medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Title: User Testing of a Diagnostic Decision Support System with Machine-assisted Chart Review to Facilitate Clinical Genomic Diagnosis

Authors

Alanna Kulchak Rahm[1]

Nephi A. Walton[2]

Lynn K. Feldman[3]

Conner Jenkins[4]

Troy Jenkins[4]

Thomas N. Person[1]

Joseph Peterson[4]

Jonathon C. Reynolds[1]

Peter N. Robinson[5]

Makenzie A. Woltz[1]

Marc S. Williams[1]

Michael M. Segal[3]

Corresponding author

Name: Alanna Kulchak Rahm

Address: Genomics Medicine Institute, Geisinger, 100 North Academy Avenue, Danville PA 17822-2620, USA

Email: akrahm@geisinger.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

AUTHOR AFFILIATIONS

1 Genomic Medicine Institute, Geisinger, Danville, Pennsylvania, USA

2 Intermountain Precision Genomics, Intermountain Healthcare, St George, Utah, USA

3 SimulConsult Inc., Chestnut Hill MA, USA

4 University of Utah, Salt Lake City, UT, USA

5 The Jackson Laboratory for Genomic Medicine and Institute for Systems Genomics, University of Connecticut, Farmington, CT, USA.

Keywords: health information systems, heuristic evaluation, cognitive walkthrough, Diagnostic Decision Support System (DDSS), Clinical Decision Support (CDS),

Word count: 2998

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

ABSTRACT

Background: There is a need in clinical genomics for systems that assist in clinical diagnosis, analysis of genomic information and periodic re-analysis of results, and can utilize information from the electronic health record to do so. Such systems should be built using the concepts of human-centered design, fit within clinical workflows, and provide solutions to priority problems.

Methods: We adapted a commercially available diagnostic decision support system (DDSS) to use extracted findings from a patient record and combine them with genomic variant information in the DDSS interface. Three representative patient cases were created in a simulated clinical environment for user testing. A semi-structured interview guide was created to illuminate factors relevant to human factors in CDS design and organizational implementation.

Results: Six individuals completed the user testing process. Tester responses were positive and noted good fit with real-world clinical genetics workflow. Technical issues related to interface, interaction, and design were minor and fixable. Testers suggested solving issues related to terminology and usability through training and infobuttons. Time savings was estimated at 30-50% and additional uses such as in-house clinical variant analysis were suggested for increase fit with workflow and to further address priority problems.

Conclusion: This study provides preliminary evidence for usability, workflow fit, acceptability, and implementation potential of a modified DDSS that includes machine-assisted chart review. Continued development and testing using principles from human-centered design and implementation science are necessary to improve technical

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

functionality and acceptability for multiple stakeholders and organizational implementation potential to improve the genomic diagnosis process.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

SUMMARY

What is already known?

- There is a need in clinical genomics for tools that assist in analysis of genomic information and can do so using information from the electronic health record.
- Such tools should be easy to use, fit within clinical workflows, and provide solutions to priority problems as defined by clinician end-users.
- Natural language processing (NLP) is a useful tool to read patient records and extract findings.

What does this paper add?

- We demonstrated the use of Human-centered design and implementation science principles in a simulated environment for assessment of a new version of a decision support tool prior to large-scale implementation.
- This study provides preliminary evidence that a clinical decision support tool with machine-assisted chart review is acceptable to clinical end-users, fits within the clinical workflow, and addresses perceived needs within the differential diagnosis process across all Mendelian genetic disorders.
- Terminology codes for DDSSs should have levels of granularity tuned to the sensitivity and specificity appropriate to its various functions, e.g., NLP versus chart documentation.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

INTRODUCTION

Clinical Decision Support (CDS) integrated into Electronic Health Records (EHRs) has long been considered a promising way to improve patient outcomes and decrease inefficiencies.[1-4] It is also recognized that CDS must be designed with the user in mind, fitting the concepts of human-centered design with computer interfaces at the individual clinician level.[1 5] Design alone, however, is insufficient to facilitate implementation. For CDS to impact care and patient outcomes, it must fit within clinician workflow and provide a solution to a priority problem for the clinician and the healthcare system.[4 6-8]

Diagnostic Decision Support Systems (DDSSs) are a key type of CDS needed in genomics to supplement a shortage of trained clinicians and address the inherent complexity of genomic diagnosis.[9 10] This complexity arises from the heterogeneous nature of genetic diseases, the variable expression in patients, and the degree of overlap in findings (i.e., signs, symptoms, and test results) among genetic conditions, sometimes differentiated only by onset age of individual findings.[11] Position statements and a systematic review note two new functions needed for DDSSs in genomics: (1) a cost-effective, regular approach to re-evaluation of patient cases in light of new findings or genetic knowledge, when testing does not immediately yield a diagnosis; and (2) developing machine-assisted chart review.[12 13] Most genomic patient records are extensive with input from by multiple clinicians, such that manual review is prohibitively time-consuming; resulting in added costs from repeated or unnecessary tests and increased risk of missed information that could have facilitated timely diagnosis. Because most of the relevant information is in unstructured clinical notes, approaches

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

such as natural language processing (NLP) are needed to automate and assist this manual process.

To address both re-evaluation and automation, we adapted a commercially available DDSS already capable of incorporating genomic sequencing data to perform automated chart review and present the information to a clinician in the form of findings obtained through structured data mining and NLP of an EHR. We then created clinical case vignettes to simulate the real-world clinical diagnostic workflow for user testing. The goal was to provide preliminary evidence of usability, perceived fit with clinical need and workflow, and potential for implementation into the real-world clinical environment.

METHODS

Setting

Development of the clinical case vignettes, simulated EHR environment, and user testing were conducted at Geisinger, a healthcare system in rural Pennsylvania.

