Comparison of a Large Language Model (LLM)-Facilitated Cognitive Debiasing Strategy versus LLM-Generated Diagnostic Feedback Alone in Musculoskeletal Specialty Care: A Randomized Controlled Trial

EMILY H. JAARSMA & DAVID RING

6 June 2025

The University of Texas at Austin
Dell Medical School

Study ID: STUDY00004831
NCT number: TBD

Hypotheses

Primary null hypothesis:

There is no difference in self-reported trust and experience with the clinician between patients who receive a Large Language Model (LLM)-facilitated cognitive debiasing intervention and those who receive LLM-generated diagnostic feedback without a debiasing component during a musculoskeletal specialty care visit.

Secondary hypotheses:

There is no difference in patients' subjective experience between those who receive the LLM-facilitated cognitive debiasing intervention and those who receive LLM-generated diagnostic feedback alone.

Authors: Emily Jaarsma, _____, David Ring, _____,
Acknowledge: Azita Yadollahi, Ali Azarpey
Target journal: TBD
IRB approval required: Yes (as per blanket protocol, see end of document)
ClinicalTrials.gov: Yes

Background and Rationale

Background

A patient's experience of discomfort and incapability is closely linked to their interpretation of bodily sensations.[1-3] Indeed, the human mind acts as an interpretative and story-telling machine that attempts to make sense of bodily sensations through automatic, often unconscious, thoughts. When an individual experiences pain or physical discomfort, they rarely conceptualize their symptoms in terms of specific medical diagnoses, pathophysiology or treatment pathways. Instead, they tend to default to broad interpretative categories such as: (1) Injury or harm that requires protection and rest; (2) Threat to valued roles and activities; or (3) the belief that symptom elimination is the sole path to recovery. While such heuristics can facilitate quick, adaptive responses to bodily signals in times of acute danger (think hand on a hot stove or stepping on a sharp object), they can also be misleading when they are not aligned with the underlying pathology.[4-7] In such cases, faulty heuristics contribute to distress, delay accurate diagnosis, and can lead to unnecessary tests and treatments.[7] The hope lies in helping individuals adopt more flexible, balanced ways of thinking—thereby promoting effective critical thinking skills and fostering an accurate, healthy understanding of symptoms.

Existing tools such as decision aids and question prompt lists are designed to facilitate value-aligned decision-making. However, preliminary evidence suggests such tools may not be sufficient to challenge rigid or unhelpful cognitive patterns.[8]

Rationale

Cognitive debiasing strategies offer a promising approach to address both clinician and patient faulty heuristics and promote flexible, reflective, critical thinking. By fostering a more adaptive and nuanced understanding of bodily signals, such strategies may have the potential to align decisions with what matters most to a patient (their values) and improve health when compared to traditional educational and care approaches.[9] Additionally, recent advances in Artificial Intelligence (AI), particularly the emergence of Large Language Models (LLMs), offer new strategies to enhance cognitive debiasing interventions. LLMs can be used to analyze and synthesize patients' verbal responses to gently reflect back their beliefs in simple, supportive language.[10] In doing so, this approach may help patients gain awareness of their own interpretive tendencies and invites consideration of alternative, less distressing explanations. LLMs can also support clinicians by flagging potentially unhelpful beliefs prior to a consultation, thereby facilitating more tailored and empathetic communication.

This randomized controlled trial aims to evaluate whether a cognitive debiasing strategy facilitated by an LLM can improve decision quality and patient experience in musculoskeletal specialty care, when compared to use of ChatGPT for diagnostic feedback only (without debiasing).

Proposed methods

Study design: Randomized Trial

Recruitment methods: New patients seeking musculoskeletal specialty care will be invited to participate in the study. After obtaining informed consent and prior to their consultation with the musculoskeletal specialist, patients will be randomized to either receive an LLM-facilitated debiasing checklist (Appendix 1) or LLM-generated diagnostic feedback. Randomization will occur via an online random number generator (*random.*org). All recruitment will be conducted by enrollers who have received one-on-one training from the lead investigator.

