Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comparing AI-generated and evidence-based recommendations for chronic pain and tDCS parameters: a descriptive qualitative study

Edson Silva-Filho

emailmeneses@gmail.com

Federal University of Rio Grande do Norte

Denise Mota Araripe Pereira Fernandes

Federal University of Paraíba

Maria Isabel França

Federal University of Rio Grande do Norte

Letícia Amaro Vieira

Federal University of Rio Grande do Norte

Maria Thereza Micussi

Federal University of Rio Grande do Norte

Rodrigo Pegado

Federal University of Rio Grande do Norte

Research Article

Keywords: artificial intelligence, transcranial direct current stimulation, chronic pain

Posted Date: May 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9270660/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Objective

To describe the qualitative concordance between AI-generated recommendations and established evidence for chronic pain management, focusing on pharmacological treatments and transcranial direct current stimulation (tDCS) parameters.

Method

This is a descriptive qualitative study used evidence-based reference standards, including clinical practice guidelines, Cochrane systematic reviews, systematic reviews with high-certainty evidence (GRADE level 1 or 2), and randomized controlled trials. Data were obtained from two AI platforms (ChatGPT Go and OpenEvidence) between January and March 2026. AI-generated recommendations were examined using qualitative content analysis and compared evidence-based standards. Concordance was classified as full agreement, partial agreement, or no agreement.

Results

Thirteen chronic pain syndromes were analyzed. Regarding pharmacological management, full concordance between AI-generated recommendations and evidence-based standards was observed in 38.5% of cases, partial concordance in 53.8%, and no concordance in 7.7%. In contrast, recommendations concerning tDCS parameters demonstrated variability, particularly with respect to electrode montage precision and session standardization.

Conclusion

AI platforms demonstrated substantial potential to support clinical decision-making in chronic pain management, showing good qualitative concordance with established evidence. AI-generated outputs do not account for individual patient characteristics, clinical context, or therapeutic prioritization. Therefore, AI should be considered a supportive tool rather than a substitute for clinical judgment.

INTRODUCTION

Artificial intelligence (AI) has been increasingly adopted across diverse areas of biomedical science, with accelerating integration into clinical and translational research. AI-based tools increasingly support diagnostic accuracy, clinical reasoning, and procedural safety, with documented advances in anesthesiology (Lopes, Rocha, & Guimarães-Pereira, 2024), neurology (D'Antoni et al., 2022), endocrinology (Shahriari et al., 2023), primary care (Gomez-Cabello et al., 2024), and orthopedics (Lex et al., 2023). Beyond diagnostic support, AI systems are also being explored as tools to inform therapeutic decision-making. In parallel, complementary therapeutic approaches such as neuromodulation have gained attention in the management

of chronic pain, raising questions about how AI-generated recommendations may align with established evidence in this domain (Silva-Filho; Pegado, 2024).

ChatGPT™ and OpenEvidence™ are AI platforms widely used to access medical information and support clinical decision-making. Both offer a user-friendly interface capable of providing rapid responses. However, a fundamental distinction exists between them. ChatGPT typically generates answers without explicit reference to verifiable evidence, which may limit the reliability of its outputs in clinical contexts. In contrast, OpenEvidence systematically grounds its responses in the best available scientific literature, increasing transparency and evidence-based support. Given the potential influence of AI in clinical decision-making, it is very important to analyze the outputs in conditions that are relevant for the general population, such as chronic pain (OpenAI, 2026; OpenEvidence, 2026).

Chronic pain affects individuals globally and represents a major public health concern. Approximately 20% of the U.S. population experiences chronic daily pain (Yong, 2022). This condition is frequently associated with comorbidities such as depression (Aaron et al., 2025). Several pharmacological approaches are considered standard treatments for chronic pain, including non-steroidal anti-inflammatory drugs (Enthoven et al., 1996) and opioids (Busse et al., 2018). In addition, transcranial direct current stimulation (tDCS), a non-invasive neuromodulation technique, has demonstrated meaningful clinical effects across various chronic pain conditions (O'Connell et al., 2018).

It is important to mention that although scientific literature provides validated, evidence-based guidance for clinicians and general population, the information generated by AI platforms may not always align with these findings (Perez et al., 2026). Furthermore, translation and language-processing issues must be carefully evaluated to prevent the introduction or amplification of inequities in healthcare (Martos et al., 2025). Clinical decision-making should rely on robust, and low-bias evidence. However, the increasing availability and convenience of AI-generated outputs pose a risk of influencing clinical reasoning without adequate critical appraisal. To date, few PubMed-indexed studies have examined the application of OpenEvidence, and none have performed direct comparisons or qualitative concordance analyses between OpenEvidence outputs, ChatGPT, and the best available evidence for pharmacological chronic pain management and tDCS parameters.

Considering theses assumptions, the present study aims to describe and analyze the qualitative concordance between AI-generated recommendations and established evidence for chronic pain management, particularly concerning pharmacological approaches and tDCS parameters.

METHOD

This descriptive qualitative study compared AI-generated recommendations with the best available evidence for chronic pain management. The evidence bases primarily included clinical guidelines and systematic reviews of randomized controlled trials. In the absence of guidelines or systematic reviews for a specific condition, individual clinical trials with low risk of bias were considered.

