Making Connections

Advancing Healthcare Research Via Consumer Mobile Devices

Eric Horvitz, MD, PhD; Ryen W. White, PhD

By numerous measures, the smartphones in our pockets and pocketbooks have more computing power than the supercomputers of the 1990s. Smartphones provide a general platform for software applications and host a suite of sensors, including location and activity sensing. Smartphones have become ubiquitous in the developed world. They are nearly always on and always nearby. As such, smartphones provide rich opportunities for healthcare research—and for promising directions in health and wellness more generally.

See Article by Nguyen et al

The article, by Nguyen et al,[1] published in this issue of *Circulation: Cardiovascular Quality and Outcomes*, highlights the untapped potential of bringing together smartphones, online survey tools, and electronic health records for healthcare research and delivery. The authors present research on tracking hospitalizations via the use of smartphones and online surveys. Identifying hospitalizations in an accurate and efficient manner could help with tracking and understanding the incidence of disease, adverse outcomes of medical interventions, and recurrence of illness. Developing a low-cost and accurate means of tracking significant hospitalizations, such as admissions for acute cardiovascular events, would enable new kinds of studies. Although visits to hospitals are tracked in electronic medical records, the approach presented by the authors offers a holistic view on hospitalizations that span multiple medical facilities and healthcare networks.

A key contribution of the work by Nguyen et al[1] is a characterization of the effectiveness of a version of geofencing—the use of location services to identify when people enter an area of interest. In a multiyear study, volunteer participants aged ≥18 years installed a special application on their phones. The application uses the global positioning system and cell tower triangulation to identify regions near hospitals, per a predetermined set of stored locations. When a participant using the application is detected within a region marked as a hospital for 4 hours or more, a questionnaire appears seeking confirmation of the candidate hospitalization. Patient records and survey data are later linked to the detected event, providing validation of a hospital visit. The authors provide measures of the specificity and sensitivity of the detections of hospitalization using this methodology.

The study used in-person and remote arms of participants who downloaded the smartphone application. The smaller in-person study included consenting patients who had been scheduled for electrophysiology and cardiac catheterization procedures. Participants in the remote arm were drawn from the Health eHeart program.[2] The accuracy of detected hospitalizations and of lengths of stay in hospitals was confirmed by survey responses, electronic mail, and a review of medical records. The authors report that 17 of 22 confirmed hospitalizations were detected for the in-person arm sensitivity of 77%. In the larger remote arm, 102 of 157 medical visits were correctly identified, demonstrating a positive predictive value of 70%. Challenges with accuracy include errors in location sensing, potential gaps in reporting, and false positives that come when people visit hospitals for their work or other reasons, such as visiting a sick friend or relative.

The newly published work comes in the context of the growth of interest and results in mobile health or mHealth,[3] and related areas of research, referred to with the terms digital disease detection[4] and infodemiology.[5] Work in these realms extends healthcare research by tapping into new streams of data available via consumer computing devices and such web services as search logs and social media. Efforts include analyses of large-scale anonymized logs of queries input to web search engines and words appearing in posts on Twitter. Studies span such topics as tracking the incidence and spread of infectious diseases,[6,7] stratification of people at risk for illness,[8,9] and identifying adverse effects of medications.[10] Logs of user activity that are collected on devices and servers as part of the normal flow of ongoing activities, and in accordance with usage agreements, may include information about the locations of users.

The work by Nguyen et al is related to previous studies of potential hospital utilizations with geofencing using logs of tens of thousands of users of mobile devices.[11–13] In previous work by our group, we used geofencing to identify when people submitting queries to search services on smartphones approach one of ≈35 000 preidentified hospital sites throughout the United States. In the work, recurrent patterns of visitation are removed as likely employees. For each candidate hospitalization, the search query log is rolled back in time to consider search terms entered over the days and hours leading up to the hypothesized hospital visit. For privacy reasons, the detailed location information is removed from the data before analysis; only the absolute distances to the identified target hospitals are considered. Statistical models built from

The opinions expressed in this article are not necessarily those of the editors or of the American Heart Association.

From Microsoft Research, Redmond, WA.