Intervention (random number generator = 1): Cognitive debiasing strategy (Appendix 1)

Control (random number generator = 0): LLM diagnosis (Appendix 2)

Inclusion Criteria:

Adults (18+)

New or return patient seeking musculoskeletal specialty care

Total combined score on the 6 feelings and thoughts items of > 10* (Appendix 3)

English-speaking

Pre-visit diagnosis of chronic, non-traumatic musculoskeletal condition

Exclusion criteria*:

Any impairment preventing completion of surveys on a tablet

**Note: participant screening for inclusion (and reasons for exclusion) will be tracked via the RedCap survey platform*

Measured variables

Response variables:

TRECS

Subjective experience measures (Appendix 3)

Explanatory variables:

Study arm

Age

Gender

Social health

Personal health agency (ANCHOR)

Unhelpful thoughts (Appendix 4)

Unhelpful feelings (Appendix 4)

Statistical analysis

Demographic and baseline characteristics will be summarized using descriptive statistics. Continuous variables will be reported as means and standard deviations (or medians and interquartile ranges if not normally distributed), and categorical variables as frequencies and percentages.

Between-group comparisons for the primary and secondary outcome measures (i.e., total scores on Trust in Physician Scale (e.g., TRECS-7), and three subjective experience questions) will be conducted using independent samples t-tests. Where appropriate, assumptions of normality and homogeneity of variance will be checked. If these are violated, non-parametric alternatives (e.g., Mann–Whitney U test) will be applied.

Bivariate analyses (i.e., Pearson's or Spearman's correlations, t-tests, or chi-square tests depending on variable types) will be used to identify potential associations between demographic/clinical factors and each outcome measure. Variables found to be statistically significant in bivariate analysis ($p < 0.10$) will be included in multivariable linear regression models to identify independent predictors of outcome scores.

Statistical significance will be set at $p < 0.05$.

Sample size calculation

A priori sample size estimation was conducted using G*Power (version 3.1) for a two-tailed independent samples t-test, comparing mean Trust and Experience in Clinician Scale (TRECs) scores between two groups (primary outcome). Assuming a medium effect size (Cohen's d = 0.5), a significance level (α) of 0.05, and 80% power (1–β), the required sample size was calculated to be 128 participants (64 per group). To account for an anticipated attrition or non-complete response rate of approximately 15%, the final target sample size was set at 150 participants (75 per group).

References

1. Vranceanu AM, Barsky A, Ring D. Psychosocial aspects of disabling musculoskeletal pain. J Bone Joint Surg Am. 2009 Aug;91(8):2014–8.

2. Vranceanu AM, Bachoura A, Weening A, Vrahas M, Smith RM, Ring D. Psychological factors predict disability and pain intensity after skeletal trauma. J Bone Joint Surg Am. 2014 Feb 5;96(3):e20.

3. Teunis T, Al Salman A, Koenig K, Ring D, Fatehi A. Unhelpful Thoughts and Distress Regarding Symptoms Limit Accommodation of Musculoskeletal Pain. Clin Orthop Relat Res. 2022 Feb 1;480(2):276–83.

4. O'Connor CM, Kortlever J, Ring D. Misinformation in Patient Handouts About Upper Extremity Conditions. J Patient Exp. 2020 Dec;7(6):1331–5.

5. Lemmers M, Versluijs Y, Kortlever JTP, Gonzalez AI, Ring D. Misperception of Disease Onset in People with Gradual-Onset Disease of the Upper Extremity. J Bone Joint Surg Am. 2020 Dec 16;102(24):2174–80.

6. Becker SJE, Briet JP, Hageman MGJS, Ring D. Death, taxes, and trapeziometacarpal arthrosis. Clin Orthop Relat Res. 2013 Dec;471(12):3738–44.

