

KANSAS JOURNAL *of* MEDICINE

TABLE OF CONTENTS

ORIGINAL RESEARCH

- 125 Zero-Shot Extraction of Structured Clinical Data from Free-Text CT Reports Using Foundational Large Language Models
Amulya Gupta, MBBS, Alex Honglo, MS-3, Quynh Nguyen, MS-3, Jake Caldwell, MS-3, Ben Choi, MS-3, Christopher J. Harvey, B.S., Jeffrey Thompson, Ph.D., Diego Mazzotti, Ph.D., Zijun Yao, Ph.D., Amit Noheria, MBBS, S.M.
- 133 Changes in Burnout and Other Types of Emotional Distress among KUSM Medical Students Between 2019 and 2025
Samuel Ofei-Dodoo, Ph.D., MPA, M.A., CPH, Rosey Zackula, M.A.

QUALITY IMPROVEMENT

- 148 Training and Assessment of Non-Technical Human Factor Skills Among Anesthesiology Residents: An Exploratory Educational Quality Improvement Initiative
Sukruta Pradhan, M.D., William Krogman, M.S., Dakota Rapp, M.D., Sawyer Schmidt, B.S., James Walker, M.D., Brian Mahoney, D.O., Blake Gorman, D.O.
- 160 Clinical Workflow Associated with Spanish-Speaking Patient Encounters at JayDoc Community Clinic-Wichita: A Quality Improvement Project
Leah C. Duncan, MS-3, Wyl Halimeh, M.D., M.S., Dania Jumpa, M.D., Carlos Schwindt, MS-4, Evie Schmitz, MS-3, Dulcinea Rakestraw, Ph.D., MPH

LETTERS TO THE EDITOR

- 167 Delayed Eyelid Masses: When to Suspect Infiltrative Mimickers
Emmanuel Vazquez, M.D.
- 170 In Reply to Vazquez
Diana N. Le, MS-4, Jennifer K. Burgoyne, M.D.

Zero-Shot Extraction of Structured Clinical Data from Free-Text CT Reports Using Foundational Large Language Models

Amulya Gupta, MBBS^{1,2*}

Alex Hongslo, MS-3^{1*}

Quynh Nguyen, MS-3¹

Jake Caldwell, MS-3¹

Ben Choi, MS-3¹

Christopher J. Harvey, B.S.^{1,2}

Jeffrey Thompson, Ph.D.^{1,3}

Diego Mazzotti, Ph.D.^{1,4}

Zijun Yao, Ph.D.⁵

Amit Noheria, MBBS, S.M.¹

¹The University of Kansas School of Medicine-Kansas City, Kansas City, KS

²Department of Cardiovascular Medicine

³Department of Biostatistics and Data Science

⁴Department of Medical Informatics

⁵Department of Electrical Engineering and Computer Science, The University of Kansas, Lawrence, KS, USA

**Both authors contributed equally to this manuscript.*



MORE ONLINE
journals.ku.edu/kjm

Corresponding Author:

Amit Noheria, MBBS, S.M.

3901 Rainbow Blvd., MS 3901

Kansas City, KS 66160

Phone: (913) 588-9600

Email: noheriaa@gmail.com

ABSTRACT

Introduction. Large language models (LLMs) are being explored for multiple applications in medical research, including medical text classification. Authors of the study evaluated the performance of three off-the-shelf LLMs in classifying unstructured free-text CT pulmonary angiography reports for pulmonary embolism (PE)-related diagnostic labels.

Methods. Authors assessed human-labeled CT reports using three LLMs (ChatGPT-4o, Llama 3.3 70B, and Llama 3.1 8B) with zero-shot simple and chain-of-thought (CoT) prompts. Classification tasks included acute PE (N = 1,025) and, among reports with PE (n = 660), bilateral PE and large PE.

Results. LLMs demonstrated high classification performance for acute PE and bilateral PE but lower performance for large PE. With simple prompts, the highest F1 scores were 97.0% for acute PE (Llama 3.1 8B; sensitivity, 95.6%; specificity, 97.3%), 92.9% for bilateral PE (ChatGPT-4o; sensitivity, 98.0%; specificity, 88.6%), and 72.8% for large PE (ChatGPT-4o; sensitivity, 82.9%; specificity, 82.2%). CoT prompting did not meaningfully improve acute PE classification but provided slight improvements in subclassification performance for bilateral and large PE.

Conclusions. Off-the-shelf LLMs can classify free-text CT reports using simple prompt engineering and may facilitate automated data extraction for medical research following appropriate evaluation of classification performance.

INTRODUCTION

In clinical research, extracting structured data from unstructured medical reports is a potential application of large language models (LLMs). Historically, various natural language processing (NLP) systems have been considered the standard for automated information extraction from free-text documentation.¹ Although these systems computationally are efficient, they lack flexibility and often require substantial

task-specific customization and preprocessing.² In contrast, LLMs offer greater semantic flexibility and contextual understanding; however, they computationally are intensive, prone to hallucinations, and may not consistently follow prompt instructions.³

We therefore investigated the utility of readily available, off-the-shelf LLMs for extracting pulmonary embolism (PE) labels from free-text computed tomography pulmonary angiography (CTPA) reports using simple zero-shot and chain-of-thought (CoT) prompting. In simple zero-shot prompting, the LLM directly is prompted to provide a yes/no classification without worked input-output examples. In CoT prompting, the LLM is prompted to first generate a description of the report and then provide a yes/no classification, also without worked examples. Previous studies have evaluated LLMs for extracting structured information from free-text clinical notes.⁴⁻⁶ The aim of our study was to assess the reproducibility of these findings and directly compare simple zero-shot and CoT prompting strategies.

METHODS

This study was approved by the Institutional Review Board (IRB) of the University of Kansas Medical Center and did not require informed consent because of its retrospective design. The manuscript was prepared in accordance with the CHART (Chatbot Assessment Reporting Tool) statement for reporting evaluations of artificial intelligence chatbots, as recommended by the EQUATOR Network.⁷ We identified CTPA reports from the University of Kansas Health System from 2010 to 2022 using a natural language search for PE-related terms in Illuminate[®] Insight[™] (Softtek Illuminate Inc., Overland Park, KS), a clinical search engine, as part of a larger project evaluating PE diagnosis and outcomes. Patients with eligible CTPA reports were randomly assigned to four medical students for chart review using a REDCap[®] form. Each chart was reviewed by only one student, and the original CTPA report text was used for adjudication.

Anonymized CTPA report text was uploaded to a HIPAA-compliant, cloud-based institutional research platform built on Databricks[®] (Databricks, Inc., San Francisco, CA). Three LLM endpoints: ChatGPT-4o (released May 2024; closed-source, developed by OpenAI PBC, San Francisco, CA), Llama 3.3 (70 billion parameters; released De-

cember 2024; open-source, developed by Meta Platforms, Inc., Menlo Park, CA), and Llama 3.1 (8 billion parameters; released July 2024; open-source, developed by Meta Platforms, Inc., Menlo Park, CA) were queried independently on October 8, 2025, in Kansas City, USA, using the same prompts. The primary classification task was acute PE (yes/no), based on the radiologist's impression (Table 1). Among reports classified as positive for acute PE, additional sub-classification tasks assessed bilateral PE (defined as acute PE findings in both lungs, whether explicitly described or implicitly indicated) and large PE (defined as lobar involvement or greater). LLM outputs were post-processed using simple string-matching rules to extract binary labels.

