Received: 17 August 2024 | Accepted: 20 September 2024

DOI: 10.1111/cts.70044

ARTICLE

Pharmacogenomic augmented machine learning in electronic health record alerts: A health system-wide usability survey of clinicians

Caroline W. Grant[1] | **Jean Marrero-Polanco**[1] | **Jeremiah B. Joyce**[2] | **Barbara Barry**[3,4] | **Ashley Stillwell**[5,6] | **Kellie Kruger**[5] | **Therese Anderson**[7] | **Heather Talley**[8] | **Mary Hedges**[9] | **Jose Valery**[9] | **Richard White**[9] | **Richard R. Sharp**[10] | **Paul E. Croarkin**[2] | **Liselotte N. Dyrbye**[11] | **William V. Bobo**[12] | **Arjun P. Athreya**[1,2]

[1]Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, Minnesota, USA

[2]Department of Psychiatry and Psychology, Mayo Clinic, Rochester, Minnesota, USA

[3]Division of Health Care Delivery Research, Mayo Clinic, Rochester, Minnesota, USA

[4]Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, Minnesota, USA

[5]Department of Family Medicine, Mayo Clinic, Scottsdale, Arizona, USA

[6]Department of Psychiatry and Psychology, Mayo Clinic, Scottsdale, Arizona, USA

[7]Department of Family Medicine, Mayo Clinic, Jacksonville, Florida, USA

[8]Department of Family Medicine, Mayo Clinic, Rochester, Minnesota, USA

[9]Department of Internal Medicine, Mayo Clinic, Jacksonville, Florida, USA

[10]Biomedical Ethics Research Program, Mayo Clinic, Rochester, Minnesota, USA

[11]Department of Medicine, University of Colorado School of Medicine, Denver, Colorado, USA

[12]Department of Behavioral Science & Social Medicine, Florida State University College of Medicine, Tallahassee, Florida, USA

Correspondence
Arjun P. Athreya, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, 200 First St. SW, Rochester, MN 55902, USA.
Email: athreya.arjun@mayo.edu

William V. Bobo, Department of Behavioral Science & Social Medicine, Florida State University, 1115 West Call Street, Tallahassee, FL 32306, USA.
Email: william.bobo@med.fsu.edu

Liselotte N. Dyrbye, Mail Stop C290, Fitzsimons Bldg., 13001 E 17th Pl. Rm #E1347, Aurora, CO 80045, USA.
Email: liselotte.dyrbye@cuanschutz.edu

Abstract

Pharmacogenomic (PGx) biomarkers integrated using machine learning can be embedded within the electronic health record (EHR) to provide clinicians with individualized predictions of drug treatment outcomes. Currently, however, drug alerts in the EHR are largely generic (not patient-specific) and contribute to increased clinician stress and burnout. Improving the usability of PGx alerts is an urgent need. Therefore, this work aimed to identify principles for optimal PGx alert design through a health-system-wide, mixed-methods study. Clinicians representing multiple practices and care settings ($N = 1062$) in urban, rural, and underserved regions were invited to complete an electronic survey comparing the usability of three drug alerts for citalopram, as a case study. Alert 1 contained a generic warning of pharmacogenomic effects on citalopram metabolism. Alerts

Caroline W. Grant and Jean Marrero-Polanco contributed equally.

Funding information
National Science Foundation, Grant/Award Number: IIS-2041339; National Institutes of Health Predoctoral T32 Fellowship, Grant/Award Number: GM144233; Harry C. and Debra A. Stonecipher Predoctoral Fellowship (CWG) at the Mayo Clinic Graduate School of Biomedical Sciences

2 and 3 provided patient-specific predictions of citalopram efficacy with varying depth of information. Primary outcomes included the System's Usability Scale score (0–100 points) of each alert, the perceived impact of each alert on stress and decision-making, and clinicians' suggestions for alert improvement. Secondary outcomes included the assessment of alert preference by clinician age, practice type, and geographic setting. Qualitative information was captured to provide context to quantitative information. The final cohort comprised 305 geographically and clinically diverse clinicians. A simplified, individualized alert (Alert 2) was perceived as beneficial for decision-making and stress compared with a more detailed version (Alert 3) and the generic alert (Alert 1) regardless of age, practice type, or geographic setting. Findings emphasize the need for clinician-guided design of PGx alerts in the era of digital medicine.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

Pharmacogenomic (PGx) alerts in the electronic health record (EHR) are known contributors to clinician stress and burnout. The comparison of the design and content of PGx alerts, and their impact on clinical decision-making and clinician stress, is understudied.

WHAT QUESTION DID THIS STUDY ADDRESS?

This study aimed to identify alert design components which either improve or reduce alert usability and to assess whether PGx alerts augmented with individualized efficacy predictions based upon genomic markers have superior usability compared with generic pharmacogenomic alerts. The study assessed preferences among three PGx alerts for citalopram in an enterprise-wide survey of a wide range of clinicians in varying practices, geographic locations, and with varying levels of experience.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

The generic PGx alert, representative of current non-individualized citalopram alerts in electronic medical records, scored in the "F" range on usability. An alternate individualized alert providing patient-specific predictions of drug efficacy was preferred. Preference may vary based on the clinician specialty and experience. Clinician suggestions for alert improvement were compiled.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

This study elucidates PGx alert design principles which may augment clinical decision-making and reduce clinician stress in the era of predictive analytics.

INTRODUCTION

Electronic health records (EHRs) increasingly contain pharmacogenomic (PGx) alerts that must be addressed by clinicians when prescribing medications.[1,2] PGx alerts warn providers of genetic variation which may impact the safety or efficacy of a prescribed drug.[3,4] Such alerts are therefore potentially indispensable tools for facilitating safe prescribing.[5] However, these alerts are also known contributors to stress and burnout,[6–8] and due to their high volume, poor design, and perceived lack of clinical relevance, they are often dismissed without action.[9] Accordingly, the EHR has been widely criticized as a barrier, rather than an aide, to care.[10] To improve EHR usability, reduce EHR-driven clinician stress, and facilitate precision treatment at the point of care, user-centered research wherein clinicians inform optimal PGx alert design is essential.[11–13]

Existing work on PGx alert design suggests that alerts should (i) detail the strength of the recommendation, (ii) share only actionable information, (iii) indicate references such as scientific literature, and (iv) design intuitive icons

for knowledge resources, among additional recommendations.[14,15] However, prior work has been limited to single centers, single specialties, or small focus groups.[12,16–22] Importantly, prior work suggests that the depth of information contained in alerts, when such information is needed, and by whom it is needed, may vary across clinical settings.[3] Thus, there is an evidence gap for comparison of PGx alert preferences across diverse geographic and clinical settings.[17,21,23,24] Such an assessment can quantify the extent to which current generic drug alert designs are broadly useful and where alternative designs and content may benefit clinicians and patients. Furthermore, machine learning/artificial intelligence (ML/AI) methods are being utilized to incorporate broader sets of PGx biomarkers to individualize predictions of drug response. In the digital medicine era, it is expected that drug alerts will incorporate these predictions for a given drug if the PGx results are available for a given patient. Therefore, it is important to study the clinician perspectives of EHR drug alerts that combine both PGx and ML/AI methods to convey not only regulatory information on side effects, but also the estimated likelihood of drug response derived from PGx-informed machine learning algorithms.