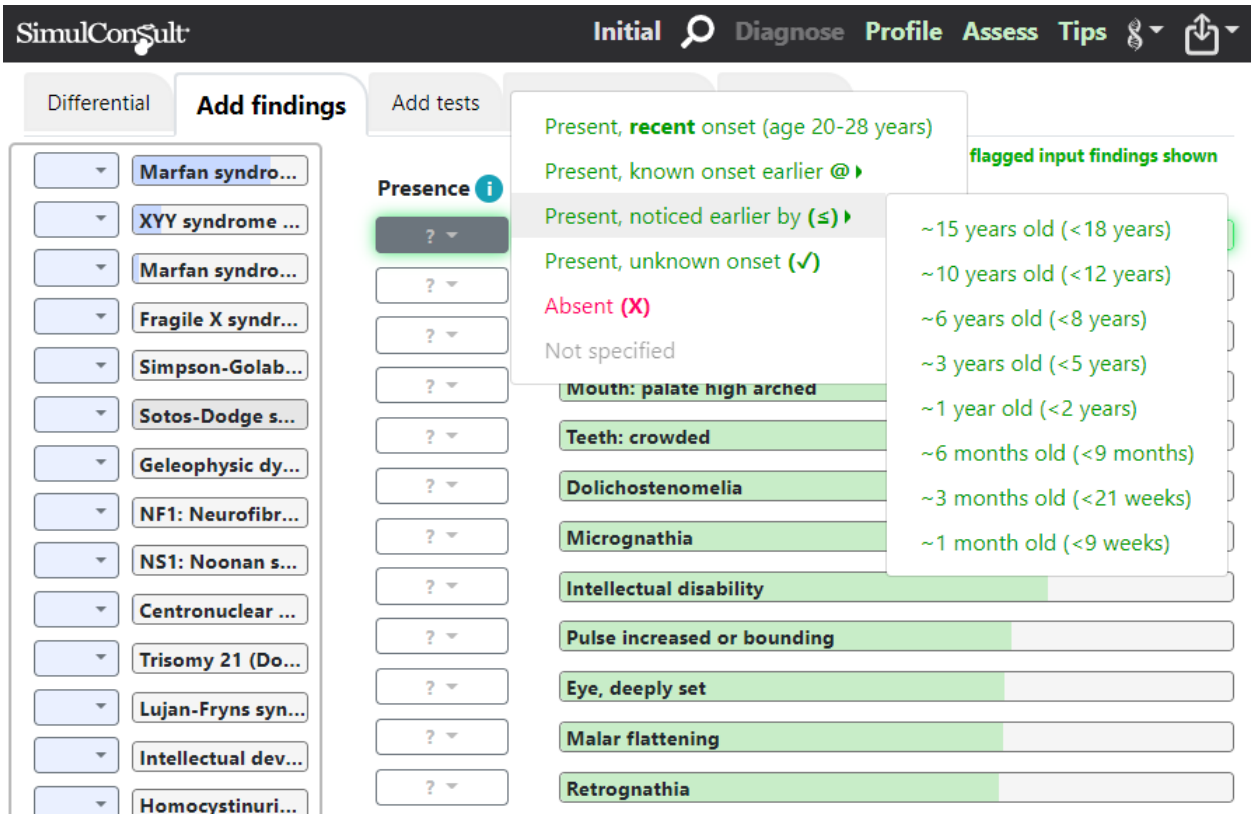
Adapting a DDSS for machine-assisted chart review of clinical findings

We adapted SimulConsult's Genome-Phenome Analyzer, as it is the one DDSS that allows for detailed analysis of clinical information, including pertinent negatives, findings onset information, and frequency and treatability of diseases. It has also been shown to be accurate and helpful in clinical diagnosis, including interpreting genomic results.[14-16] Described in detail elsewhere,[11 14 15] SimulConsult correlates annotated variant call files (VCFs) with patient-specific clinical and family history information; and the underlying algorithms include age-dependent Bayesian pattern-matching and computational

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

metrics of usefulness and pertinence. SimulConsult also generates a Patient Summary for saving interim patient findings and a customizable genomic Return of Results (RoR) report shown in previous research to be effective for facilitating standardized communication for patients and referring clinicians.[17-20] When clinicians enter findings, the DDSS returns a ranked list of candidate diseases and suggestions of other findings to check, ranked by usefulness in narrowing the differential diagnosis in a way that accounts for cost and treatability; thus facilitating the iterative approach of information gathering in diagnosis.[21 22] For each finding, presence (with onset age) or absence can be specified (Figure A).

Figure A. SimulConsult main interface showing ranked list of candidate diseases and guidance for entering finding presence (or absence) with onset age.



medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

We used the Logica platform to create a simulated EHR and the cTAKES tool with the Unified Medical Language System (UMLS) module[23] for NLP of patient notes. Steps in adaption included (1) mapping DDSS findings to Human Phenotype Ontology (HPO) and UMLS codes, including creation of hundreds of new HPO terms resulting in creation of new UMLS concepts, (2) using results from NLP analysis of EHR notes to flag "Mentions" of the findings used by the DDSS, and (3) augmenting the DDSS's interface to present the flagged findings with contextual information needed to clinically assess the information (Table 1).