7. Ring D, Leopold SS. Editorial: The Goal is Health, Not Surgery. Clin Orthop Relat Res. 2025 Apr 1;483(4):551–3.

8. Kortlever JTP, Ring D, Schuurman AH, Coert JH, Vagner GA, Reichel LM. Patient Perspectives on Decision Making for Carpal Tunnel Syndrome. J Hand Surg. 2019; 44(11): 940-946.

9. Cordero R, Ponce H, Ring D, Tonn M, Ramtin S. A Debiasing Checklist Has the Potential to Improve Efficiency, Effectiveness, and Outcomes in Claims of Work Injury. J Occup Environ Med. 2025; 67(6): 419-423.

10. Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt JN, Laleh NG, et al. The future landscape of large language models in medicine. Commun Med. 2023 Oct 10;3(1):141.

Appendix 1: Debiasing strategies Survey

Part 1 - "What do you believe?"
Patients are asked to reflect on and respond to questions about their beliefs regarding their symptoms. Responses are collected as verbatim transcribed text using a tablet or laptop. Questions include:

- How many people experience these symptoms at some time during their life?
- What's usually behind these symptoms?
- What percentage of people can stick to their daily routine and achieve what they want in spite of these symptoms?
- How often do these symptoms completely change someone's life?
- If these symptoms are not diagnosed correctly, how serious are the consequences? What could they be?

Part 2 - LLM-facilitated summary of current beliefs
The patient's responses are input into ChatGPT using a structured prompt (Appendix 5). ChatGPT generates a brief, patient-facing summary of the patient's current symptom beliefs, written in simple, supportive language. This summary is shared with the patient.

Part 3 - "Reflect on your beliefs"
Patients are invited to reflect on their earlier responses and the LLM generated summary. Again, their answers are transcribed verbatim via iPad or laptop. Reflection prompts include:

- Considering this summary of your current beliefs about your body and its sensations: Have you considered the opposite?
- Is there a downside to your current beliefs about your body? What?
- Could there be another, less worrisome explanation for your current symptoms that you haven't considered?
- What circumstances and emotions might be influencing your thinking about this?

Part 4- Sharing with clinician

The patient's reflection responses are analyzed by ChatGPT using a second prompt that asks the model to identify potential unhelpful beliefs in supportive language (Appendix 6). This output is provided to both the patient and the clinician, ahead of the consultation.

Appendix 2: LLM-Facilitated Diagnostic Feedback (Control Arm)

Part 1 – What is happening?

Patients are asked to describe their presenting musculoskeletal symptom. Responses are collected as verbatim transcribed text using a tablet or laptop and input into ChatGPT using a structured prompt (Appendix 7). A likely diagnosis and brief description is generated by ChatGPT. This output is chared with both the patient and the clinician, ahead of the consultation.

Appendix 3: Subjective Experience

Questions provided to both control and intervention arm

Quantitative scale 0 (strongly disagree) to 100 (strongly agree) for each question

1. The computer gave me an accurate description/summary
2. The computer put me in a healthy mindset regarding my symptoms
3. The computer increased my confidence to manage this on my own

Appendix 4: Questions of Unhelpful Thoughts and Feelings

Five-point Likert scale between 1 (Strongly disagree) and 5 (Strongly agree) (scores range from 6-30)

1. I feel I can't stand it anymore.
2. I keep thinking about how much it hurts.
3. My problem makes me feel awful and it overwhelms me.

1. My problem has put my body at risk for the rest of my life.
2. Pain always means I have injured my body.
3. I will never be happy again as long as I have pain.

**Point on Likert scale corresponds to scoring system for each question. Scores for individual questions are then added to create a score between 3- 15 for each part (total score 6-30)*

Appendix 5: ChatGPT Prompt 1

Generate a short, simple summary of what this person appears to believe or feel about their condition.

Write **directly to the patient**, not the clinician.

