User-Centered Design for Antidepressant Prediction Models

Maia Jacobs

Computer Science + Preventive Med.

Northwestern University

Northwestern

Personalized & Adaptive Tech for Health (NU-PATH Lab)

Patient

How can we design tools that account for individuals' changing health needs?

Community

How can we integrate mental health support within community organizations?

Clinic

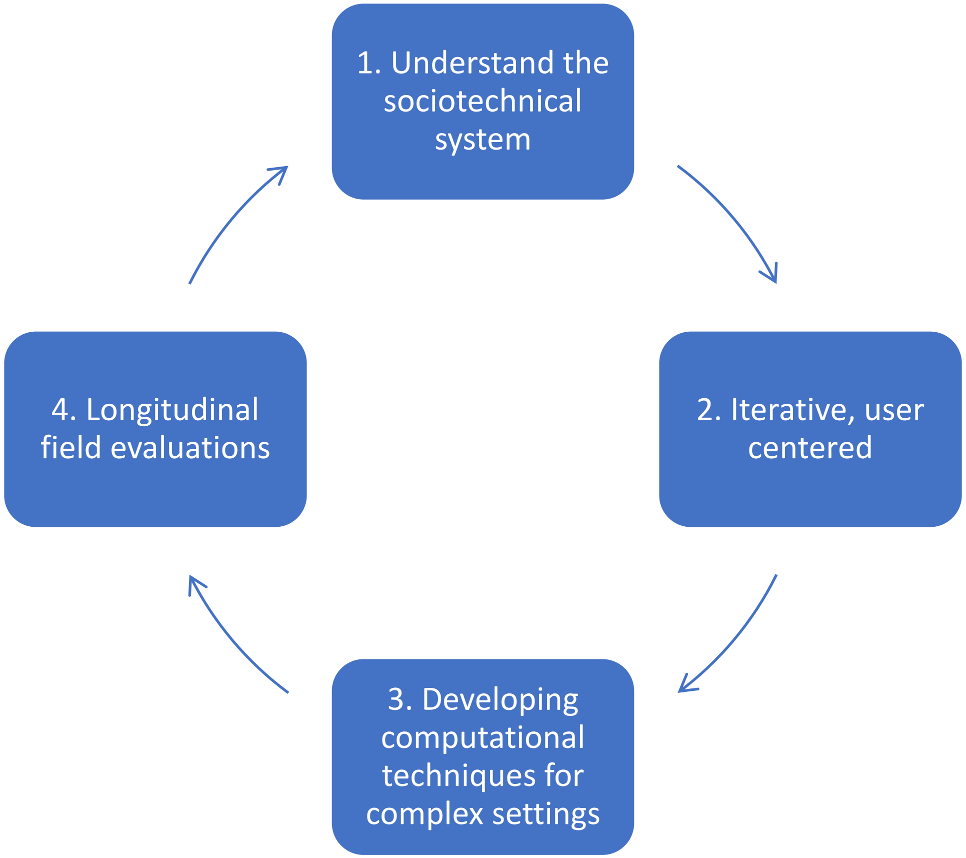
How should we design for decision-support tools embodying machine learning models

Why technology fails

Study	Favorable response to CDSS	Unfavorable response to CDSS	CDSS Description
Bergman & Fors (2005) [15]	Can save time and provide structure	Not suitable to workflow and there is the risk of becoming dependent	CDSS for medical diagnosis of psychiatric diseases
Curry & Reed (2011) [16]	Concept was supported	Interference with workflow and questionable validity	Prompts for adhering to diagnostic imaging guidelines
Gadd et al (1998) [18]	Easy to use, limits the need for data entry, accurate, and relevant	Benefits are lost because it takes so long to use	Internet-based system that interactively presents clinical practice guidelines at point of care
Johnson et al (2014) [19]	Longitudinal acceptance behavior, perceived ease of use, and perceived usefulnes	Computer literacy, user satisfaction, and general optimism	Clinical reminders and alerts for patients with asthma, diabetes, hypertension, and hyperlipidemia
Rosenbloom et al (2004) [20]	Can improve efficiency and quality of care; enhances education	Senior physicians did not think it was necessary	CDSS for computerized order entry system
Rousseau et al (2003) [21]	Use of "active" CDSS can bridge the gap between own practice and best practice	Clinicians found it to be difficult to use and unhelpful clinically	CDSS for chronic disease in general practice
Shibl et al (2013) [22]	Performance expectancy, usefulness, and effort expectancy	Trust in CDSS and need for the system	No specified CDSS; responses based on past and present experiences with multiple CDSSs
Sousa et al (2015) [23]	Belief that the suggestions were good for the patient	Low confidence in the evidence	CDSS for nursing care plan
Terraz et al (2005) [24]	Ease of use and easy access to information	Information that is presented is already known	Guidelines for colonoscopies
Wallace et al (1995) [25]	Can improve patient outcomes	Alerts are ignored because there is not enough time to dedicate to forming an appropriate response	CDSS to standardize administration of supplemental oxygen
Zheng et al (2005) [17]	Improves performance leading to better care, easy to use, and efficient	Iterative advisories, lack of relevance, a lot of data entry, and disruptive	Clinical reminders for chronic diseases and preventive care

Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: Critical analysis. J Med Internet Res. 2018;20(4).

Research Process



Common Methods:

1. **Qualitative + Participatory design studies with stakeholders**
2. **In-lab user studies**
3. Field trials

Understanding Implementation Barriers

?

Factorial experiment with clinicians

- To study how ML interface design influences behavior
  - *Surprise: Years of of clinical training won't overcome an inaccurate recommendation!*

User-Centered design

- To involve healthcare experts directly in the development of new tools
  - *Surprise: No participants initiated discussions about trust in ML models*

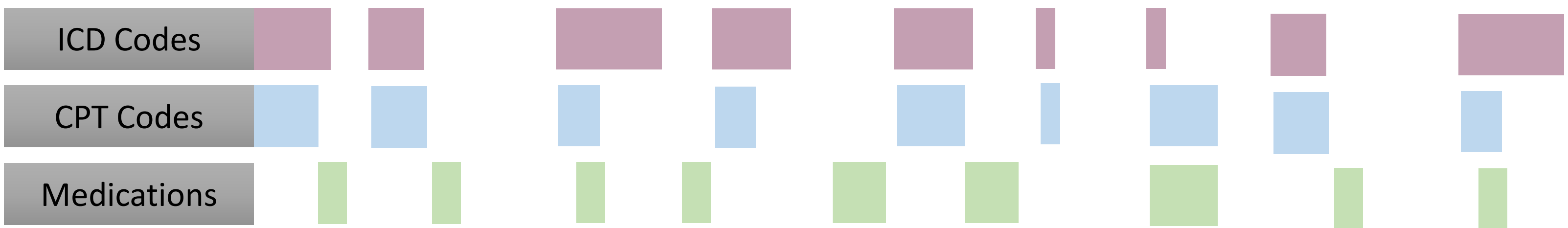
Why study antidepressant recommendations?

Hard problem:

- 2/3 of patients don’t reach remission after first antidepressant trial
- 1/3 of the patients do not remit despite up to four antidepressant trials

> “ *Oftentimes I'll have the choice of four or five different medications and not any great reason to choose among them.*

ML Outputs



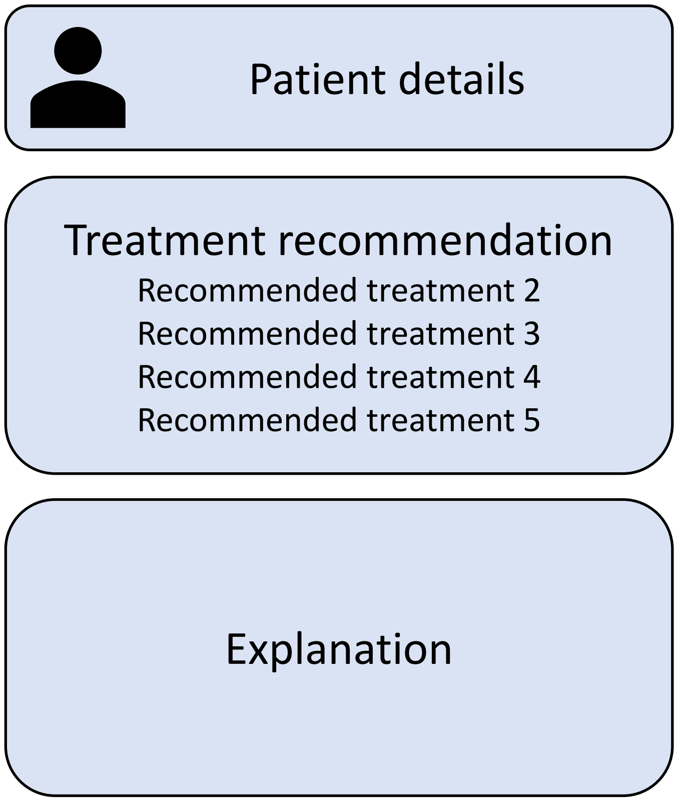
Stability score: The probability of continued use of the same medication for at least 3 months[1]