Addressing these evidence gaps, we conducted a survey to understand clinician preferences among three proposed PGx alerts for citalopram, a commonly prescribed antidepressant. In the survey, the usability of an alert containing generic pharmacogenomic information (representative of currently implemented alerts) was compared with two versions of alerts containing patient-specific drug response profiles (i.e., predicted efficacy based on genomic and clinical information). Quantitative analyses of alert preference were supplemented with qualitative analysis to contextualize preference. Citalopram was selected as a case study as it triggers a large number of PGx alerts in the EHR due to its high prescribing prevalence across multiple specialties,[4] its metabolism by CYP2C19 (subjecting it to multiple potential gene-drug and drug–drug interactions), its dose-dependent effects on cardiac conduction (increasing the risk of prolonging the QT interval and associated ventricular arrhythmias). Indicating an approaching translation to practice in the coming years, several machine learning methods have utilized a broader set of PGx biomarkers (going beyond single genes) and demonstrated improved predictability of response to citalopram and additional antidepressants.[25–30]

This study hypothesized that clinicians would prefer individualized alerts (based on patient genotypes) over a generic alert (non-individualized warning of CYP2C19-based variation in citalopram efficacy or side effects). Alert preference was hypothesized to vary by clinician specialty and age, which are associated with overall EHR satisfaction.[31] Finally, higher usability was hypothesized to correspond to reduced perceived stress, captured by qualitative data. These results, collected from a broad array of healthcare providers, may be used to inform organizational strategies aimed at reducing EHR-related stress and improving patient care.

MATERIALS AND METHODS

Study design and setting

An explanatory mixed-methods study design (Figure 1) was used to obtain feedback from academic and community-based physicians, physician assistants, and advanced practice nurses employed by Mayo Clinic across multiple sites (a health-system-wide study). As an explanatory analysis, qualitative data (electronic text-based responses) were collected to help explain and build upon quantitative findings (System's Usability Scale[32] ratings). Quantitative and qualitative data were collected following the 'Strengthening the Reporting of Observational Studies

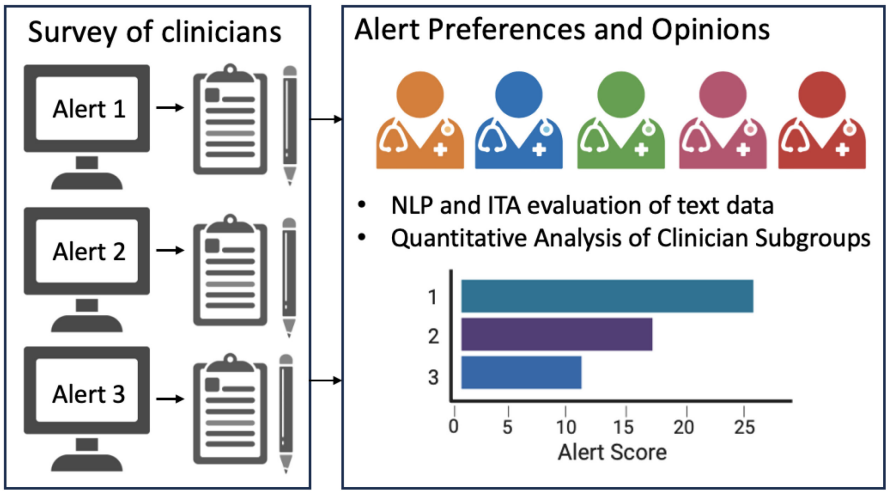

FIGURE 1 Conceptual overview of the study. NLP: Natural language processing; ITA: Inductive thematic analysis.

in Epidemiology' (STROBE)[33] and 'Consolidated criteria for reporting qualitative research' COREQ[34] guidelines, respectively. Invited participants ($N=1062$) included those from clinical sites located in the Upper Midwest ($N=778$), Florida ($N=205$), and Arizona ($N=79$) (see Table S1). The study was approved by the Mayo Clinic Institutional Review Board.

Study population

Administrative records were used to identify clinicians with prescribing privileges in outpatient primary care, hospital settings, or specialty outpatient mental health clinics, who were then invited to participate in the current study via email. Invited clinicians worked in family medicine ($N=681$), internal medicine ($N=167$), psychiatry ($N=147$), and community internal medicine ($N=67$) practices. Informed consent was implied by survey completion.

Survey

The electronic survey was designed by the Mayo Clinic Survey Research Center and delivered via email through Qualtrics (Qualtrics, Provo, UT). It contained a clinical vignette (Figure 2a) that prompted study participants to prescribe citalopram. Afterward, clinicians were presented with a sequence of three PGx alerts (Figure 2b–d). *Alert 1* was a generic warning of potential altered citalopram efficacy or side effects by CYP2C19 variation. This alert is representative of those currently implemented in electronic medical record systems. *Alert 2* was a simple, individualized alert displaying predicted citalopram efficacy as determined by a hypothetical genome-guided AI algorithm. Such an alert represents a potential implementation of existing research-based algorithms which provide individual-level outcome predictions using genomics and additional patient data.[26] *Alert 3* was a more detailed, individualized alert that was similar to Alert 2 but also included a list of top genomic predictors of citalopram efficacy identified in prior research.[25,26,35] For each participant, the generic alert (Alert 1) was presented first. To determine whether AI-enhanced individualized alerts had higher usability than the generic alert, and if there were differences in their level of usability, alerts 2 and 3 were subsequently presented in random order. Comparing each individualized alert against the generic alert enables assessment of whether changing the current alert design and content

