Patient Experience and Trust When Viewing Original Medical Notes versus Large Language Model (LLM)-Generated Plain Language Summaries

NCT ID: Not yet assigned

The University of Texas at Austin
Dell Medical School

Hypotheses

Primary hypothesis:

There are no factors associated with patient ratings of trust and experience with the clinician (TRECS), including whether patients viewed their original medical record entry or an LLM-generated plain language summary based on a curated version of the record (where PHI is removed) prior to the visit and other patient factors (demographics, mental health clusters).

Secondary hypotheses:

There are no factors associated with emotional responses to viewing one's own medical record entry, including whether patients view the original entry or an LLM-generated plain language summary based on a curated version of the record (where PHI is removed) prior to the visit and other patient factors (demographics, mental health clusters).

Qualitative hypothesis:

What themes does an LLM identify in patient verbatim comments regarding viewing their medical record entry or an LLM plain language summary?

Authors: Niek Parser, BSc; Zane Ahlfinger, BSc; David Ring, MD, PhD; Emily Jaarsma, MD
Target journal: TBD
IRB approval: Approved (STUDY00008863)

Background and Rationale

Background

As patient access to electronic medical records becomes more widespread, understanding how individuals emotionally respond to their clinical documentation has become increasingly important. A qualitative study analyzing 600 medical encounter notes authored by 138 clinicians identified distinct patterns of language reflecting clinicians' attitudes toward patients. Five categories of negative language were described (questioning credibility, disapproval, stereotyping, labeling patients as "difficult," and unilateral decision-making), alongside six categories of positive language (compliments, approval, self-disclosure, minimizing blame, personalization, and collaborative decision-making).[1] Prior research has demonstrated that access to clinical notes ('open notes') can influence patient experience, with potential benefits including improved understanding, greater trust, enhanced perceived quality of care, and increased engagement in self-care and health management. These benefits appear particularly pronounced with patients with lower education attainment or those from ethnic minority groups. Despite this, fewer than half of clinicians routinely discuss shared notes with patients during visits.[2–5]

Rationale

Language choices within clinical documentation may meaningfully shape how patients perceive their diagnosis, their relationship with healthcare providers, and their emotional well-being. For example, in a survey study of 100 healthy companions of orthopedic hand surgery patients, participants evaluated 19 commonly used medical terms and their synonyms using the Self-Assessment Manikin (SAM). Terms such as "pain" was rated more negatively than alternatives such as "discomfort" or "ache", while "rupture" elicited a more negative response than "tear" or

"defect". These findings suggest that frequently used clinical terminology may also unintentionally elicit negative emotional reactions.[6] Patients may be particularly vulnerable to language-related distress when reading their own medical records. The tone, complexity, and structure of these notes, especially in surgical specialties, may impact patient trust and overall comfort, and, in some cases, may unintentionally erode trust or intensify anxiety. Emerging evidence suggests that plain-language summaries might mitigate these effects. In a small study of 20 participants, artificial intelligence was used to generate plain-language medical notes, which were perceived as more useful, supportive of the patient-clinician relationship, and empowering for patients' understanding in their health.[7] Although the importance and benefits of accessible visit notes have been well established, there remains limited empirical evidence regarding the optimal tone, structure, and language of medical documentation from the patient perspective.

Proposed methods

Study design:
Randomized Controlled Trial

Recruitment methods:
Prior to specialist consultation, return patients will be recruited to participate in the study.

Randomization:
Each participant will be equally randomized (1:1) to review their original medical record entry from their most recent clinical visit or an LLM-simplified version of their medical record. Randomization will occur using an online random number generator (*random.org*).

Intervention:
Original medical records from patients will be simplified using an LLM. This LLM will focus on clarifying the notes and avoiding technical jargon.

Measurement:
Prior to a consultation, both the intervention group as the non-intervention group will complete an online questionnaire that includes demographic information, a validated measure of trust in their clinician (e.g., the TRECS scale) and a Likert-based scale assessing emotional responses.