Dropout score: The probability of early treatment discontinuation following prescription while staying in the same health system[2]

Personalized treatment recommendations

1. Hughes MC, Pradier MF, Ross AS, McCoy TH, Perlis RH, Doshi-Velez F. Assessment of a Prediction Model for Antidepressant Treatment Stability Using Supervised Topic Models. JAMA Netw Open. 2020;3(5):e205308.
2. Pradier MF, McCoy TH, Hughes M, Perlis RH, Doshi-Velez F. Predicting treatment dropout after antidepressant initiation. Transl Psychiatry. 2020;10(1):1–8.

MDD Study 1



1. **Which antidepressant medication would you be most likely to prescribe in this situation?**
2. How confident are you with this decision? (1-5)
3. How frequently do you prescribe this medication? (1-5)
4. To what extent did the recommendations help you to make your treatment decision? (1-5)

MDD Study 1

Patient details

Treatment recommendation

Recommended treatment 2
Recommended treatment 3
Recommended treatment 4
Recommended treatment 5

Explanation

1. **Which antidepressant medication would you be most likely to prescribe in this situation?**
2. How confident are you with this decision? (1-5)
3. Ho
4.

What makes a recommendation incorrect?

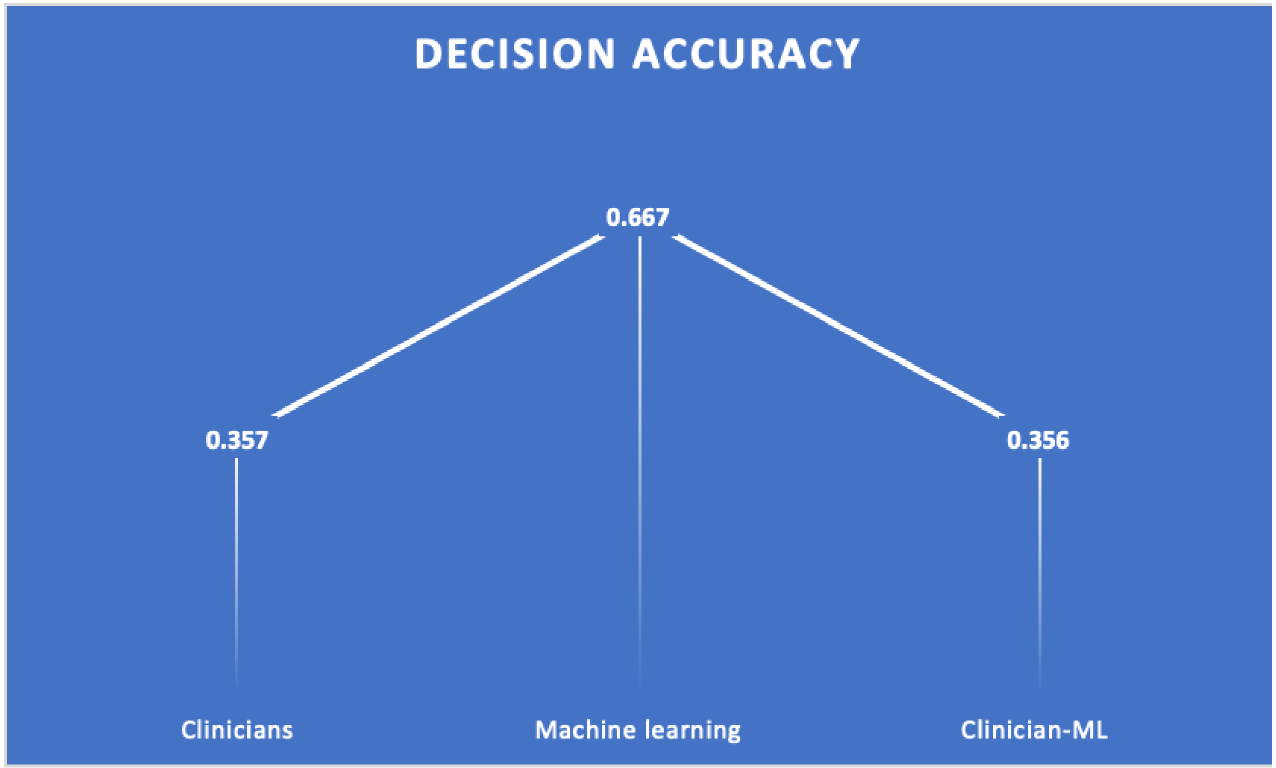
Existing tolerability & safety standard:

- Fatigue → avoid sedating (Mirtazapine)
- Suicidality → favor drugs safer in overdose (avoid bupropion)

MDD Study 1

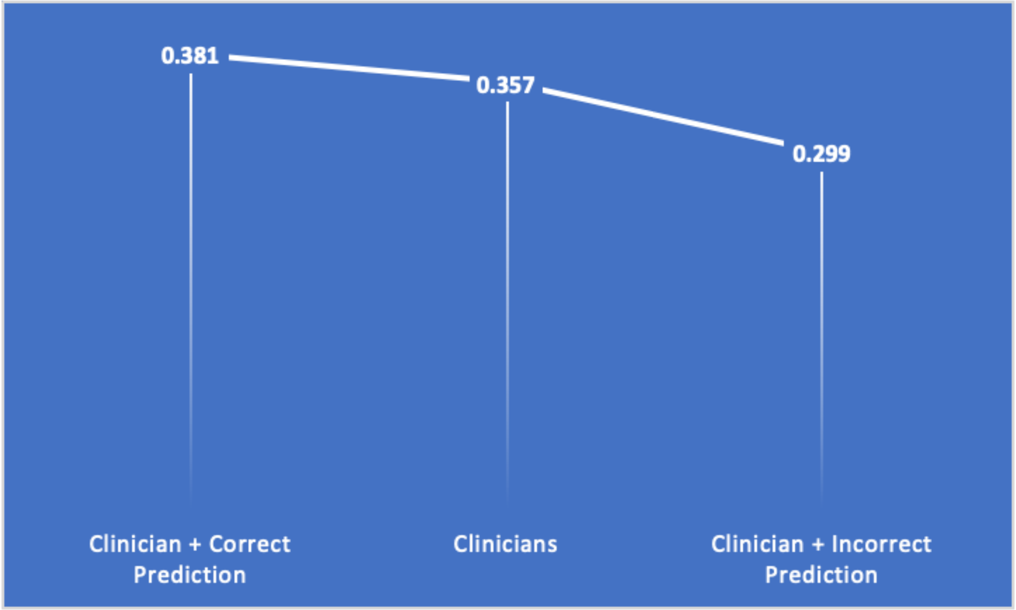
Age (SD)	42.52 (9.28)
Medical specialty (%)	
Psychiatry	195 (88.64)
Primary Care	18 (8.18)
Other	7 (3.18)
Years of experience prescribing antidepressants (IQR)	10 (7–15)
Machine-learning familiarity (%)	
Extremely familiar	45 (20.45)
Very familiar	51 (23.18)
Moderately familiar	30 (13.64)
Slightly familiar	54 (24.55)
Not familiar at all	40 (18.18)

An algorithm performing better than a person will not always lead to improved decision-making



(accuracy based on concordance with psychopharmacology experts)

Clinicians, even with years of training, are at risk of over-trusting imperfect tools



MDD Study 1

Patient details

Treatment recommendation

Recommended treatment 2
Recommended treatment 3
Recommended treatment 4
Recommended treatment 5

Explanation

1. **Which antidepressant medication would you be most likely to prescribe in this situation?**
2. How confident are you with this decision? (1-5)
3. How frequently do you prescribe this medication? (1-5)
4. To what extent did the recommendations help you to make your treatment decision? (1-5)

Explanation style matters

1. Baseline (no treatment recommendation)

2. Treatment recommendation only (no explanation)

3. Placebo

Why are these therapies being recommended?