CoT prompting was evaluated using the same three LLM endpoints. The prompts instructed the LLM to provide an explanation of the PE-related findings in the CTPA report followed by a binary conclusion. The output from this initial query was provided to a second LLM query, which was instructed to extract a single-word binary outcome from the preceding output.

For each classification task, we constructed voting ensemble models by aggregating the binary outputs from all three LLMs separately for simple and CoT prompting. Ensemble performance was evaluated at different voting thresholds based on the number of LLMs producing a positive classification.

We also evaluated the performance of the LLMs against a rule-based natural language processing (NLP) approach using regular-expression (Regex) string matching (detailed in the Supplementary Material, available online at journals.ku.edu/kjm). The matching rules were predefined based on domain expertise and were not iteratively refined after initial specification to avoid overfitting and ensure a fair comparison with the zero-shot LLM approaches.

LLM performance was evaluated using sensitivity, specificity, precision, accuracy, and F1 score, with human-labeled data serving as the reference standard. All analyses were performed using R[®] (version 4.4.1; R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Patient charts for 1,106 CTPA reports were manually reviewed for acute PE. Of these, 81 were excluded because of equivocal or undetermined findings, leaving 1,025 reports

Table 1. Prompts provided to the large language model for each report.

Task	Simple Prompt	Chain-of-Thought Prompt
Acute PE (Yes/No)	Evaluate the following CT scan report to determine if the patient has acute pulmonary embolism or not. This report is from a single patient only. Based on the contents of the clinical note, answer “yes” or “no” whether the report strongly suggest acute pulmonary embolism. Make sure to only return a single word answer, “yes” or “no” without any explanations, words, symbols. If the report is suggestive of chronic pulmonary embolism without any new acute PE, then conclude as “no”.	Evaluate the following CT scan report to determine if the patient has acute pulmonary embolism or not. This report is from a single patient only. Based on the contents of the clinical note, give a clear explanation about the description of pulmonary embolism related findings. Then conclude with a “yes” or “no” answer to whether acute pulmonary embolism is present or not. If the report is suggestive of chronic pulmonary embolism without any new acute PE, then conclude as “no”.
Bilateral PE (Yes/No)	Evaluate the following CT scan report of a patient with acute pulmonary embolism to determine whether the patient’s report suggests bilateral PE i.e. acute PE findings are present in both the lungs. This report is from a single patient only. Based on the contents of the clinical note, answer “yes” or “no” whether the report strongly suggest that acute PE is bilateral in this patient. Make sure to only return a single word answer, “yes” or “no” without any explanations, words, symbols. Do not include findings related to chronic PE in making this decision.	Evaluate the following CT scan report of a patient with acute pulmonary embolism to determine whether the patient’s report suggests bilateral PE i.e. acute PE findings are present in both the lungs. This report is from a single patient only. Based on the contents of the clinical note, give a clear explanation about the description of pulmonary embolism related findings. Then conclude with a “yes” or “no” answer to whether acute PE findings are present in both lungs or not. Do not include findings related to chronic PE in making this decision.
Extent of PE (Large/Small)	Evaluate the following CT scan report of a patient with acute pulmonary embolism to determine whether the patient’s report suggests large PE defined as affecting the lobar arteries, right or left pulmonary arteries, or being a saddle PE. The embolism should be classified as small if it is segmental or sub-segmental. If the affected arteries are not provided but the report categorically mentions it to be small, then answer it to be small. In case findings are suggestive of both small and large PE, answer it to be large. Based on the contents of the clinical note, answer “large” or “small” for acute PE in this patient. Make sure to only return a single word answer, “large” or “small” without any explanations, words, symbols. Do not include findings related to chronic PE in making this decision.	Evaluate the following CT scan report of a patient with acute pulmonary embolism to determine whether the patient’s report suggests large PE defined as affecting the lobar arteries, right or left pulmonary arteries, or being a saddle PE. The embolism should be classified as small if it is segmental or sub-segmental. If the affected arteries are not provided but the report categorically mentions it to be small, then answer it to be small. In case findings are suggestive of both small and large PE, answer it to be large. Based on the contents of the clinical note, give a clear explanation about the description of pulmonary embolism related findings. Then conclude with a “large” or “small” answer for acute PE. Do not include findings related to chronic PE in making this decision.

for analysis. Acute PE was present in 660 (64.4%) reports; among the 642 reports with non-missing bilateral labels, 300 (46.7%) were bilateral and among the 660 reports with acute PE, 187 were classified as large PE. Manually extracted yes/no labels for bilateral PE were missing for 18 (2.7%) patients, who were excluded from the bilateral PE analysis.

Acute PE (Yes/No)

As shown in Table 2, with simple prompting, Llama 3.1 8B achieved the numerically highest F1 score (97.0%), followed by ChatGPT-4o (96.3%) and Llama 3.3 70B (95.5%) for acute PE classification. With CoT prompting, performance declined across all three models (F1 scores: ChatGPT-4o,

95.6%; Llama 3.3 70B, 94.7%; Llama 3.1 8B, 88.8%). The rule-based NLP approach had an F1 score of 75.8%.

Bilateral PE (Yes/No)

With simple prompting, ChatGPT-4o achieved the numerically highest F1 score (92.9%), followed by Llama 3.3 70B (90.3%) and Llama 3.1 8B (84.5%). With CoT prompting, ChatGPT-4o and Llama 3.3 70B showed slight improvements in performance (F1 scores, 93.6% and 93.8%, respectively), whereas Llama 3.1 8B showed no change (84.5%). The rule-based NLP approach had an F1 score of 66.7%.

Table 2. Performance of large language models in CTPA report classification tasks for pulmonary embolism labels.

Classification Task and Model	Simple Prompting					Chain-of-Thought Prompting				
	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)	F1-score (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)	F1-score (%)
Acute PE (N = 1025)										
Rule-based NLP	78.0	49.9	73.8	68.0	75.8	-	-	-	-	-
ChatGPT-4o	93.8	98.1	98.9	95.3	96.3	92.4	98.4	99.0	94.5	95.6
Llama 3.3 70B	92.1	98.6	99.2	94.4	95.5	90.8	98.4	99.0	93.5	94.7
Llama 3.1 8B	95.6	97.3	98.4	96.2	97.0	91.1	74.5	86.6	85.2	88.8
Bilateral Acute PE (n = 642)										
Rule-based NLP	81.7	44.4	56.3	61.8	66.7	-	-	-	-	-
ChatGPT-4o	98.0	88.6	88.3	93.0	92.9	97.0	90.9	90.4	93.8	93.6
Llama 3.3 70B	99.0	82.2	83.0	90.0	90.3	98.0	90.4	89.9	93.9	93.8
Llama 3.1 8B	98.0	70.2	74.2	83.2	84.5	98.7	69.3	73.8	83.0	84.5
Large PE (n = 660)										
Rule-based NLP	63.1	76.3	51.3	72.6	56.6	-	-	-	-	-
ChatGPT-4o	82.9	82.2	64.9	82.4	72.8	84.0	83.1	66.2	83.3	74.1
Llama 3.3 70B	80.2	82.9	64.9	82.1	71.8	79.1	84.4	66.7	82.9	72.4
Llama 3.1 8B	87.2	62.4	47.8	69.4	61.7	82.4	68.5	50.8	72.4	62.9

Extent of PE (Large/Small)

Classification performance was lowest for large PE compared with the other two classification tasks. With simple prompting, ChatGPT-4o had the highest F1 score (72.8%), followed by Llama 3.3 70B (71.8%) and Llama 3.1 8B (61.7%). With CoT prompting, there were slight numerical improvements in performance across all three models (F1 scores: ChatGPT-4o, 74.1%; Llama 3.3 70B, 72.4%; Llama 3.1 8B, 62.9%). The rule-based NLP approach had an F1 score of 56.6%.