Additionally, participants will be invited to respond to a prompt regarding how it felt to review the material. We will automatically transcribe verbatim text into a tablet and use an LLM to identify themes.

Inclusion Criteria:
Adult (18-89)
Seeking musculoskeletal specialty care
English language
Return patient to clinic

Exclusion criteria:
Any impairment precluding completion of a survey on a tablet

Measured variables
Response variables:
TRECS 7[8]

Emotional responses to viewing one's own medical record entry measured via a sliding scale from 0 to 100 anchored at the ends with:
Sad / happy
Worried / at ease
Your doctor is caring / your doctor is not caring
Uncomfortable / comfortable
Affects communication with clinician / does not affect communication with clinician

Themes identified by an LLM in verbatim text regarding reading the material.

Explanatory variables:

Age
Gender (male/female/non-binary)
Type of insurance (Medicare, Medicaid, Private, Military, Self-pay (no health insurance), MAP, Other)
Annual Household Income (Less than $15,000, From $15,000 to $29,999, From $30,000 to $49,999, From $50,000 to $99,999, More than $100,000, Prefer not to answer)
Level of education (Elementary school, High school, Some college, Graduated college, Post-college graduate degree (e.g., PhD), Other)
Employment status (employed, unemployed, disabled, other (retired, student))
Marital status (Single, married or partner, other (divorced, separated, windowed)
The patient has/has not viewed their medical record before

Specialty of original visit: Arthroplasty, hand/upper extremity, shoulder/sports, spine
Clinical: diagnosis, pain severity (VAS/NRS),
Condition Location: LE, UE, Spine

Perceived Health Status: very poor / poor / fair / good / very good
Brief Agency (3 Q)[9]
Brief Social Health (3Q)[10]
Questions of Unhelpful Thoughts and Distress[11]

Statistical analysis
Descriptive statistics:
Demographic and baseline characteristics of participants will be summarized using descriptive statistics. Continuous variables will be reported as means ± standard deviations or medians with interquartile ranges if not normally distributed. Categorical variables will be summarized as

frequencies and percentages. Subgroups with fewer than 15 participants will be consolidated for analysis.

Cluster analysis: To identify natural groupings of participants based on psychosocial measures, an unsupervised k-means cluster analysis will be conducted using variables including interpersonal trust, unhelpful thoughts, distress, social health, and personal health agency. The optimal number of clusters will be determined using the elbow method and silhouette coefficients. Once clusters are identified, mean TRECS-7 scores and emotional response scores will be compared across clusters using ANOVA or Kruskal-Wallis tests, depending on data distribution. Post-hoc pairwise comparisons between clusters will be performed using chi-square tests or appropriate nonparametric alternatives, with applied to control for multiple comparisons.

Bivariate Analyses: Bivariate associations between demographic/clinical characteristics and primary or secondary outcomes will be evaluated using appropriate tests depending on variable type:

- Continuous outcomes: independent samples t-tests or Mann-Whitney U tests (if non-normal)
- Categorical outcomes: chi-square tests or Fisher's exact tests
- Correlations between continuous variables: Pearson's or Spearman's correlation coefficient

Variables showing potential association with outcomes at $p < 0.10$ in bivariate analyses will be considered for inclusion in multivariable models.

Multivariable analysis: Multivariable linear regression models will identify independent predictors of trust, experience, and emotional response outcomes. Explanatory variables will include demographics, specialty of visit, psychosocial clusters, and prior experience viewing medical records.

All quantitative analyses will be performed using **Stata version 18.0** (StataCorp, College Station, TX)

Sample size calculation:

A priori sample size estimation was conducted using G*Power (version 3.1) for multiple linear regression. Assumptions included:

- Medium effect size ($f^2 = 0.15$, corresponding roughly to Cohen's $d \approx 0.5$ for continuous outcomes)
- $\alpha = 0.05$
- Power = 0.80
- Number of predictors = 10 (demographics, specialty, psychosocial clusters, prior record viewing)

Under these assumptions, the minimum required sample size is 118 participants to detect a medium effect. To account for potential incomplete surveys or attrition (~10–15%), the target recruitment is **~135 participants** (approximately 67-68 participants per study arm), which will provide adequate power to detect meaningful differences in TRECS-7 and emotional response outcomes, as well as enable robust cluster analysis.

