Using a Machine Learning Algorithm to Predict Online Patient Portal Utilization: A Patient Engagement Study

Ahmed U. Otokiti,[1] Colleen M. Farrelly,[2] Leyla Warsame,[3] Angie Li[4]

1. Icahn School of Medicine at Mount Sinai Hospital, Internal Medicine and Informatics Department, New York, NY 10029, USA

2. Staticlysm, LLC, Palmetto Bay, FL 33157, USA

3. Geisinger Health Systems, Internal Medicine and Clinical Informatics Department, Danville, PA 17821, USA

4. University at Buffalo, Department of Biomedical Informatics, Buffalo, NY 14203, USA

Abstract

Objective: There is a low rate of online patient portal utilization in the U.S. This study aimed to utilize a machine learning approach to predict access to online medical records through a patient portal.

Methods: This is a cross-sectional predictive machine learning algorithm-based study of Health Information National Trends datasets (Cycles 1 and 2; 2017-2018 samples). Survey respondents were U.S. adults (≥18 years old). The primary outcome was a binary variable indicating that the patient had or had not accessed online medical records in the previous 12 months. We analyzed a subset of independent variables using k-means clustering with replicate samples. A cross-validated random forest-based algorithm was utilized to select features for a Cycle 1 split training sample. A logistic regression and an evolved decision tree were trained on the rest of the Cycle 1 training sample. The Cycle 1 test sample and Cycle 2 data were used to benchmark algorithm performance.

Results: Lack of access to online systems was less of a barrier to online medical records in 2018 (14%) compared to 2017 (26%). Patients accessed medical records to refill medicines and message primary care providers more frequently in 2018 (45%) than in 2017 (25%).

Discussion: Privacy concerns, portal knowledge, and conversations between primary care providers and patients predict portal access.

Conclusion: Methods described here may be employed to personalize methods of patient engagement during new patient registration.

Abbreviations: American Medical Informatics Association (AMIA), area under the curve (AUC), body mass index (BMI), electronic health record (EHR), Health Information National Trends Survey (HINTS), information technology (IT), National Cancer Institute (NCI), Veteran health (VA)

Introduction

Patient engagement is a set of behaviors that foster active patient involvement in care, thereby increasing motivation and self-determination to become an active player in the healthcare journey [1]. These behaviors increase compliance, improve health outcomes, and overall public health and reduce cost [1-3]. Health IT solutions can serve as a means to increase patient engagement, as online patient portals have been shown to increase patient engagement and personalized care [4,5].

Online patient portals are web-based applications tethered to a patient's EHR that allow secure access to health data. Through this portal, patients can view lab results, medication history, and discharge summaries, and they can securely message their physicians, request prescription refills, and schedule appointments [6].

The meaningful use Stage 2 incentive mandated by the Health Information Technology for Economic and Clinical Health Act in 2009 was a significant driver for increase in patient portal offerings by healthcare institutions across the nation [7]. Despite the significant investment in online portals, these sites continue to experience a low rate of adoption/use by patients, which hinders the potential benefits of patient engagement and its public health impact [8-10].

The most significant positive factors associated with higher portal use include higher education level, female gender, Caucasian ethnicity, Internet access, higher income, patients not on Medicaid insurance, and patient trust in the healthcare provider and system [7,8,11]. The most significant negative factors associated with lower patient portal use include privacy and security concerns and user friendliness [12,13].

Machine learning is gaining popularity in healthcare due to the ability of this method to process complex nonlinear relationships between predictors and yield stable predictions [14]. This approach has been used to predict outbreaks, suicide risk among Army personnel, and intrusion detection within EHR systems [15,16].

Several prior studies have analyzed patient behavior regarding health technology usage and its impact on patient health [17-21]. One study which employed the random forest algorithm found that health-related Internet searches predicted patient healthcare utilization [22]. These findings suggest that understanding patient interactions with medical technology may help providers offer better care and be proactive in making decisions about online patient engagement tools.

This study aimed to determine which patients are most likely to utilize online portals at patient registration and to build a predictive model that could be used to create a short survey to support real-time decision support. As interaction terms likely exist between factors and because the model is high dimensional, we choose to use machine learning models to parse out factors and groups of factors associated with online portal utilization. Patients who opt for a technology-based platform may benefit from other types of engagement with technology beyond patient portals, including text messaging or automated calls. Machine learning algorithms can identify important predictors of portal usage, as well as provide robust predictions to flag those most likely to benefit from portal

usage versus those who may engage better with alternative channels. To our knowledge, no previous studies have utilized a machine learning algorithm to predict patients to utilize patient portals as a patient engagement tool.