As shown in Supplementary Tables 1 and 2 (available online at journals.ku.edu/kjm), combining the outputs of the three LLMs in voting ensemble models resulted in slight absolute improvements in performance. For the simple-prompt voting ensemble, the highest F1 scores were 97.6% for acute PE (≥ 1 vote), 93.3% for bilateral PE (≥ 3 votes), and 73.2% for large PE (≥ 3 votes). For the CoT-prompt voting ensemble, the highest F1 scores were 95.3% for acute PE (≥ 2 vote), 94.5% for bilateral PE (≥ 2 votes), and 72.7% for large PE (≥ 3 votes).

DISCUSSION

We evaluated the performance of three publicly available, off-the-shelf LLMs for extracting PE-related diagnostic labels from free-text CTPA reports. The key findings of our study are as follows. First, all tested models, including the smaller and computationally efficient Llama 3.1 8B, demonstrated high classification performance for the relatively straightforward task of identifying acute PE. Second, performance declined across models for more nuanced classification tasks, including identification of bilateral PE and large PE. Although Llama 3.1 8B maintained relatively strong performance for bilateral PE, it had the lowest F1 scores among the three models for both subclassification tasks, suggesting that larger or better-tuned models may be more reliable for more nuanced information extraction. Third, CoT prompting did not improve performance for the straightforward primary classification task but may provide modest improvements for more nuanced classifications. Fourth, voting ensemble models that aggregated binary outputs from multiple LLMs yielded only marginal improvements in overall performance; however, adjusting the

voting threshold provided a means of tuning the balance between sensitivity and specificity.

This analysis builds upon several recent studies evaluating LLMs for structuring free-text clinical data.⁸⁻¹² As reported in these studies, although LLMs relatively are easy to deploy and can achieve high classification performance, they may produce inconsistent outputs that are difficult to predict and systematically mitigate.¹³ LLMs are also prone to hallucinations, in which they generate information or labels that are unsupported by the input or assign classifications when the available information is ambiguous.¹⁴

LLMs are based on transformer architectures that use contextual representations to predict subsequent tokens in a sequence.¹⁵ Unlike conventional rule-based NLP approaches, LLMs do not rely solely on explicitly defined terms or patterns to assign classifications. This flexibility may contribute to their ability to recognize contextual relationships in free-text reports but also makes their outputs less transparent and more probabilistic. Moreover, in most LLM endpoints, the prompt instructions and input text, such as clinical reports, are provided as part of a single input sequence. This may make LLM-based systems sensitive to irregularities in free-text input, including unexpected formatting, special characters, or potentially adversarial instructions embedded within the input.

Nonetheless, our findings demonstrate good empirical performance of zero-shot LLMs for free-text classification, particularly for relatively straightforward tasks such as acute PE diagnosis. Based on these observations, we present a general framework for binary label classification using LLMs in Figure 1.

CoT prompting is a technique in which the model is prompted to generate intermediate reasoning steps before providing a final answer, which may improve classification performance and provide additional information for reviewing the basis of a classification.¹⁶ In our dataset, performance for the primary classification of acute PE was lower with CoT prompting than with simple prompting. The decline was most pronounced for the smallest LLM, Llama 3.1 8B, suggesting that smaller models may be less suitable for CoT prompting in this setting. In contrast, CoT prompting produced modest improvements in performance for the subclassification tasks of bilateral and large PE.

