Schema-Constrained, Clinician-Verified LLM Abstraction to Improve Documentation Completeness for CNS Trial Eligibility Screening: A Real-World Methodology Pilot

Submitter Tanveer Padder

Affiliation Health synergy clinical research.com

SUBMISSION DETAILS

I agree to provide poster pdf for attendee download. Yes

What is the Methodological Issue Being Addressed? CNS clinical trials face high screen-fail rates and protocol deviations, partly because routine psychiatric documentation is incomplete for trial-critical variables — undocumented medication adequacy, limited suicide-risk capture, absent mania or psychosis screening — which can cause inappropriate enrollment or rejection of eligible candidates. Documentation quality varies by care setting. A reproducible structured-screening abstraction may improve the completeness and consistency of eligibility documentation before formal screening.

Introduction Because LLM-based clinical extraction and patient–trial matching already exist, the methodological focus here is not AI itself but a trial-specific, schema-constrained abstraction under mandatory clinician verification. Using real outpatient and inpatient notes, we tested whether structured abstraction improves the completeness, consistency, and trial relevance of eligibility assessment versus original documentation, and compared baseline quality and post-structuring gain between settings. Under a fixed pre-specified schema, the model extracted existing content or flagged items as missing, was barred from clinical inference, and every output was clinician-reviewed; it organizes and flags information rather than imputing clinical facts.

Methods One hundred fifty real, de-identified psychiatric encounters were abstracted (75 outpatient specialty/community; 75 inpatient acute admissions); no simulated cases. Each encounter was reviewed twice — as originally documented and after schema-constrained, clinician-verified structuring. Twelve pre-specified domains were scored 0 to 2 (24-point maximum) using an operationalized rubric with explicit anchor definitions; this Documentation Completeness Score is investigator-developed for methodological comparison, not psychometrically validated. Because the pilot targeted feasibility and effect size rather than reliability, a single trained rater scored all cases, with a powered two-reviewer reliability sub-study (Cohen's kappa) pre-registered next. Sixteen trial-relevant safety and eligibility flags were applied; preliminary eligibility was classified as Likely eligible, Excluded, or Unclear. Within-case change used paired t-tests (primary) and Wilcoxon signed-rank tests (sensitivity); domain-level completeness used exact McNemar tests; outpatient-versus-inpatient contrasts used Welch t-tests and Mann–Whitney tests; means and proportions are reported with 95% confidence intervals (Student's t or bootstrap for means, Wilson for proportions). Cohen's kappa could not be computed because no second-reviewer data were available. Extraction fidelity (structured fields vs source text) was clinician-confirmed on every output.

Results Across 150 encounters, mean Documentation Completeness Score increased from 16.4/24 (SD 3.2) to 20.7/24 (SD 1.4); mean within-case change was +4.3 points (95% CI 3.9–4.7; paired $t(149)=24.1$, $p<0.001$; Wilcoxon $p<0.001$; Cohen's $dz=1.97$). Baseline completeness was higher inpatient than outpatient (18.6 vs 14.3; Welch $p<0.001$), while post-structuring gain was larger outpatient (+5.8 vs +2.8; $p<0.001$). Median trial-critical missing elements per case fell from 5 to 3 ($p<0.001$); most domains improved (McNemar $p<0.001$). Suicide-risk documentation showed the clearest setting contrast: complete in 0% of outpatient and 100% of inpatient notes at baseline, rising to 87% outpatient after structuring. No note (0/150) contained a validated structured suicide-risk tool or complete validated symptom-severity scale, so these domains did not change. Preliminary eligibility also differed by setting: outpatient, 16 Likely eligible, 40 Excluded, 19 Unclear; inpatient, 0 Likely eligible, 72 Excluded, 3 Unclear — consistent with acute inpatient admissions being largely trial-ineligible.

Conclusion In this methodology pilot of 150 real psychiatric encounters, schema-constrained, clinician-verified abstraction improved documentation completeness without adding clinical facts, with the largest gain where documentation was sparsest: outpatient care. Inpatient admission notes were more complete at baseline; acute inpatient cases were overwhelmingly trial-ineligible. Unchanged symptom-severity and absent structured suicide-risk-tool domains define the ceiling of retrospective structuring and point to the need for prospective capture. Findings are preliminary and limited by single-rater scoring, addressed prospectively with a pre-specified two-reviewer protocol; estimates are effect-size and feasibility signals, not reliability claims. Results support multi-site expansion, prospective C-SSRS and symptom-scale capture, and formal time-and-motion measurement before any inference about screening accuracy.

Co-Authors

Tanveer Padder[1], Asim Nisar[1]

[1] Health synergy clinical research.com

Keywords

Keywords
AI-assisted screening;
Eligibility precision;
Structured abstraction
Screen-fail reduction
Pre-screening workflow

Guidelines I have read and understand the Poster Guidelines

Disclosures No conflicts of interest to disclose.
This is a methodology pilot, not a clinical efficacy evaluation. AI-assisted structuring is not a

diagnostic device and does not replace clinician judgment.

Related Tables and Supporting Materials ISCTM_Poster_2page_reviewer_aligned.pdf