https://doi.org/10.47102/annals-acadmedsg.2026422

Preliminary clinician usability and acceptability evaluation of an AI-powered toxicology clinical decision support tool

Adeline Su-Yin Ngo[1] *MBBS*, Lokesh Krishnaji Kolhe[1] *MBBS*, Bryon Frankie Hon Khi Chong[1] *MBBS*, Ee Xion Tan[2] *PhD*, Annette Yingcui Lin[1] *MBBS*, Vaikunthan Rajaratnam[3] *MBBS*

Dear Editor,

Acute toxicological presentations are complex and time-critical, requiring clinicians to integrate toxicokinetics, evolving clinical manifestations, and antidote recommendations under marked time pressure. This complexity increases cognitive load and contributes to variability in decision-making, particularly where specialist toxicology expertise is not immediately available.

Artificial intelligence (AI)-enabled clinical decision support systems (CDSS) aim to augment clinician performance through evidence-informed recommendations. Such systems may standardise care and improve workflow efficiency. However, successful integration into routine practice depends on clinician acceptance, usability, and trust in system outputs.[1] Even technically robust systems often fail to achieve sustained adoption when perceived as cumbersome, poorly integrated, or misaligned with clinical reasoning.[2]

The System Usability Scale (SUS) is a widely validated measure of perceived usability and provides a standardised benchmark for user experience.[3] Beyond usability, both clinician trust and perceived usefulness are recognised as important factors associated with AI-driven decision-support adoption, particularly in high-risk clinical settings. Despite expanding interest in AI applications in healthcare, evidence evaluating clinician-centred outcomes, including usability, trust, and perceived usefulness within toxicology remains limited. Existing evaluations have focused primarily on technical performance rather than clinician interaction and usability.[4,5]

An AI-powered toxicology clinical decision support tool (ToxC) was implemented using Google's NotebookLM platform with a manually curated corpus of 19 sources selected for authority, currency, and relevance to acute toxicology, including toxicology textbooks, clinical guidelines, journal literature, and local protocols. This study evaluated clinician-reported usability and acceptability of ToxC and examined subgroup differences and associations among usability, the study-specific trust score, perceived usefulness, and satisfaction.

This was a cross-sectional evaluation of clinician interaction with ToxC using convenience sampling. Healthcare professionals managing acute toxicological presentations received instructions on using ToxC before personally entering poisoning scenarios or questions based on clinical experience or hypothetical situations, reviewing the generated responses, and asking follow-up questions as desired. They subsequently completed a structured electronic questionnaire comprising the 10-item SUS; a 5-item study-specific trust scale assessing (1) recommendation relevance, (2) cognitive-load reduction, (3) real-life trust, (4) clinical alignment, and (5) willingness to recommend the tool; and global ratings of usefulness and satisfaction. Data were analysed using descriptive statistics, subgroup analyses, correlations, and univariable linear regression. Descriptive usability and acceptability findings were the main focus of the evaluation; correlations and univariable regressions were treated as exploratory analyses of relationships among clinician-reported measures. The study was approved by the National Healthcare Group Domain Specific Review Board (2025-1761).

A total of 95 clinicians completed the evaluation. Most respondents were aged 31–40 years, practised in emergency medicine, and were junior doctors. The mean SUS score was 75.3 (95% confidence interval [CI] 71.6–79.0), corresponding to a "Good" grade on the Bangor adjective scale and significantly exceeding the conventional acceptability threshold of 68.[6,8] The composite trust score was 4.39 (95% CI 4.27–4.51), while mean usefulness and satisfaction scores were both 4.43. More than 90% of respondents rated usefulness and satisfaction highly. Recommendation relevance, alignment with clinical reasoning, and willingness to recommend the tool were also rated highly. No statistically significant subgroup differences were

The Annals is an open access journal, allowing non-commercial use under CC BY-NC-SA 4.0.

[1] Department of Emergency Medicine, Woodlands Hospital, Singapore
[2] Department of Data Science and Artificial Intelligence, Faculty of Engineering and Technology, Sunway University, Selangor, Malaysia
[3] Department of Orthopaedic Surgery, Khoo Teck Puat Hospital, Singapore
Correspondence: Dr Adeline Su-Yin Ngo, Department of Emergency Medicine, Woodlands Hospital, 17 Woodlands Drive 17, Singapore 737628.
Email: adeline.ngo@nhghealth.com.sg
Accepted: 4 November 2026
Published Online First: 11 November 2026

Table 1. Summary of clinician characteristics, usability, trust, and acceptability outcomes of the AI-powered toxicology clinical decision support tool (n=95).