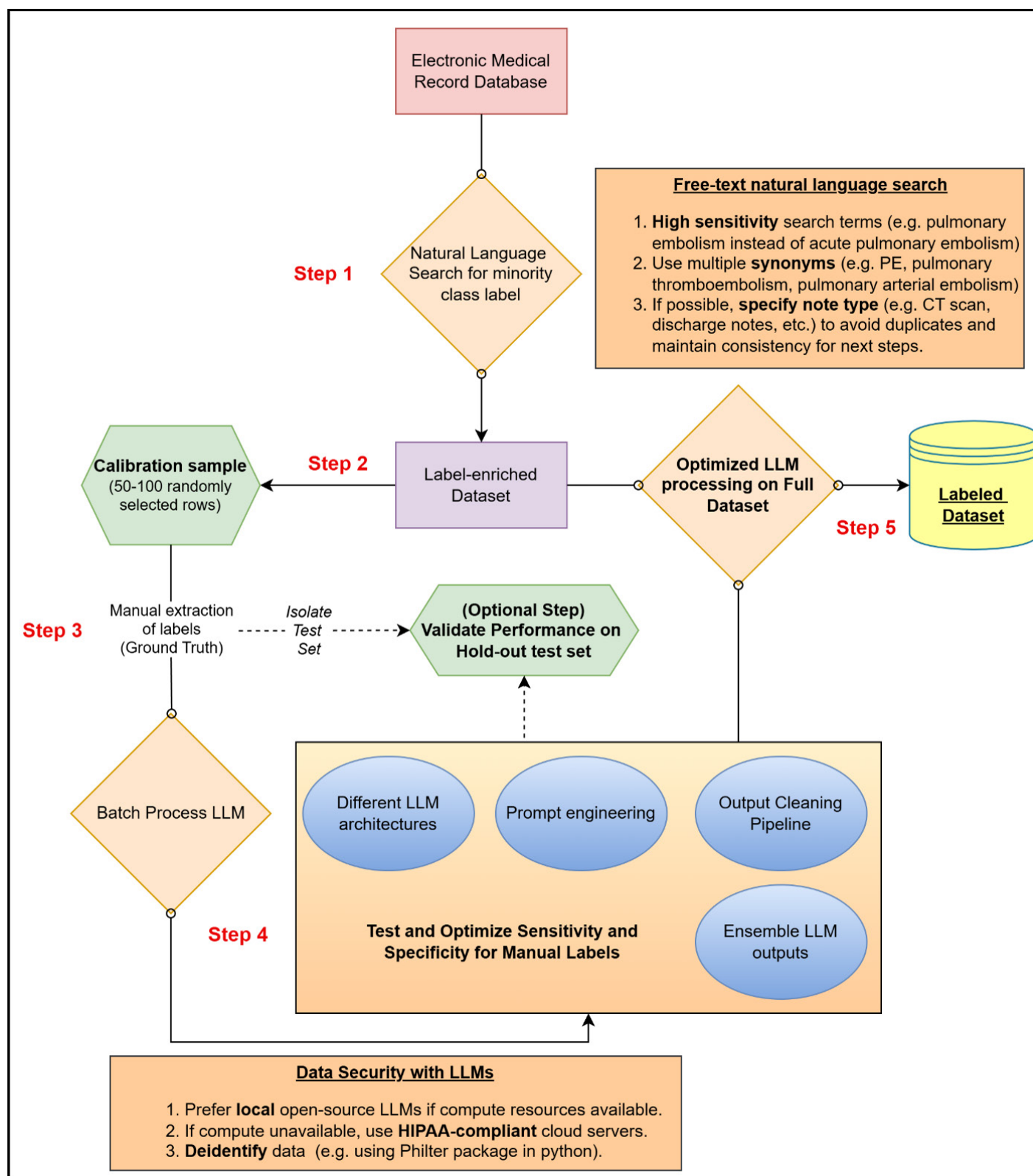


Figure 1. A general outline for binary label classification of free-text clinical reports using a simple large language model and natural language search framework.

Alternatives to LLMs include rule-based NLP approaches. However, these approaches can be brittle to minor variations in language and may therefore have limited sensitivity or specificity.^{17,18} Machine-learning and hybrid NLP approaches are other alternatives for text classification and can achieve high classification performance but generally require substantial numbers of training samples and extensive testing before deployment.^{19,20} Fine-tuning LLMs can improve classification performance but is resource-intensive, requires labeled datasets, and produces task-specific models that may be less practical for smaller-scale research projects.²¹

Limitations

First, we used label-enriched reports that were prefiltered using a natural language search for PE-related terms. This sampling approach may have resulted in a study population that differs from the broader population of CTPA reports encountered in clinical practice and may have affected observed classification performance. Second, each case was manually labeled by a single reviewer from a group of four medical students without independent cross-verification. In addition, reviewers may have considered broader clinical information from the patient charts beyond the CTPA reports when assigning labels, whereas the LLMs were provided only with the CTPA report text. This difference between the human labeling process and the LLM input may have introduced misclassification and affected estimates of model performance. Third, we did not evaluate the effects of varying LLM hyperparameters or prompt formulations on performance. Finally, larger studies involving more diverse clinical reports and classification tasks are needed to evaluate the generalizability of our findings.

CONCLUSIONS

Off-the-shelf LLMs effectively can classify free-text CTPA reports into diagnostic labels using simple zero-shot prompts. In this study, larger models, Llama 3.3 70B and ChatGPT-4o, generally demonstrated higher observed performance than the smaller Llama 3.1 8B model, particularly for more nuanced classification tasks. Further research is warranted to evaluate the utility of LLMs for medical text classification across diverse clinical datasets, prompting strategies, model architectures, and hybrid NLP approaches.

ARTICLE INFORMATION

Received April 22, 2026; Accepted for publication September 8, 2026; Published online October 16, 2026, *Kans J Med* 2026 Sep-Oct; 19:125-132. <https://doi.org/10.17161/kjm.vol19.25396>.

Disclosures: ChatGPT was used solely for grammatical and stylistic refinement after the initial draft was prepared by the authors. The authors reviewed and approved the final manuscript and take full responsibility for the accuracy, integrity, and attribution of its content. The authors report no relevant financial conflicts of interest.

REFERENCES

1. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: A systematic review. *J Am Med Inform Assoc*. 2016; 23(5):1007–15. Epub 2016 Feb 5. PMID: 26911811.
2. Carrell DS, Schoen RE, Leffler DA, et al. Challenges in adapting existing clinical natural language processing systems to multiple, diverse health care settings. *J Am Med Inform Assoc*. 2017; 24(5):986–991. PMID: 28419261.
3. Lu Z, Peng Y, Cohen T, Ghassemi M, Weng C, Tian S. Large language models in biomedicine and health: Current research landscape and future directions. *J Am Med Inform Assoc*. 2024; 31(9):1801–1811. PMID: 39169867.
4. Joseph E, Vallee P, Perennec T, et al. Development and assessment of a pipeline for extracting structured data from free-text medical reports using a large language model. *JCO Clin Cancer Info*. 2026; 10(1): e2500133. PMID: 41707099.
5. Garcia-Carmona AM, Prieto ML, Puertas E, Beunza JJ. Leveraging large language models for accurate retrieval of patient information from medical reports: Systematic evaluation study. *JMIR AI*. 2025; 4:e68776. PMID: 40608403.
6. Reichenpfader D, Müller H, Denecke K. A scoping review of large language model based approaches for information extraction from radiology reports. *NPJ Digit Med*. 2024 Aug 24; 7(1):222. PMID: 39182008.
7. Huo B, Collins G, Chartash D, et al. Reporting guideline

- for chatbot health advice studies: the Chatbot Assessment Reporting Tool (CHART) statement. *BMJ Med.* 2025; 4(1):e001632. PMID: 40761518.
8. Le Guellec B, Lefèvre A, Geay C, et al. Performance of an open-source large language model in extracting information from free-text radiology reports. *Radiol Artif Intell.* 2024; 6(4):e230364. PMID: 38717292.
 9. Abdullah A, Kim ST. Automated radiology report labeling in chest X-ray pathologies: Development and evaluation of a large language model framework. *JMIR Med Inform.* 2025; 13:e68618. PMID: 40153539.
 10. Hu D, Liu B, Zhu X, Lu X, Wu N. Zero-shot information extraction from radiological reports using ChatGPT. *Int J Med Inform.* 2024; 183:105321. PMID: 38157785.
 11. Huang J, Yang DM, Rong R, et al. A critical assessment of using ChatGPT for extracting structured data from clinical notes. *NPJ Digit Med.* 2024; 7(1):106. PMID: 38693429.
 12. Burford KG, Itzkowitz NG, Ortega AG, Teitler JO, Rundle AG. Use of generative AI to identify helmet status among patients with micromobility-related injuries from unstructured clinical notes. *JAMA Netw Open.* 2024; 7(8):e2425981. PMID: 39136946.
 13. Shah SV. Accuracy, consistency, and hallucination of large language models when analyzing unstructured clinical notes in electronic medical records. *JAMA Netw Open.* 2024; 7(8):e2425953. PMID: 39136951.
 14. Bijoy Das A, Ahmed S, Sakib SK. Hallucinations and key information extraction in medical texts: A comprehensive assessment of open-source large language models. *arXiv e-prints.* 2025:arXiv:2504.19061. doi:10.48550/arXiv.2504.19061.
 15. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *arXiv e-prints.* 2017:arXiv:1706.03762. doi:10.48550/arXiv.1706.03762.
 16. Miao J, Thongprayoon C, Suppadungsuk S, Krisanapan P, Radhakrishnan Y, Cheungpasitporn W. Chain of thought utilization in large language models and application in nephrology. *Medicina (Kaunas).* 2024; 60(1):148. PMID: 38256408.
 17. Rashedi S, Bukhari S, Krishnathan D, et al. Optimizing the accuracy of natural language processing tools for pulmonary embolism detection through integration with claims data: The PE-EHR+ study. *Thromb Haemost.* Epub 2026 Feb 9. PMID: 41605431.
 18. Somani SS, Honarvar H, Narula S, et al. Development of a machine learning model using electrocardiogram signals to improve acute pulmonary embolism screening. *Eur Heart J Digit Health.* 2021; 3(1):56-66. PMID: 35355847.
 19. Hassan SM, Mylvaganam R, Didebulidze T, et al. Leveraging hybrid natural language processing techniques for large-scale pulmonary embolism identification. *JACC Adv.* 2025; 4(11 Pt 2):101845. Epub 2025 May 21. PMID: 41350101.
 20. Chen MC, Ball RL, Yang L, et al. Deep learning to classify radiology free-text reports. *Radiology.* 2018; 286(3):845-852. Epub 2017 Nov 13. PMID: 29135365.
 21. Shankar SV, Dhingra LS, Aminorroaya A, et al. Automated transformation of unstructured cardiovascular diagnostic reports into structured datasets using sequentially deployed large language models. *Eur Heart J - Digit Health.* 2025; 6(4):783-796. PMID: 40703108.