Data from HINTS was used for our analysis. HINTS is a nationally representative survey that has been administered by the NCI since 2003 [23]. The HINTS survey and data collection program was set up to monitor changes in the rapidly evolving field of health communication. It collects nationally representative data about the public's use of cancer-related information and serves as a test bed for researchers to evaluate new theories in health information and communication. The data can also be used to help understand how adults use different communication channels to obtain health information [23,24]. Two cycles of HINTS data were utilized in our analysis: HINTS-5 Cycle 1 (2017) and HINTS-5 Cycle 2 (2018). Although HINTS is funded by the NCI with the primary goal of evaluating health communication theories in cancer patients, only 504 individuals out of 3,285 survey participants (15.3%) in HINTS 5 Cycle 1 were diagnosed with cancer [25].

Materials and Methods

Study Design and Setting

This was a predictive analytic study using data from two iterations of the HINTS survey. The survey for both HINTS cycles utilized in this study was disseminated via mail to the participants. More information on the survey mailing protocol, data collection, data cleaning/editing, and handling of incomplete/invalid data can be found on the NCI HINTS website [23].

Study Participants

Survey respondents were sampled from the U.S. population ($\geq$ 18 years old). A two-stage sampling method was utilized: stage one was a stratified sample of residential addresses and stage two sampling was the selection of one adult from each sampled residential address. The same sampling methodology was utilized for HINTS 5 Cycle 1 and 2. More information about sampling methodology of the HINTS survey can be found on the NCI HINTS website [23].

The sample sizes of both iterations were as follows: HINTS 5 Cycle 1: of the 3,285 respondents, 97% of the surveys were completely filled out (November 2017); HINTS 5 Cycle 2: of the 3,504 total respondents, 98% of the surveys were completely filled out (November 2018). These two iterations were chosen because they were the most recent at the time of our analysis, had uniformity of survey collection, and had captured similar variables of interest.

Study Variables

Target Variable/Outcome Variable

Access to online medical records or patient portals was the target or outcome variable. The survey question was, "How many times did you access your medical record in the past 12 months?" (HINTS 5 Cycle 1: question D4; HINTS 5 Cycle 2: question D6). We recoded the response as >1 for "accessed online medical record" and <1 for "did not access online medical record."

Labels/Predictor Variables

A total of 51 initial predictor variables were added based on domain knowledge and a literature search of previously identified significant determinants of online portal use (Table 1) [5,7,13,26,27].

The following variables were re-coded:

a. BMI re-coded to an ordinal variable from a continuous variable for clinical significance (BMI: <25, 25-30, 30-40, >40)
b. Chronic medical condition: any one of the following: hypertension, heart condition, lung disease, and arthritis
c. Anxiety/depression: re-coded as an independent variable

A random forest-based Boruta method was used for variable selection after initial variable inclusion based on domain knowledge and a literature review; a total of 39 variables were selected (Table 1).

Table 1: Variables

Initial variables before the Boruta algorithm	Final variables analyzed after the Boruta algorithm
Demographics	**Demographics**
1. Age	1. Age
2. Education level	2. Education level
3. Race/ethnicity	3. Race/ethnicity
4. Marital status	4. Occupation status
5. Occupation status	5. English language proficiency
6. English language proficiency	6. Sexual orientation
7. Sexual orientation	7. Annual household income
8. Total persons in the household	**Looking for Health Information**
9. Gender	8. Trust health information from newspapers/magazines
10. Rent or own a house	9. Trust health information from the Internet
11. Annual household income	10. Trust health information from charitable organizations
Looking for Health Information	11. Trust health information from religious organizations
12. Trust health information from newspapers/magazines	12. If there were a strong need to get information about your health, where would you go first?
13. Trust health information from the Internet	
14. Trust health information from charitable organizations	