Data were obtained from two artificial intelligence platforms (ChatGPT Go and OpenEvidence) between January and March 2026. To identify reference standards for first-line pharmacological treatment of chronic pain syndromes, structured searches were conducted in PubMed and the Cochrane Library.

Search strategies combined condition-specific descriptors (burns; constipation; chikungunya virus infection; chronic kidney failure; Interstitial cystitis; diabetic neuropathy; fibromyalgia; headache; temporomandibular joint disorders; low back pain; phantom limb pain; plantar fasciitis; spinal cord diseases with the descriptor “chronic pain” using the Boolean operator AND. For neuromodulation interventions, only studies evaluating tDCS as a sole treatment for chronic pain were included. To identify AI recommendations, standardized prompts were submitted to both platforms requesting the first line pharmacological treatment for chronic pain syndrome in adults. The same procedure was applied to obtain recommended tDCS parameters.

Evidence hierarchy followed a predefined order: (1) international or national clinical guidelines; (2) Cochrane systematic reviews; (3) systematic reviews with high certainty of evidence according to GRADE 1 or 2; and (4) randomized clinical trials with low risk of bias according to validated risk-of-bias assessment tools.

AI-generated recommendations were analyzed through structured qualitative content analysis and compared against evidence-based reference standards. Agreement was classified using a predefined framework: (1) full agreement, when the main therapeutic strategy was consistent with the evidence; (2) partial agreement, when the overall strategy aligned, but specific components differed; and (3) no agreement, when recommendations diverged from the evidence base. Two independent reviewers performed the classification and all the procedures. Discrepancies were resolved by consensus, with arbitration by a third researcher if necessary. Each investigator independently queried the AI platforms using identical prompts. Extracted data were organized in standardized comparison tables.

RESULTS

Treatment recommendations were categorized according to predefined chronic pain syndromes derived from evidence-based frameworks, including pharmacological and neuromodulation-based interventions. Table 1 shows partial to full convergence in several pharmacological domains, with divergences primarily related to therapeutic hierarchy and scope.

Table 1
Comparison of first-line pharmacological treatment recommendations for chronic pain syndromes from AI outputs and evidence-based literature.

Chronic pain syndrome	Guidelines/ Systematic review	ChatGPT	OpenEvidence	Qualitative concordance
Burns	Opioids, paracetamol, non-hormonal anti-inflammatory, cognitive/behavioral therapies, hypnosis, and virtual reality (Romanowski et al., 2020)	Analgesia (paracetamol or non-hormonal anti-inflammatory or opioids for severe pain) + topical antimicrobial.	Antibiotics pomade for superficial burns and silver sulfadiazine for deeper burns	Partial concordance. ChatGPT captured the main analgesic approach but omitted multimodal strategies. OpenEvidence prioritized topical antimicrobials, diverging from the guideline's analgesic focus.
Constipation	Osmotic laxatives (Ihara et al., 2025)	Osmotic laxative or bulk-forming agent	Osmotic laxatives	Full concordance. AI outputs aligned with first-line osmotic laxative therapy recommended in the guideline.
Chikungunya fever	Paracetamol, opioids, and hydration (Simon et al., 2015)	Paracetamol	Paracetamol or opioids	Partial concordance. AI outputs reflected paracetamol use but incompletely represented the guideline's inclusion of opioids for severe cases.
Chronic renal failure	Renin–angiotensin system inhibitor + SGLT2 inhibitor (Levin et al., 2013)	Angiotensin-converting enzyme inhibitors or angiotensin receptor blockers	Renin–angiotensin system inhibitor + SGLT2 inhibitor + finerenone	Partial concordance. ChatGPT omitted SGLT2 inhibitors, while OpenEvidence aligned with the recommended combination but expanded therapy with finerenone.
Interstitial Cystitis	Third and fourth-generation quinolones and pivmecillinam (Kim et al., 2020)	Nitrofurantoin	Nitrofurantoin, trimethroprim-sulfamethoxazole, fosfomicyn, and pivmecillinam	Partial concordance. AI outputs favored nitrofurantoin, whereas the reference highlighted quinolones and

Chronic pain syndrome	Guidelines/ Systematic review	ChatGPT	OpenEvidence	Qualitative concordance
				pivmecillinam; OpenEvidence partially overlapped by including pivmecillinam.
Diabetic neuropathies	Pregabalin. venlafaxine, duloxetine, amitriptyline, gabapentin, valproate, opioids, and capsaicin (Bril et al., 2011)	Duloxetine and pregabalin	Gabapentinoids, serotonin-norepinephrine reuptake inhibitors, tricyclic antidepressants or sodium channel blockers	Full concordance. AI outputs identified core first-line pharmacological classes consistent with the guideline, particularly gabapentin and serotonin–norepinephrine reuptake inhibitors, with only minor variation in scope.
Fibromyalgia	Tricyclic antidepressants (Goldenberg, Burckhardt, Crofford, 2004)	Duloxetine and amitriptyline	Serotonin-norepinephrine reuptake inhibitors, gabapentinoid, and tricyclic antidepressants	Full concordance. Both AI platforms incorporated tricyclic antidepressants, consistent with the reference, while expanding additional evidence-supported pharmacological classes.
Headache	Triptans and aspirin-acetaminophen-caffeine, ibuprofen and paracetamol (Sico et al., 2024)	Paracetamol and Nonsteroidal anti-inflammatory	Paracetamol and Nonsteroidal anti-inflammatory	Partial concordance. AI recommendations reflected NSAIDs and paracetamol but omitted triptans emphasized in the guideline.
Temporomandibular joint	Non-steroidal anti-inflammatory (Busse et al., 2023)	Nonsteroidal anti-inflammatory	Nonsteroidal anti-inflammatory, cyclobenzaprine, amitriptyline, and gabapentin	Full concordance. AI outputs aligned with guideline-recommended NSAIDs, with minor expansion in scope.