Keywords: *large language models; artificial intelligence; natural language processing; diagnosis; electronic health records*

Changes in Burnout and Other Types of Emotional Distress Among KUSM Medical Students Between 2019 and 2025

Samuel Ofei-Dodoo, Ph.D., MPA, M.A.,
CPH^{1,2,3}

Rosey Zackula, M.A.^{1,4}

¹The University of Kansas School of
Medicine-Wichita

²Department of Internal Medicine

³Academic and Student Affairs

⁴Office of Research

Received July 23, 2026; Accepted for
publication October 1, 2026; Pub-
lished October 16, 2026, Kans J Med
2026 Sept-Oct; 19:133-147. <https://doi.org/10.17161/kjm.vol19.25804>.

Conflict of Interest Disclosure: Samuel Ofei-Dodoo is the Editor-in-Chief and Managing Editor for *The Kansas Journal of Medicine*. Rosey Zackula is an Editorial Board member of *The Kansas Journal of Medicine*.

Corresponding Author:

Samuel Ofei-Dodoo, Ph.D., MPA, M.A.,
CPH, The University of Kansas School of
Medicine-Wichita
1010 N. Kansas St., Wichita, KS 67214
sofeidodoo@kumc.edu

ABSTRACT

Background. Burnout among medical students is associated with depression, fatigue, suicidal ideation, and impaired academic performance. Authors evaluated changes in burnout and well-being among medical students at the University of Kansas School of Medicine (KUSM) between 2019 and 2025 and explored students' perspectives on promoting wellness.

Methods. A repeated cross-sectional, anonymous online survey was administered to all KUSM medical students in 2025. Measures included burnout, depression, fatigue, quality of life, general health, suicidal ideation, and open-ended wellness questions. Quantitative findings were compared with a similar 2019 survey. Bivariate analyses, multivariable logistic regression, and thematic analysis were performed.

Results. Of 808 eligible students, 295 participated (36.5%) and 288 completed the survey (response rate, 35.6%). Overall burnout prevalence remained similar to 2019 (43.4% vs. 48.0%), although depersonalization decreased significantly (16.3% vs. 26.5%; $p = 0.002$). Depression remained common (40.4%), while fatigue worsened significantly ($p = 0.018$). Overall, 7.6% of respondents reported suicidal ideation during the previous 12 months. Burnout was significantly associated with depression, fatigue, and suicidal ideation (all $p < 0.001$). Third-year students had the highest rates of psychological distress and more than twice the odds of screening positive for depression than first-year students (aOR, 2.13; 95% CI, 1.04-4.36). Qualitative analysis identified 10 themes, with students attributing burnout primarily to insufficient time, limited flexibility, excessive workload, and reduced autonomy.

Conclusions. Burnout remains highly prevalent among medical students and is closely associated with depression, fatigue, and suicidal ideation. Students emphasized the need for system-level interventions that increase flexibility, protected time, and autonomy to improve learner well-being.

INTRODUCTION

Medical students in the United States face intense academic demands, prolonged study hours, and high-stakes evaluations, all of which contribute to substantial stress and emotional strain.¹⁻³ A large body of literature has consistently documented high rates of burnout, depression, anxiety, and other forms of psychological distress in this population.⁴⁻⁶ For example, a 2019 study of the University of Kansas School of Medicine (KUSM) students found that 48% reported manifestations of burnout, 46% screened positive for depression, and 44.6% experienced high levels of fatigue within the past week.⁴ Burnout, characterized by emotional exhaustion, depersonalization, and a reduced sense of personal accomplishment,⁷ has been reported in more than half of U.S. medical students in some national surveys.^{8,9} These findings underscore that psychological distress is not an isolated issue but a pervasive and persistent concern across medical education.

Importantly, the consequences of emotional distress extend beyond individual well-being. Burnout and related conditions have been associated with impaired academic performance, reduced empathy, increased risk of professional misconduct, and lower long-term career satisfaction, all of which may ultimately compromise patient care and the sustainability of the physician workforce.¹⁰⁻¹² Thus, addressing medical student wellness is not only an educational priority but also a critical component of maintaining a high-quality health care system.

In response to these concerns, the Liaison Committee on Medical Education (LCME) now requires medical schools to implement structured wellness and support programs.¹³ In alignment with this mandate, KUSM has continued to support student well-being through a multifaceted approach that includes access to counseling and Student Health services, academic and personal coaching, small-group learning and peer support, wellness and recreational activities, and social and community-building opportunities. These initiatives reflect an institutional commitment to addressing students' emotional and physical well-being. Most of these initiatives have been in place prior to the 2025 data collection. However, despite widespread implementation of such programs, there remains limited evidence regarding their effectiveness in improving student well-being over time within individual institutions. Therefore, this

follow-up study was conducted to evaluate the changes in student wellness associated with these wellness initiatives at KUSM by comparing 2025 survey findings with baseline data from the 2019 study.⁴

METHODS

Study Design, Setting, and Participants

This repeated cross-sectional survey study included 808 medical students across the three campuses of KUSM. The 2025 survey used methods similar to those employed in the study published in 2019,⁴ with the addition of two questions: (1) an item assessing suicidal thoughts and (2) an open-ended question asking students to share resources and/or activities they believe would help promote wellness among medical students.

A sample size of 250 participants was determined to provide adequate statistical power (>0.95) to detect significant relationships among variables, assuming 2 degrees of freedom, an effect size of 0.30, and a significance level of $p < 0.05$.¹⁴ The study was approved by the University of Kansas Medical Center Institutional Review Board (IRB).

The KUSM has three campuses: Kansas City, Salina, and Wichita. The Kansas City campus is a large, urban academic medical center, with class sizes of 175 students in each of the first two years and 125 in the third and fourth years. The Wichita campus, located in an urban/suburban setting, is community-based, with 28 students in each of the first two years and 75 in the third and fourth years (this includes 45 students who complete their first two years in Kansas City and then transfer to Wichita for their clinical training). The Salina campus, situated in a town of 46,889, enrolls eight students per class. Each campus offers distinct clinical experiences and educational opportunities. While the curricular goals, objectives, and assessment methods are standardized across all campuses, the student experience differs.^{4,15}

Study Measures

The survey included items on demographic information (age, gender, year in training, and campus location), burnout, symptoms of depression, suicidal ideation, fatigue, quality of life, and self-reported general health.