Table 1. Adaptations made to existing DDSS to create GPACSS

Adaptation	Component	Approach
Overall design	SMART-on-FHIR enabled **EHR**	• Logica platform (https://www.logicahalth.org/; formerly Health Services Platform Consortium; HSPC)
	Archive	• Custom archive stores key files • RESTful interface.
Coordination and communication	**User interface**	• SMART-on-FHIR application (GPACSS FHIR app client, Figure B). • Interface allows user access to DDSS directly from patient record. • Choice to launch with no findings or with findings previously saved
	Coordination	• GPACSS "Coordinator" API saves the NLP output • Matching of UMLS codes in NLP output to DDSS findings • Send the matched flagged findings to the DDSS at launch (Figure B)
Natural language processing	**Extraction of findings**	• NLP: open source Apache cTAKES 4.0 [23] • cTAKES default modules to handle sentence boundary detection, tokenization, normalization, tagging parts of speech, recognizing named entities, and negation. • cTAKES pre-trained module to recognize UMLS concepts in text
	Mapping in	• DDSS findings mapped within the DDSS to one or

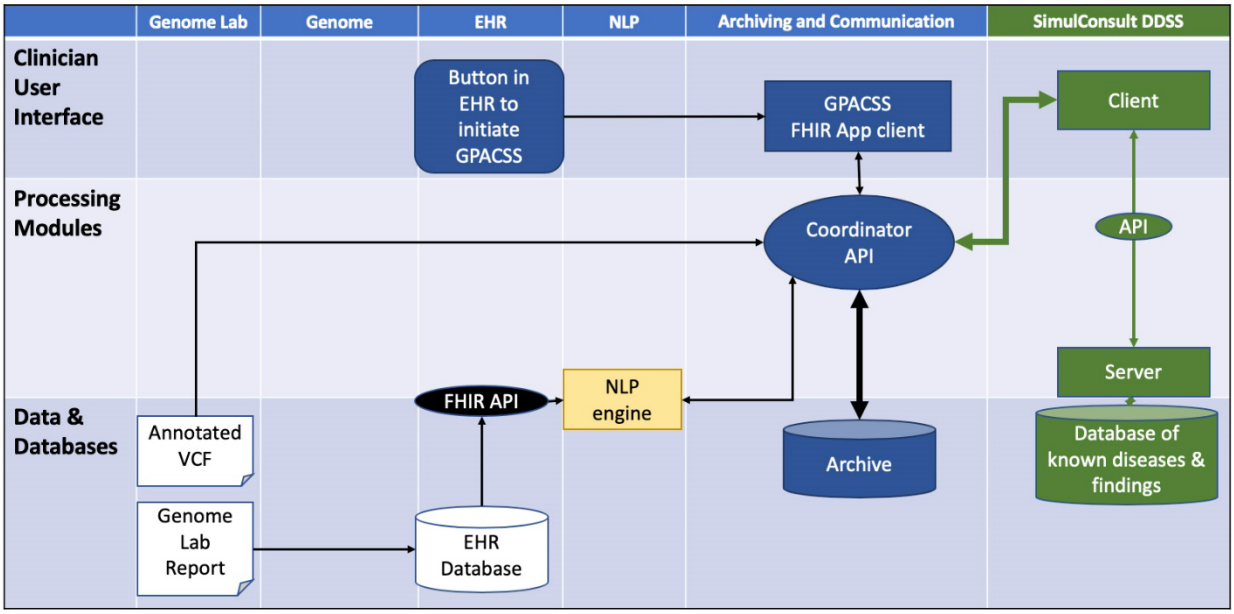
medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

	DDSS	more Unified Medical Language System (UMLS) and Human Phenotype Ontology (HPO) codes • Mapping strategy minimizes false negatives in term capture while tolerating false positives (identifying information unrelated or irrelevant to the diagnostic process).
	Display in DDSS	• Findings identified by NLP display a flag icon • Clicking the flag enables viewing of metadata

The architecture of the resulting prototype, called the Genotype-Phenotype Archiving and Communication System with SimulConsult (GPACSS), is shown in Figure B.

FIGURE B: Architecture of the Genotype-Phenotype Archiving and Communication System with SimulConsult (GPACSS)

The key components are the coordination / archiving system (blue), the DDSS (green) and the NLP (yellow).



Clinician review of the flagged findings created from the automated findings search using NLP is facilitated through flag icons (Figure C). Through this “machine-assisted”

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

chart review, the clinician reviews flagged findings and decides whether and how to specify presence (with a particular onset) or absence (or omit) as shown in Figure A. The mapping of DDSS findings to multiple UMLS concepts was chosen to minimize false negatives in concept identification; relying on the user decisions about findings and the limited set of UMLS concepts to minimize false positives (Table 2).

FIGURE C: Flagged findings with EHR text display for DDSS

A finding having a flag icon indicates that information was found in the EHR. Clicking the flag shows the various mentions of the flagged finding.