Chronic pain syndrome	Guidelines/ Systematic review	ChatGPT	OpenEvidence	Qualitative concordance
Low back pain	Nonsteroidal anti-inflammatory and opioids (Cochrane Back and Neck Group et al., 2023)	Nonsteroidal anti-inflammatory	Nonsteroidal anti-inflammatory	Partial concordance. NSAIDs were correctly identified, but opioids cited in the reference were not included.
Phantom limb	Ketamine, morphine, epidural anesthesia and gabapentin (McCormick et al., 2014)	Amitriptyline	Gabapentin, oral morphine and ketamine	Partial concordance. OpenEvidence reflected core recommended agents, while ChatGPT diverged by prioritizing amitriptyline.
Plantar fasciitis	Botulinum toxin A (Acosta-Olivo et al., 2022)	Nonsteroidal anti-inflammatory	Nonsteroidal anti-inflammatory and paracetamol	No concordance. AI recommendations emphasized NSAIDs rather than botulinum toxin A highlighted in the reference.
Spinal cord	Pregabalin, gabapentin, and amitriptyline (Loh et al., 2022)	Pregabalin	Gabapentin, pregabalin, Serotonin-norepinephrine reuptake inhibitors, and tricyclic antidepressant	Full concordance. AI recommendations aligned with core guideline-supported neuropathic pain agents.

Across the 13 chronic pain syndromes analyzed, full concordance between AI-generated recommendations and evidence-based standards was observed in 38.5% of cases (5/13), partial concordance in 53.8% (7/13), and no concordance in 7.7% (1/13). Divergences were predominantly related to therapeutic hierarchy and scope rather than complete disagreement with the underlying pharmacological strategy.

In contrast, the analysis of tDCS parameters revealed greater variability. Unlike pharmacological recommendations, which are structured around drug classes, neuromodulation protocols depend on precise technical specifications, including cortical target, montage configuration, intensity, and number of sessions. Although AI outputs frequently reproduced general stimulation characteristics, inconsistencies were observed in montage precision and session standardization. Table 2 illustrates this broader dispersion across neuromodulation parameters.

Table 2
Evidence on tDCS parameters (montage, duration, intensity, electrode size, number of sessions, and qualitative concordance) for the management of chronic pain syndromes.

Chronic pain syndrome	Guidelines/ Systematic review/clinical trials	ChatGPT	OpenEvidence	Qualitative concordance
Burns	Montage: Anodal M1 Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 1–10 sessions (Thibaut et al., 2021)	Montage: Anode over primary motor cortex (M1) contralateral to pain; cathode over contralateral supraorbital area Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 5–10 sessions (commonly 5 consecutive daily sessions)	Montage: Anode over M1 or DLPFC Duration: 20 minutes Intensity: 1–2 mA Electrode size: 25–35 cm² Number of sessions: 5–20 daily or near-daily sessions	Partial concordance. Core parameters were aligned, but OpenEvidence broadened cortical targets and session range beyond the reference.
Constipation	No studies	Montage: Anode over left dorsolateral prefrontal cortex (F3); cathode over right supraorbital area Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 10–20 sessions (commonly 5	Montage: Anode over primary motor cortex (M1) Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 5–20 daily or near-daily sessions[a]	No concordance (absence of evidence). No clinical studies support tDCS for constipation; AI platforms generated protocol specifications without reference evidence.

[a] There are no direct studies of tDCS for constipation-related chronic pain in the medical literature. The optimal parameters are extrapolated from studies in abdominal pain syndromes and chronic pain population.

Chronic pain syndrome	Guidelines/ Systematic review/clinical trials	ChatGPT	OpenEvidence	Qualitative concordance
		sessions/week for 2–4 weeks)		
Chikungunya fever	Montage: M1 using the C3/Fp2 montage Duration: 20 min Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 10 consecutive sessions (Cavalcante et al., 2024)	Montage: Anode over primary motor cortex (M1) contralateral to predominant pain; cathode over contralateral supraorbital area Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm^2 Number of sessions: 5–10 sessions (commonly 5 consecutive daily sessions)	Montage: Anode over the primary motor cortex (M1), typically C3/Fp2 Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm (about 1.15 ft)² Number of sessions: 6–10 sessions, delivered on consecutive or nonconsecutive days	Full concordance. Core montage and stimulation parameters were aligned, with minor variation in session scheduling.
Chronic kidney disease	Montage: Anode over C3 and cathode over contralateral supraorbital area Duration: 20 min Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 15 sessions, 3 times per week (Mendes et al., 2025)	Montage: Anode over primary motor cortex (M1) contralateral to predominant pain; cathode over contralateral supraorbital area Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 5–10	Montage: Anode over the primary motor cortex (M1), typically C3 or M1/Sp2 Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 10–15 sessions, delivered three times per week or nonconsecutively over 4–5 weeks.	Partial concordance. Core parameters were consistent, but session number and frequency differed from the referenced protocol.