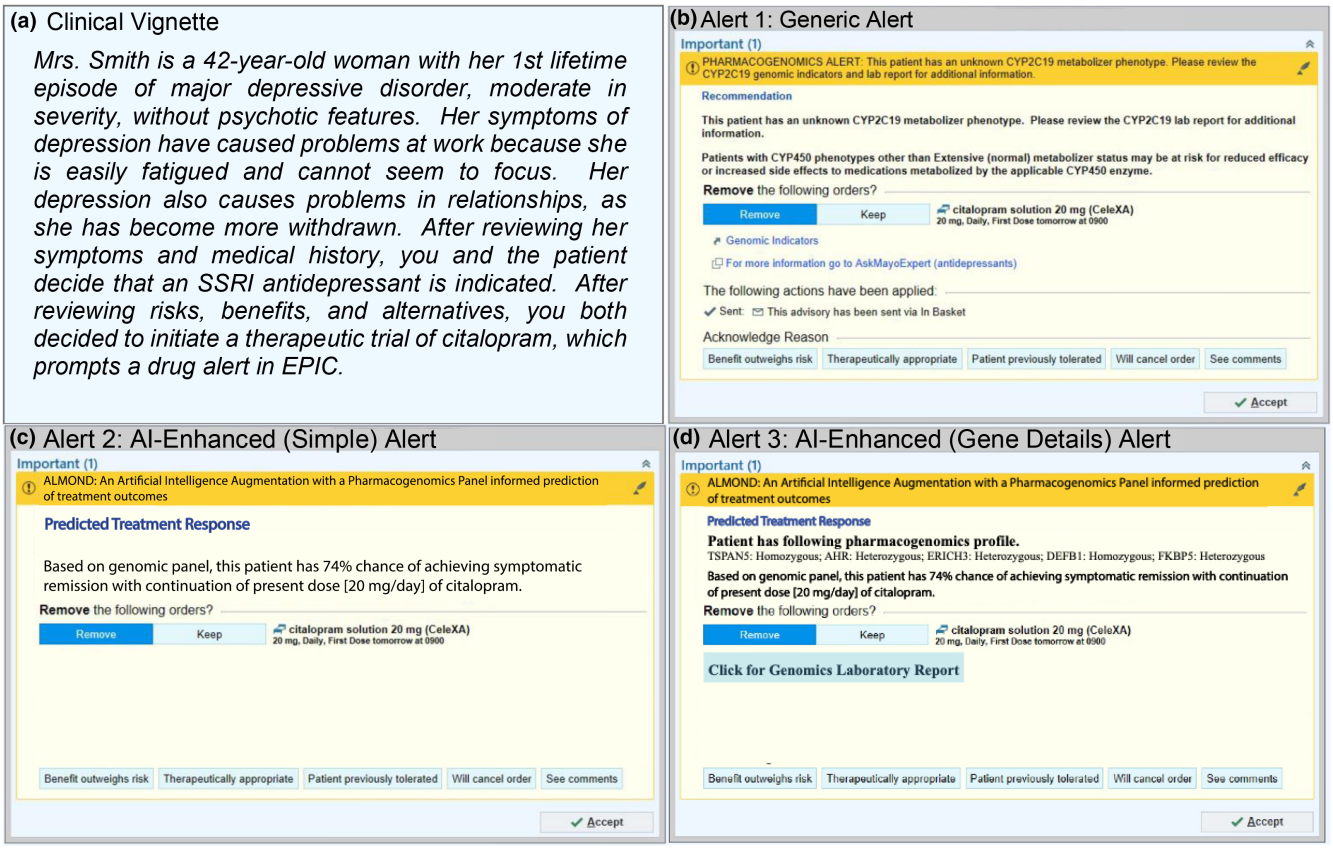

FIGURE 2 Clinical vignette and pharmacogenomic alerts. (a) Clinical vignette, designed by study team clinicians; (b) Generic Alert; (c) AI-Enhanced (simple) Alert; (d) AI-Enhanced (gene details) Alert. Top genomic predictors all represent pharmacodynamic markers.

may benefit clinicians. After viewing each alert, participants were asked to rate the alert's usability with the System Usability Scale (SUS), a validated 10-item questionnaire providing a score (0–100), with higher values indicating greater usability.[36] Participants were also asked to provide free-form text responses collected using an investigated-developed survey about how each drug alert would impact (i) decision to prescribe citalopram and (ii) work-related stress. For example, participants answered: "Please indicate whether Drug Alert 1 would improve or worsen your level of work stress, as compared to having no drug alert" (additional survey questions: Supplementary Methods in Appendix S1). Additional survey items captured age, sex, race, site, patient population characteristics (medically underserved community vs. other), work unit, occupational role, clinical specialty, years in specialty, estimated number of patients cared for per day, and the estimated number of patients presenting with anxiety or depression per day.

Data analysis

Quantitative data

SUS scores, the main measure of user preference, were compared using repeated measures analysis of variance (RM-ANOVA) with alert type (Alert 1, 2, or 3) as a within-subject factor. Post hoc pairwise comparisons were made using *t*-tests with Bonferroni-corrected *p*-values.

As clinician characteristics may contribute to differences in alert preferences,[31] preference was compared across professional and demographic covariates using RM-ANOVA (categorical variables) and linear mixed-effect models (continuous and ordinal variables), modeling SUS score as a function of alert and each individual covariate. Covariates included age, sex, years in practice, employment status (full-time, part-time), site (urban, rural, underserved) unit (e.g., hospital, outpatient), role (e.g., physician, advanced practice nurse), and estimated daily number of patients with depression (Supplementary Methods in Appendix S1).

To test the hypothesis that alert preference depends on both clinical specialty and age, multinomial logistic regression was used to model alert preference as a function of the interaction between specialty and age. Using the resulting regression model, predicted probabilities for alert preference were calculated across the span of observed ages (24–75 years in this dataset) within specialties (family medicine, internal medicine, and psychiatry). These probabilities were plotted for visual comparison of predicted alert preference by age and specialty.

Qualitative data

Inductive thematic analysis (ITA) and natural language processing (NLP) were used to understand clinician perceptions of alert usability. Input text to ITA and NLP analyses included, for each alert individually, clinician responses to questions regarding their (i) decision to prescribe citalopram and (ii) work-related stress (see Supplementary Methods in Appendix S1 for questions). ITA was used to understand content and design factors which contribute to or mitigate stress, given the prevalence of stress and burnout resulting from alerts in the EHR. Given that the code "Suggestions for improvement" arose from ITA analysis, ITA enabled the aggregation of clinician suggestions for alert improvement. Given the potential limitations of ITA as a researcher-dependent process, an NLP-based topic modeling approach was also used as a supporting analysis to determine whether overlapping latent topics could be detected rapidly from the free-form text in a data-driven manner.[37] NLP analyses yielded summarized topics discussed in clinician responses.

Inductive Thematic Analysis (ITA)

ITA was carried out using standard protocol.[38] For each alert, two independent reviewers (CWG & JMP) read clinician responses. Then, reviewers generated thematic codes from the responses, grouped codes into broader themes, and labeled themes by sentiment. This process was conducted iteratively until final codes and themes were established by each independent reviewer for each alert. Then, reviewers compared codes and themes and arrived at a consensus with the help of a third reviewer (APA) to resolve any conflicts. Suggestions for improving alerts were summarized by frequency.

NLP Topic modeling and BART summarization results

The number of latent topics in survey text-based responses to decision-making and stress-related questions for each alert was computationally determined using the perplexity metric.[37] Then, topic modeling, an unsupervised NLP approach for identifying latent topics within a text corpus, was performed using Latent Dirichlet Allocation (LDA).[37] Documents relating to each topic were identified based on

the LDA posterior probabilities of words per topic. For concise interpretation of LDA results, the top documents with the highest posterior probabilities ($N=5$) for each topic were summarized with abstractive summarization using a bidirectional and autoregression transformer (BART) model (Supplementary Methods in Appendix S1). Agreement between NLP-generated topics and ITA-generated themes was assessed (Supplementary Methods in Appendix S1).

Analyses were implemented in R v4.0.3[39] using RStudio v2022.02.3+492[40] and Python v3.12[41] using PyCharm Build #PC-223.8214.51.

RESULTS

Study population

Thirty-three percent ($N=351$) of invited clinicians responded, 305 of whom provided complete survey data and were included in analyses (Demographics: Table 1; Sample inclusion: Figure S1). Most responders worked in outpatient or ambulatory non-mental health specialties (76%). Clinicians from rural and medically underserved areas[42,43] comprised 20% of the sample. The majority (63%) practiced in the department of family medicine.