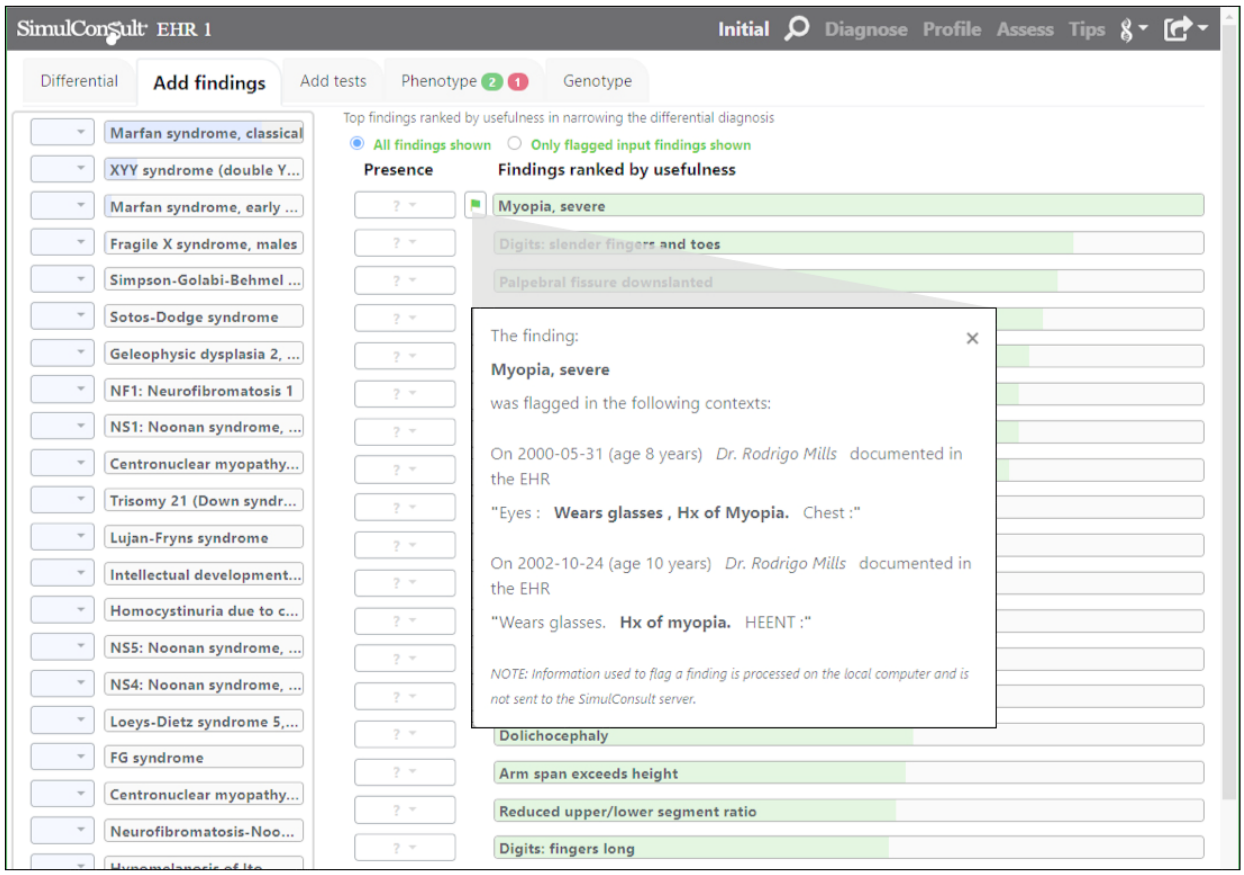


Table 2. Solutions for Minimizing False Positives and Negatives Identified Through NLP and DDSS by Clinician Review

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

False Negative/Positive Problem	Solution Included in GPACSS
Minimizing false negatives on NLP flagging of findings	• Include parent and child codes (e.g. finding of intellectual disability in DDSS includes codes for developmental delay and particular types of intellectual disability)
Minimizing false positives through the DDSS Usefulness metric	• Use DDSS usefulness algorithm[24] to display flagged findings; thus prioritizing data of greater relevance and de-prioritizing data of low relevance for clinician review
Minimizing false positives through clinician verification	• Use flag icon to indicate findings identified through NLP (Figure C) • Clinician clicks the flag icon to display information needed to assess reliability, presence or absence, and onset • Information displayed from the EHR includes date of chart note, observer identity, and 3 sentences of chart note (sentence with finding plus preceding and subsequent sentence)

Creating simulated cases

Three cases of increasing complexity were created using real but de-identified clinical phenotypic and time course data from medical notes of Geisinger patients with known genetic diagnoses (Supplemental Table 1). Cases were selected for conditions of varying complexity yet relatively common in the context of rare disease and where diagnosis might be difficult using phenotype alone. Simulated cases were created by research assistants trained in capturing information from the EHR, supervised by a practicing Geisinger clinician certified in genetics and informatics. The three final cases were reviewed by a second Geisinger physician certified in genetics and informatics prior to user testing.

Case vignettes for the test scenarios assumed that some patient characterization was previously noted by the clinician and genomic results were now available and could be

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

interpreted with clinical information available in the EHR (Supplemental Figure A). For the 3 cases, a total of 5 findings were used as initial information before the genomic results, with 3 (one per case) being flagged findings identified through NLP. This created a "near live"[25] experience within the simulated EHR for user testing while limiting the expense and time of EHR integration during this preliminary phase.

User Testing Methods

Participants: GPACSS is both a DDSS and communication tool to facilitate utilization of genomic and phenotypic information available in the EHR by all clinicians to improve patient care within a healthcare system. Therefore, we purposively selected primary testers from Geisinger staff representative of current end users of the genome-phenome analyzer. Because a limited number of individuals at Geisinger regularly engage in utilizing genomic information for differential diagnosis, we followed guidance recommending 3 - 5 evaluators for preliminary usability testing.[26] A group of secondary testers (inclusive of a pilot tester) with other roles in the genetic testing and interpretation process were purposively selected for potential broader utilization in the healthcare system.

Testing Sessions: At the beginning of each session, testers viewed a 4-minute training video (https://simulconsult.com/videogpacss) beginning from saved patient findings, then importing a VCF, and review of flagged findings to make a diagnosis and create a customizable patient-friendly RoR report.

A semi-structured interview guide was created to elucidate factors relevant to human factors in CDS design (information, interaction, interface)[1 5 27] and organizational

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

implementation (acceptability, perceived need, feasibility, workflow fit).[28] We utilized a think aloud[25] approach where testers were asked to verbalize thoughts while using the GPACSS prototype with the interviewer asking questions as needed and at key points in the testing to create a cognitive walkthrough with heuristic evaluation.[26 29] Testers were invited via direct contact from study staff and provided a description of the study. At the beginning of each session, study staff reviewed a study information sheet and obtained verbal consent to participate. Test sessions lasted 2 hours and testers received a $100 gift card. The user testing protocol was reviewed and approved by the Geisinger IRB.

An experienced interviewer (AKR) and observer (MAW) from Geisinger worked with each tester to imagine using GPACSS for each test scenario. The interview and process were piloted with a cancer genetic counselor reviewing one test vignette. At the end of the session, testers were asked a series of study-specific questions using a 0-10 rating scale (hard to easy) to rate the overall usefulness, satisfaction, and navigation. Transcripts were created from the audio portion of each session and the computer screen was video recorded to capture tester movement through GPACSS.