Correspondence to Eric Horvitz, MD, PhD, Microsoft Corporation, One Microsoft Way, Redmond, WA 98052. E-mail horvitz@microsoft.com; and Ryen W. White, PhD, Microsoft Corporation, One Microsoft Way, Redmond, WA 98052. Email: ryenw@microsoft.com

(*Circ Cardiovasc Qual Outcomes.* 2017;10:e003573. DOI: 10.1161/CIRCOUTCOMES.117.003573.)

***Circ Cardiovasc Qual Outcomes* is available at http://circoutcomes.ahajournals.org**

DOI: 10.1161/CIRCOUTCOMES.117.003573

the resulting aggregate data sets are then used to predict if and when future healthcare utilizations will occur in advance of detected visits, given the terms input to search engines (eg, chest pain) over time. Follow-on work[12] pursues additional insights about the nature of a hospitalization and a patient's changing information needs as expressed in the terms input to search engines before, during, and after the candidate visits. Findings include a rise in use of technical medical terms during and following a visit.

In the absence of a link between location data collected by smartphones and confirmatory evidence provided by medical records, we cannot be certain that patients visited a hospital facility. Reasons for this may include inaccurate geofencing, inaccurate global positioning system tracking, or trouble disambiguating between hospitals and proximal facilities (eg, those immediately next door or located above or below). Researchers in computer science have explored opportunities to refine location accuracy, including methods for enhanced place labeling that uses time-stamped location measurements, combined with signals from other sources (demographics, timing, visits from others, and visit sequences) to assign meaningful labels to locations in an automated manner.[14,15] However, location signals do not yet come with the requisite precision to identify hospital visits with certainty.

The studies of Nguyen et al involve a relatively small cohort compared with the related work on utilization using large-scale mobile logs. However, the study brings to the fore a critically important direction required for the maturation of mHealth: making connections between evidence and inferences drawn from nontraditional digital streams and clinical data and reporting about ground truth. Efforts like the study by Nguyen et al that align traditional electronic health records data and survey tools with noisy, nonvalidated digital streams are important for the maturation of mHealth. Although such studies may be more costly and cumbersome (and thus smaller in scale), they are important for validating and integrating the nontraditional digital streams of health-related information drawn from devices.

Beyond performing validations by making links to traditional medical data, we must address other challenges with drawing data from computing devices and online services. Using data provided in the course of daily life for inferences raises concerns about disclosure, consent, and privacy. The leveraging of nonmedical data, such as logs of locations and of the terms used in web search over time, for healthcare research and delivery brings to the fore interesting and engaging questions about privacy and use of data—and of evolving views in the United States and Europe on leveraging nonmedical data for medical purposes.[16] As with all studies leveraging personal data, a key consideration is protecting the privacy of individuals via careful containment of data, use of anonymization methods, and using statistical aggregations that cloak personally identifiable information. Researchers need to be clear with participants about the type of data that is recorded and the reasons for the data capture. Methods explored in studies on privacy include uses of encryption[17,18] and mechanisms that could provide patients with understandings and control of privacy risks and tradeoffs.[19,20] Some approaches to ensuring the privacy of sensitive data, such as location information, maintain the sensitive data within devices and rely more centrally on client-side analyses for services.

To date, the methods presented by Nguyen et al are best viewed as a complement rather than as a replacement for established means of recording hospitalizations, such as insurance claims and electronic medical records. However, the methods are promising, and the work highlights an important direction with aligning data from smartphones and electronic medical records. A wealth of opportunities lay ahead in mHealth, including new tools for health research and myriad applications of sensing and intervention for health and wellness. However, sensing and inferences drawn from the use of mobile devices in the wild may be fraught with uncertainty and incompleteness, especially when they are based on noisy signals and interpretations. Making connections to traditional sources of clinical data, as demonstrated in the study of Nguyen et al is important in helping to realize the full potential of mHealth.

Disclosures

None.