Domain	Outcome	Result
Study participants	Total participants	95 clinicians
	Age distribution	21–30 years: 16 (16.8%) 31–40 years: 43 (45.3%) >40 years: 36 (37.9%)
	Clinical discipline	Emergency medicine: 76 (80.0%) Internal medicine: 3 (3.2%) Intensive care: 1 (1.1%) Other specialties: 15 (15.8%)
	Professional role	Junior doctors (MO/CA/RP): 50 (52.6%) Residents: 9 (9.5%) Senior doctors: 36 (37.9%)
System usability (SUS)	Mean composite SUS score	75.3 (95% CI 71.6–79.0)
	Interpretation	"Good" usability (Bangor adjective scale); exceeded the conventional acceptability threshold of 68
	Distribution of SUS scores	≥68 (acceptable): 65/95 (68.4%) ≥80 (good–excellent): 45/95 (47.4%) ≥90 (excellent): 27/95 (28.4%)
Individual usability perceptions (SUS)	Highest-rated positive items	Would like to use the system frequently (4.62) Most people would learn to use it quickly (4.54) Easy to use (4.49) Functions well integrated (4.47) Confident using the system (4.15)
	Lowest-rated negative items[a]	Need support to use the system (2.94); needed to learn a lot before getting started (2.51); unnecessarily complex (2.37); too much inconsistency (2.19); cumbersome to use (2.17)
Study-specific trust evaluation[b]	Composite trust score	4.39 (95% CI 4.27–4.51)
	Relevant recommendations	Mean 4.56; 93/95 (97.9%) rated ≥4
	Reduced cognitive load	Mean 4.31; 84/95 (88.4%) rated ≥4
	Trust recommendations in real-life scenarios	Mean 4.24; 81/95 (85.3%) rated ≥4
	Recommendations aligned with clinical reasoning	Mean 4.32; 80/95 (84.2%) rated ≥4
	Would recommend the tool to colleagues	Mean 4.55; 89/95 (93.7%) rated ≥4
Overall evaluation	Perceived usefulness	Mean 4.43 (95% CI 4.30–4.56); 89/95 (93.7%) rated ≥4
	Overall satisfaction	Mean 4.43 (95% CI 4.28–4.58); 87/95 (91.6%) rated ≥4
Relationships between study outcomes	Pearson correlations	Positive correlations were observed among SUS, the study-specific trust score, usefulness, and satisfaction (r=0.684–0.812; all P<0.001)
Exploratory univariable associations with clinician-reported outcomes	Trust score–satisfaction association	β=0.694 (95% CI 0.580–0.808); R^2=0.659; P<0.001
	Trust score–perceived usefulness association	β=0.625 (95% CI 0.526–0.724); R^2=0.626; P<0.001
	SUS–satisfaction association	β=0.040 (95% CI 0.032–0.048); R^2=0.530; P<0.001
	SUS–perceived usefulness association	β=0.036 (95% CI 0.028–0.044); R^2=0.468; P<0.001

Table 1. Summary of clinician characteristics, usability, trust, and acceptability outcomes of the AI-powered toxicology clinical decision support tool (n=95). (Cont'd)

Domain	Outcome	Result
Subgroup analyses	Professional role	Senior doctors demonstrated numerically higher SUS scores than junior clinicians; differences were not statistically significant (P>0.70)
	Clinical discipline	No statistically significant differences in SUS scores across emergency medicine, internal medicine, or other specialties (P>0.64)
	Overall findings	No statistically significant subgroup differences were detected; small subgroup sizes limited inference.

β: unstandardised regression coefficient; CA: clinical associate; CI: confidence interval; MO: medical officer; R^2: coefficient of determination; RP: resident physician; SUS: System Usability Scale

[a] Negative SUS items are reverse-scored within the composite SUS score; lower raw values on these negatively worded items indicate more favourable usability.

[b] The study-specific trust score comprised items assessing recommendation relevance, cognitive-load reduction, trust in real-life scenarios, alignment with clinical reasoning, and willingness to recommend ToxC. These items represent related clinician-reported perceptions and overlap conceptually with concurrently assessed usefulness and satisfaction.

detected, although small subgroup sizes limited inference. Exploratory analyses showed positive correlations among SUS, the study-specific trust score, usefulness, and satisfaction (r=0.684–0.812; all P<0.001). In exploratory univariable linear regression, the trust score was associated with satisfaction (R^2=0.659, P<0.001) and perceived usefulness (R^2=0.626, P<0.001); SUS was also associated with satisfaction (R^2=0.530, P<0.001) and perceived usefulness (R^2=0.468, P<0.001) (Table 1).