Analysis

Two Geisinger coders (MAW, JCR) viewed each user test session recording, read transcripts, and created a codebook of themes identified across sessions. Transcripts were coded and the corresponding quotes were organized into a matrix using the 3 categories of CDS components (information, interface, and interaction) identified by Miller et al,[1] and categories of acceptability, perceived need, feasibility, and workflow fit according to Rogers' Diffusion of Innovations in organizations constructs.[28] Coders

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

analyzed transcripts independently and reviewed for agreement with discrepancies resolved by the primary author.

RESULTS

Three clinicians currently using genomic information to diagnose patients participated as primary testers: a pediatric geneticist (orders exomes daily), internal medicine physician (orders 4-5 exomes per month), and a pediatric genetic counselor. Three additional clinicians participated as secondary testers; representing broader usability within the healthcare system: the pilot tester (cancer genetic counselor), a laboratory director (conducts variant interpretation), and a laboratory genetic counselor (conducts variant analysis).

GPACSS Usability: Human Factors of CDS Design

Overall impression of the prototype was positive. Testers raised general issues relevant to human factors in CDS design[1 5].

Interface: Testers liked the flagged findings (Figure C), the contextual information for each mention in the EHR, and the rank ordering of flagged findings by usefulness. The visualization of the evolving differential diagnosis and the automated RoR report for sharing with patients and referring clinicians, including the ability to save and access this report from the EHR were also appreciated.

The interface was noted to be complex, but testers stated this was expected due to the inherent complexity of genetic diagnosis and that they anticipated a learning curve to develop proficiency. Placement, positioning, and the multiple presentation layers (text

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

and graphics in the interface)[1] were well liked. In particular, the "Assess diagnosis" display was noted as valuable because it made transparent the logic used by the DDSS in comparing patient findings to information about the disease. Of note, each tester interpreted differently the meaning of the graphical bars and shading, however, this did not hinder their ability to make the diagnosis, and the bar itself was appreciated as a design feature. To help with interpretation, more labeling was suggested (Table 3).

Table 3. GPACSS Usability: Human Factors of CDS Design and Organizational Implementation Factors Through Tester Experiences*

Human Factors of CDS Design	Interface	*"More training would be good… unless I was doing it all the time for all of my patients, every step, I might not realize that some of the features are available…"* [Tester 3]
		"These bars are different lengths, so I assume it's having something to do with frequencies... so I'm not sure why this part is purple....if there were something [on the assess diagnosis tab] that said this is 100% over here and this is 0% over here, that would kind of help, if I knew that that was the case... I'm not sure what these other colors are referring to." [Tester 5]
	Interaction	*"To me, the green bar in it shows me they are confident that this genetic variant aligns with the phenotypic markers that we have identified. I don't necessarily know how far the bars will tell me they're confidence in pathogenic versus VUS."* [Tester 3]
		"It's going to take a lot to learn. A lot of clicking back and forth and it's not super intuitive but I get it. So, the report gets generated and that becomes part of the record. I can see how that can be helpful because it has now particular phenotypical diagnosis and even genetic finding" [Tester 2]
	Information	*"The term pertinent gene zygosity is not something I would normally make part of my lexicon…I have a general sense of the term zygosity but I can't remember the last ten years using that term in any of my discussions in clinical care or genetics in some of the cases I found what's their zygosity"* [Tester 2]
		"But the variant severity score doesn't mean anything to me personally. To me it's easier to know, if you verify

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

		know the true classification they are giving it… Pathogenic, likely pathogenic VUS benign." [Tester 3]
		"I think this one is nice [the 'Mention' displayed in a flagged finding]. Whenever someone says it had been noticed earlier by, it's nice when someone is talking about their niece or nephew, or like a proband cousin, they are saying they had myopia and I remember them having glasses before they were 5 years old." [Tester 3]
Organizational Implementation Factors	Acceptability	*"I would use it most of the time. To me, this is the frontier of genomic medicine and I look at my role as not only taking care of a patient but figuring out how we make genomics part of everyday medical practice. The useful things in the chart, genetics people can now get to right away*. [Tester 2]
		"…Typing them up, writing the summary [of all the patient findings in the chart]. If I could see what's been flagged in the chart, see what has not actively been flagged and decide do I need to go back and look at it or not. It would save my time" [Tester 3]
		"I think the interface is really good, in that you have that ability to explore those variants that may or may not make it on the reports that we get now, so you can drill deeper if you want. [Tester 5]
	Perceived Need	*"…The report is a great idea for highlighting why you think it's [the care instructions] important, [in] a standard format… The average primary care physician that gets the genetic testing reports, says I don't know what this means at all. I think this [the Prognosis Table] is a step towards making it more understandable."* [Tester 2]
		"*Everything's there [in the chart] and the question is how easy is it to find. I'm sure if you're a malpractice lawyer you get very good at pulling stuff out of these charts and asking why didn't you see that. Yet I can't look at everything."* [Tester 2]
		"This is stuff that you are doing anyway… you could make your note a lot shorter and just refer to that document [the automated Summary] … I like the idea that you can explore. Clinical genetics now is limited on time." [Tester 5]
	Workflow Fit	*"It's nice because it helps guide me… it's very easier for me to realize that Prader–Willi is associated with narcolepsy...."* [Tester 3]
		"I think the nice part is I don't have to go searching myself to find all the signs and symptoms associated with

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

		it and potentially miss something, that I may not know is a less common finding or feature of the condition. That actually could be beneficial for a provider or for us to give to the testing lab, to say these are all symptoms that we see, and then analyzing the data" [Tester 3]

**Comments from Primary User-testers only (testers experienced with differential diagnosis of genetic conditions through sequencing): n=3; pediatric genetic counselor, pediatric geneticist, internist ordering 4-5 exomes in the past month*

Interaction: Testers were thoughtful and purposeful using GPACSS. Notably, in case 3 (the most complex case), one primary tester did not immediately choose the top diagnosis offered by GPACSS. Supported by the data displayed, the tester indicated that to make a definitive diagnosis they would next evaluate for the second-ranked disease – as that condition had a test that was easy and accurate and the condition was also more treatable – indicating utilization of the DDSS as intended and consistent with clinical diagnostic decision-making.