Quantitative study of alert preference

Mean SUS scores for Alerts 1, 2, and 3 were 56 (standard error (se): 1.24), 76 (se 1.11), and 62 (se 1.28), respectively ($F_{1.9,575.5}=117.1$, $p<0.0001$). The simple AI-enhanced alert (*Alert 2*) scored significantly better than the gene-detailed AI-enhanced alert (b) ($t=12.1$; Bonferroni-corrected $p<0.0001$), which was significantly better than the generic alert (*Alert 1*) ($t=3.91$; Bonferroni-corrected $p=3.5e\text{-}4$) (Figure 3a). An inverse relationship was observed between SUS score and years in specialty ($p=0.023$), age ($p=0.045$), and the estimated number of patients with mood or anxiety disorders evaluated per clinic day ($p=0.009$). Alert preferences did not differ significantly by sex, race, site, population census site characteristics, unit, occupational role, or specialty (Tables S2 and S3).

When jointly considering specialty and age, alert preference decreased for the generic alert (Alert 1) with increasing age across specialties (Figure 3b–d). Internists increasingly preferred the simple AI-enhanced alert (Alert 2) with age, while family medicine clinicians increasingly preferred the detailed AI-enhanced alert (Alert 3) with age. Psychiatrists remained relatively consistent in predicted preference for the two AI-enhanced alerts (Table S4).

Qualitative study of alert preference

Inductive Thematic Analysis (ITA)

ITA revealed 8, 8, and 7 themes for the generic (Alert 1), AI-enhanced (simple) (Alert 2), and AI-enhanced (detailed) (Alert 3) alerts, respectively (Tables S5–S7). Stress-related codes that emerged (including "Alert increases Stress" and "Alert decreases stress") mapped to themes regarding added workflow burden, knowledge gap, and confidence in prescribing. For all three alerts, clinicians reported both positive and negative impacts on stress. For example, for Alert 2, clinicians reported the following mixed sentiments:

> *More steps and clicks always adds more stress to the information overload of the EMR.*
> Study participant #115, attending physician, family medicine, male age forties, medically underserved community

> *This gives a number to the information which is more helpful, causing less stress.*
> Study participant #266, resident physician, family medicine, female age twenties, medically underserved community

For Alert 3, clinicians also reported both positive and negative influences of the provided genotypes on stress:

> *I do not know all the genotypes expressed in the alert, so I might be a little more stressed given this lack of knowledge.*
> Study participant #229, resident physician, psychiatry, male age thirties, Minnesota

> *Alert 3 reduces stress by increasing the detail which enhances clinical reasoning.*
> Study participant #178, resident physician, internal medicine, male age thirties, Florida

ITA also enabled the aggregation of suggestions for alert improvement. The main suggestion for improving all alerts was to provide alternative pharmacotherapy recommendations and comparisons (Table S8). Clinicians also requested shorter alerts and fewer button clicks for all alerts. For Alerts 2 and 3, clinicians emphasized the need for educational resources for interpreting the results and model inputs. Finally, clinicians requested that alerts be available earlier in the appointment—for example, in the problem list or pharmacy support tabs.

TABLE 1 Demographics.

	N	Percent
Sex (Female/Male)	176/129	58/42%
Age (median [Min, Max])	39 [24,75]	N/A
Race (White)	233	76%
Years in role (median [Min, Max])	6 [0,44]	N/A
Full-time employment status	265	87%
Site		
Urban: Rochester, MN	118	39%
Urban: Jacksonville, FL	63	21%
Rural and Medically Underserved: Mayo Clinic Health System	60	20%
Urban and Medically Underserved: Mayo Clinic Health System	38	20%
Urban: Phoenix/Scottsdale AZ	26	8%
Unit		
Outpatient/ambulatory, non-mental health specialty	241	76%
Hospital, medical and/or surgical	61	20%
Outpatient, mental health specialty	38	13%
Hospital, psychiatric	36	12%
Emergency department	2	1%
Role		
Physician, Attending	148	49%
Physician, Resident/Fellow	77	25%
Advanced Practice Nurse (master's degree)	34	11%
Physician Assistant	25	9%
Doctor of Nursing Practice (DNP)	21	7%
Specialty		
Family Medicine	192	63%
Psychiatry	62	20%
Internal Medicine	41	13%
Other	10	3%
Number of Patients with Depression/Anxiety in a Day (median [Min, Max])		
<1	13	4%
1 to 2	76	25%
3 to 5	168	55%
6 to 10	38	13%
11+	10	3%

Note: Demographics. Full-time employment status is defined as ≥36 h per week. Specific race demographics: East-Asian ($N=17$); South-Asian ($N=11$); Hispanic ($N=11$); Black ($N=6$); Middle Eastern ($N=6$); Other ($N=6$); Prefer not to respond ($N=27$).

Topic modeling and BART summarization results

The optimal number of latent topics in clinician responses determined by perplexity calculations was 7 for the generic alert (Alert 1) and 6 for each AI-enhanced alert (Alerts 2, 3) (Figure S2). Across alerts, NLP topics validated ITA themes (Results, Figures S3–S5). Top documents for each topic, identified in topic modeling, were summarized via the BART algorithm, which provided a simple, comprehensive summarization of topics Figure 4. Five of the seven topics (71%) for the generic alert (Alert 1) had negative sentiments, including interpretation challenges, actionability issues, visual complexity, and premature evidence supporting PGx alerts for citalopram. Participants also expressed mixed sentiments about Alert 1, noting that the alert may have safety benefits but at the cost of increased time-related stress. Example clinician responses capturing these sentiments include:

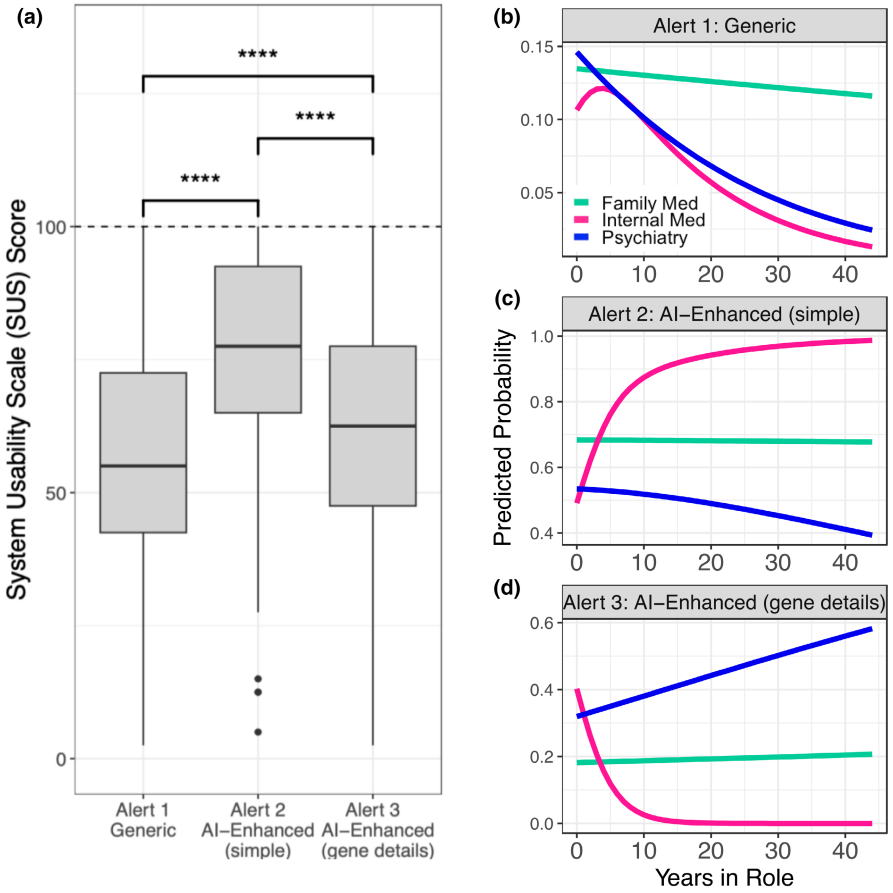

FIGURE 3 Alert scores and alert preference by specialty and age. (a) Bonferroni-corrected significance of post hoc pairwise *t*-tests following repeated measures analysis of variance (ANOVA). ****$P < 0.0001$. (b) Predicted probabilities of alert preference by age and specialty category for Alert 1 (Generic alert), (c) Alert 2 (AI-Enhanced [simple]), and (d) Alert 3 (AI-Enhanced [gene details]). Multinomial logistic regression analyses utilized the generic alert was designated as the referent outcome. Probabilities were modeled based on fit extracted from multinomial logistic regression and applied to a synthetic dataset spanning the observed age range (24–75 years) for each specialty.

In a 20 min visit that is dedicated to ONLY depression I probably leave 1-2 min to actually do the prescription and orders. In a typical visit where we deal with 5-7 different topics, I often have no dedicated time to do prescriptions and have to do it while talking about something else and I can't read any of these pop ups in that context.

Participant #173, attending physician, family medicine, male, age forties, medically underserved community

The reality is I don't know enough about pharmacogenomics to interpret this. I would then have to refer to MTM for consult on best options, which delays treatment. I would likely just switch to another drug, which may pop up a similar alert. At that point I would go away from any SSRI, or just click something to get it to go away and prescribe anyway, thus defeating the purpose of the alert.

Participant #299, attending physician, family medicine, male, age forties, medically underserved community

In contrast, four of the six topics (67%) for the simple AI-enhanced alert (Alert 2) had positive sentiments, including decision-making utility, stress reduction, and user-friendly design:

There is not much fluff in the alert, therefore it is more likely to be read compared to the first one. It also has the information presented in an easy to understand format and the specifics of my patient that are pertinent to medication decision making.

Participant #57, resident/fellow physician, neurology, female age twenties, Florida

This gives me more confidence about what I am prescribing and helps dictate next steps better. It also helps me set expectations with the patient.

Participant #190, resident/fellow physician, family medicine, female, age twenties, medically underserved community

However, clinicians also noted that Alert 2 may discourage citalopram prescriptions in favor of drugs without

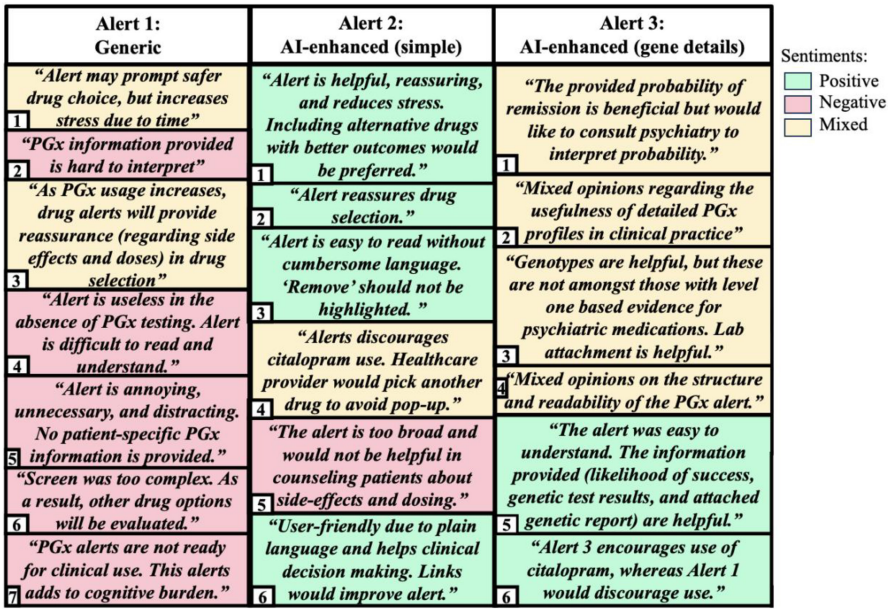

FIGURE 4 NLP summarized topics. Topics were determined from the extraction of top documents ($N = 5$) in topic modeling and summarized using the bidirectional and autoregressive transformers (BART). In abstractive summarization, the output text may contain words and phrases that did not appear in the source text yet succinctly convey the same meaning. BART-summarized text output was reviewed and refined by two independent human interpreters. Color represents topic sentiment as determined independently by two independent human interpreters: Green, positive sentiment; pink, negative sentiment; yellow, mixed sentiment.

pop-ups, lacks utility for side effects and dosing counseling, and may appear too late in the appointment:

> *74% chance seems high likelihood of remission, but compared to what? Is there a different option that is ever higher likelihood? What if it was 24%, and I'm not getting that information until I've already discussed the treatment options. The alert/information support is coming too late to do anything valuable other than make me feel better/more confident in a decision I already made.*
>
> Participant #4, attending physician, internal medicine, female, age forties, Minnesota

> *I'd still prefer to hyperlink to the evidence that made the statement. By suggesting an outcome, you might not get a behavior change if that outcome cannot be substantiated—the inquisitive nature of providers need to be given a way to confirm with evidence.*
>
> Participant #83, attending physician, family medicine, male age fifties, medically underserved community

Lastly, four of the six (67%) topics for Alert 3 (AI-enhanced gene details) had mixed sentiment topics regarding the usefulness of genetic details in clinical practice, structure/readability of the alert, and evidence supporting the listed genomic biomarkers:

> *Same alert with distracting superfluous information for most consumers...takes additional time to digest and this an impediment to patient care.*
>
> Participant #21, attending physician, family medicine, male age forties, Arizona

> *Bringing up all those genes will confuse and frustrate the clinician. If the clinician can't explain it to the patient, that is another frustration. We are taught in med school not to order a lab we cannot explain...I can explain p450 to patient but have no idea what these other genes are.*
>
> Study participant #5, attending physician, psychiatry, male, age fifties, Minnesota

DISCUSSION

This study demonstrated that clinicians found a simplified AI-enhanced alert (Alert 2) more usable than a lengthier version containing details of pharmacogenomic predictors (Alert 3). As expected, both AI-enhanced alerts were preferred over a generic alert (Alert 1), a non-individualized warning of CYP2C19-based

variation in citalopram efficacy or side effects. Greater usability coincided with clinician reports that the predicted efficacy percentage was helpful in Alert 2 (the preferred alert) for improving perceived stress compared with the generic alert. The qualitative analyses in this study added important context to the numerical survey responses. Taken together, these findings indicate that alerts which provide a quantitative likelihood of drug response may improve EHR-driven clinician stress in the era of digital medicine.