The principal contribution of this preliminary evaluation is the description of clinician reported usability and acceptability of ToxC. Associations among the questionnaire measures were exploratory. The study-specific trust score incorporates recommendation relevance, cognitive-load reduction, real-life trust, clinical alignment, and willingness to recommend, several of which overlap conceptually with perceived usefulness and satisfaction. This overlap, together with collection of all measures through the same self-report questionnaire at the same time, may inflate the observed associations. These results therefore do not establish trust as an independent factor influencing acceptance or demonstrate predictive or causal relationships.

These findings support previous evidence that usability and clinician perceptions are important considerations in AI-enabled clinical decision support.[8,9] High ratings for recommendations relevance, clinical alignment, usefulness, and satisfaction reflect favourable clinician perceptions following interaction with ToxC but do not establish clinical effectiveness.

This study has several limitations. The evaluation was conducted in a simulated rather than live clinical environment, and behavioural outcomes such as decision change, time-to-decision, and patient outcomes were not assessed. All outcomes were self-reported and may not reflect actual usage behaviour. The study-specific trust score incorporates several clinician-reported perceptions that are conceptually related to usefulness and satisfaction, which should be considered when interpreting the observed associations. The cross-sectional design does not capture longitudinal adoption or changes in clinician perceptions with repeated use. Convenience sampling and the predominance of emergency medicine clinicians may limit generalisability to other disciplines and healthcare settings, while small subgroup sizes for non-emergency medicine disciplines limit inference from subgroup analyses. Accuracy, completeness, and safety of ToxC-generated responses were not evaluated in this study. Accordingly, the present findings alone do not establish readiness for clinical deployment, and these dimensions require separate evaluation to complement the present usability and acceptability findings. Finally, the absence of an active comparator precludes assessment of the relative benefit of ToxC over existing clinical resources.

In conclusion, this preliminary evaluation demonstrated favourable clinician-reported usability and acceptability of ToxC, with good system usability and high ratings of trust, perceived usefulness, and satisfaction. Accuracy, completeness, and safety of ToxC-generated responses were not evaluated. Therefore, these findings do not support clinical deployment. Further evaluation of these dimensions is required before considering clinical use.

Ethics statement

The study was approved by the National Healthcare Group Domain Specific Review Board (2025-1761).

Declaration

No funding was received for this study. The authors declare that they have no affiliations or financial

involvement with any commercial organisation with a direct financial interest in the subject or materials discussed in the manuscript. The authors used ChatGPT (OpenAI) to assist with language editing, improving clarity and grammar during the preparation of this manuscript. All content was reviewed, verified, and approved by the authors, who take full responsibility for the accuracy and integrity of the manuscript.

Keywords: *artificial intelligence; clinical decision support; clinical toxicology; digital health; emergency medicine*

REFERENCES

1. Bowens FM, Frye PA, Jones WA. Health information technology: integration of clinical workflow into meaningful use of electronic health records. Perspect Health Inf Manag 2010;7:1d.
2. Miller A, Moon B, Anders S, et al. Integrating computerized clinical decision support systems into clinical work: A meta-synthesis of qualitative research. Int J Med Inform 2015;84:1009-18.
3. Lewis JR. The System Usability Scale: Past, Present, and Future. Int J Hum Comput Interact 2018;34:577-90.
4. Chary MA, Boyer E, Burns M. Diagnosis of acute poisoning using explainable artificial intelligence. Comput Biol Med 2021;134:104469.
5. Zellner T, Romanek K, Rabe C, et al. ToxNet: an artificial intelligence designed for decision support for toxin prediction. Clin Toxicol (Phila) 2023;61:56-63.
6. Brooke J. SUS: A "Quick and Dirty" Usability Scale. In: Jordan PW, Thomas B, McClelland IL, et al., editors. Usability Evaluation in Industry. London: Taylor & Francis; 1996:207-12.
7. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008; 24:574-94.
8. Knop M, Weber S, Mueller M, et al. Human factors and technological characteristics influencing the interaction of medical professionals with artificial intelligence-enabled clinical decision support systems: literature review. JMIR Hum Factors 2022;9:e28639.
9. Kawamoto K, Houlihan CA, Balas EA, et al. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005;330:765.