System.09's predictions are **based on the patient's ICD-9 codes.**

4. Rule-based explanation

Why are these therapies being recommended?

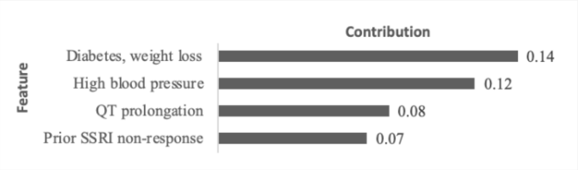
The following **rules** had the highest contributions to system.15's predictions:

1. *If concern for QT prolongation, favor Sertraline, avoid Citalopram*
2. *If avoiding weight gain, favor weight loss, favor Bupropion, avoid Mirtazapine*
3. *If concern for increased blood pressure, avoid SNRI's*
4. *If lack of response to Paroxetine, avoid SSRI's*

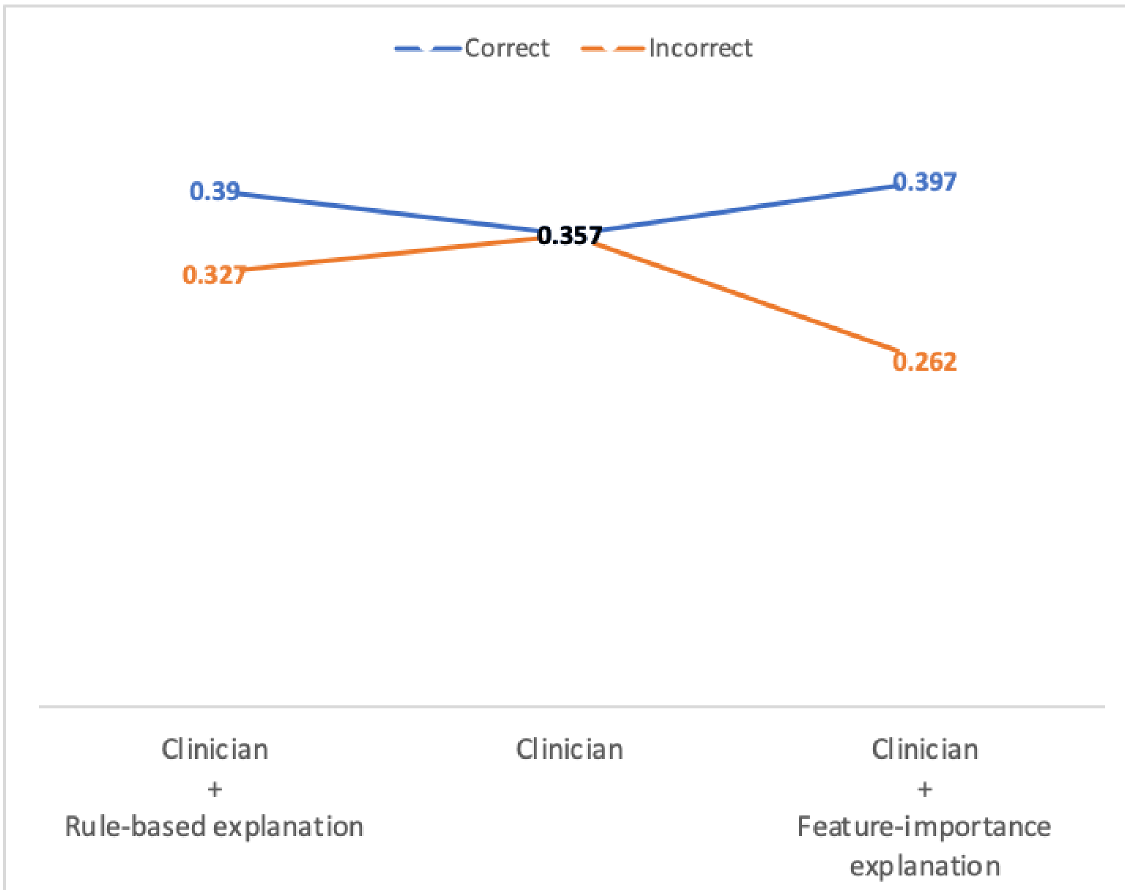
5. Feature-based explanation

Why are these therapies being recommended?