This study supplemented ITA with NLP analysis of qualitative, free-text survey responses that enabled rapid identification of major themes in a data-driven manner. This helped support and validate codes and themes derived from human reviewers, which are subject to human interpretation. NLP-derived topics overlapped with ITA-derived themes, indicating that relatively rapid NLP techniques serve as useful tools for analyzing survey responses. NLP found that sentiments regarding Alert 2 were largely positive and centered around its patient-specificity, high actionability, and user-friendliness. NLP also uncovered perceived shortfalls in Alert 2, including inadequate information on inputs (genomic, clinical variables), insufficient guidance on predicting adverse events, appearance too late in the appointment, and a necessity for providing alternate antidepressant recommendations. This aligns with current wisdom that physicians strongly resist suggestions to avoid actions when alternatives are not offered, even when the actions are counterproductive.[44] Additional research is needed, as most existing algorithms are derived from pools of patients taking one of several antidepressants, and there is a scarcity of robust models predicting side effects.[25–27,45]

The relatively lower overall preference for Alert 3 (the gene details alert), along with several participants expressing the need for a pharmacogenomics consult to interpret the alert, highlight the potential need for tailored education or easy-to-navigate consultative features. This will be especially critical as the complexity of genomic information increases. In one effort to enhance education surrounding new gene-drug associations, organizations such as the Clinical Pharmacogenetic Implementation Consortium (CPIC) evaluate the quality and actionability of gene-drug associations.[46] CPIC also provides a corresponding language for pharmacogenomic alerts, for example, suggesting alternative dosing dependent on CYP2C19 phenotypes.[47] Clinician responses demonstrate greater familiarity with CYP450 genes compared with pharmacodynamic genes in alerts 2 and 3. Developing and implementing language in CPIC for multicombinatorial algorithms which include pharmacodynamic markers would be one path toward facilitating ongoing education around novel gene-drug associations and the basics of AI-based tools. Such efforts are imperative, as numerous participants reported that they would simply avoid prescribing drugs with PGx pop-ups due to limited education. This behavior change may be attributed to a discomfort utilizing information that is incompletely understood in the context of limited appointment time, as supported by a separate investigation which showed that providing PGx test results to clinicians who have limited PGx experience reduces the prescription of medications with predicted drug–gene interactions, regardless of whether patients possess the risk genotypes.[48] Several government and academic institutions, as well as consortia (e.g., PharmGKB) offer online PGx education courses.

While most clinicians preferred a simplified alert, however, some favored the inclusion of genetic details. Preference may be specific to age and specialty. With higher age, internists increasingly preferred the simplified version, while later career family medicine clinicians reported a preference for the detailed alert. Psychiatrists, in contrast, showed higher preference across all ages for the detailed alert. This may be due to increased awareness by psychiatrists of pharmacogenomic variations in the efficacy of antidepressants, which they regularly prescribe. Alternatively, it may be an unfamiliarity with the listed pharmacogenomic markers that psychiatrist experts may be less willing to trust than clinicians with other expertise. Future work should aim to clarify these associations. Likely, no singular alert may satisfy the span of divergent clinician needs across healthcare systems and clinical domains.[49] Rather, as medicine becomes increasingly intertwined with analytics, individualized medicine efforts must consider both patient- and clinician-specific characteristics when determining the best implementation of algorithm-based alerts.[50]

This study has limitations. Sample size limitations prevented further analysis of the divergence in alert preference across specialties by age. Although ours was a multisite study, all sites were part of Mayo Clinic, and therefore, may not be fully representative of other healthcare settings. Confirmatory evaluation is needed with validation beyond a single healthcare enterprise. Additionally, only citalopram PGx alerts were evaluated presently. Future studies may extend the investigation to additional drugs with different indications and side effect profiles. Likely, these will replicate themes presently uncovered, such as the desirability of concise alerts, with links to Supporting information, and benefits of patient-specific genetic profiles. Future studies can move beyond clinician sentiments to evaluate how each alert impacts clinician performance (e.g., accuracy of prescribing decisions or correct interpretation of information), which were not presently investigated. Additionally, NLP and

ITA outputs were interpreted by two reviewers; universal agreement in interpretations cannot be guaranteed. Finally, while this work investigated utility of PGx alerts at the point of prescribing, it has been suggested that clinicians be alerted if a PGx test shows an actionable variant, prompting medication review and corrective action, if necessary, before the start of a clinical encounter.[3] The total volume of drug alerts encountered in clinical practice may be expected to be composed of alerts that are both accurate and applicable (or perhaps even appropriately individualized) to a given patient, which may then result in cancellation or modification of the order. On the other hand, appropriate alerts are often mixed with a high number of alerts with limited to no applicability to the patient being seen at the point of care. Prior work has shown that fewer alerts encountered per day were associated with increased alert salience, possibly as a result of lower alert fatigue, a higher concentration of appropriate alerts, or both.[51] Our work sought instead to identify a quantum of information provided in citalopram drug alerts deemed useful by clinicians—an important design consideration that may bear on the risk of alert fatigue. However, this work does not address possible associations of the cognitive/manual steps needed for corrective actions when such alerts are encountered with the risk of alert burden, nor does it address the potential but foreseeable benefits of specifying individualized and actionable alternatives coupled with an efficient interface for order modification.

In conclusion, this study revealed that clinicians prefer a simplified, patient-specific alert with optional links to Supporting information among three genome-guided alerts for citalopram. Prior work established that the optimal PGx alert design contains only actionable information, and the present study builds upon this principle to demonstrate that patient-specific probabilities of drug efficacy and additional highlighted suggestions are ways to improve actionability. Continued engagement with clinical end-users and enhanced clinician education on emerging EHR-based AI tools is imperative to their successful integration into clinical workflows. Pursuant to this goal, NLP-enhanced analytics may become indispensable methods to efficiently analyze qualitative information from digital surveys of healthcare providers. In the era of digital medicine, clinicians will be increasingly exposed to electronic drug alerts to facilitate safe prescribing, which may also escalate the risks of alert fatigue and burnout. Alert design must be carefully considered such that alerts augment, rather than hinder, clinical care. Accordingly, design preferences elucidated presently should be considered as part of organizational strategies to reduce clinician EHR-related stress and to improve patient care.

AUTHOR CONTRIBUTIONS

C.W.G., J.M.-P., and J.B.J. wrote the manuscript. A.P.A., L.N.D., and W.V.B. designed the research. A.S., B.B. K.K., T.A., H.T., M.H., J.V., R.W., P.E.C., L.N.D., A.P.A., and W.V.B. performed the research. C.W.G., J.M.P., and J.B.J. analyzed the data. R.R.S. contributed analytical tools.

ACKNOWLEDGMENTS

The study team would like to acknowledge the staff of the Mayo Clinic Survey Research Center for the development, launch, and data management of the survey presented in this work.

FUNDING INFORMATION

This work is supported in part by a National Science Foundation award IIS-2041339 (PI – APA, LND, WVB, and RS), National Institutes of Health Predoctoral T32 Fellowship GM144233 (JMP), and Harry C. and Debra A. Stonecipher Predoctoral Fellowship (CWG) at the Mayo Clinic Graduate School of Biomedical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, National Science Foundation, or other funding sources.