15. Trust health information from television

16. Trust health information from religious organizations

17. If there were a strong need to get information about your health, where would you go first?

Overall Health

18. In general, what is your state of health?

19. Body mass index

20. Chronic medical conditions: diabetes mellitus, hypertension, heart disease, lung disease, rheumatologic

21. Chronic medical condition: depression/anxiety

Your Healthcare

22. Health insurance from employer?

23. Health insurance bought directly from insurance company?

24. Medicare

25. Medicaid

26. Military healthcare/TRICARE

27. VA

28. Indian health services

29. Health insurance other

Medical Research and Records

30. Who offered you online access to your medical records: healthcare provider?

31. Who offered you online access to your medical records: insurance company?

32. How many times have you accessed online medical record in the last 12 months?

Your Healthcare

13. Health insurance from employer?

14. Health insurance bought directly from insurance company?

15. Medicare

16. Medicaid

17. Military health care/TRICARE

18. Indian health services

Medical Research and Records

19. Who offered you online access to your medical records: healthcare provider?

20. Who offered you online access to your medical records: insurance company?

21. How many times have you accessed your online medical record in the last 12 months?

22. How confident are you about the safety and confidentiality of your electronic medical record?

23. Have you ever kept information from your health care provider because of privacy concerns?

Internet Use

24. Internet use through broadband

25. Internet use through a cellular network

26. Internet use through a wireless network

27. Internet use through a computer at home

28. Internet use through a computer at work

29. Internet use on a mobile device (cell phones, tablet, etc.)

30. In the past 12 months, have you looked for medical information for yourself?

31. In the past 12 months, have you used the Internet to communicate with a healthcare provider's office?

33. How confident are you about safety and confidentiality of your electronic medical record?

34. Have you ever kept information from your healthcare provider because of privacy concerns?

Internet Use

35. Internet use through broadband

36. Internet use through a cellular network

37. Internet use through a wireless network

38. Internet use through; computer at home

39. Internet use through a computer at work

40. Internet use on a mobile device (cell phones, tablet, etc.)

41. In the past 12 months, have you looked for medical information for yourself?

42. In the past 12 months, have you used the Internet to communicate with a health care provider's office?

43. In the past 12 months, have you used the Internet to view your test results?

44. Do you have a tablet?

45. Do you have a smart phone?

46. Do you have a wellness app on your phone or tablet?

47. Has your tablet or smartphone helped you make health decisions?

48. In the last 12 months, have you used other electronic devices to monitor your health?

32. In the past 12 months, have you used the Internet to view your test results?

33. Do you have a tablet?

34. Do you have a smart phone?

35. Do you have a wellness app on your phone or tablet?

36. Has your tablet or smartphone helped you make health decisions?

37. Have you visited a social networking site in the last 12 months?

38. Have you watched a health-related video on YouTube in the last 12 months?

39. Have you sent or received a text message from a health care provider in the last 12 months?

49. Have you visited a social networking site in the last 12 months?	
50. Have you watched a health-related video on YouTube in the last 12 months?	
51. Have you sent or received a text message from a health care provider in the last 12 months?	

Machine Learning Approach/Statistical Analysis

Since there are known limitations for some statistical algorithms and notable issues with the reproduction or generalization of clinical and social science study results, we decided to use more robust methodologies, including multiple supervised machine learning approaches; we also used Cycle 2 as a replication population upon which to compare our initial Cycle 1 results to ensure replicability across populations [28,29]. Thus, Cycle 1 was partitioned for use in variable selection, model training, and initial testing of the trained models, and Cycle 2 was saved for replication of Cycle 1 test sample results.

Often, especially with linear regression, either only one data collection step is used to validate a model, leading to generalization problems on other sets of data collected on similar populations, or the model is trained on one population and tested on another. Both are statistically problematic methods in creating a model [29]. One study applied multiple-sampling approaches with pooling (how this study set up the methodology) and was able to replicate >90% of the problematic samples noted in one of the prominent replication studies suggesting that most clinical paper results do not generalize properly [29].

Unsupervised Learning

To determine which subgroups of patients did not choose to access online health records, we clustered two samples of patients who did not access online health records (NotAccessed_ConcernedPrivacy and NotAccessed_NoInternet) using k-means clustering on the data from Cycle 1 and Cycle 2. The number of clusters was determined using the elbow on both Cycles [30]. Results were compared between Cycles to understand how behaviors changed over time.

To identify the types of records accessed by patients who did choose to access online health records, we clustered four variables (RecordsOnline_RefillMeds, RecordsOnline_RequestCorrection, RecordsOnline_MessageHCP, and RecordsOnline_AddHealthInfo) on record-accessing patients from Cycles 1 and 2. The main

groups that appeared in both clustering results were compared across Cycle 1 and Cycle 2 to understand how usage changed over time.

Supervised Learning

We used stratified sampling to split Cycle 1 data into three parts: variable selection training sample, model training sample, and model test sample. To select variables, we used the Boruta algorithm, which statistically tests a random forest model to select statistically significant variables (set to $p<0.05$). This allowed us to identify main effects as well as interaction terms related to our outcome.

To identify main effects and interaction terms separately for clinical evaluation, we fit two supervised learning models in R (logistic regression for main effects and evolved tree model for complex interaction effects) [32,33]. The evolved tree model, fit using evtree in R, allowed us to visualize complex interaction terms that are common in medical data.