The following **patient features** had the highest contributions to system.12's predictions:



Feature importance explanations considered more helpful, but reduced accuracy

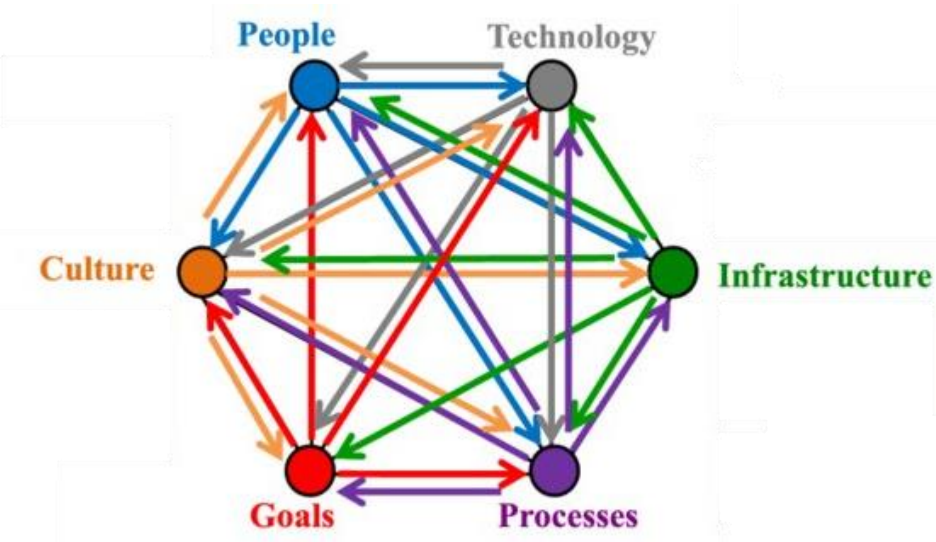


Takeaways

1. Incorrect ML recommendations may adversely impact clinician treatment selections
2. Explanations are insufficient for addressing overreliance on imperfect ML algorithms
3. Findings challenge the common assumption that clinicians interacting with ML tools will perform better than either clinicians or ML algorithms individually.

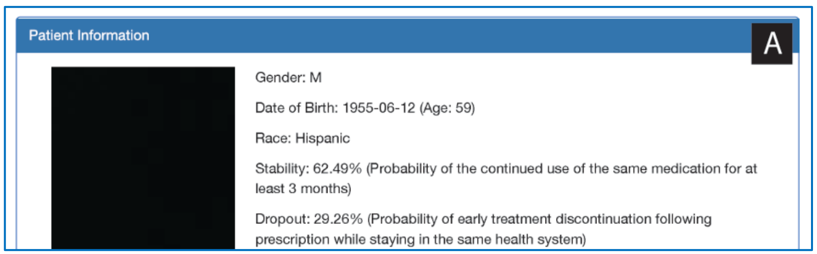
Jacobs, M., Pradier, M. F., McCoy, T. H., Perlis, R. H., Doshi-Velez, F., & Gajos, K. Z. (2021). How machine-learning recommendations influence clinician treatment selections: the example of antidepressant selection. *Translational psychiatry*, *11*(1), 1-9

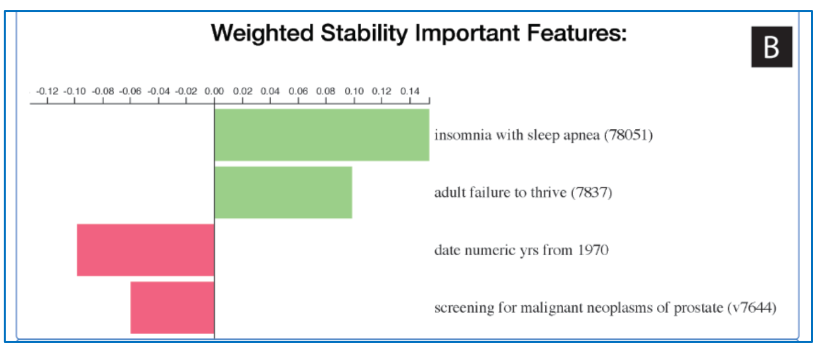
MDD Study 2: Co-design DSTs with clinicians for use in primary care

