

## Appendix: Summary of Best Papers Selected for the 2021 Edition of the IMIA Yearbook, Section Human Factors and Organizational Issues

Adler-Milstein J, Zhao W, Willard-Grace R, Knox M, Grumbach K

**Electronic health records and burnout: Time spent on the electronic health record after hours and message volume associated with exhaustion but not with cynicism among primary care clinicians**

*J Am Med Inform Assoc* 2020;27(4):531-8

Primary care physicians have among the highest rate of burnout of all clinical specialties; further, these physicians have self-reported after hours work as a key factor for feelings of burnout. This study compared self-reported and objective measures of electronic health record (EHR) work and proficiency and then correlated them with burnout measures for primary care providers to better understand and improve factors driving burnout. To do this, the study used self-reported measures of EHR use and proficiency, EHR supplied metrics for after-hours EHR use, messaging volume, and proficiency, and Maslach Burnout Inventory exhaustion and cynicism subscale responses for 87 primary care providers in an urban academic medical center. The results showed that the subjective, self-reported perceived EHR use time was correlated with objective measures of EHR use after-hours and on unscheduled days, message volume, and proficiency metrics. There was not a relationship between the subjective and objective measures of EHR proficiency. About one-third (34%) of providers reported high cynicism and 51% reported high emotional exhaustion. Those providers in the top 2 quartiles of EHR use after hours had 4.78 (95% confidence interval [CI], 1.1-20.1;  $P=0.04$ ) and 12.52 (95% CI, 2.6-61;  $P=0.002$ ) greater odds of high exhaustions. Similarly, clinicians in the top quartile of message volume had 6.17 greater odds of high exhaustion (95% CI, 1.1-41;  $P = .04$ ).

No objective measures were associated with high cynicism. These results are important since they fill an important gap of correlating self-reported EHR use with widely reported objective EHR metrics of use, as well as correlating objective measures of EHR use with components of burnout. Since prior studies have correlated self-reported EHR use time with burnout, there was a need to validate this correlation with objective measures of use time, particularly as interventions are designed to address EHR burden. Interestingly, after hours EHR use was associated with exhaustion, but not cynicism, suggesting that providers were feeling overwhelmed by work but did not develop cynicism as a response. The authors were hopeful that interventions designed to reduce the EHR burden (i.e., after hours work and message volume) could help reduce these providers' exhaustion. It is possible, however, that providers who have developed cynicism are those who have potentially inappropriately cut short their EHR use as a coping mechanism. Further studies are needed to further illuminate the relationship between cynicism and EHR burden in order to design effective interventions.

Brewer LC, Fortuna KL, Jones C, Walker R, Hayes SN, Patten CA, Cooper LA

**Back to the future: Achieving health equity through health informatics and digital health**

*JMIR Mhealth Uhealth* 2020;8(1):e14512

This paper called attention to the serious issues of health equity and its systemic effects and unintended consequences. These issues were brought to the fore during the pandemic, with some technology advances and uses creating further inequities that need to be addressed. The authors presented a set of principles and two example projects, contextually tailored, which aimed to address health inequities by community involvement, co-design approaches, and strategically merged health services research with community-based participatory research for innovation and development. The primary contribution of the paper was a set of resilience strategies that could be supported by health information technology in future work. The

first project, FAITH! (Fostering African American Improvement in Total Health), described an intervention based on mHealth and face to face church-based health education and social support, especially directed at cardiovascular disease which has double the mortality rate in African Americans compared to whites, and a higher incidence. They used iterative formative design processes with a team of clinicians, technologists, behavioral and social scientists, and made use of patient preferences, such as using spiritual verses and messaging. This led to high apps ratings and acceptability, usability, and satisfaction. Blood pressure, diet, and physical activity all significantly improved, and the app had a 98% retention rate, remarkable for a health app. Tailored visual study results were also fed back to the community. The second project, Peertech CBPR partnership, addressed premature mortality in people with serious mental illness (SMI) such as bipolar disorder, major depressive disorder, and schizophrenia. The partnering academic-community team identified this as a major health disparity. This led to co-creation of a smartphone app: Peer- and Technology-Supported Self-Management Training (PeerTECH), which carried out simultaneous management of mental and chronic health conditions in patients aged over 60. The approach consisted of equal partnership between patients, certified peer specialists (CPS), leaders, and scientists from idea conception, defining of research questions, intervention development and usability testing extending to dissemination. Results included statistically significant psychiatric self-management (on the Illness Management and Recovery Scale (IMRS) ( $p<0.001$ ) and improvements in medical self-management, hope, quality of life, and empowerment. The authors summarized best practices for strategic design and implementation of digital health interventions for the marginalized:

1. Increase recruitment and retention of diverse populations throughout R&D; assess differential responses/outcomes of technologies; mitigate preferential access;
2. Leverage established stakeholders and trusted social networks;
3. Understand the social context of potential

end users and populations;

4. Integrate community engagement through user-centered design or participatory design;
5. Gain an understanding of community partner technology infrastructure for capacity building to support and strengthen community-based health informatics interventions; and
6. Plan the appropriate amount of time and resources to devote to community engagement processes for intervention development and sustainability.