We then evaluated our logistic regression model and evolved tree model on the Cycle 1 test set by measuring the AUC, false positive and negative rates, and accuracy. For the logistic regression model only, we used the Akaike information criterion, which measures the goodness of fit balanced with the number of variables included in the logistic regression model [34]. Evaluation was replicated on the Cycle 2 sample to assess reproducibility of model performance across time periods.

Results

Variable Selection

After the five runs of the Boruta algorithm on our first Cycle 1 sample, we looked at which variables were not selected by any of the selection runs and identified the following: MaritalStatus, TotalHousehold, SelfGender, RentOrOwn, TrustTelevision, GeneralHealth, BMIOver25, ChronicMedicalFlag, MedConditions_Depression, HealthIns_VA, HealthIns_Other, and OtherDevTrackHealth. These were discarded from the subsequent training and test sets (Figure 1).
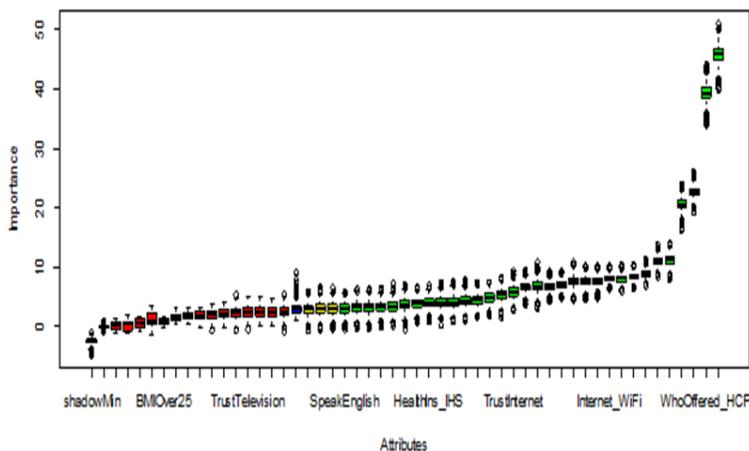
Figure 1: Variable output of the Boruta algorithm

Unsupervised Learning Results

For patients who did not access their medical records, the k-means model for both Cycle 1 and Cycle 2 selected the optimum number of clusters as 4 (all possible combinations of the two variables, giving 100% of the variance accounted for in the k-means models). Access issues generally decreased between Cycle 1 and Cycle 2, suggesting that access to the Internet declined as a significant barrier to usage over time (Table 2).

Table 2: Unsupervised learning results: k-means model for both Cycle 1 and Cycle 2 for those who did not access their medical records

Not Accessed Subgroup	Cycle 1 Percent	Cycle 2 Percent
Privacy Only	17%	13%
Access and Privacy	10%	4%
Other	57%	74%
Access Only	16%	10%

For patients who accessed their online records, the optimal clustering for Cycle 1 included 5 clusters (~60% of variance accounted for), with major groups including a large subset of patients who mainly refilled medication and messaged primary care providers, a small subset who performed every task online, and a large subset that rarely used online portals for any tasks. The best k-means model for Cycle 2 comprised six cluster groups, including three groups of interest from Cycle 1 results. The number of patients who refilled medications and messaged primary care providers increased dramatically between cycles, suggesting a common use of online medical records (Table 3).

Table 3. Unsupervised learning results: k-means model for both Cycle 1 and Cycle 2 for those who accessed their online medical records

Main Accessed Subgroup	Cycle 1 Percent	Cycle 2 Percent
Rare Usage	49%	35%
Refill Meds and Message Primary Care	25%	45%

Every Task Online	3%	4%

Supervised Learning Results

For the logistic regression model, we found that the model selected in Cycle 1 training data did not generalize to Cycle 2 data (with the test set AUC falling from 84% to 55% between cycles). Thus, we discarded our results as not reproducible or useful as a clinical decision model. However, the evolved tree model (Figure 2) was reproducible between cycles with AUC falling marginally from 85% to 81% between cycles, see AUC of cycle 1 data in Figure 3 and AUC of cycle 2 data in Figure 4. Significant predictors of online portal usage, according to this model, included privacy concerns, a proactive offering of access to online portals by primary care providers, and prior use of the portal to check test results (Figure 2).
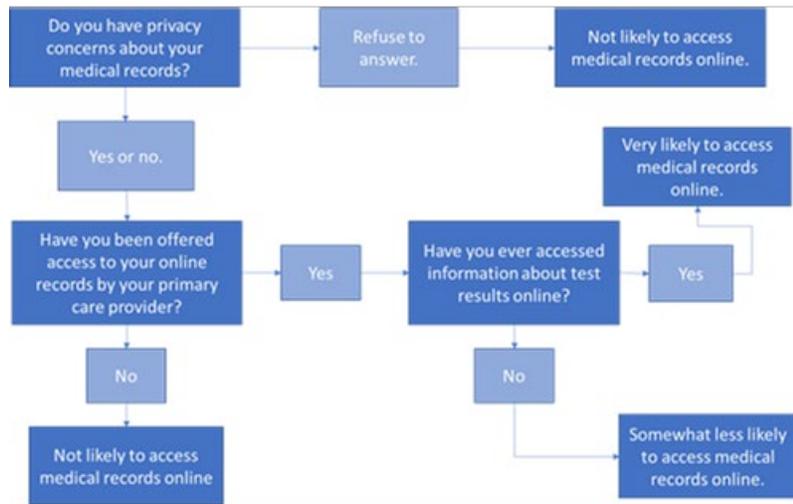