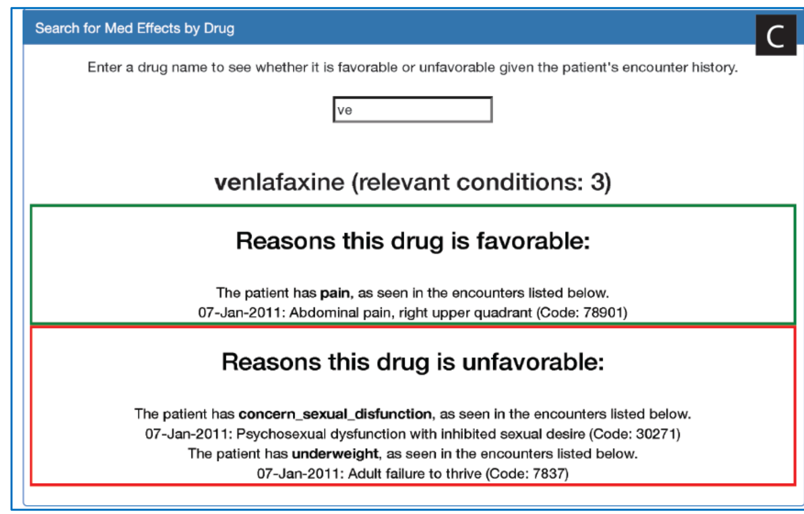


- Focused on primary care
- Consider the social, technical, and organizational issues
- 14 co-design sessions with primary care providers

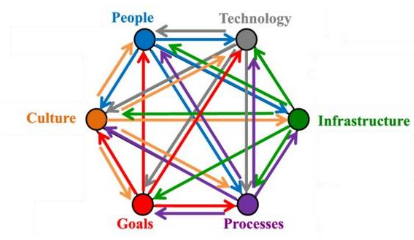
Initial Prototype







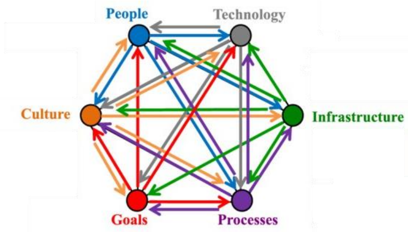
How DSTs may better engage with the healthcare system in to support complex treatment decisions



1. Engage patients in the decision-making process

“Having an option like, patient is also worried about this and that. You can click on the two major side effects and then based on that, a specific drug will come up.”

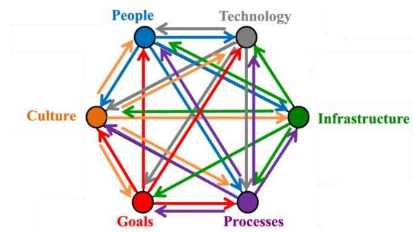
How DSTs may better engage with the healthcare system in to support complex treatment decisions



1. **Engage patients in the decision-making process**
2. **Show a path forward**

"If there is a lower stability and higher dropout that it would be important to then involve more of a care team… I would say, let me have so-and-so in my clinic call you in two weeks."