References

1. Nguyen KT, Olgin JE, Pletcher MJ, Ng M, Kaye L, Moturu S, Gladstone RA, Malladi C, Fann AH, Maguire C, Bettencourt L, Christensen MA, Marcus GM. Smartphone-based geofencing to ascertain hospitalizations. *Circ Cardiovasc Qual Outcomes*. 2017;10:e003326. doi: 10.1161/CIRCOUTCOMES.116.003326.
2. Health eHeart Study. https://www.health-eheartstudy.org. Accessed February 20, 2017.
3. Estrin D, Sim I. Health care delivery. Open mHealth architecture: an engine for health care innovation. *Science*. 2010;330:759–760. doi: 10.1126/science.1196187.
4. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection—harnessing the Web for public health surveillance. *N Engl J Med*. 2009;360:2153–2155. doi: 10.1056/NEJMp0900702pmid:19423867.
5. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res*. 2009;11:e11. doi: 10.2196/jmir.1157.
6. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. *PLoS One*. 2013;8:e83672. doi: 10.1371/journal.pone.0083672.
7. Sadilek A, Kautz H, Silenzio V. Predicting disease transmission from geo-tagged micro-blog data. *Proceedings of the Twenty-Sixth AAAI National Conference on Artificial Intelligence*. 2012:136–142.
8. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. 2013;3267–3276. doi: 10.1145/2470654.2466447.
9. Paparrizos J, White RW, Horvitz E. Screening for pancreatic adenocarcinoma using signals from web search logs: feasibility study and results. *J Oncol Pract*. 2016;12:737–744. doi: 10.1200/JOP.2015.010504.
10. White RW, Harpaz R, Shah NH, DuMouchel W, Horvitz E. Toward enhanced pharmacovigilance using patient-generated data on the internet. *Clin Pharmacol Ther*. 2014;96:239–246. doi: 10.1038/clpt.2014.77.
11. White R, Horvitz E. From web search to healthcare utilization: privacy-sensitive studies from mobile data. *J Am Med Inform Assoc*. 2013;20:61–68. doi: 10.1136/amiajnl-2011-000765.
12. Yang SH, White RW, Horvitz E. Pursuing insights about healthcare utilization via geocoded search queries. *Proceedings of the Thirty-Sixth Annual ACM SIGIR International Conference on Research and Development in Information Retrieval*. 2013;993–996. doi: 10.1145/2484028.2484147.
13. Agarwal V, Han L, Madan I, Saluja S, Shidham A, Shah NH. Predicting hospital visits from geo-tagged Internet search logs. *Proceedings of the AMIA Joint Summits on Translational Science*. 2016;15–24; PMCID: PMC5001755.

14. Krumm J, Rouhana D, Chang MW. Placer++: semantic place labels beyond the visit. *Proceedings of the International Conference on Pervasive Computing and Communications*. 2015;11–19. doi: 10.1109/percom.2015.7146504.
15. Yang D, Li B, Cudré-Mauroux P. POIsketch: semantic place labeling over user activity streams. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. 2016;2697–2703.
16. Horvitz E, Mulligan D. Policy forum. Data, privacy, and the greater good. *Science*. 2015;349:253–255. doi: 10.1126/science.aac4520.
17. Benaloh J, Chase M, Horvitz E, Lauter K. Patient controlled encryption: ensuring privacy of electronic medical records. *Proceedings of the ACM Workshop on Cloud Computing Security*. 2009:103–114. doi: 10.1145/1655008.1655024.
18. Bos JW, Lauter K, Naehrig M. Private predictive analysis on encrypted medical data. *J Biomed Inform*. 2014;50:234–243. doi: 10.1016/j.jbi.2014.04.003.
19. Singla A, Horvitz E, Kamar E, White RW. Stochastic privacy. *Proceedings of the Twenty-Eighth AAAI National Conference on Artificial Intelligence*. 2014:152–158.
20. Krause A, Horvitz E. A utility-theoretic approach to privacy in online services. *J Artif Intell Res*. 2010;39:633–662. doi: 10.1613/jair.3089.

Key Words: Editorials ■ electronic health records ■ hospitalization ■ smartphones ■ software

Circulation
Cardiovascular Quality and Outcomes

Making Connections: Advancing Healthcare Research Via Consumer Mobile Devices

Eric Horvitz and Ryen W. White

Circ Cardiovasc Qual Outcomes. 2017;10:
doi: 10.1161/CIRCOUTCOMES.117.003573
Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 1941-7705. Online ISSN: 1941-7713