, all other cardiovascular diseases, liver disease, unintentional injury, drug overdose, suicide, and homicide (Hobbs et al., 2016). It is different from standard morality categorizations, because it involves more young age death causes and fewer old age morality causes.

**Study Considerations**

**Strengths**

**Big.** A major advantage of using online big data is that we can build a large sample and access vast information to support the results. **Always-on.** As Salganik (2017) mentioned in *Bit By Bit*, online big data could also enable the study of real-time measure and get data be collected every second. This is a longitudinal observational study and users usually update their pages very often, so “big data” could help to record those changes and keep up with current situations. Also, always-on big data enables the study of unexpected events. Observational study by using big data could derive some innovative solutions or help to better understanding new phenomena. For this research, I hope it could shed lights on the online social interaction effects on physical health. **Non-reactive.** Users won’t react to our measurement, because we will not inform them that their Instagram photos and health situations would be collected for the purpose of this study.

**Limitations.**

**Incomplete.** For this study, online big data tends to be missing users’ demographics and behaviors on other platforms. I have limited accesses to users’ complete backgrounds, so that we are not able to control their socioeconomic status. Also, I can’t rule out the possibility that people received less tags or use Instagram less often may spend more time on Snapchat, which is a new visual-basic social media. **Sensitivity.** Department of Public Health have detailed information about individual health situation. If this information becomes public, then it may potentially lead to personal loss (e.g., embarrassment) and business loss (e.g., loss of job). Sensitivity is part of reason that big data sources are inaccessible. **Inaccessible.** I need Instagram to grant my access to users’ data, including demographic information, photos they post, and tags received. Those are personal data trails which may be restricted by companies and health information are controlled by government. Therefore, these data will not be immediately available to researchers at universities. However, the aim of this study is to find a new way related to individual physical health, so I will show this importance to Instagram company and Department of Public Health and promise not to share my data with other researchers. **Dirty.** During data collection, I may include data that don’t reflect real actions of interest to me. For instance, with more advertisement showed up on Instagram, tags may not only be restricted for a relationship, which means that it could become a commercial way to “remind” you of new products. Thus, I plan to use advanced crawling algorithms to collect clean data and perform simple exploratory analysis after that. **Privacy for users.** The privacy of the users on social media is an important issue in the era of online big data. Instagram is a visual data source, where we could get users’ location category preferences, users’ activities for different social sources, and their demographic profiles. A long trail of digital footprints can be detected by social media companies and be purchased by advertisers, so that they could predict your next actions and promote products to you. Finally, this observational study may only indicate an association between Instagram use and morality risk. Although we may get many relationships between these two factors, we still cannot say that spurring users to receive tags on Instagram will increase their longevity, because we have not provided an experiment study to prove a causal relationship here. Third, the privacy of the users on social media is an important issue in the era of online big data. Instagram is a visual data source, where we could get users’ location category preferences, users’ activities for different social sources, and their demographic profiles. A long trail of digital footprints can be detected by social media companies and be purchased by advertisers, so that they could predict your next actions and promote products to you.

**Reference**

Rainie H, Wellman B, (2012) Networked: The New Social Operating System (MIT Press,

Cambridge, MA.

Salganik, Matthew J. 2017. Bit by Bit: Social Research in the Digital Age. Princeton, NJ: Princeton University Press. Open review edition.

William R. Hobbs, Moira Burked, Nicholas A. Christakis, and James H. Fowler, (2016). Online

social integration is associated with reduced mortality risk.