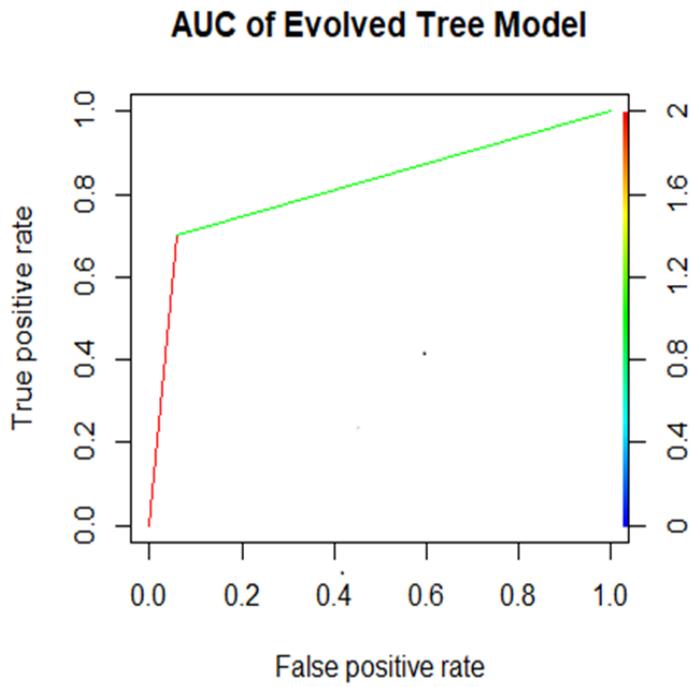Figure 2: Decision tree diagram of the supervised learning method
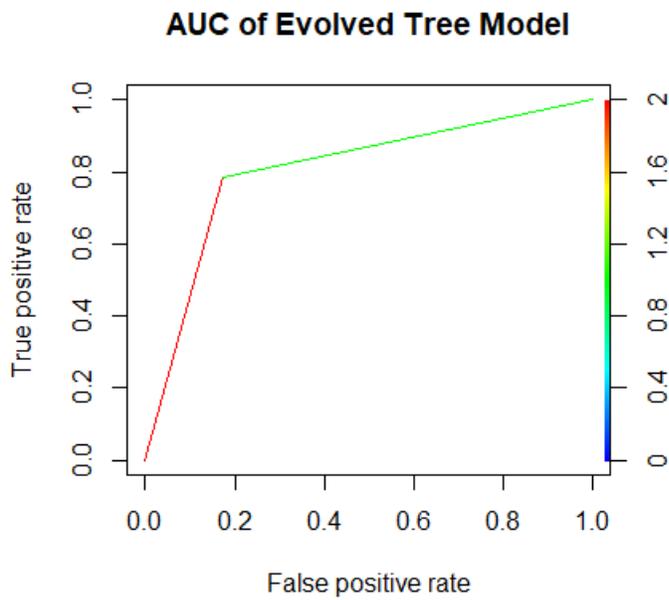
Figure 3: Evolved tree model AUC for cycle 1

Figure 4: Evolved tree model AUC for cycle 2

Discussion

This study sought to identify predictive factors that determine online patient portal use using machine learning methodology. We found that previous use of online portals is a positive predictor of online portal usage, as well as the offering of online portals by primary care providers.

The 2009 Health Information Technology for Economic and Clinical Health Act and meaningful use facilitated the creation and availability of online patient portals; however, there has been a low adoption rate among patients. Studies have shown that, although organizations have created portals and provided patients with log-in information, patients did not utilize the portals. However, providers that encourage portal use by tasking patients with items to complete or helping patients with the initial log in improves usage [9,35]. Irizarry et al. (2015) found that provider endorsement and engagement with patient portals positively affected patient portal utilization [5,7].