Do **not** use subheadings. Keep the tone **non-judgmental, clear, and supportive**.

You may **gently read between the lines** to identify patterns or themes, but do **not** simply repeat the patient's responses word-for-word. Do **not** reference or challenge any diagnosis.
The goal is to help the patient become more aware of how they are currently thinking about their symptoms.

Do **not** try to fix or reframe the beliefs—just summarize them neutrally and concisely.

Focus only on what this patient has said during this current session—do not refer to any prior interactions or assumptions.

Appendix 6: ChatGPT Prompt 2

The following is a patient's verbatim reflection about their health and symptoms. Please carefully read their responses and gently identify any signs of emotional distress (e.g., worry, fear, hopelessness) or patterns of thinking that may be unhelpful (e.g., catastrophizing, rigid beliefs, overgeneralization).

Present your response **in a kind and supportive tone**, as if you're speaking directly to the patient. Use plain language, and aim to **raise awareness rather than correct**. Do not challenge their beliefs, diagnosis, or experience. Instead, **invite them to consider alternative perspectives** or gently highlight how certain thoughts might be making things feel more difficult.

Please avoid medical jargon or overly technical language. The goal is to help the patient reflect on how their thoughts might be influencing how they feel and encourage more flexible, balanced thinking.

Then, provide a concise summary (in a separate paragraph) for the clinician, identifying any key unhelpful beliefs, emotional distress indicators, or cognitive biases you observed. Include suggestions that may help guide more tailored and empathetic communication during the consultation.

Appendix 7: ChatGPT Prompt 3 (Control arm)

"When a hypothetical patient presents their symptoms, adhere to the following simplified guidelines in your analysis and response:

1. Purpose and Role

- Provide educational insight by exploring the most likely hypothetical diagnoses based on recent evidence-based clinical research.
- Responses should focus on academic and informative perspectives, not substitute professional medical care.

2. Provide the Most Likely Diagnosis First (MAIN Objective)

- Clearly state the most likely or accurate diagnosis based on the provided information.
- State the most likely diagnosis confidently, without hesitation or disclaimers.

3. Adapt to Health Literacy Levels

- Assume the patient has average or below-average health literacy with limited familiarity with medical terminology.
- Use plain language to ensure accessibility and understanding.

4. Critical Thinking on Patient Expression

- Consider the possibility that patients may **catastrophize their symptoms**, using phrases like "I can't take it anymore" or "It hurts like crazy."
- Integrate this into your critical thinking by:
    - Exploring both physical and psychological factors influencing symptom perception.
    - Balancing empathy with objective analysis.

5. Emphasis on Recent Research

- Incorporate findings from the latest clinical studies (2018–2024), including systematic reviews and meta-analyses from reputable sources such as JAMA, The Lancet, NEJM, or Cochrane reviews.
- Highlight emerging research that challenges traditional guidelines, when applicable.

6. Detailed Clinical Context

- Use patient-specific details, such as age, gender, medical history, and lifestyle, to tailor your response.

7. Comparison Between Guidelines and Research

- Where relevant, compare mainstream clinical guidelines with the latest research, highlighting differences and their implications.

8. Structured Approach with Citations

- Provide a clear, structured response that includes reasoning and evidence from recent clinical research (2018–2024).
- Cite reputable sources such as Cochrane reviews, JAMA, The Lancet, or NEJM. If specific citations are unavailable, reference the general source of evidence.

Output Restriction.

- Respond with **only the most likely diagnosis**, without additional reasoning or elaboration unless explicitly requested.
- Avoid differential diagnoses, pathophysiology, or management details.

When the patient's symptoms are provided, strictly adhere to these guidelines while forming your response. This is a new patient, do not expand on the previous patient query.

