Discovering relationships in consumer use of personal health information: A causal modeling approach

Ryan H. Sandefer, PhD[1], Stuart M. Speedie, PhD, FACMI[2]
[1]College of St. Scholastica, Duluth, MN; [2]University of Minnesota, Minneapolis, MN

Introduction

Patient and family engagement has become a central element in the delivery of quality healthcare services.[1-3] Patient engagement has been called a "blockbuster drug."[4] It Multiple initiatives, including the EHR Incentive Program's patient access objectives, the Blue Button initiative's expansion to connect patients to billing information, and the Clinical Laboratory Improvement Amendment's changes to allow patients direct access to laboratory data, have been implemented in the past several years.[5]

In an effort to engage patients more effectively with information regarding diagnoses, conditions, medications and other information, healthcare organizations have begun using information technology (IT) to provide patients access to their personal health information.[6,7] Research has shown that approximately thirty percent of physicians engage in secure messaging with patients and twenty-four percent provide patients access online access to medical records.[8] It is estimated that in 2013, 12.58% of U.S. adults communicated with a physician using the internet (up from 8.44% and 9.11% in 2008, respectively) and 10.98% of U.S. adults tracked health information using the internet.[9]

We are interested in the relationships of dependency and conditional independency between demographic, socioeconomic and health-related variables and personal healthcare management (PHM). PHM refers to the individual use of internet-based technology to access personal health information or communicate with healthcare professionals regarding personal health information. PHM was defined as the individuals use of health-related information to email healthcare providers, using a computer to schedule a healthcare appointment, and using a computer for requesting a prescription refill. Understanding cause and effect relationships is of primary interest in the field of health informatics.[10,11] Significant progress has been made in the development of algorithms aimed at representing the causal relationships from existing datasets.[12-14] Observational data from electronic health records and other sources is increasingly rapidly, and employing methods aimed at identifying causal relationships from these data provide informatics professionals an opportunity to better predict, explain, and develop consumer engagement techniques aimed at improving health.[11,15]

Objectives

The goal of this study was to develop a deeper understanding of the relationship between predictor variables and PHM use by developing model that represents the causal structure of personal healthcare management through systematic feature selection and use of a causal discovery algorithm.

Materials And Methods

National Health Interview Series (NHIS) was used for this analysis and included 605,001 respondents between 2009-2014.[16] We assessed individual use of PHM. Use of PHM was regarded as positive if individuals responded "Yes" to communicating with a healthcare provider using email, scheduled an appointment with a healthcare provider on the Internet, or refilled a prescription on the Internet. Our analysis included eighty-four individual characteristics as covariates. The covariates were demographic, socioeconomic, and health-related factors.

We used a three-step approach to analyzing the data as follows. The first step in the process was focused on selecting the NHIS features most predictive of PHM using a Markov Blanket Fan Search (MBFS) through the Tetrad software application. The MBFS is a form of feature selection; it is an algorithm that employs the principles of the PC Algorithm to search for a specific node's Markov Blanket [17]. The second step of the process used the Markov Blanket of PHM to construct a Bayesian model network structure that explains the causal relationship between the features most predictive of PHM. The process involved constructing a Directed Acyclic Graph (DAG) using the Weka 3.3. The final step of the process involved evaluating the model's performance.

Results

Of the 84 demographic, socioeconomic, and health-related variables included in the initial sample, the MBFS identified six variables within the Markov Blanket of PHM, including two technology related items (online chat groups and health-related internet search), two socioeconomic variables (poverty level and education level), and two

health-related variables (delayed appointment and flu shot). The factors identified through the feature selection are supported by previous research.[18,19] A DAG was produced through the Bayesian network construction—the DAG is the predictive model used for evaluating PHM use in this study. The resulting model includes seven attributes and nine edges. Poverty level is causal to both education level and PHM. Education level is causal to PHM and flu shot. PHM is causal to online chat group, health-related internet search, and delayed appointment. Health-related internet search is also causally linked to delayed appointment. The model correctly classified 4607 (76.8%) of instances with a kappa statistic of 0.536.

Conclusions
We have developed a Bayesian Network model that predicts PHM use among US adults. The model's parameters were learned through feature selection using the Markov Blanket Fan Search algorithm. The feature selection identified six variables within the Markov Blanket of PHM, and two variables (poverty and education) were determined to be causally related to PHM. The model has the ability to predict PHM use accurately. These findings demonstrate the importance of using social determinants of health for explaining consumer use of health information technology.

References

1. Kohn LT, Corrigan JM, Donaldson MS. Crossing the quality chasm: a new health system for the 21st century. *Washington, DC: Committee on Quality of Health Care in America, Institute of Medicine.* 2001.
2. Triberti S, Barello S, Graffigna G, Riva G, Candelieri A, Archetti F. Evaluating patient engagement and user experience of a positive technology intervention. Paper presented at: Walter de Gruyter GmbH2016.
3. Carman KL, Dardess P, Maurer M, et al. Patient and family engagement: a framework for understanding the elements and developing interventions and policies. *Health Affairs.* 2013;32(2):223-231.
4. Dentzer S. Rx for the 'blockbuster drug'of patient engagement. *Health Affairs.* 2013;32(2):202-202.
5. Patel V, Barker W, Siminerio E. Disparities in Individuals' Access and Use of Health IT in 2013: ONC Data Brief, no.26. 2015.
6. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. *New England Journal of Medicine.* 2010;363(6):501-504.
7. Ahern DK, Woods SS, Lightowler MC, Finley SW, Houston TK. Promise of and potential for patient-facing technologies to enable meaningful use. *American journal of preventive medicine.* 2011;40(5):S162-S172.
8. Furukawa MF, King J, Patel V, Hsiao C-J, Adler-Milstein J, Jha AK. Despite substantial progress in EHR adoption, health information exchange and patient engagement remain low in office settings. *Health Affairs.* 2014;33(9):1672-1679.
9. Ford EW, Hesse BW, Huerta TR. Personal Health Record Use in the United States: Forecasting Future Adoption Levels. *Journal of medical Internet research.* 2016;18(3).
10. Kalisch M, Mächler M, Colombo D, Maathuis MH, Bühlmann P. Causal inference using graphical models with the R package pcalg. *Journal of Statistical Software.* 2012;47(11):1-26.
11. Kleinberg S, Hripcsak G. A review of causal inference for biomedical informatics. *Journal of biomedical informatics.* 2011;44(6):1102-1112.
12. Pearl J. Causality: models, reasoning and inference. *Econometric Theory.* 2003;19:675-685.
13. Spirtes P, Glymour CN, Scheines R. *Causation, prediction, and search.* MIT press; 2000.
14. Glymour CN, Cooper GF. *Computation, causation, and discovery.* Aaai Press; 1999.
15. Sauer B, VanderWeele TJ. Use of directed acyclic graphs. 2013.
16. Minnesota Population Center and State Health Access Data Assistance Center. 2012; Version 5.0:Integrated Health Interview Series. Available at.
17. Ramsey J. A PC-Style Markov Blanket for High Dimensional Datasets. 2006.
18. Sandefer RH, Khairat S, Pieczkiewicz DS, Speedie SM. Using Publicly Available Data to Characterize Consumers Use of Email to Communicate with Healthcare Providers. Paper presented at: MEDINFO 2015: EHealth-enabled Health: Proceedings of the 15th World Congress on Health and Biomedical Informatics2015.
19. Sandefer RH, Westra BL, Khairat SS, Pieczkiewicz DS, Speedie SM. Determinants of Consumer eHealth Information Seeking Behavior. Paper presented at: AMIA Annual Symposium Proceedings2015.