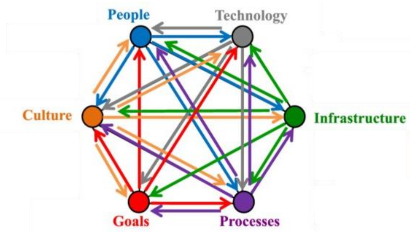
How DSTs may better engage with the healthcare system in to support complex treatment decisions



1. **Engage patients in the decision-making process**
2. **Show a path forward**
3. **Consider resource constraints: trust in the technology will not be decided at each decision point.**

"I don't know if you necessarily need to get into super nitty-gritty details"

How DSTs may better engage with the healthcare system in to support complex treatment decisions



1. **People: Engage patients in the decision-making process**

2. **Processes: Show a path forward**

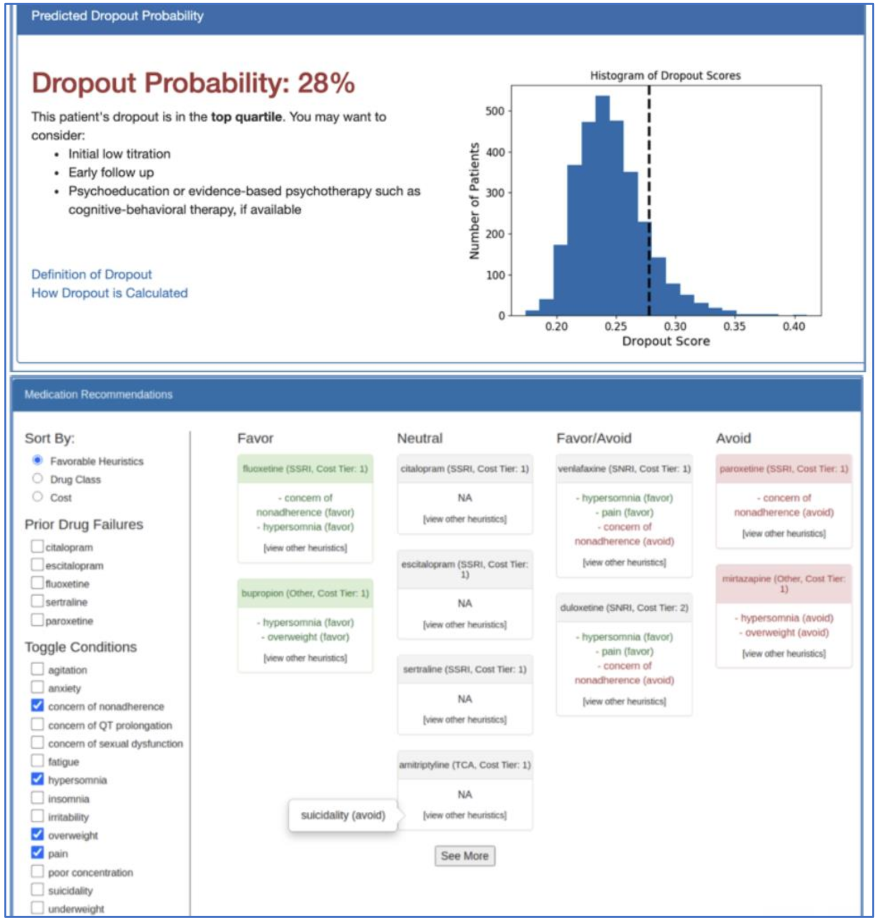
3. **Resource constraints: Do not require trust in the technology to be decided at each decision point**

4. **Domain Knowledge: Adapt designs for instances in which model output contrasts with existing domain knowledge**

Prototype Redesign

Our recommendations:

- ➢ Create multi-user systems
- ➢ Connect to healthcare processes
- ➢ Design for resource constraints
- ➢ Adapt for contrasting information



Jacobs, M., He, J., Pradier, M. F., Lam, B., Ahn, A. C., McCoy, T. H., ... & Gajos, K. Z. (2021). Designing AI for Trust and Collaboration in Time-Constrained Medical Decisions: A Sociotechnical Lens. In CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3411764.3445385

Steps Forward

- Designing with patients: Can we give patients a greater voice in treatment decisions?
- Designing for instances in which the model output diverges from existing domain knowledge.
- How do clinicians use AI tools during patient encounters?

Thank you

Harvard Data Science Initiative

Intelligent Interactive Systems Group, Harvard HCI

Beth Israel Deaconess Medical Center

Massachusetts General Hospital

Krzysztof Z. Gajos, Finale Doshi-Velez, Melanie Fernandez Pradier, Jeffrey He, Roy Perlis, Thomas McCoy, Barbara Lam, Andrew Ahn

Contact:
Maia.Jacobs@northwestern.edu
https://maiajacobs.com
@maia_jacobs