CONFLICT OF INTEREST STATEMENT

Dr Bobo has received grant or research support from the Mayo Clinic, NIH, National Science Foundation, the Watzinger Foundation, and the Myocarditis Foundation. He has contributed chapters to UpToDate concerning the pharmacological management of bipolar depression. Dr Croarkin has received grant or research support from the Mayo Clinic, NIH, National Science Foundation, Brain and Behavior Foundation, and Pfizer, Inc. He has received a research grant for a multicenter trial from Neuronetics, Inc. He has received in-kind equipment support from Neuronetics, Inc. and Magventure, Inc. for investigator-initiated studies funded by NIH. He received in-kind support (supplies and genotyping) from AssureRX for an investigator-initiated study. He has served on the advisory boards for Engrail Therapeutic, Inc. and Sunovion, Inc. He served as a consultant to Meta Platforms, Inc. Dr Dyrbye has received royalties from CWS, Inc., for the Well-being Index licensed by the Mayo Clinic outside the submitted work. Dr White has received grant or research support from the Mayo Clinic, NIH, Pfizer, Inc., and PCORI. No other disclosures were reported. Dr Barry receives research support from the Mayo Clinic, Anumana, Inc., the US FDA, and the NSF. All other authors declared no competing interests in this work.

DATA AVAILABILITY STATEMENT

The code is available on request. Data requests can be made to Mayo Clinic Ventures and will be reviewed per Mayo Clinic data management policies.

PREVIOUS PRESENTATION

None.

ORCID

Caroline W. Grant https://orcid.org/0000-0001-9342-2138
Jean Marrero-Polanco https://orcid.org/0000-0003-2232-0788
Jeremiah B. Joyce https://orcid.org/0000-0002-4367-6380
Barbara Barry https://orcid.org/0000-0002-0755-6939
Ashley Stillwell https://orcid.org/0000-0003-0409-8618
Therese Anderson https://orcid.org/0000-0003-3228-0388
Mary Hedges https://orcid.org/0000-0002-3487-916X
Jose Valery https://orcid.org/0000-0003-2671-7854
Richard White https://orcid.org/0000-0002-1133-7819
Richard R. Sharp https://orcid.org/0000-0001-5441-2084
Paul E. Croarkin https://orcid.org/0000-0001-6843-6503
Liselotte N. Dyrbye https://orcid.org/0000-0002-7820-704X
William V. Bobo https://orcid.org/0000-0002-5674-3079
Arjun P. Athreya https://orcid.org/0000-0001-9764-1768

REFERENCES

1. Jankovic I, Chen JH. Clinical decision support and implications for the clinician burnout crisis. *Yearb Med Inform*. 2020;29:145-154. doi:10.1055/s-0040-1701986
2. Chaparro JD, Beus JM, Dziorny AC, et al. Clinical decision support stewardship: best practices and techniques to monitor and improve interruptive alerts. *Appl Clin Inform*. 2022;13:560-568. doi:10.1055/s-0042-1748856
3. Carter JL, Critchlow J, Jackson S, et al. Pharmacogenomic alerts: developing guidance for use by healthcare professionals. *Br J Clin Pharmacol*. 2022;88:3201-3210. doi:10.1111/bcp.15234
4. Wang L, Scherer SE, Bielinski SJ, et al. Implementation of preemptive DNA sequence-based pharmacogenomics testing across a large academic medical center: the Mayo-Baylor RIGHT 10K Study. *Genet Med*. 2022;24:1062-1072. doi:10.1016/j.gim.2022.01.022
5. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med*. 2020;3:17. doi:10.1038/s41746-020-0221-y
6. Gregory ME, Russo E, Singh H. Electronic health record alert-related workload as a predictor of burnout in primary care providers. *Appl Clin Inform*. 2017;8:686-697. doi:10.4338/ACI-2017-01-RA-0003
7. Gardner RL, Cooper E, Haskell J, et al. Physician stress and burnout: the impact of health information technology. *J Am Med Inform Assoc*. 2019;26:106-114. doi:10.1093/jamia/ocy145
8. Co Z, Holmgren AJ, Classen DC, et al. The tradeoffs between safety and alert fatigue: data from a national evaluation of hospital medication-related clinical decision support. *J Am Med Inform Assoc*. 2020;27:1252-1258. doi:10.1093/jamia/ocaa098
9. Carli D, Fahrni G, Bonnabry P, Lovis C. Quality of decision support in computerized provider order entry: systematic literature review. *JMIR Med Inform*. 2018;6:e3. doi:10.2196/medinform.7170
10. Melnick ER, Sinsky CA, Dyrbye LN, et al. Association of perceived electronic health record usability with patient interactions and work-life integration among US physicians. *JAMA Netw Open*. 2020;3:e207374. doi:10.1001/jamanetworkopen.2020.7374
11. Nguyen KA, Patel H, Haggstrom DA, Zillich AJ, Imperiale TF, Russ AL. Utilizing a user-centered approach to develop and assess pharmacogenomic clinical decision support for thiopurine methyltransferase. *BMC Med Inform Decis Mak*. 2019;19:194. doi:10.1186/s12911-019-0919-4
12. St Sauver JL et al. Integrating pharmacogenomics into clinical practice: promise vs reality. *Am J Med*. 2016;129:1093-1099.e1091. doi:10.1016/j.amjmed.2016.04.009
13. Muhiyaddin R, Elfadl A, Mohamed E, et al. Electronic health records and physician burnout: a scoping review. *Stud Health Technol Inform*. 2022;289:481-484. doi:10.3233/SHTI210962
14. Khelifi M, Tarczy-Hornoch P, Devine EB, Pratt W. Design recommendations for pharmacogenomics clinical decision support systems. *AMIA Jt Summits Transl Sci Proc*. 2017;2017:237-246.
15. Tutty MA, Carlasare LE, Lloyd S, Sinsky CA. The complex case of EHRs: examining the factors impacting the EHR user experience. *J Am Med Inform Assoc*. 2019;26:673-677. doi:10.1093/jamia/ocz021
16. Nishimura AA, Shirts BH, Salama J, Smith JW, Devine B, Tarczy-Hornoch P. Physician perspectives of CYP2C19 and clopidogrel drug-gene interaction active clinical decision support alerts. *Int J Med Inform*. 2016;86:117-125. doi:10.1016/j.ijmedinf.2015.11.004
17. Devine EB, Lee CJ, Overby CL, et al. Usability evaluation of pharmacogenomics clinical decision support aids and clinical knowledge resources in a computerized provider order entry system: a mixed methods approach. *Int J Med Inform*. 2014;83:473-483. doi:10.1016/j.ijmedinf.2014.04.008
18. Caraballo PJ, Hodge LS, Bielinski SJ, et al. Multidisciplinary model to implement pharmacogenomics at the point of care. *Genet Med*. 2017;19:421-429. doi:10.1038/gim.2016.120
19. Herr TM, Peterson JF, Rasmussen LV, Caraballo PJ, Peissig PL, Starren JB. Pharmacogenomic clinical decision support design and multi-site process outcomes analysis in the eMERGE network. *J Am Med Inform Assoc*. 2019;26:143-148. doi:10.1093/jamia/ocy156
20. Rosenman MB, Decker B, Levy KD, Holmes AM, Pratt VM, Eadon MT. Lessons learned when introducing pharmacogenomic panel testing into clinical practice. *Value Health*. 2017;20:54-59. doi:10.1016/j.jval.2016.08.727
21. Stanek EJ, Sanders CL, Taber KAJ, et al. Adoption of pharmacogenomic testing by US physicians: results of a nationwide survey. *Clin Pharmacol Ther*. 2012;91:450-458. doi:10.1038/clpt.2011.306
22. Goodspeed A, Kostman N, Kriete TE, et al. Leveraging the utility of pharmacogenomics in psychiatry through clinical decision support: a focus group study. *Ann Gen Psychiatry*. 2019;18:13. doi:10.1186/s12991-019-0237-3
23. Johansen Taber KA, Dickinson BD. Pharmacogenomic knowledge gaps and educational resource needs among physicians in selected specialties. *Pharmgenomics Pers Med*. 2014;7:145-162. doi:10.2147/PGPM.S63715
24. Frye MA, Nemeroff CB. Pharmacogenomic testing for antidepressant treatment selection: lessons learned and