The importance of this paper is in addressing a current timely issue, with an unprecedented degree of success in retention and disease control effectiveness, serving as an example of how to conduct such design successfully.

**Reading Turchioe M, Grossman LV, Myers AC, Baik D, Goyal P, Masterson Creber RM**

**Visual analogies, not graphs, increase patients' comprehension of changes in their health status**

**J Am Med Inform Assoc 2020;27(5):677-89**

The aim of this paper was to understand which visualizations are most effective in aiding patients to self-monitor their health status, via the use of patient-reported outcome (PRO) measures in hospitalized patients, with a focus on objective comprehension. 40 hospitalized patients were included to compare four visualization conditions: (1) text-only, (2) text plus visual analogy, (3) text plus number line, and (4) text plus line graph assessed objective comprehension using the International Organization for Standardization protocol. Secondary outcomes included response times, preferences, risk perceptions, and behavioral intentions. Sixty-three percent correctly comprehended the text-only condition and 60% comprehended the line graph condition, compared with

83% for the visual analogy and 70% for the number line ( $P < 0.05$ ) conditions. The results supported using visual analogies rather than text to display longitudinal PROs but the authors cautioned against relying on graphs (known high prevalence of inadequate graph literacy). Discrepancies between comprehension and preferences suggested factors other than comprehension influenced preferences. Future researchers should assess comprehension rather than preferences to guide presentation decisions. This paper was the best example of a continued emerging emphasis on visualization for patients and clinicians to improve comprehension and workflow.

**Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, Janda M, Lallas A, Longo C, Josep Malvehy J, Paoli J, Puig S, Rosendahl C, Soyer HP, Zalaudek I, Kittler H**

**Human-computer collaboration for skin cancer recognition**

**Nat Med 2020;26(8):1229-34**

Image-based artificial intelligence (AI) has the potential to improve visual diagnostic accuracy in healthcare. Recent studies in dermatology have shown that the accuracy of AI for identifying skin lesions was equivalent to or better than human experts in controlled experimental studies; further, human-AI cooperation can improve accuracy even more, but studies have not determined the best ways to incorporate AI into clinical workflows for improving diagnostic accuracy in real healthcare settings. This study explored the impact of different representations of AI based clinical decision support for identifying skin lesions on the accuracy of clinical diagnoses. The first representation presented the probability that the lesion was in one of seven diagnostic categories: four malignant and three benign. The second collapsed the seven categories into two: malignant and benign. The third used im-

age retrieval to present similar images with known diagnoses and the fourth presented previously collected probabilities for each of the seven diagnoses as determined by 511 human raters. A total of 302 human raters from 41 countries diagnosed batches of images both without decision support and then with one type of decision support. The results showed that the first representation improved the accuracy of human raters from 63.6% to 77.0% (increase of 13.3%, 95% CI 11.5% to 15.2%;  $P = 4.9 \times 10^{-35}$ , two-sided paired t-test,  $t = 14.5$ , d.f. = 301;  $n = 302$  raters). No improvement was observed for the second or third representation, with some improvement for the fourth. They observed an inverse relationship between the net gain accuracy increase due to AI and rater experience; more inexperienced raters more frequently changed their initial ratings after viewing the AI results than experts did. After establishing a positive impact of AI on diagnostic accuracy, the study tested the impact of errors in AI on diagnostic accuracy. These errors could be due to models classifying images outside the domain of images used to train the model, something which is a big concern for AI researchers. They intentionally showed raters incorrect multiclass probabilities, which decreased the providers' accuracy back to where it was without AI. The paper was very thorough and presented several other analyses such as identifying two different diagnostic categories that benefitted most of clinical decision support based on AI, measuring the tendency of raters to change their minds, and determining how background features affected both the AI and the human rater's classification. Based on these multiple analyses, this paper justifiably advocates for studying human computer collaboration in real world clinic settings when evaluating the performance of AI for any and all healthcare applications, not just those in dermatology.