The University of Texas at Austin

Office of Research Support & Compliance
Institutional Review Board
P.O. Box 7426, Campus Code A3200
Austin, Texas 78713
T: 512-232-1543 F: 512-471-8873
Email: irb@austin.utexas.edu
www.research.utexas.edu/ors

EXEMPT DETERMINATION

August 24, 2023

FWA # 00002030

David Ring
1601 TRINITY ST BLDG B
AUSTIN, TX 78712

+1 512 495 5067
david.ring@austin.utexas.edu

Dear David Ring:

On 8/23/2023, the IRB reviewed the following submission:

Type of Review:	Initial Study
Special Determinations:	None
Title:	Patient factors and PROM/PREM/ and PRAM
Investigator:	David Ring
IRB ID:	STUDY00004831
Funding:	None
Grant ID:	None
IND, IDE, or HDE:	None
Approval Date:	8/23/2023
Documents Reviewed:	• All_questions.docx, Category: Other; • HRP - UT902 - Template IRB Proposal Exempt Submission Apr2023.docx, Category: IRB Protocol; • mental_health_service.pdf, Category: Other; • PROMIS Bank v1.0 Emotional Distress Depression.pdf, Category: Other; • PROMIS Bank v1.0 Physical Function.pdf, Category: Other; • PROMIS Bank v1.1 - Pain Interference.pdf, Category: Other; • PROMIS Bank v1.2 - Physical Function - Mobility.pdf, Category: Other;

Office of Research Support & Compliance
Institutional Review Board
P.O. Box 7426, Campus Code A3200
Austin, Texas 78713
T: 512-232-1543 F: 512-471-8873
Email: irb@austin.utexas.edu
www.research.utexas.edu/ors

	• PROMIS Bank v1.2 - Physical Function - Mobility.pdf, Category: Other; • PROMIS Bank v1.2 - Upper Extremity.pdf, Category: Other; • PROMIS SF v1.1 - Global Health.pdf, Category: Other; • PROMIS v1 0_ED-Depression_Item Bank_SPA_26Nov12.pdf, Category: Other; • PROMIS v1 1_Pain Interference_Item Bank_SPA_08Jan2014.pdf, Category: Other; • PROMIS v1 2_Physical Function_Upper Extremity_SPA_08Jan2014.pdf, Category: Other; • PROMIS v1.0-1.1_Global_SPA_15Feb12.pdf, Category: Other; • PROMIS v1.2 - Physical Function - SF 20a.pdf, Category: Other; • PROMIS v1.2_Physical Function_Mobility_SPA_05Nov2015.pdf, Category: Other; • PROMIS v1.2_Physical Function_SF 20a_SPA_02Nov2015.pdf, Category: Other; • Verbal Consent.pdf, Category: Consent Form;

The IRB determined that this protocol meets the criteria for exemption from IRB review under 45 CFR 46.104 (2)(ii) Tests, surveys, interviews, or observation (low risk).

In conducting this protocol, you are required to follow the requirements listed in the Investigator Manual (HRP-103) and UT IRB Policies and Procedures (HRP-UT1000), which can be found by navigating to the IRB Library, General tab, within UTRMS-IRB.

Ongoing IRB review and approval by this organization is not required. This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. Modifications that involve a change in PI, increase risk, or otherwise affect the exempt category or the criteria for exempt determination must be submitted as a modification. Investigators are strongly encouraged to contact the IRB staff to describe any changes prior to submitting an amendment/modification.

If you have any questions, contact the RSC by phone at 512 -232-1543 or via e-mail at irb@austin.utexas.edu.

Sincerely,

The University of Texas at Austin

Office of Research Support & Compliance
Institutional Review Board
P.O. Box 7426, Campus Code A3200
Austin, Texas 78713
T: 512-232-1543 F: 512-471-8873
Email: irb@austin.utexas.edu
www.research.utexas.edu/ors

Institutional Review Board

University of Texas at Austin

cc:
David Ring (PI), Sina Ramtin (Primary Contact), Sina Ramtin (Proxy)