roadmap forward. *Neuropsychopharmacology*. 2024;49:282-284. doi:10.1038/s41386-023-01667-4
25. Athreya A, Iyer R, Neavin D, et al. Augmentation of physician assessments with multi-omics enhances predictability of drug response: a case study of major depressive disorder. *IEEE Comput Intell Mag*. 2018;13:20-31. doi:10.1109/MCI.2018.2840660
26. Athreya AP, Neavin D, Carrillo-Roa T, et al. Pharmacogenomics-driven prediction of antidepressant treatment outcomes: a machine-learning approach with multi-trial replication. *Clin Pharmacol Ther*. 2019;106:855-865. doi:10.1002/cpt.1482
27. Joyce JB, Grant CW, Liu D, et al. Multi-omics driven predictions of response to acute phase combination antidepressant therapy: a machine learning approach with cross-trial replication. *Transl Psychiatry*. 2021;11:513. doi:10.1038/s41398-021-01632-z
28. Shelton RC, Parikh SV, Law RA, et al. Combinatorial pharmacogenomic algorithm is predictive of citalopram and escitalopram metabolism in patients with major depressive disorder. *Psychiatry Res*. 2020;290:113017. doi:10.1016/j.psychres.2020.113017
29. Chanfreau-Coffinier C, Hull LE, Lynch JA, et al. Projected prevalence of actionable pharmacogenetic variants and level a drugs prescribed among US veterans health administration pharmacy users. *JAMA Netw Open*. 2019;2:e195345. doi:10.1001/jamanetworkopen.2019.5345
30. US Food and Drug Administration. *FDA Drug Safety Communication: Abnormal heart rhythms associated with high doses of Celexa (citalopram hydrobromide)*. 2017. https://www.fda.gov/drugs/drug-safety-and-availability/fda-drug-safety-communication-abnormal-heart-rhythms-associated-high-doses-celexa-citalopram.
31. Menachemi N, Powers TL, Brooks RG. Physician and practice characteristics associated with longitudinal increases in electronic health records adoption. *J Healthc Manag*. 2011;56:183-197; discussion 197-188, 198.
32. Kortum PT, Bangor A. Usability ratings for everyday products measured with the system usability scale. *Int J Human Computer Int*. 2013;29:67-76. doi:10.1080/10447318.2012.681221
33. Uhlig K, Menon V, Schmid CH. Recommendations for reporting of clinical research studies. *Am J Kidney Dis*. 2007;49:3-7. doi:10.1053/j.ajkd.2006.10.012
34. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care*. 2007;19:349-357. doi:10.1093/intqhc/mzm042
35. Athreya AP, Iyer R, Wang L, Weinshilboum RM, Bobo WV. Integration of machine learning and pharmacogenomic biomarkers for predicting response to antidepressant treatment: can computational intelligence be used to augment clinical assessments? *Pharmacogenomics*. 2019;20:983-988. doi:10.2217/pgs-2019-0119
36. Brooke J. Sus: a quick and dirty usability. *Usability Evaluation in Industry*. Vol 189; 1996:189-194.
37. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *J Machine Learn Res*. 2003;3:993-1022.
38. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3:77-101.
39. *R: A Language and Environment for Statistical*. R Foundation for Statistical Computing; 2021.
40. *RStudio: Integrated Development for R RStudio*. PBC; 2022.
41. *Python 3 Reference Manual*. CreateSpace; 2009.
42. Health Resources & Service Administration. *Find Shortage Areas*. https://data.hrsa.gov/tools/shortage-area
43. Health Resources & Service Administration. *Rural Health Grants Eligibility Analyzer*. https://data.hrsa.gov/tools/rural-health
44. Bates DW, Kuperman GJ, Wang S, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. *J Am Med Inform Assoc*. 2003;10:523-530. doi:10.1197/jamia.M1370
45. Iniesta R, Hodgson K, Stahl D, et al. Antidepressant drug-specific prediction of depression treatment outcomes from genetic and clinical variables. *Sci Rep*. 2018;8:5530. doi:10.1038/s41598-018-23584-z
46. Relling MV, Klein TE, Gammal RS, Whirl-Carrillo M, Hoffman JM, Caudle KE. The clinical pharmacogenetics implementation consortium: 10 years later. *Clin Pharmacol Ther*. 2020;107:171-175. doi:10.1002/cpt.1651
47. Bousman CA, Stevenson JM, Ramsey LB, et al. Clinical pharmacogenetics implementation consortium (CPIC) guideline for CYP2D6, CYP2C19, CYP2B6, SLC6A4, and HTR2A genotypes and serotonin reuptake inhibitor antidepressants. *Clin Pharmacol Ther*. 2023;114:51-68. doi:10.1002/cpt.2903
48. Oslin DW, Lynch KG, Shih MC, et al. Effect of pharmacogenomic testing for drug-gene interactions on medication selection and remission of symptoms in major depressive disorder: the PRIME care randomized clinical trial. *JAMA*. 2022;328:151-161. doi:10.1001/jama.2022.9805
49. McCoy AB, Thomas EJ, Krousel-Wood M, Sittig DF. Clinical decision support alert appropriateness: a review and proposal for improvement. *Ochsner J*. 2014;14:195-202.
50. Liu S, Kawamoto K, del Fiol G, et al. The potential for leveraging machine learning to filter medication alerts. *J Am Med Inform Assoc*. 2022;29:891-899. doi:10.1093/jamia/ocab292
51. Dexheimer JW, Kirkendall ES, Kouril M, et al. The effects of medication alerts on prescriber response in a pediatric hospital. *Appl Clin Inform*. 2017;8:491-501. doi:10.4338/ACI-2016-10-RA-0168

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Grant CW, Marrero-Polanco J, Joyce JB, et al. Pharmacogenomic augmented machine learning in electronic health record alerts: A health system-wide usability survey of clinicians. *Clin Transl Sci*. 2024;17:e70044. doi:10.1111/cts.70044