Privacy concerns are a negative a predictor of patient portal use [12,13]. News of recent data breaches does little to instill confidence in how institutions protect health information and how accessible it is to unauthorized entities [36]. Communicating institutional safety measures to secure patient privacy could improve patient trust [37]. Anthony et al. (2018) recommended that providers play a role in improving trust in portals by addressing privacy concerns directly with patients [9].

Our study indicates that access to the Internet is not as significant of a barrier as described in previous studies. The AMIA released a statement in 2018 that "broadband access is or will become a social determinant of health;" [38] however, with greater access to smartphones, a socioeconomic divide in Internet access is no longer a strong predictor of portal use [8,39]. Additionally, other populations, such as seniors, now have improved Internet access [8]. Nambisan (2017) postulated that use of the Internet for health information seeking is a better predictor of portal use rather than access to the Internet [40]. However, even with the minimal digital divide, health literacy, computer literacy, and care preferences may continue to represent barriers to patient portal utilization [7,39].

According to our study, online portals are most commonly used to refill medications and message primary care providers. Patel et al. showed that more than half of patients who access their online portals use it to perform health-related tasks and to communicate with their healthcare providers [9]. The Institute of Medicine identified patient-provider communication as a core focus in improving patient outcomes. Secure messaging augments clinical encounters by providing asynchronous communication between providers and patients [41].

Our study has shown that the prospect of utilizing a machine learning model to predict patient engagement via patient portals is promising. This technique may be scaled up to a clinical decision support tool as a user-friendly web interface or app to predict IT engagement patterns for clinic registration of new patients. Further research and validation of the model in a real ambulatory setting is necessary prior to implementation of such a tool.

Limitations

Although this study is a novel attempt to implement a machine learning approach for patient portal utilization, including the clustering method that provided additional insight, it is not without

limitations. First, the cross-sectional design of the HINTS survey does not allow inferences of causality. Secondly, the variables in the survey are subject to individual interpretations of the survey questions by the respondents in addition to any response bias that may be present.

Limitations of k-means clustering include assumptions about outliers (that groups are even-sized and non-overlapping). Most real-world data will violate this to some extent. In addition, generally, evolved trees are not the most stable learners; therefore, it is possible that other tree models can be used. However, our results were consistent across partitions of data, and statistical testing on the validation sample confirmed that the model was robust.

Conclusions

The tree model produced more consistent prediction accuracy across cycles than the regression model. It also identified privacy and data protection concerns (negative predictors) and proactive patient portal access offering by physicians (positive predictors) as the most significant determinants of patient portal use. Our unsupervised learning algorithm identified a fairly consistent cluster of patients who did not use online portals due to privacy concerns across both cycles of data. Among patients who used online portals, there was a consistent cluster of patients across cycles that used the online portal for medication refills and to message their primary care provider.

Our results showed that machine learning algorithms can be used to identify factors associated with online portal use. These methods may be employed in a clinical decision support tool during new patient registration to personalize methods of patient engagement. The variables identified by our model corresponded with the characteristics of online portal users identified by previous studies [5,8]. We recommend asking patients about privacy concerns and proactively offering patients a way to access their records online or providing an alternative (text messaging, automated call, etc.) based on their response to questions asked during registration.

Financial Disclosure

No Financial Disclosures.

Competing Interests

No Competing Interests.

Data Availability

The data set used and analyzed for the study are available for free on the U.S. Department of Health and National Cancer Institute website: https://hints.cancer.gov/

References

1. Graffigna G, Barello S, Bonanomi A, Riva G. 2017. Factors affecting patients' online health information-seeking behaviours: The role of the Patient Health Engagement (PHE). *Model.*

Patient Educ Couns. 100(10), 1918-27. Epub May 2017. doi:https://doi.org/10.1016/j.pec.2017.05.033. PubMed

2. Laurance J, Henderson S, Howitt PJ, Matar M, Al Kuwari H, et al. 2014. Patient engagement: four case studies that highlight the potential for improved health outcomes and reduced costs. *Health Aff (Millwood).* 33(9), 1627-34. doi:https://doi.org/10.1377/hlthaff.2014.0375. PubMed

3. James J. "Health Policy Brief: Patient Engagement," Health Affairs [Internet]. 2013 February [cited April 23, 2021] Available from: https://www.healthaffairs.org/do/10.1377/hpb20130214.898775/full/healthpolicybrief_86.pdf

4. Reed ME, Huang J, Brand RJ, Neugebauer R, Graetz I, et al. 2019. Patients with complex chronic conditions: Health care use and clinical events associated with access to a patient portal. *PLoS One.* 14(6), e0217636. doi:https://doi.org/10.1371/journal.pone.0217636. PubMed

5. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res. 2015;17(6):e148. doi: https://doi.org/10.2196/jmir.4255