[a] There are no direct studies of tDCS for constipation-related chronic pain in the medical literature. The optimal parameters are extrapolated from studies in abdominal pain syndromes and chronic pain population.

Chronic pain syndrome	Guidelines/ Systematic review/clinical trials	ChatGPT	OpenEvidence	Qualitative concordance
		sessions (commonly 5 consecutive daily sessions)		
Interstitial Cystitis	There is only a study protocol (Dutra et al., 2021)	Montage: Anode over primary motor cortex (M1) contralateral to predominant pelvic pain; cathode over contralateral supraorbital area Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 5–10 sessions (commonly 5 consecutive daily sessions)	Montage: Anode over the primary motor cortex (M1) Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 5–20 consecutive daily sessions	No concordance (absence of evidence). No clinical studies support tDCS for interstitial cystitis; AI platforms generated protocol specifications without reference evidence.
Diabetic neuropathies	Montage: Anode over C3 and the cathode over the contralateral supraorbital area (Fp2). Duration: 20 minutes Intensity: 2mA Electrode size: 35 cm² Number of sessions: 5 consecutive days (Ferreira et al., 2020)	Montage: Anode over primary motor cortex (M1) contralateral to the most painful area; cathode over contralateral supraorbital area Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm²	Montage: Anode over the primary motor cortex (M1, contralateral to the most painful side) or dorsolateral prefrontal cortex (DLPFC/F3) Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 10–12 sessions, typically	Partial concordance. Core M1 parameters were aligned, but session number and target scope were expanded beyond the reference.

[a] There are no direct studies of tDCS for constipation-related chronic pain in the medical literature. The optimal parameters are extrapolated from studies in abdominal pain syndromes and chronic pain population.

Chronic pain syndrome	Guidelines/ Systematic review/clinical trials	ChatGPT	OpenEvidence	Qualitative concordance
		Number of sessions: 5–10 sessions (commonly 5 consecutive daily sessions)	delivered three times per week	
Fibromyalgia	Montage: Anode over the left primary motor cortex C3 or dorsolateral prefrontal cortex and the cathode over the right supraorbital area Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 5 to 10 consecutive daily sessions (Moshfeghinia et al., 2023)	Montage: Anode over left primary motor cortex (M1); cathode over right supraorbital area Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 10–20 sessions (commonly 5 sessions/week for 2–4 weeks)	Montage: Anode over left M1 or left dorsolateral prefrontal cortex (anode over C3, with the cathode over the contralateral supraorbital area. Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 15–20 sessions, delivered daily or near-daily	Partial concordance. Core M1/DLPFC parameters were reproduced, but session dosing was broadened beyond the referenced framework.
Headache	Montage: Anode over the primary motor cortex (M1) Duration: 20 minutes Intensity: 1–2 mA Electrode size: 25–35 cm² Number of sessions: 10–20 sessions (Martelletti et al., 2013)	Montage: Anode over primary motor cortex (M1) or left dorsolateral prefrontal cortex; cathode over contralateral supraorbital area Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm²	Montage: Anode M1 (C3/C4) or Anodal DLPFC (F3) Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 5–28 sessions, delivered daily or several times per week	Partial concordance. Core M1 parameters were reproduced, but session dosing and target scope showed variability.

[a] There are no direct studies of tDCS for constipation-related chronic pain in the medical literature. The optimal parameters are extrapolated from studies in abdominal pain syndromes and chronic pain population.

Chronic pain syndrome	Guidelines/ Systematic review/clinical trials	ChatGPT	OpenEvidence	Qualitative concordance
		Number of sessions: 5–10 sessions (commonly 5 consecutive daily sessions)		
Temporomandibular joint	Montage: Anode over C3 or C4 region contralateral to the painful temporomandibular joint, and the cathode over the opposite supra orbital region Duration: 20 minutes Intensity: 2 mA Electrode size: not informed Number of sessions: 5 consecutive days (Oliveira et al., 2015)	Montage: Anode over primary motor cortex (M1) contralateral to pain; cathode over contralateral supraorbital area Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 5–10 sessions (commonly 5 consecutive daily sessions)	Montage: Anode over the primary motor cortex (C3 or C4, contralateral to the most painful side), with the cathode over the contralateral supraorbital area. Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 5–20 daily sessions	Full concordance. Core M1 contralateral parameters were aligned, with broader session ranges.
Low back pain	Montage: Anode over the primary motor cortex C3 (contralateral to the most painful side), and cathode over the contralateral supraorbital area Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 1–15 daily sessions	Montage: Anode over primary motor cortex (M1) contralateral to predominant pain (or left M1 for bilateral pain); cathode over contralateral supraorbital area Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm²	Montage: Anode over the primary motor cortex (M1) or DLPFC for affective/cognitive symptoms and the dorsal anterior cingulate cortex for pain-related affective distress. The anode is typically placed over M1 (C3 or C4, contralateral to the most painful side), with the cathode over the contralateral supraorbital area.	Partial concordance. Core parameters were aligned; additional cortical targets were introduced beyond the referenced review.