References

1. Park J, Saha S, Chee B, Taylor J, Beach MC. Physician Use of Stigmatizing Language in Patient Medical Records. *JAMA Netw Open*. 2021;4(7):e2117052. doi:10.1001/jamanetworkopen.2021.17052

2. Delbanco T, Walker J, Bell SK, et al. Inviting Patients to Read Their Doctors' Notes: A Quasi-experimental Study and a Look Ahead. *Ann Intern Med*. 2012;157(7):461-470. doi:10.7326/0003-4819-157-7-201210020-00002

3. Walker J, Leveille S, Bell S, et al. OpenNotes After 7 Years: Patient Experiences With Ongoing Access to Their Clinicians' Outpatient Visit Notes. *J Med Internet Res*. 2019;21(5):e13876. doi:10.2196/13876

4. Gerard M, Chimowitz H, Fossa A, Bourgeois F, Fernandez L, Bell SK. The Importance of Visit Notes on Patient Portals for Engaging Less Educated or Nonwhite Patients: Survey Study. *J Med Internet Res*. 2018;20(5):e9196. doi:10.2196/jmir.9196

5. Bell SK, Mejilla R, Anselmo M, et al. When doctors share visit notes with patients: a study of patient and doctor perceptions of documentation errors, safety opportunities and the patient–doctor relationship. *BMJ Qual Saf*. 2017;26(4):262-270. doi:10.1136/bmjqs-2015-004697

6. Vranceanu AM, Elbon M, Adams M, Ring D. The emotive impact of medical language. *Hand N Y N*. 2012;7(3):293-296. doi:10.1007/s11552-012-9419-z

7. Bala S, Keniston A, Burden M. Patient Perception of Plain-Language Medical Notes Generated Using Artificial Intelligence Software: Pilot Mixed-Methods Study. *JMIR Form Res*. 2020;4(6):e16670. doi:10.2196/16670

8. Brinkman N, Looman R, Jayakumar P, Ring D, Choi S. Is It Possible to Develop a Patient-reported Experience Measure With Lower Ceiling Effect? *Clin Orthop*. 2025;483(4):693-703. doi:10.1097/CORR.0000000000003262

9. Santesso N, Rader T, Wells GA, et al. Responsiveness of the Effective Consumer Scale (EC-17). *J Rheumatol*. 2009;36(9):2087-2091. doi:10.3899/jrheum.090363

10. Brinkman N, Broekman M, Teunis T, Choi S, Ring D, Jayakumar P. A New Measure of Quantified Social Health Is Associated With Levels of Discomfort, Capability, and Mental and General Health Among Patients Seeking Musculoskeletal Specialty Care. *Clin Orthop*. 2025;483(4):647-663. doi:10.1097/CORR.0000000000003394

11. Teunis T, Al Salman A, Koenig K, Ring D, Fatehi A. Unhelpful Thoughts and Distress Regarding Symptoms Limit Accommodation of Musculoskeletal Pain. *Clin Orthop*. 2022;480(2):276-283. doi:10.1097/CORR.0000000000002006

Appendix 1: Prompts

Intervention Group
Prompt for generating the plain language summary based on a curated version of the record:

- Delete PHI (personal health information, including name, date of birth, address, phone number)
- Delete all macros (appointment note, diagnoses, billing information, minutes spent in visit)
- Delete PT notes
- Replace **normal** findings (PE, imaging) for 'Exam otherwise normal'
- Ask ChatGPT the following: 'In one brief paragraph please summarize this note using plain language'

Intervention and Control Group
Emotional responses to viewing one's own medical record entry measured via a sliding scale from 0 to 100 anchored at the ends with:

- Sad / happy
- Worried / at ease
- Your doctor is caring / your doctor is not caring
- Uncomfortable / comfortable
- Affects communication with clinician / does not affect communication with clinician

Qualitative questions (to be recorded via speech to text dictation, no audio recording):
How did you feel reading your medical records before the visit with the clinician? (Please elaborate)

Appendix 2: Trust and Experience with the Clinician Scale (TRECS7)[8]

The Trust and Experience with the Clinician Scale (TRECS-7) is a validated 7-item scale that measures patients' trust in and experience with their clinician during a medical consultation. Designed to minimize ceiling effects, it enables more sensitive detection of variation in patient experience across different clinical interactions. Each of 7 statements is scored from 0-4 (strongly disagree, disagree, neutral, agree, strongly agree), resulting in a total score between 0 and 28. Higher scores indicate greater perceived trust in the clinician.

Source: https://pubmed.ncbi.nlm.nih.gov/39466401/

The clinician seems concerned about me and my family.

a) Strongly disagree

b) Disagree
c) Neutral
d) Agree
e) Strongly agree

The clinician seemed to know all important information about my medical history.

a) Strongly disagree
b) Disagree
c) Neutral
d) Agree
e) Strongly agree

The clinician would tell me if a mistake was made about my treatment.

a) Strongly disagree
b) Disagree
c) Neutral
d) Agree
e) Strongly agree

I trust the clinician so much I will always try to follow their advice.

a) Strongly disagree
b) Disagree
c) Neutral
d) Agree
e) Strongly agree

I would trust the clinician with my life.

a) Strongly disagree
b) Disagree
c) Neutral
d) Agree
e) Strongly agree

The clinician would only ever think about what is best for me.

a) Strongly disagree
b) Disagree
c) Neutral
d) Agree
e) Strongly agree

I will have complete trust in the clinician.

a) Strongly disagree

b) Disagree
c) Neutral
d) Agree
e) Strongly agree

Appendix 3: Brief Personal Health Agency Questions[9]

Health agency will be measured using a brief 3-item version of the Effective Consumer Scale (EC-17). Patients rate their agreement with each statement on a 5-point scale from 0 (strongly disagree) to 4 (strongly agree), resulting in a total score ranging from 0 to 12. Higher scores indicate a greater ability to self-manage health (health agency).
Source: https://pubmed.ncbi.nlm.nih.gov/19738218/

1. I can set realistic goals about the management of my disease.
2. I can organize my life to act on decisions about how to manage my disease.
3. I feel confident in making decisions about my health.

Appendix 4: Brief Social Health Questions[10]

Social health will be assessed using a brief, 3-item short form of a validated Social Health Questionnaire. Patients rate their agreement with each statement on a 5-point scale from 0 (strongly disagree) to 4 (strongly agree), yielding a total score between 0 and 12. Higher scores indicate greater social disadvantage.

Source: https://pubmed.ncbi.nlm.nih.gov/39915110/

I am worried about losing my housing or have already lost my housing
0) Strongly disagree
1) Disagree
2) Neutral
3) Agree
4) Strongly agree

Within the last 12 months, I was worried that my food would run out before I got money to buy more
0) Strongly disagree
1) Disagree
2) Neutral
3) Agree
4) Strongly agree

I have trouble paying my electricity bill

0) Strongly disagree
1) Disagree
2) Neutral
3) Agree
4) Strongly agree

Appendix 5: Questions of Unhelpful Thoughts and Distress[11]

Unhelpful thoughts and feelings will be assessed using a validated 6-item scale, adapted from the Pain Catastrophizing Scale and the Negative Pain Thoughts Questionnaire. The scale is divided into two parts, each containing 3 statements. Patients rate their agreement with each statement on a 5-point scale from 0 (strongly disagree) to 4 (strongly agree), yielding a total score ranging from 0 to 12 for each part. Higher scores reflect greater levels of unhelpful thoughts and emotional distress, respectively.

Source: https://pmc.ncbi.nlm.nih.gov/articles/PMC8747479/

Distress

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.

Unhelpful Thoughts

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.