Testers initially expressed concern around "too many clicks" and "click fatigue" but noted as they progressed through the cases that the clicking was unavoidable and necessary. For example, they saw value in taking the time to correctly specify onset information (which requires clicking and cognitive load in the DDSS), as this is part of the genetic diagnostic process. "Cognitive Load" in DDSS testing refers to additional thinking required to interact with the tool, and the general recommendation is to minimize this in CDS design.[1] Testers who commented on the cognitive load required to review flagged findings and choose age of onset noted the cognitive load as similar to completing this task without GPACSS.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Information: Testers appreciated resources such as the hover feature that revealed synonyms to findings and requested even more hovers and infobuttons. Confusion over some terminology occurred, notably "zygosity" and "severity score," when reviewing the genomic variants; as only some testers located the explanatory resource for these terms.

The fact that the EHR "Mentions" displayed in flagged findings were sometimes triggered by parent or by child concepts was noticed by all testers, and some stated the findings used in the DDSS were not as granular as they were expecting. Regardless, testers recognized and emphasized the importance of being able to review the "Mention" information from the EHR and manually adjust for any false positives and false negatives from the NLP process.

GPACSS Usability: Organizational implementation factors

Acceptability: For the primary testers, satisfaction averaged 8.5 out of 10 (range 8-9.5) and navigation ease averaged 8 out of 10 (range 7.5-9). All 3 felt GPACSS would save time throughout the clinical process, with one primary tester estimating it at 30-50%. Specific value in time saved was noted for chart review by all testers.

Perceived need: The RoR report and detailed prognosis table[20] generated in each scenario was highly valued for being standardized and for its ability to communicate complex genetic information to patients and other clinicians (Table 3). The RoR report was also noted as an improvement over current laboratory reports; with one tester stating it was "where the most utility would be" [Tester 4].

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Testers exhibited learning and familiarity with GPACSS as they progressed through the testing session; appreciating the DDSS assistance as each vignette increased in complexity; noting *"It takes it [clinical diagnosis and diagnostic thinking] to a higher level". [Tester 2]*. Primary testers expressed readiness to adopt the tool in clinical practice; and one (pediatric geneticist) suggested GPACSS could also serve as a differential diagnosis training tool for medical students and residents in their clinic.

Two secondary testers (lab director and variant analyst) expressed enthusiasm that GPACSS could fill a need for in-house sequencing laboratories because full EHR data would be available during sequence interpretation. These testers also hypothesized that the ability to periodically re-analyze an existing VCF in minutes using GPACSS would improve the diagnosis rate over time.

Workflow Fit: The three primary testers noted that the GPACSS process as tested fit with their clinical workflow diagnosing genetic conditions. As an added benefit, they described how using GPACSS also helped them learn about diseases and associated findings with which they were less familiar (Table 3).

The three secondary testers questioned GPACSS fit with a clinical genetic testing workflow in which only a report with variants labeled as to pathogenicity and association with a condition (implying a clinical diagnosis) is received from an external lab. However, they did identify value and possible workflow fit for situations with uncertainty as to the diagnosis after sequencing or where flagged findings and the usefulness ranking would allow clinicians to review the EHR with flagged findings in light of the genomic information to make the diagnosis.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

DISCUSSION

We provide preliminary evidence through user testing in a simulated real-world clinical workflow that the combination of NLP with a CDS tool optimized to support the clinical process of differential diagnosis may address the needs of those involved in this complex task. Such assessment of fit is critical if CDS is to fulfil the promise of standardizing and improving care.[1 4 5 8]

Technical issues related to the interface and interaction of CDS design were minor and fixable; as were issues with design layout. Despite initial remarks on the number of clicks and cognitive load, testers acknowledged these as necessary to the genetic diagnosis process and no different than without the DDSS. Other issues related to terminology and usability could be solved and evaluated in future usability studies through a combination of training, added infobuttons, and experience using GPACSS. Some of the technical gaps noted and additions requested by testers are addressed within GPACSS, however, the 4-minute training video was created to provide enough instruction only to facilitate user testing. These results therefore provide direction for training and ongoing reference materials for future implementation.

For CDS to be acceptable and implemented by clinicians and organizations, it must fit with the real-world workflow and must present a solution to a perceived need.[5 28] All primary testers identified ways GPACSS added such value and fit and noted ways GPACSS filled multiple needs in their diagnostic workflow. Workflow fit was highest among primary testers but opportunities for workflow fit were described by all testers. GPACSS was also noted as acceptable for implementation by all testers regardless of individual issues identified and suggestions for technical improvements.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

LIMITATIONS

To facilitate user testing of GPACSS in the context of clinical workflow prior to full integration and implementation, simulations of the real-world were required. Because this study used the Logica EHR simulation, benefits or drawbacks of GPACSS in a production EHR could not be directly observed. Also, full annotations for the causal variants were not included in the variant table for the simulated patients limiting full assessment of the value of the DDSS in variant interpretation. This impacted the understanding of the "severity score" by all testers, as the annotation information that would have been provided for a real patient was not included for the simulated cases. Finally, the generic cTAKES NLP using the UMLS concepts found only 20 of the 30 (67%) pertinent positive concepts within the test cases that a pediatric neurologist (MMS) identified manually. This was sufficient for GPACSS to generate the correct differential diagnosis for user testing, as further enrichment of the generic NLP to improve detection and avoid false positives was out of scope for this preliminary user testing[30]. Subsequent automated search for UMLS terms for flagging and addition of a separate stage of text search enrichment for terms missed by the NLP such as "tall" improved NLP yield to 30 of 30 (100%).