[a] There are no direct studies of tDCS for constipation-related chronic pain in the medical literature. The optimal parameters are extrapolated from studies in abdominal pain syndromes and chronic pain population.

Chronic pain syndrome	Guidelines/ Systematic review/clinical trials	ChatGPT	OpenEvidence	Qualitative concordance
	(Alwardat et al., 2020)	Number of sessions: 5–10 sessions (commonly 5 consecutive daily sessions)	Duration: 20 minutes per session is standard in clinical trials and meta-analyses. Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 5–12 sessions, daily or several times per week	
Phantom limb	Montage: Anode over the contralateral M1 to the affected limb and the cathode contralateral supraorbital area Duration:15–20 minutes Intensity: 1–2 mA Electrode size: 25–35 cm² Number of sessions: 1–10 sessions (Navarro-López et al., 2022)	Montage: Anode over primary motor cortex (M1) contralateral to the amputated limb; cathode over contralateral supraorbital area Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 5–10 sessions (commonly 5 consecutive daily sessions)	Montage: Anode over the contralateral primary motor cortex (M1) to the affected limb Duration: 15–20 minutes per session Intensity: 1–2 mA Electrode size: 25–35 cm² Number of sessions: 5 consecutive daily sessions	Full concordance. AI outputs reproduced the referenced protocol parameters without substantive deviation.
Plantar fasciitis	Montage: Anode over the primary motor cortex (M1) leg area Duration: 20 minutes Intensity: 2 mA	Montage: Anode over primary motor cortex (M1) contralateral to the affected foot; cathode over contralateral	Montage: Anode over the primary motor cortex (M1) leg area contralateral to the symptomatic foot Duration: 20 minutes	Full concordance. AI outputs reproduced from the referenced M1 leg-area protocol with minimal

[a] There are no direct studies of tDCS for constipation-related chronic pain in the medical literature. The optimal parameters are extrapolated from studies in abdominal pain syndromes and chronic pain population.

Chronic pain syndrome	Guidelines/ Systematic review/clinical trials	ChatGPT	OpenEvidence	Qualitative concordance
	Electrode size: 25–35 cm² Number of sessions: 5 consecutive daily sessions (Tedeschi, 2024)	supraorbital area Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 5–10 sessions (commonly 5 consecutive daily sessions)	Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 5 consecutive daily sessions	variation in session range.
Spinal cord	Montage: Anode over M1 (contralateral to pain), cathode over contralateral supraorbital area Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 1–20 sessions (Mehta et al., 2015)	Montage: Anode over primary motor cortex (M1) contralateral to the most painful area; cathode over contralateral supraorbital area Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions 5–10 sessions (commonly 5 consecutive daily sessions)	Montage: Anode over the primary motor cortex C3 or C4, contralateral to the most painful side, with the cathode over the contralateral supraorbital area. Duration: 20 minutes Intensity: 2 mA Electrode size: 25–35 cm² Number of sessions: 5–20 daily sessions	Full concordance. AI outputs aligned with montage and dosing parameters within the referenced session range.

[a] There are no direct studies of tDCS for constipation-related chronic pain in the medical literature. The optimal parameters are extrapolated from studies in abdominal pain syndromes and chronic pain population.

Across neuromodulation conditions, AI-generated outputs demonstrated high consistency in core stimulation parameters, particularly primary motor cortex target (anodal M1), duration (20 minutes), intensity (2 mA), and electrode size (25–35 cm²). Variability was concentrated in session number and frequency, with several outputs expanding dosing beyond referenced protocols. Complete divergence

occurred only in the absence of empirical evidence. Overall, concordance was stronger for technical stimulation architecture than for protocol dosing precision.

DISCUSSION

This study examined the agreement between AI-generated recommendations and established evidence for the management of chronic pain, focusing on both pharmacological treatment and tDCS parameters. Across conditions, agreement was more frequent when recommendations relied on well-established first-line drug classes. In contrast, discrepancies emerged in scenarios requiring therapeutic prioritization, multimodal integration, or careful interpretation of guideline hierarchies.

These findings suggest that AI systems tend to perform more reliably in domains where pharmacotherapy follows stable and linear decision pathways, while struggling in contexts that require hierarchical reasoning, contextual adaptation, and nuanced interpretation of the evidence. Consistent with this pattern, higher pharmacological alignment was observed in conditions with well-established and guideline-concordant treatment frameworks, including chronic constipation (Ihara et al., 2025), diabetic neuropathy (Bril et al., 2011), fibromyalgia (Goldenberg et al., 2004), temporomandibular disorders (Busse et al., 2023), and neuropathic pain after spinal cord injury (Loh et al., 2022). In these contexts, AI outputs reproduced the core therapeutic backbone with minimal distortion. On the other hand, in burns (Romanowski et al., 2020), chikungunya-related pain (Simon et al., 2015), chronic kidney disease (Levin et al., 2013), interstitial cystitis (Kim et al., 2020), headache (Sico et al., 2024), low back pain (Cochrane Back and Neck Group et al., 2023), and phantom limb pain (McCormick et al., 2014), recommendations frequently captured central drug classes, but incompletely reflected the hierarchy, scope, or contextual nuance embedded in the literature. The clearest divergence was observed in plantar fasciitis, where botulinum toxin A was emphasized yet absent from AI-generated recommendations.