Burnout. We measured burnout using two validated single-item measures of emotional exhaustion and deper-

sonalization adapted from the Maslach Burnout Inventory (MBI-22).¹⁶ The emotional exhaustion item (“I feel burned out from my work/education”) and depersonalization item (“I’ve become more callous toward people since I became a medical student”) have demonstrated strong correlations with their respective MBI-22 domains and have been used in previous studies of medical students.^{4,8,9,16}

Students rated each item on a 7-point Likert scale ranging from 0 (*Never*) to 6 (*Every day*). Scores were categorized as low, moderate, or high using established cutoffs.^{8,9} Higher scores indicate greater burnout. Consistent with prior studies, students with high scores in emotional exhaustion and/or depersonalization were classified as having at least one manifestation of burnout.^{8,9}

Symptoms of Depression. Symptoms of depression were assessed using the 2-item Patient Health Questionnaire (PHQ-2), a validated screening tool for depression with reported sensitivity ranging from 72% to 91% and specificity ranging from 67% to 85%, depending on the cutoff score used.^{17,18} The PHQ-2 asks respondents how often, during the past 2 weeks, they have experienced (1) feeling down, depressed, or hopeless and (2) little interest or pleasure in doing usual activities.

Suicidal Ideation. Suicidal ideation was assessed using a single-item screening question: “During the past 12 months, have you had thoughts of taking your own life?” Single-item measures commonly are used in large survey-based studies to screen for suicidal ideation and have demonstrated utility for population-level surveillance, although they are intended for screening purposes and not diagnostic evaluation.¹⁹

Fatigue. Fatigue was assessed using a standardized linear analog scale ranging from 0 (*as bad as it can be*) to 10 (*as good as it can be*), consistent with the approach described by West, et al.²⁰ Students rated their level of fatigue during the past week, and scores of 5 or lower were classified as indicating high fatigue.⁹

Quality of Life and General Health. Quality of Life and General Health were assessed using two single-item measures. Overall quality of life was measured using a global item from the World Health Organization Quality of Life (WHOQOL)-BREF instrument, which has been validated for use among medical students.²¹ Participants rated their overall quality of life on a 5-point Likert scale ranging from 1 (*Very poor*) to 5 (*Very good*).

General health was assessed using a single item adapted from the WHOQOL-BREF: “How satisfied are you with your health?”²¹ Responses were recorded on a 5-point Likert scale ranging from 1 (*Very dissatisfied*) to 5 (*Very satisfied*).

Data collection

Study data were collected and managed using Research Electronic Data Capture (REDCap®) tools hosted at the University of Kansas Medical Center.^{22,23} Each student received an email invitation containing a link to the 13-item survey. Participation was voluntary, and responses were anonymous. To prevent multiple submissions from the same individual, REDCap’s participant list feature was used to distribute a unique survey link to each student. Once a participant accessed the survey through the link, the link became invalid, ensuring that each participant could submit only one response. Data were collected from October 15, 2025, through December 31, 2025.

Statistical Analysis

Standard descriptive statistics were used to characterize the study population and estimate the prevalence of burnout among the respondents. Differences between groups were assessed using one-way analysis of variance or Kruskal-Wallis tests for continuous variables, chi-square tests for categorical variables, and correlation analyses as appropriate. Multivariable logistic regression was performed to identify factors associated with burnout, symptoms of depression, and fatigue. Covariates included year in training, age, gender, and campus location. All analyses were two-sided with a significance level of $\alpha = 0.05$. Analyses were conducted using IBM® SPSS® Statistics for Windows, version 29.0 (IBM Corp., Armonk, NY, USA). No adjustment was made for multiple comparisons; secondary analyses should be interpreted as exploratory.

We used immersion-crystallization approach²⁴ to analyze the content of the open-ended responses regarding resources and/or activities the respondents believed would help promote wellness among medical students.

The study followed STROBE (Strengthening the Reporting of Observational Studies in Epidemiology)²⁵ and CHERRIES (The Checklist for Reporting Results of Internet E-Surveys)²⁶ reporting guidelines.

RESULTS

Demographic Information

Among the 808 KUSM medical students invited to participate in the 2025 survey, 295 completed at least part of the questionnaire, yielding a participation rate of 36.5%. Of these, 288 completed the entire survey, resulting in an overall response rate of 35.6% and a survey completion rate of 97.6%. The 2025 sample had a margin of error of 4.6% at the 95% confidence level, indicating acceptable precision for estimating characteristics of the overall KUSM medical student population.²⁷

Among the 288 respondents who completed the survey, 171 (59.8%) were female, 239 (86.0%) were younger than 30 years, and 83 (29.0%) were first-year medical students (Table 1). Compared with respondents to the 2019 survey, the 2025 cohort differed significantly with respect to age group ($\chi^2[4] = 18.03$, $p = 0.001$) and campus ($\chi^2[2] = 15.17$, $p < 0.001$). No significant differences were observed between survey cohorts in gender or year in medical school.

Comparisons of burnout, depressive symptoms, fatigue during the previous week, quality of life, and general health between the 2025 and 2019 survey respondents are summarized in Table 2a. Overall burnout prevalence was similar between survey years (43.4% in 2025 vs. 48.0% in 2019). Likewise, rates of emotional exhaustion did not differ significantly. However, the prevalence of high depersonalization was significantly lower in 2025 than in 2019 (16.3% vs. 26.5%; $\chi^2[1] = 9.20$, $p = 0.0024$).

Although fewer students screened positive for depression in 2025 than in 2019 (40.4% vs. 45.9%), the difference was not statistically significant. In contrast, fatigue was significantly worse in 2025, with a lower mean (SD) fatigue score than in 2019 (5.5 [2.2] vs. 5.9 [2.1]; $t[610] = 2.37$, $p = 0.0180$). Mean quality of life and general health remained relatively stable across survey years, with most students reporting good or very good quality of life and being satisfied or very satisfied with their general health.

Table 2b presents 2025 rates of burnout, depression, sui-

Table 1. Demographic characteristics of respondents, comparing 2019 and 2025 respondents.

Characteristics	2025 Responders (N = 288)	2019 Responders (N = 379)	χ^2 (df)	P value
Gender, no. (%)			2.27 (2)	0.321
Male	114 (39.9)	166 (45.5)		
Female	171 (59.8)	197 (54.0)		
Other	1 (0.3)	2 (0.5)		
Missing	2	14	-	-
Age, years				
Median age	26 (23 to 51)	25.0	-	-
Age group, no. (%)			18.03 (4)	0.001
20-24	67 (24.1)	140 (38.9)		
25-29	172 (61.9)	182 (50.6)		
30-34	26 (9.4)	28 (7.8)		
35-39	10 (3.6)	5 (1.4)		
40 or more	3 (1.1)	5 (1.4)		
Missing	10	19	-	-
Year in medical school, no. (%)			2.59 (3)	0.459
First-year	83 (29.0)	91 (24.7)		
Second-year	62 (21.7)	96 (26.0)		
Third-year	71 (24.8)	96 (26.0)		
Fourth-year	70 (24.5)	86 (23.3)		
Missing	2	10	-	-
Campus location			15.17 (2)	<0.001
Kansas City	161 (56.3)	259 (70.8)		
Wichita	111 (38.8)	98 (26.8)		
Salina	14 (4.9)	9 (2.5)		
Missing	2	13	-	-

Table 2a. Comparing 2025 and 2019 respondents on overall burnout, depression, fatigue, quality of life, and general health.

