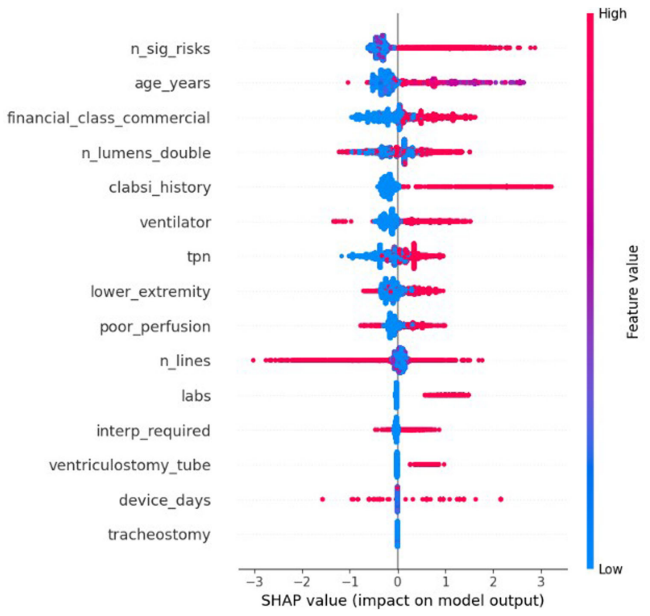
Risk Factor	Tunneled Lines Model	Nontunneled Lines Model
Diaphoresis	X	X
Drainage	X	
Ecchymosis	X	X
Edema	X	X
Erythema	X	X
Leak		X
Nonocclusive dressing		X
Not flushing	X	X
Not intact	X	X
Phlebitis		X
Pulled	X	
Soiled dressing	X	X



Exploratory data analysis was completed using RStudio integrated development environment to inform model development. Separate models were created for nontunneled lines and tunneled lines. Prediction models were developed using Python coding language and Databricks data intelligence platform. Under-sampling was used to account for uneven distribution between CLABSI and non-CLABSI cases. Logistic regression, random forest, and XGBoost models were created. Candidate models were compared using area under the curve (AUC), sensitivity, and specificity. **Result:** Using the XGBoost model for nontunneled line, the test AUC was 0.831, demonstrating strong discriminative power. This model achieved a sensitivity of 0.828, indicating its ability to correctly identify a high proportion of actual CLABSI cases, and a specificity of 0.751, reflecting its capacity to correctly identify non-CLABSI cases. Similarly, the tunneled line XGBoost model yielded an AUC of 0.845, showcasing excellent predictive performance. Its sensitivity was 0.833 and specificity was 0.668. In both models, the number of modifiable risks was the most significant factor (Tables 1-3). **Conclusion:** The CLABSI prediction model enables real-time identification and intervention to support proactive prevention, leveraging these performance characteristics to improve patient care. While predictive modeling has limitations due to EHR documentation, clinician expertise and frontline team engagement remain essential for deploying resources to highest risk patients.

Antimicrobial Stewardship & Healthcare Epidemiology 2026;6(Suppl. S1):s189–s190
doi:10.1017/ash.2026.10736

Fig 1

IP Validation Round	Intent (N = 0, Y = 5)	Relevance (N = 0, Y = 5)	Accuracy (N = 0, Y = 5)	Completeness (1-5)	Consistency (1-5)	Overall
1	5.0 (100%)	4.2 (84%)	3.2 (64%)	3.8 (76%)	4.5 (90%)	84%
2	5.0 (100%)	4.5 (90%)	3.1 (62%)	3.6 (72%)	4.4 (88%)	82%

Fig 2

IP Validation Round	Intent (1-4)	Accuracy (1-4)	Completeness (1-4)	The response helped me achieve my intended goal (1-4)	This content is ready to be deployed to end users (1-4)	Overall
3	3.9 (99%)	3.8 (94%)	3.7 (93%)	3.7 (92%)	3.6 (89%)	93%

Fig 3

Nurse Manager Validation Round	Intent (1-5)	Trust Accuracy of Response (1-4)	Completeness (1-5)	The response helped me achieve my intended goal (1-5)	Ease of Use (1-4)	Overall
1	4.7 (94%)	3.2 (79%)	4.5 (91%)	4.3 (86%)	3.8 (95%)	89%

Fig 4

Queries	Optimal	Suboptimal/ No Harm	Suboptimal/ Possible Harm
29	20	1	8

Presentation Type:
Poster Presentation
Subject Category: Technology
Validation of an Internal Infection Control Chatbot‡

Molly Steele[1], Vera Chu[1], Palak Patel[2], Patricia Zuccaro[3] and David Zhang[4]
[1]University of Chicago Medicine; [2]University of Chicago; [3]The University of Chicago Medical Center and [4]University of Chicago Medical Center

Background: Effective infection prevention and control is integral to the delivery of safe patient care, and with ever increasing demands on their attention, systems to help navigate policy-based guidance can support infection preventionists (IPs) and their healthcare colleagues. Here we present the validation process and results for an internal infection control chatbot. **Methods:** Eleven frequently asked questions were selected to evaluate the performance of ChatUCM, an AI-powered chatbot. Each question was posed five times and response quality was scored by intent recognition, relevance, accuracy, completeness, and consistency. A category pass threshold was set at 75%. IP feedback was incorporated in a ChatUCM update and response quality was scored again. Reference document edits and real-time prompt modifications were made based on post-update scores. A final round of scoring was performed using similar validation metrics. Two IPs scored a subset of five out of the eleven frequently asked questions to assess interrater reliability. Six nursing managers were invited to ask ChatUCM any questions and score its responses by intent recognition, their ability to trust response accuracy, whether their question was fully answered, the achievement of intended goals, and ease of use. 29 questions were submitted. An IP reviewer also scored these ChatUCM responses as optimal, suboptimal/ no harm, and suboptimal/ possible harm. **Results:** The impact of the update on scores was variable. ChatUCM performed well on the same 7 out of 11 questions before and after validation round 1 (Fig. 1). After the final round of retraining, ChatUCM passed all categories with high interrater reliability for our frequently asked questions (Fig. 2). Response quality was less predictable outside of these eleven questions (Fig 3, Fig 4). To ensure its safe use, multiple safety checks were created, including disclaimer language, citation links, education, and a query report to facilitate audits. We encountered multiple surprises while validating ChatUCM. The first update resulted in significant improvements, such as the ability to digest data in table and flow

‡The abstract titles in this supplement have been corrected since original publication. A notice detailing the changes has been published at https://doi.org/10.1017/ash.2026.10802.

https://doi.org/10.1017/ash.2026.10737 Published online by Cambridge University Press