The neuromodulation analysis revealed a distinct pattern of structured technical reproduction. Across burns (Thibaut et al., 2021), chikungunya arthritis (Cavalcante et al., 2024), chronic kidney disease (Mendes et al., 2025), interstitial cystitis (Dutra et al., 2021), diabetic neuropathy (Ferreira et al., 2020), fibromyalgia (Moshfeghinia et al., 2023), chronic headache (Martelletti et al., 2013), temporomandibular disorders (Oliveira et al., 2015), low back pain (Alwardat et al., 2020), phantom limb pain (Navarro-López et al., 2022), plantar fasciitis (Tedeschi, 2024), and spinal cord injury–related neuropathic pain (Mehta et al., 2015), AI platforms consistently reproduced canonical stimulation architecture: anodal motor cortex targeting, 20-minute sessions, 2 mA intensity, and conventional electrode sizing. Divergences were rarely anatomical, instead, they emerged in session dosing and frequency standardization, where AI outputs tended to broaden treatment exposure beyond the exact trial parameters. Notably, in constipation and interstitial cystitis, where no empirical studies exist, both systems generated complete protocol specifications despite absence of reference evidence. This suggests that AI systems internalize stable procedural templates and extrapolate them across conditions, even when empirical grounding is limited.

A contrast between pharmacological and neuromodulation domains is highlighted. In pharmacological recommendations, divergence frequently reflected incomplete reconstruction of therapeutic hierarchy, prioritization, and multimodal reasoning, particularly in burns (Romanowski et al., 2020), chronic kidney

disease (Levin et al., 2013), headache (Sico et al., 2024), low back pain (Cochrane Back and Neck Group et al., 2023), and phantom limb pain (McCormick et al., 2014). AI systems reliably retrieved drug classes but inconsistently reproduced guideline logic and sequencing. One possible explanation is that pharmacological decision-making operates within a context-dependent and hierarchically organized framework, in which choices are continuously modulated by comorbidities, risk profiles, and treatment tolerability. In contrast, neuromodulation protocols rely more heavily on standardized parameter sets and procedural logic, making them inherently more reproducible by AI systems.

It is important to emphasize that there is currently no consensus regarding the comparative accuracy of AI-based platforms. Some investigations have reported superior performance of ChatGPT in specific contexts, such as heart disease issues (Hajj et al., 2025), whereas other studies have demonstrated greater accuracy with OpenEvidence (Nihal et al., 2026). These discrepancies are not unexpected, given variations in study design, domain specificity, evaluation of metrics, and the rapid evolution of these technologies. Despite such heterogeneity, when used appropriately and with critical appraisal, AI tools may support and potentially enhance clinical decision-making processes (Hurt et al., 2025).

This study has several limitations that should be acknowledged. Although we attempted to categorize treatments according to specific pain syndromes, important clinical nuances must be considered, including patient age, sex, and stage or severity of disease progression. These factors may substantially influence therapeutic selection, response to treatment, and overall clinical outcomes. Also, AI-based platforms may generate different responses depending on how and when the prompt is formulated. Currently, there is no standardized descriptor to characterize or control this variability. Evidence-based medicine should not be understood as a mere catalogue of therapeutic options, but rather as a structured and methodologically rigorous process involving critical appraisal, hierarchical evaluation of evidence, and clinical judgment. In chronic pain management, decisions are rarely linear. They depend on prioritization, context, safety, and integration of multiple strategies. While tools that efficiently retrieve and organize information may facilitate this process, they do not replace the interpretive reasoning and contextual decision-making that characterize clinical expertise.

CONCLUSION

This study showed that AI platforms have significant potential to support clinical decision-making. Although the systems demonstrated good qualitative concordance with established evidence, the findings suggest that AI more easily reproduces well-defined pharmacological categories and tDCS parameters than it reflects the structured reasoning required in evidence-based care. The differences observed in this study likely represent not only isolated inaccuracies, but also the complexity inherent to clinical decision-making. When used appropriately, such systems may strengthen scientific practice by encouraging comparison, verification, and critical appraisal of primary evidence.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Not applicable

Competing interests

The authors declare no competing interests.

Funding

This research received no funding.

Authors' contributions

Edson Silva-Filho and Denise Mota Araripe Pereira Fernandes conceptualized the study, performed data analysis, and revised the manuscript. Maria Isabel França analyzed the data and revised the manuscript. Letícia Amaro Vieira revised the manuscript. Maria Thereza Micussi and Rodrigo Pegado coordinated the study and revised the manuscript. All authors have read and approved the final version of the manuscript.

Acknowledgements

Not applicable.