Measures	2025 Respondents	2019 Respondents	<i>t</i> or χ^2 (<i>df</i>)	<i>P</i> value
Burnout, no (%)				
Present	125 (43.4)	182 (48.0)	1.41 (1)	0.2356
Not present	163 (56.6)	197 (52.0)		
Emotional Exhaustion, no (%)				
High	116 (40.3)	160 (42.2)	0.25 (1)	0.6146
Low	172 (59.7)	219 (57.8)		
Depersonalization, no (%)				
High	47 (16.3)	99 (26.1)	9.20 (1)	<0.001
Low	241 (83.7)	258 (73.9)		
Depression, no (%)				
Screened positive	115 (40.4)	174 (45.9)	2.05 (1)	0.1527
Screened negative	170 (59.6)	205 (54.1)		
Fatigue*				
Mean (SD)	5.5 (2.2)	5.9 (2.1)	2.37 (610)	0.0180
Level of Fatigue over the past week, no (%)				
High degree	150 (52.3)	169 (45.2)	3.26 (1)	0.0711
Low degree	137 (47.7)	205 (54.8)		
Quality of Life				
Mean (SD)	4.1 (0.7)	4.0 (0.8)	1.72 (600)	0.0860
Overall quality of life, no (%)				
Very poor	0 (0.0)	3 (0.8)	6.67 (4)	0.1541
Poor	8 (2.8)	23 (6.1)		
Neither poor nor good	31 (10.9)	37 (9.8)		
Good	168 (58.9)	208 (54.9)		
Very good	78 (27.4)	107 (28.2)		
General Health				
Mean (SD)	3.6 (1.0)	3.7 (1.0)	12.28 (665)	0.2030
General health satisfaction, no (%)				
Very dissatisfied	0 (0)	6 (1.6)	6.99 (4)	0.1361
Dissatisfied	56 (19.6)	68 (17.9)		
Neither dissatisfied nor satisfied	44 (15.4)	46 (12.1)		
Satisfied	144 (50.5)	192 (50.7)		
Very satisfied	41 (14.4)	66 (17.4)		

*Scored as 0 (*as bad as it can be*) to 10 (*as good as it can be*).

cidal thoughts, fatigue, quality of life, and general health by year in medical school. None of these outcomes differed significantly across class years. Nevertheless, burnout remained common throughout medical school, affecting 38.6% of first-year, 45.2% of second-year, 46.5% of third-year, and 42.9% of fourth-year students. Similar patterns were observed for depression and fatigue, with third-year students reporting the highest prevalence of high fatigue (58.6%).

Overall, 7.6% (*n* = 22) of respondents reported having thoughts of taking their own life during the previous 12 months. By campus, suicidal ideation was reported by 7.5%

(*n* = 12) of respondents from the Kansas City campus, 7.2% (*n* = 8) from the Wichita campus, and 14.3% (*n* = 2) from the Salina campus. Rates by year in medical school were 7.3%, 8.2%, 9.9%, and 5.7% among first-, second-, third-, and fourth-year students, respectively. Although these differences were not statistically significant, third-year students consistently exhibited the numerically highest rates of burnout, depression, suicidal ideation, and fatigue. Despite these findings, students across all years generally rated their quality of life and overall health favorably.

As shown in Table 3a, students with manifestations of burnout were significantly more likely than those without

Table 2b. Burnout, depression, fatigue, quality of life, and general health among 2025 respondents.

Measures	First Year (n = 83 [28.8])	Second Year (n = 62 [21.5])	Third Year (n = 71 [24.7])	Fourth Year (n = 70 [24.3])	t or χ^2 (df)	P value
Burnout index; no (%)^a						
High emotional exhaustion	31 (37.3)	26 (41.9)	30 (42.3)	27 (38.6)	0.55 (3)	0.9080
High depersonalization	8 (9.6)	13 (21.0)	15 (21.1)	10 (14.3)	5.16 (3)	0.1610
Burnout ^b	32 (38.6)	28 (45.2)	33 (46.5)	30 (42.9)	1.14 (3)	0.7680
Depression, no (%)						
Screened positive for depression	27 (32.5)	27 (45.0)	33 (47.1)	26 (37.1)	4.283 (3)	0.2320
Suicidal Thoughts in the Past 12 Months	6 (7.3)	5 (8.2)	7 (9.9)	4 (5.7)	0.886 (3)	0.8290
Fatigue						
Mean (standard deviation)(SD) ^c	5.7 (1.9)	5.4 (2.2)	5.0 (2.2)	5.8 (2.3)	1.684 (3)	0.1710
Level of fatigue in the past week; no (%)						
Low	44 (53.0)	29 (46.8)	29 (41.4)	35 (50.0)	2.195 (3)	0.5330
High	39 (47.0)	33 (53.2)	41 (58.6)	35 (50.0)		
Quality of Life						
Mean (SD)	4.2 (0.6)	4.1 (0.7)	4.0 (0.7)	4.2 (0.6)	1.095 (3)	0.3520
Overall quality of life, no (%)						
Poor	2 (2.4)	1 (1.6)	3 (4.2)	1 (1.5)		
Neither poor nor good	5 (6.0)	11 (18.0)	9 (12.7)	5 (7.4)	8.973 (9)	0.4400
Good	52 (62.7)	31 (50.8)	43 (60.6)	42 (61.8)		
Very good	24 (28.9)	18 (29.5)	16 (22.5)	20 (29.4)		
General Health						
Mean (SD)	3.5 (1.0)	3.5 (1.0)	3.5 (1.0)	3.9 (0.9)	2.402 (3)	0.0680
General health satisfaction, no (%)						
Dissatisfied	17 (21.0)	13 (21.0)	18 (25.7)	8 (11.4)		
Neither dissatisfied nor satisfied	17 (21.0)	13 (21.0)	7 (10.0)	7 (10.0)	12.790 (9)	0.1720
Satisfied	38 (46.9)	28 (45.2)	34 (48.6)	42 (60.0)		
Very satisfied	9 (11.1)	8 (12.9)	11 (15.7)	13 (18.6)		

^aWe assessed burnout using the single-item measures for emotional exhaustion and depersonalization adapted from the full Maslach Burnout Inventory.

^bHigh score on Emotional Exhaustion and/or Depersonalization subscale of the Maslach Burnout Inventory.

^cScores ranged from 0 to 10, low mean fatigue score suggests higher/worsened fatigue.

Note: Missing medical student level; n (%): 2 (0.7).

Table 3a. Relationship of depression, fatigue, and suicidal thoughts with burnout among 2025 responders.

