Published online: 2020-08-21

Appendix: Content Summaries of Selected Best Papers for the IMIA Yearbook 2020 Section "Clinical Information Systems"

Gordon WJ, Wright A, Aiyagari R, Corbo L, Glynn RJ, Kadakia J, Kufahl J, Mazzone C, Noga J, Parkulo M, Sanford B, Scheib P, Landman AB

Assessment of employee susceptibility to phishing attacks at US health care institutions

JAMA Netw open 2019;2(3):e190393

The current paper from Gordon *et al*., picks up on an important topic from the field of data security and investigates the susceptibility of healthcare employees to phishing attacks in the US. The recent past has shown that attackers have increasingly targeted healthcare organizations, not only with substantial economic impact but also with a strong influence on patient treatment. The authors illustrate multiple examples ranging from partial unavailability of systems up to a two-week complete shutdown of systems. The authors have carried out an investigation to get an insight into the reasons why employees of healthcare organizations fall victims of phishing campaigns. The investigation is based on a retrospective, multicenter quality improvement study that included six US healthcare organizations that represent the entire spectrum of care and a range of US geographies. All organizations have an information security program in place. The respective organizations have carried out phishing simulations in their facilities in the past, based on vendor- or custom-developed software tools.

Data about the phishing attacks were collected from the different institutions, and emails were classified according to their content in three categories: office-related, personal, or information technology-related. Several statistical values were calculated, such as the click rates, median click rates, and odds ratios (with 95% CI). Correlation was, amongst others, computed for the year, number of campaigns, email category, and season. In total, the data set included 95 campaigns with emails sent from 2011 to 2018. The overall click rate across all institutions and campaigns was 14.2%, although the authors observed considerable differences in the click rate of institutions ranging from 7.4% to 30.7%. The authors found out that repeated phishing campaigns were associated with decreased odds of clicking on a subsequent phishing email.

Interestingly, the year is not significantly associated with the click rate. Further, emails that were related to the personal email category had a significantly higher probability of being clicked. The same is true for seasons, both the spring and summer seasons were associated with higher click rates. In models adjusted for several potential confounders, including year, the institutional campaign number, institution, and email category, the odds of clicking on a phishing email were 0.511 lower for six to ten campaigns at an institution and 0.335 lower for more than ten campaigns at an institution. The study could well demonstrate that the healthcare domain compares well to other industries and that employees benefit from education, training, and that experiences made from other simulated phishing campaigns can help employees to stay aware. In addition, the healthcare domain has some particularities that make it especially vulnerable to attacks such as turnover of employees, endpoint complexity, or information system interdependence. It is therefore inevitable that all participants in the healthcare domain understand these security risks, particularly as safe and effective health care delivery involves more and more information technology.

Hill BL, Brown R, Gabel E, Rakocz N, Lee C, Cannesson M, Baldi P, Loohuis LO, Johnson R, Jew B, Maoz U, Mahajan A, Sankararaman S, Hofer I, Halperin E

An automated machine learning-based model predicts postoperative mortality using readily-extractable preoperative electronic health record data

Br J Anaesth 2019;123(6):877–86

The majority of surgical complications is associated with a small group of high-risk patients. Often these patients would substantially benefit from early identification of their high-risk of potential complications, as proactive, early interventions can help reduce or even avoid perioperative complications. Existing approaches to this problem either require a clinician to review a patient's chart such as the American Society of Anesthesiologists (ASA) physical status classification or lack specificity. The work of Hill *et al*., is dedicated to the investigation of a machine learning approach that uses readily available patient data for the prediction of certain risks and takes changing patient conditions into account. Data from 53,097 surgical patients (2.01% mortality rate) who underwent general anesthesia between 2013 and 2018 were collected from the perioperative data warehouse at UCLA Health to populate a series of 4,000 distinct measures and metrics. In the next step, classification models to predict in-hospital mortality as a binary outcome were trained (model endpoint) and the outcome for a subset of patients was checked with trained clinicians. For the actual creation and training of models, four different classification models, logistic regression, ElasticNet, random forests, and gradient boosted trees were evaluated. The performance of the created models was then compared with existing clinical risk scores such as the ASA score, POSPOM score, and Charlson comorbidity score. The mean value of the area under the receiver operating characteristic (AUROC) curve (95% CI, 1,000 predictions) was used to compute the performance. When using the ASA status or the Charlson comorbidity score as the only input features, the linear models (logistic regression, ElasticNet) outperform the non-linear models (random forest, XGBoost). However, for the other feature sets, the non-linear models outperform the linear models. In particular, the random forest has the highest AUROC compared with the other models. The authors were able to show that a fully automated preoperative risk prediction score can better predict in-hospital mortality than the ASA score, the POSPOM score, and the Charlson comorbidity score. Unlike previously developed models, the results also indicate that the inclusion of the ASA score in the model did not improve the predictive ability. Another advantage of such an automated model is that it allows for the

continuous recalculation of risk longitudinally over time. However, the authors also state several limitations of the study such that the incidence of mortality in the testing set was less than 2%, implying that a model that blindly reports 'survives' every time will have an accuracy greater than 98%. Nonetheless, the model outperforms current major models in use.

Shen N, Bernier T, Sequeira L, Strauss J, Silver MP, Carter-Langford A, Wiljer, D

Understanding the patient privacy perspective on health information exchange: A systematic review

Int J Med Inform 2019;125:1–12

The exchange of health information and the ability to share information regarding the patient and treatment has become an essential element in the effective and efficient provision of healthcare services. On the other side, these developments have also led to increasing concerns by patients not being able to properly control these information flows. Although privacy concerns are often quoted in publications regarding the exchange of health information, they are seldom investigated with regard to their influence on the patient-provider relationship in healthcare. The paper of Shen *et al*., focuses on an in-depth exploration of the patient's perspective towards privacy in the context of health information exchange. For this reason, the authors have conducted a systematic review, which was based on PRISMA and aimed at providing a conceptual synthesis of the patient privacy perspective and its associated antecedents and outcomes; identify gaps in the APCO model (Antecedent, Privacy Concern, Outcomes macro-model) and describe the current state of privacy research. The APCO model was developed by the authors in advance to the present study. Major databases were queried for empirical studies focused on patient/public privacy perspectives in the context of HIT that were published between 2015 and 2017. Data was extracted based on the elements of the APCO model, and subsequently, new elements were added if outside the APCO model. The authors found 39 quantitative, 15 qualitative, and five mixed-methods studies that were relevant. The analysis of the antecedent factors with regard to their influence on patient privacy concerns showed a mixed picture, and an evident positive or negative influence was often not deducible, or the number of studies was low. The authors assumed income, political ideology, and quality of care as being agreed upon studies. The same is true for privacy concerns related to outcome factors, where the authors assume a willingness to share, protective behaviors, benefits, and risks, as agreed by the studies found. Summarizing, the patient privacy perspective seems to be of dynamic and nuanced meaning that is strongly dependent on its context. So, it is difficult to characterize the patient perspective, although privacy concerns, as such, are found in many studies and are expressed as a serious concern. This may also be because many studies have analyzed the concept of privacy only as a peripheral topic and have distilled privacy into a single question. The authors plead that future research needs to place greater emphasis on understanding how antecedent factors can alleviate privacy concerns, build trust, and empower patients. In addition, they claim that building a base of evidence on the actual effects of privacy concerns will help to reduce value-laden discussions and normative assumptions.