6. Health IT. gov [Internet]. [cited 2020 January 26] Available from: https://www.healthit.gov/faq/what-patient-portal

7. Anthony DL, Campos-Castillo C, Lim PS. 2018. Who isn't using patient portals and why? evidence and implications from a national sample of US adults. *Health Aff (Millwood).* 37(12), 1948-54. doi:https://doi.org/10.1377/hlthaff.2018.05117. PubMed

8. Hong YA, Jiang S, Liu PL. 2020. Use of patient portals of electronic health records remains low from 2014 to 2018: results from a national survey and policy implications. *Am J Health Promot.* 34(6), 677-80. Epub Feb 2020. doi:https://doi.org/10.1177/0890117119900591. PubMed

9. Patel V, Johnson C. " Individuals' Use Of Online Medical Records And Technology For Health Needs," ONC Data Brief [Internet]. 2018 April [cited January, 26 2020] Available from: https://www.healthit.gov/sites/default/files/page/2018-03/HINTS-2017-Consumer-Data-Brief-3.21.18.pdf

10. Han HR, Gleason KT, Sun CA, Miller HN, Kang SJ, et al. 2019. Using patient portals to improve patient outcomes: systematic review. *JMIR Human Factors.* 6(4), e15038. doi:https://doi.org/10.2196/15038. PubMed

11. Lyles CR, Sarkar U, Ralston JD, Adler N, Schillinger D, et al. 2013. Patient-provider communication and trust in relation to use of an online patient portal among diabetes patients: The Diabetes and Aging Study. *J Am Med Inform Assoc.* 20(6), 1128-31. Epub May 2013. doi:https://doi.org/10.1136/amiajnl-2012-001567. PubMed

12. Baldwin JL, Singh H, Sittig DF, Giardina TD. 2017. Patient portals and health apps: Pitfalls, promises, and what one might learn from the other. *Healthc (Amst)*. 5(3), 81-85. Epub Oct 2016. doi:https://doi.org/10.1016/j.hjdsi.2016.08.004. PubMed

13. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, et al. 2018. Use and the users of a patient portal: cross-sectional study. *J Med Internet Res*. 20(9), e262. doi:https://doi.org/10.2196/jmir.9418. PubMed

14. Kuhn M, Johnson K. Applied predictive modeling. Springer; 2013.

15. Gupta S, Hanson C, Gunter CA, Frank M, Liebovitz DM, et al. Modeling and detecting anomalous topic access [abstract]. 2013 IEEE International Conference on Intelligence and Security Informatics. doi:https://doi.org/10.1109/ISI.2013.6578795

16. Boxwala AA, Kim J, Grillo JM, Ohno-Machado L. 2011. Using statistical and machine learning to help institutions detect suspicious access to electronic health records. *J Am Med Inform Assoc*. 18(4), 498-505. doi:https://doi.org/10.1136/amiajnl-2011-000217. PubMed

17. Davis Giardina T, Menon S, Parrish DE, Sittig DF, Singh H. 2014. Patient access to medical records and healthcare outcomes: a systematic review. *J Am Med Inform Assoc*. 21(4), 737-41. Epub Oct 2013. doi:https://doi.org/10.1136/amiajnl-2013-002239. PubMed

18. Mold F, Ellis B, de Lusignan S, Sheikh A, Wyatt JC, et al. 2012. The provision and impact of online patient access to their electronic health records (EHR) and transactional services on the quality and safety of health care: systematic review protocol. *Inform Prim Care*. 20(4), 271-82. doi:https://doi.org/10.14236/jhi.v20i4.17. PubMed

19. Ross SE, Lin CT. 2003. The effects of promoting patient access to medical records: a review [Corrected and republished from: J Am Med Inform Assoc. 2003 May-Jun;10] [3] [:294. doi:10.1197/jamia.m1147]. *J Am Med Inform Assoc*. 10(2), 129-38. PubMed https://doi.org/10.1197/jamia.M1147

20. Ross SE, Todd J, Moore LA, Beaty BL, Wittevrongel L, et al. 2005. Expectations of patients and physicians regarding patient-accessible medical records. *J Med Internet Res*. 7(2), e13. doi:https://doi.org/10.2196/jmir.7.2.e13. PubMed

21. Steele AJ, Denaxas SC, Shah AD, Hemingway H, Luscombe NM. 2018. Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease. *PLoS One*. 13(8), e0202344. doi:https://doi.org/10.1371/journal.pone.0202344. PubMed

22. Agarwal V, Zhang L, Zhu J, Fang S, Cheng T, et al. 2016. Impact of predicting health care utilization via web search behavior: a data-driven analysis. *J Med Internet Res*. 18(9), e251. doi:https://doi.org/10.2196/jmir.6240. PubMed