References

1. Aaron RV, Ravyts SG, Carnahan ND, Bhattiprolu K, Harte N, McCaulley CC, Dudeney J. (2025). Prevalence of depression and anxiety among adults with chronic pain: a systematic review and meta-analysis. JAMA Netw Open, 8(3), e250268.
2. Acosta-Olivo C, Simental-Mendia LE, Vilchez-Cavazos F, Pena-Martinez VM, Elizondo-Rodiguez J, Simental-Mendia M. Clinical efficacy of botulinum toxin in the treatment of plantar fasciitis: a systematic review and meta-analysis of randomized controlled trials. Arch Phys Med Rehabil. 2022;103(2):364–71.
3. Alwardat M, et al. Is transcranial direct current stimulation (tDCS) effective for chronic low back pain? A systematic review and meta-analysis. J neural transmission v. 2020;127:1257–70.
4. Bril V, England J, Franklin GM, Backonja M, Cohen J, Del Toro D, Zochodne D. Evidence-based guideline: treatment of painful diabetic neuropathy: report of the American Academy of Neurology, the American Association of Neuromuscular and Electrodiagnostic Medicine, and the American Academy of Physical Medicine and Rehabilitation. Pm&r. 2011;3(4):345–52.

5. Busse JW, Casassus R, Carrasco-Labra A, Durham J, Mock D, Zakrzewska JM, Agoritsas T. (2023). Management of chronic pain associated with temporomandibular disorders: a clinical practice guideline. BMJ, 383.
6. Busse JW, Wang L, Kamaleldin M, Craigie S, Riva JJ, Montoya L, Guyatt GH. Opioids for chronic noncancer pain: a systematic review and meta-analysis. JAMA. 2018;320(23):2448–60.
7. Cavalcante AFL, Holanda JSCB, Passos JOS, de Oliveira JMP, Morya E, Okano AH, Pegado R. Anodal tDCS over the motor cortex improves pain but not physical function in chronic chikungunya arthritis: Randomized controlled trial. Annals Phys Rehabilitation Med. 2024;67(4):101826.
8. Cochrane Back and Neck Group, Group CM, Cashin AG. Wand, B. M..
9. D'Antoni F, Russo F, Ambrosio L, Bacco L, Vollero L, Vadalà G, Denaro V. Artificial intelligence and computer aided diagnosis in chronic low back pain: A systematic review. Int J Environ Res Public Health. 2022;19(10):5971.
10. Dutra LRDV, Silva-Filho E, Oliveira MC, Tavares BNP, Pegado R, Micussi MTBAC. Transcranial direct current electrical stimulation for the treatment of interstitial cystitis: A study protocol. Eur J Obstet Gynecol Reproductive Biology. 2021;262:198–202.
11. Enthoven WT, Roelofs PD, Deyo RA, van Tulder MW, Koes BW, Cochrane Back and Neck Group. &. (1996). Non-steroidal anti-inflammatory drugs for chronic low back pain. Cochrane Database Syst Reviews, 2016(8).
12. Ferreira G, Silva-Filho E, de Oliveira A, de Lucena C, Lopes J, Pegado R. Transcranial direct current stimulation improves quality of life and physical fitness in diabetic polyneuropathy: a pilot double blind randomized controlled trial. J Diabetes Metabolic Disorders. 2020;19(1):327–35.
13. Goldenberg DL, Burckhardt C, Crofford L. Management of fibromyalgia syndrome. JAMA. 2004;292(19):2388–95.
14. Gomez-Cabello CA, Borna S, Pressman S, Haider SA, Haider CR, Forte AJ. Artificial-intelligence-based clinical decision support systems in primary care: a scoping review of current clinical implementations. Eur J Invest Health Psychol Educ. 2024;14(3):685–98.
15. Hajj J, Mdaihly M, Kassab J, Miyasaka R, Harb SC. (2025). Evaluating ChatGPT-4o and OpenEvidence for structural heart disease support: Insights from tricuspid valve therapies. Struct Heart, 100696.
16. Hurt RT, Stephenson CR, Gilman EA, Aakre CA, Croghan IT, Mundi MS, Edakkanambeth Varayil J. The use of an artificial intelligence platform OpenEvidence to augment clinical decision-making for primary care physicians. J Prim Care Community Health. 2025;16:21501319251332215.
17. Ihara E, Manabe N, Ohkubo H, Ogasawara N, Ogino H, Kakimoto K, Higuchi K. Evidence-based clinical guidelines for chronic constipation 2023. Digestion. 2025;106(1):62–89.
18. Kim DK, Kim JH, Lee JY, Ku NS, Lee HS, Park JY, Cho KS. Reappraisal of the treatment duration of antibiotic regimens for acute uncomplicated cystitis in adult women: a systematic review and network meta-analysis of 61 randomised clinical trials. Lancet Infect Dis. 2020;20(9):1080–8.
19. Levin A, Stevens PE, Bilous RW, Coresh J, De Francisco AL, De Jong PE, Winearls CG. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int supplements. 2013;3(1):1–150.