This simulated EHR and user testing were a necessary first step and provide data to guide implementation of GPACSS. NLP improvements and additional beta testing within an actual EHR, in real-world clinical workflows, with real patient results, and in real-world clinical workflows will be necessary to fully assess individual user-level and

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

organizational-level facilitators and barriers to use, implementation, and impact on clinical care. Such studies are currently in progress.

CONCLUSIONS

This study provides preliminary evidence for the usability, workflow fit, acceptability, and implementation potential of a DDSS that includes machine-assisted chart review. Overall, responses suggest the GPACSS prototype is usable based on technical CDS and human-centered design criteria, addresses perceived clinical need, and has good fit within the real-world clinical workflow of genetic testing and diagnosis. Further development is needed to improve usability for multiple clinical stakeholders and organizational implementation.

Contributors: All authors reviewed the final manuscript for submission and contributed to the study as follows:

AKR contributed to the study design, conducted user testing sessions, and led the analysis, interpretation, and writing of manuscript.

NAW contributed to the study design, created the simulated clinical environment and the de-identified patient records, contributed to the NLP testing and adaptation of the DDSS and manuscript revisions.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

TNP was the lead developer on the NLP and the integration with the SimulConsult API.

CJ, TJ, TNP, and JP contributed to NLP testing and creation of the simulated clinical environment and manuscript revisions

MAW contributed to project management, user testing recruitment, qualitative data analysis and interpretation and review of manuscript.

JCR contributed to project management, IRB applications, NLP testing and creation of the simulated clinical environment, and manuscript writing

MSW contributed to study design, simulated clinical environment, NLP testing and adaptation of the DDSS, user testing design and recruitment, data analysis, interpretation and manuscript writing

LKF contributed to project management, study design, and manuscript revisions

PNR assisted in the tagging of DDSS findings with HPO codes and created the new codes when needed.

MMS designed and implemented the adaptation of the clinical decision support tool to interact with the NLP, designed and implemented the strategy for HPO and UMLS coding, did the manual chart review to identify findings that should have been recognized by NLP, and contributed to the study design, and manuscript revision

FUNDING: This study was supported by the National Human Genome Research Institute of the National Institutes of Health under the Small Business Innovation Research (SBIR) Program, Award Number 1R43HG010322-01 (principal investigator:

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

MMS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

COMPETING INTERESTS LKF and MMS have an ownership stake in SimulConsult. MMS is PI of the NIH-funded Small Business Innovation Research (SBIR) grant. LKF and MMS provided the DDSS software to Geisinger for this research, including the modifications to enable the NLP and flagged findings, but were not involved in the user testing evaluation and analysis.

All other authors declare no competing interests

PATIENT CONSENT FOR PUBLICATION: not required

ETHICS APPROVAL: This study was reviewed and approved by the Geisinger Institutional Review Board (IRB)

PROVENANCE AND PEER REVIEW: Not commissioned; externally peer reviewed

DATA AVAILABILITY STATEMENT: Data are available upon reasonable request.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure Legends:

Figure A. SimulConsult main interface showing ranked list of candidate diseases and guidance for entering finding presence (or absence) with onset age.

FIGURE B: Architecture of the Genotype-Phenotype Archiving and Communication System with SimulConsult (GPACSS)

The key components are the coordination / archiving system (blue), the DDSS (green) and the NLP (yellow).

FIGURE C: Flagged findings with EHR text display for DDSS

A finding having a flag icon indicates that information was found in the EHR. Clicking the flag shows the various mentions of the flagged finding.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

REFERENCES

1. Miller K, Mosby D, Capan M, et al. Interface, information, interaction: a narrative review of design and functional requirements for clinical decision support. *Journal of the American Medical Informatics Association* 2017;25(5):585-92. doi: 10.1093/jamia/ocx118

2. Magrabi F, Ammenwerth E, McNair JB, et al. Artificial Intelligence in Clinical Decision Support: Challenges for Evaluating AI and Practical Implications. *Yearbook of medical informatics* 2019;28(1):128-34. doi: 10.1055/s-0039-1677903 [published Online First: 2019/04/26]

3. Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems: a systematic review. *Ann Intern Med* 2012;157(1):29-43. doi: 10.7326/0003-4819-157-1-201207030-00450 [published Online First: 2012/07/04]

4. Moxey A, Robertson J, Newby D, et al. Computerized clinical decision support for prescribing: provision does not guarantee uptake. *J Am Med Inform Assoc* 2010;17(1):25-33. doi: 10.1197/jamia.M3170 [published Online First: 2010/01/13]

5. Miller K, Capan M, Weldon D, et al. The design of decisions: Matching clinical decision support recommendations to Nielsen's design heuristics. *International journal of medical informatics* 2018;117:19-25. doi: 10.1016/j.ijmedinf.2018.05.008 [published Online First: 5/21/2018]

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

6. Moja L, Liberati EG, Galuppo L, et al. Barriers and facilitators to the uptake of computerized clinical decision support systems in specialty hospitals: protocol for a qualitative cross-sectional study. *Implementation science : IS* 2014;9:105. doi: 10.1186/s13012-014-0105-0 [published Online First: 2014/08/29]

7. Liberati EG, Ruggiero F, Galuppo L, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. *Implementation science : IS* 2017;12(1):113. doi: 10.1186/s13012-017-0644-2 [published Online First: 2017/09/17]