23. National Cancer Institute [Internet]. [cited 2020 January 26] Available from: https://hints.cancer.gov/about-hints/learn-more-about-hints.aspx

24. Assari S, Khoshpouri P, Chalian H. 2019. Combined effects of race and socioeconomic status on cancer beliefs, cognitions, and emotions. *Healthcare (Basel)*. 7(1), 17. doi:https://doi.org/10.3390/healthcare7010017. PubMed

25. Miri S. The target population for health IT solutions: The Health Information National Trends Survey (HINTS 2017) [abstract]. 2019 American Medical Informatics Association Clinical Informatics Conference. Atlanta, GA, May, 1 2019, AMIA.

26. Grossman LV, Masterson Creber RM, Benda NC, Wright D, Vawdrey DK, et al. 2019. Interventions to increase patient portal use in vulnerable populations: a systematic review. *J Am Med Inform Assoc*. 26(8-9), 855-70. doi:https://doi.org/10.1093/jamia/ocz023. PubMed

27. Zhao JY, Song B, Anand E, et al. Barriers, Facilitators, and Solutions to Optimal Patient Portal and Personal Health Record Use: A Systematic Review of the Literature. AMIA Annu Symp Proc 2017; 2017: 1913-1922. 2018/06/02.

28. Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JP. 2011. Public availability of published research data in high-impact journals. *PLoS One*. 6(9), e24357. doi:https://doi.org/10.1371/journal.pone.0024357. PubMed

29. Gilbert DT, King G, Pettigrew S, Wilson TD. 2016. Comment on "Estimating the reproducibility of psychological science". *Science*. 351(6277), 1037. doi:https://doi.org/10.1126/science.aad7243. PubMed

30. Steinley D. 2006. K-means clustering: a half-century synthesis. *Br J Math Stat Psychol*. 59(Pt 1), 1-34. doi:https://doi.org/10.1348/000711005X48266. PubMed

31. Kursa MB, Rudnicki WR. 2010. Feature selection with the Boruta package. *J Stat Softw*. 36(11), 1-13. doi:https://doi.org/10.18637/jss.v036.i11.

32. Nelder JA, Wedderburn RW. 1972. Generalized linear models. *J R Stat Soc [Ser A]*. 135(3), 370-84. doi:https://doi.org/10.2307/2344614.

33. Papagelis A, Kalles DGA. Tree: genetically evolved decision trees [abstract]. 2000 Proceedings 12th IEEE Internationals Conference on Tools with Artificial Intelligence ICTAI. doi: https://doi.org/10.1109/TAI.2000.889871

34. Sakamoto Y, Ishiguro M, Kitagawa G. Akaike information criterion statistics. Springer Netherlands; 1986.

35. Powell KR. 2017. Patient-perceived facilitators of and barriers to electronic portal use: a systematic review. *Comput Inform Nurs*. 35(11), 565-73. doi:https://doi.org/10.1097/CIN.0000000000000377. PubMed

36. Hossain MM, Hong YA. Trends and characteristics of protected health information breaches in the United States. AMIA Annu Symp Proc. 2020;2019:1081-1090. Published 2020 Mar 4.

37. Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. J Am Med Inform Assoc. 2011 Dec;18 Suppl 1(Suppl 1):i8-12. doi: https://doi.org/10.1136/amiajnl-2011-000473. PMID: 22071530.

38. AMIA. AMIA Responds to FCC Notice on Broadband-Enabled Health Technology. American Medical Informatics Association; 2017 [cited 2022 December]; Available from: https://amia.org/public-policy/public-comments/amia-responds-fcc-notice-broadband-enabled-health-technology.

39. Graetz I, Huang J, Muelly ER, Fireman B, Hsu J, et al. 2020. Association of mobile patient portal access with diabetes medication adherence and glycemic levels among adults with diabetes. *JAMA Netw Open*. 3(2), e1921429. doi:https://doi.org/10.1001/jamanetworkopen.2019.21429. PubMed

40. Nambisan P. 2017. Factors that impact Patient Web Portal Readiness (PWPR) among the underserved. *Int J Med Inform*. 102, 62-70. Epub Mar 2017. doi:https://doi.org/10.1016/j.ijmedinf.2017.03.004. PubMed

41. Wallwiener M, Wallwiener CW, Kansy JK, Seeger H, Rajab TK. 2009. Impact of electronic messaging on the patient-physician interaction. *J Telemed Telecare*. 15(5), 243-50. doi:https://doi.org/10.1258/jtt.2009.090111. PubMed