20. Lex JR, Di Michele J, Koucheki R, Pincus D, Whyne C, Ravi B. (2023). Artificial intelligence for hip fracture detection and outcome prediction: a systematic review and meta-analysis. JAMA Netw open, 6(3), e233391.
21. Loh E, Mirkowski M, Agudelo AR, Allison DJ, Benton B, Bryce TN, Mehta S. The CanPain SCI clinical practice guidelines for rehabilitation management of neuropathic pain after spinal cord injury: 2021 update. Spinal Cord. 2022;60(6):548–66.
22. Lopes S, Rocha G, Guimarães-Pereira L. Artificial intelligence and its clinical application in anesthesiology: a systematic review. J Clin Monit Comput. 2024;38(2):247–59.
23. Martelletti P, Jensen RH, Antal A, Arcioni R, Brighina F, de Tommaso M, May A. Neuromodulation of chronic headaches: position statement from the European Headache Federation. J Headache Pain. 2013;14(1):86.
24. Martos M, Fields B, Finlayson SG, Hartell N, Kim T, Larimer E, Lion KC. (2025). Accuracy of Artificial Intelligence vs professionally translated discharge instructions. JAMA Netw Open, 8(9), e2532312.
25. McCormick Z, Chang-Chien G, Marshall B, Huang M, Harden RN. Phantom limb pain: a systematic neuroanatomical-based review of pharmacologic treatment. Pain Med. 2014;15(2):292–305.
26. Mehta S, McIntyre A, Guy S, Teasell RW, Loh E. Effectiveness of transcranial direct current stimulation for the management of neuropathic pain after spinal cord injury: a meta-analysis. Spinal Cord. 2015;53(11):780–5.
27. Mendes EF, Silva-Filho E, Holanda JB, Matias MGL, Lima TC, Costa S, Pegado R. Transcranial direct current stimulation over motor cortex improves pain in end stage renal disease a randomized controlled trial. Sci Rep. 2025;15(1):21943.
28. Moshfeghinia R, Shekouh D, Mostafavi S, Hosseinzadeh M, Bahadori AR, Abdollahifard S, Razmkon A. The effects of transcranial direct-current stimulation (tDCS) on pain intensity of patients with fibromyalgia: a systematic review and meta-analysis. BMC Neurol. 2023;23(1):395.
29. Navarro-Lopez V, Manuel DEL, Fernandez-Vazquez D, Fernandez-Gonzalez P, Carratala-Tejada M, Molina-Rueda F. Transcranial direct current stimulation in the management of phantom limb pain: a systematic review of randomized controlled trials. Eur J Phys Rehabil Med. 2022;58(5):738.
30. Nihal A, Yang K, Pavlidakey PG. OpenEvidence is superior to GPT in skin cancer guideline queries. Int J Dermatol. 2026;65(2):338–9.
31. O'Connell NE, Marston L, Spencer S, DeSouza LH, Wand BM. Non-invasive brain stimulation techniques for chronic pain. Volume 3. Cochrane database of systematic reviews; 2018.
32. Oliveira LB, Lopes TS, Soares C, Maluf R, Goes BT, Sá KN, Baptista AF. Transcranial direct current stimulation and exercises for treatment of chronic temporomandibular disorders: a blind randomised-controlled trial. J Rehabil. 2015;42(10):723–32.
33. OpenAI. ChatGPT (GPT-5.2-Go) [Internet]. San Francisco (CA): OpenAI. 2026. Available from: https://chat.openai.com/
34. OpenEvidence. OpenEvidence platform [Internet] 2026. Available from: https://www.openevidence.com/
35. Perez Rivera LR, Gursky AK, Elmer N, Boyd CJ, Karp NS. Evaluating the Quality and Reliability of Large Language Models for Plastic Surgery Patient Education: A Comparative Analysis of ChatGPT and

OpenEvidence. Aesthetic Surg J. 2026;46(2):160–7.

36. Romanowski KS, Carson J, Pape K, Bernal E, Sharar S, Wiechman S, Joe V. American Burn Association guidelines on the management of acute pain in the adult burn patient: a review of the literature, a compilation of expert opinion, and next steps. J Burn Care Res. 2020;41(6):1129–51.
37. Shahriari MH, Sabbaghi H, Asadi F, Hosseini A, Khorrami Z. Artificial intelligence in screening, diagnosis, and classification of diabetic macular edema: A systematic review. Surv Ophthalmol. 2023;68(1):42–53.
38. Sico JJ, Antonovich NM, Ballard-Hernandez J, Buelt AC, Grinberg AS, Macedo FJ, Sico JJ, Antonovich NM, Ballard-Hernandez J, Buelt AC, Grinberg AS, Macedo FJ, Ford AW, et al. 2023 US Department of Veterans Affairs and US Department of Defense Clinical Practice Guideline for the Management of Headache. Ann Intern Med. 2024;177(12):1675–94.
39. Silva-Filho E, Pegado R. ChatGPT and transcranial direct current stimulation for chronic pain. Innovations Clin Neurosci. 2024;21(1–3):61.
40. Simon F, Javelle E, Cabie A, Bouquillard E, Troisgros O, Gentile G, Wendling D. French guidelines for the management of chikungunya (acute and persistent presentations). November 2014. Med et maladies infectieuses. 2015;45(7):243–63.
41. Tedeschi R. Transcranial direct current stimulation for chronic foot pain: A comprehensive review. Eneurologicalsci. 2024;35:100498.
42. Thibaut A, Shie VL, Ryan CM, Zafonte R, Ohrtman EA, Schneider JC, Fregni F. A review of burn symptoms and potential novel neural targets for non-invasive brain stimulation for treatment of burn sequelae. Burns. 2021;47(3):525–37.
43. Yong RJ, Mullins PM, Bhattacharyya N. Prevalence of chronic pain among adults in the United States. Pain. 2022;163(2):e328–32.