8. Sutton RT, Pincock D, Baumgart DC, et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ digital medicine* 2020;3:17. doi: 10.1038/s41746-020-0221-y [published Online First: 2020/02/13]

9. Williams MS. Perspectives on what is needed to implement genomic medicine. *Molecular genetics & genomic medicine* 2015;3(3):155-9. doi: 10.1002/mgg3.135 [published Online First: 2015/06/02]

10. Shyr C, Kushniruk A, van Karnebeek CD, et al. Dynamic software design for clinical exome and genome analyses: insights from bioinformaticians, clinical geneticists, and genetic counselors. *J Am Med Inform Assoc* 2016;23(2):257-68. doi: 10.1093/jamia/ocv053 [published Online First: 2015/06/29]

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

11. Segal MM, Williams MS, Gropman AL, et al. Evidence-based decision support for neurological diagnosis reduces errors and unnecessary workup. *Journal of child neurology* 2014;29(4):487-92. doi: 10.1177/0883073813483365 [published Online First: 2013/04/12]

12. Welch BM, Kawamoto K. Clinical decision support for genetically guided personalized medicine: a systematic review. *J Am Med Inform Assoc* 2013;20(2):388-400. doi: 10.1136/amiajnl-2012-000892 [published Online First: 2012/08/28]

13. Grebe TA, Khushf G, Chen M, et al. The interface of genomic information with the electronic health record: a points to consider statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine* 2020 doi: 10.1038/s41436-020-0841-2

14. Segal MM, George R, Waltman P, et al. Clinician-centric diagnosis of rare genetic diseases: performance of a gene pertinence metric in decision support for clinicians. *Orphanet J Rare Dis* 2020;15(1):191-91. doi: 10.1186/s13023-020-01461-1

15. Segal MM, Abdellateef M, El-Hattab AW, et al. Clinical pertinence metric enables hypothesis-independent genome-phenome analysis for neurologic diagnosis. *Journal of child neurology* 2015;30(7):881-8. doi: 10.1177/0883073814545884 [published Online First: 2014/08/27]

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

16. Issa MY, Chechlacz Z, Stanley V, et al. Molecular diagnosis in recessive pediatric neurogenetic disease can help reduce disease recurrence in families. *BMC medical genomics* 2020;13(1):68. doi: 10.1186/s12920-020-0714-1 [published Online First: 2020/05/15]

17. Williams MS, Williams JL, Rahm A, et al. Evaluating the Impact of a User-Friendly Report of Genetic Testing Results for Parents of Children with Genetic Disorders. Washington, DC: Patient-Centered Outcomes Research Institute (PCORI), 2018.

18. Williams JL, Rahm AK, Zallen DT, et al. Impact of a Patient-Facing Enhanced Genomic Results Report to Improve Understanding, Engagement, and Communication. *Journal of genetic counseling* 2017 doi: 10.1007/s10897-017-0176-6 [published Online First: 2017/12/06]

19. Williams JL, Rahm AK, Stuckey H, et al. Enhancing genomic laboratory reports: A qualitative analysis of provider review. *American journal of medical genetics Part A* 2016;170(5):1134-41. doi: 10.1002/ajmg.a.37573 [published Online First: 2016/02/05]

20. Segal MM, Rahm AK, Hulse NC, et al. Experience with Integrating Diagnostic Decision Support Software with Electronic Health Records: Benefits versus Risks of Information Sharing. *EGEMS (Washington, DC)* 2017;5(1):23. doi: 10.5334/egems.244 [published Online First: 2018/06/23]

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

21. Kassirer JP. Teaching clinical medicine by iterative hypothesis testing. Let's preach what we practice. *The New England journal of medicine* 1983;309(15):921-3. doi: 10.1056/nejm198310133091511 [published Online First: 1983/10/13]

22. Eddy DM, Clanton CH. The art of diagnosis: solving the clinicopathological exercise. *The New England journal of medicine* 1982;306(21):1263-8. doi: 10.1056/nejm198205273062104 [published Online First: 1982/05/27]

23. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc* 2010;17(5):507-13. doi: 10.1136/jamia.2009.001560 [published Online First: 2010/09/08]

24. Segal MM. 2004 patent 6,754,655.

25. Richardson S, Mishuris R, O'Connell A, et al. "Think aloud" and "Near live" usability testing of two complex clinical decision support tools. *International journal of medical informatics* 2017;106:1-8. doi: 10.1016/j.ijmedinf.2017.06.003 [published Online First: 2017/09/06]

26. Khajouei R, Zahiri Esfahani M, Jahani Y. Comparison of heuristic and cognitive walkthrough usability evaluation methods for evaluating health information systems. *J Am Med Inform Assoc* 2017;24(e1):e55-e60. doi: 10.1093/jamia/ocw100 [published Online First: 2016/08/09]

medRxiv preprint doi: https://doi.org/10.1101/2020.08.21.20179580; this version posted March 26, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

27. Nielsen J. Enhancing the explanatory power of usability heuristics. . *Proc ACM CHI'94 Conf (Boston, MA, April 24-28)* *https://wwwnngroupcom/articles/ten-usability-heuristics/* 1994:152-58.

28. Rogers EM. Innovation in Organizations. Diffusion of Innovations. 5th ed. New York, NY: Simon & Schuster, Inc 2003:402 - 35.

29. Khajouei R, Farahani F. A combination of two methods for evaluating the usability of a hospital information system. *BMC Med Inform Decis Mak* 2020;20(1):84. doi: 10.1186/s12911-020-1083-6 [published Online First: 2020/05/06]

30. Demner-Fushman D, Mork JG, Shooshan SE, et al. UMLS content views appropriate for NLP processing of the biomedical literature vs. clinical text. *Journal of biomedical informatics* 2010;43(4):587-94. doi: 10.1016/j.jbi.2010.02.005 [published Online First: 02/10]