Measures	Burnout			χ^2 (df)	P value	Phi
	Present	Not present	Total			
Depression, no (%)				41.32 (1)	<0.001	0.38
Screened positive	76 (61.8)	39 (24.1)	115 (40.4)			
Screened negative	47 (38.2)	123 (75.9)	170 (59.6)			
Total	123 (100)	162 (100)	285 (100)			
Level of fatigue in the past week, no (%)				26.67 (1)	<0.001	0.31
High degree	87 (69.6)	63 (38.9)	150 (52.3)			
Low degree	38 (30.4)	99 (61.1)	137 (47.7)			
Total	125 (100)	162 (100)	287 (100)			
Suicidal thoughts in the last 12 months, no (%)				14.47 (1)	<0.001	0.23
Yes	18 (14.6)	4 (2.5)	22 (7.7)			
No	105 (85.4)	159 (97.5)	264 (92.3)			
Total	123 (100.0)	163 (100.0)	286 (100.0)			

burnout to screen positive for depression (61.8% vs. 24.1%; $\chi^2[1] = 41.32$, $p < 0.001$), report high fatigue during the previous week (69.6% vs. 38.9%; $\chi^2[1] = 26.67$, $p < 0.001$), and report suicidal thoughts during the previous 12 months (14.6% vs. 2.5%; $\chi^2[1] = 14.47$, $p < 0.001$).

Multivariable logistic regression analyses are presented in Table 3b. Covariates for each model were selected based on those used in the 2019 study.⁴ No demographic or training-related factors were significantly associated with burnout. Given other variables in the model, year in medical school and gender were the only significant predictors. Compared with first-year students, third-year students had more than twice the odds of screening positive for depression (adjusted odds ratio [aOR], 2.13; 95% confidence interval [CI], 1.04-4.36). Conversely, male students had approximately 44% lower odds of screening positive for depression than female students (aOR, 0.56; 95% CI, 0.32-0.94). Similar results were shown for males when modeling fatigue: male students had approximately half the odds of reporting high fatigue compared with female students (aOR, 0.57; 95% CI, 0.32-0.89). Age and campus location were not significantly associated with burnout, depression, or fatigue in any of the adjusted models.

Figure 1 illustrates the relationship between year in medical school and burnout, emotional exhaustion, and depersonalization. Although differences by class year were not statistically significant, a consistent curvilinear pattern was observed across several outcomes. Rates of burnout, depression, and fatigue were lowest among first-year students, peaked during the third year, and declined during the fourth year, whereas ratings of quality of life and gen-

eral health remained relatively stable throughout medical school. Suicidal ideation was lowest in the fourth year.

Open-ended Results

A total of 127 respondents provided open-ended feedback on resources and activities they believed would promote wellness among medical students. Thematic analysis identified 10 major themes, which are presented in Table 4 in descending order based on the frequency with which each theme was identified in respondents' comments. Across these themes, students consistently emphasized that burnout stems less from a lack of wellness resources than from insufficient time, flexibility, and autonomy. The most frequently identified recommendations included protected time off, more flexible absence policies, improved access to mental health services, reduced administrative burden, and practical support to enable students to prioritize their well-being.

DISCUSSION

This study provides an updated assessment of burnout and well-being among medical students at KUSM six years after the institution's previous wellness survey. Although the overall prevalence of burnout remained largely unchanged between 2019 and 2025 (48.0% vs. 43.4%), several important findings emerged. First, depersonalization declined significantly, suggesting modest improvement in one dimension of burnout. Second, fatigue worsened despite relatively stable rates of burnout and depression. Third, students experiencing burnout were substantially more likely to report

Table 3b. Multivariate analyses of factors independently associated with burnout, depression, and fatigue among 2025 respondents.

Dependent variables	Independent variables	n	OR _{unadj}	OR _{adj}	95% CI	P value
Burnout (n = 274) c-index = 0.587 AIC = 236.103 Concordant pairs = 57.8% Discordant pairs = 40.5% Mean VIF = 1.092	Year in training					
	Fourth year	67	1.31	1.62	(0.74, 3.69)	0.1971
	Third year	70	1.37	1.66	(0.80, 3.84)	0.1585
	Second year	61	1.29	1.43	(0.68, 3.11)	0.3170
	First year	76	1.0 (Reference)			-
	Age (for each additional year older)	274	0.97	0.95	(0.53, 1.18)	0.1638
	Male vs. Female	111/163	0.84	0.85	(0.51, 1.44)	0.5162
	Campus location					
	Salina	13	0.621	.494	(0.14, 1.69)	0.2604
	Wichita	107	0.707	.722	(0.43, 1.21)	0.2127
	Kansas City	154	1.0 (Reference)			-
Screened positive for depression (n = 272) c-index = 0.606 AIC = 213.486 Concordant pairs = 58.7% Discordant pairs = 37.5% Mean VIF = 1.092	Year in training					
	Fourth year	67	1.37	1.35	(0.64, 2.87)	0.4327
	Third year	69	1.99	2.13	(1.04, 4.36)	0.0386
	Second year	60	1.77	1.78	(0.86, 3.65)	0.1185
	First year	76	1.0 (Reference)			-
	Age (for each additional year older)	272	1.01	1.00	(0.93, 1.07)	0.9580
	Male vs. Female	110/162	0.58	0.56	(0.32, 0.94)	0.0267
	Campus location					
	Salina	13	1.04	0.886	(0.268, 2.93)	0.8423
	Wichita	106	0.857	0.857	(0.28, 3.29)	0.5618
	Kansas City	153	1.0 (Reference)			-
High level of fatigue in the past week (n = 273) c-index = 0.604 AIC = 256.288 Concordant pairs = 59.9% Discordant pairs = 39.0% Mean VIF = 1.092	Year in training					
	Fourth year	67	1.09	1.09	(0.53, 2.23)	0.8244
	Third year	69	1.54	1.67	(0.83, 3.36)	0.1478
	Second year	61	1.16	1.18	(0.59, 2.37)	0.6355
	First year	76	1.0 (Reference)			-
	Age (for each additional year older)	273	0.99	0.99	(0.92, 1.06)	0.7842
	Male vs. Female	111/162	0.59	0.57	(0.32, 0.89)	0.0180
	Campus location					
	Salina	13	0.64	0.50	(0.15, 1.61)	0.2429
	Wichita	106	0.88	0.79	(0.47, 1.32)	0.3631
	Kansas City	154	1.0 (Reference)			-

Note: Regression formula for predicting the odds of Burnout:

$\hat{y} = 0.381 + 0.483(\text{Fourth year}) + 0.505(\text{Third year}) + 0.360(\text{Second year}) - 0.055(\text{Age}) - 0.165(\text{Male}) + 0.705(\text{Kansas City}) + 0.379(\text{Wichita})$.

Example: For a Third-year student, who is 26 years of age, female, from Kansas City, the formula would be:

$\hat{y} = 0.381 + 0.483(0) + 0.505(1) + 0.360(0) - 0.055(26) - 0.165(0) + 0.705(1) + 0.379(0)$.

$\hat{y} = 0.161$; where the odds of burnout would be $e^{0.161}$ or 1.17.

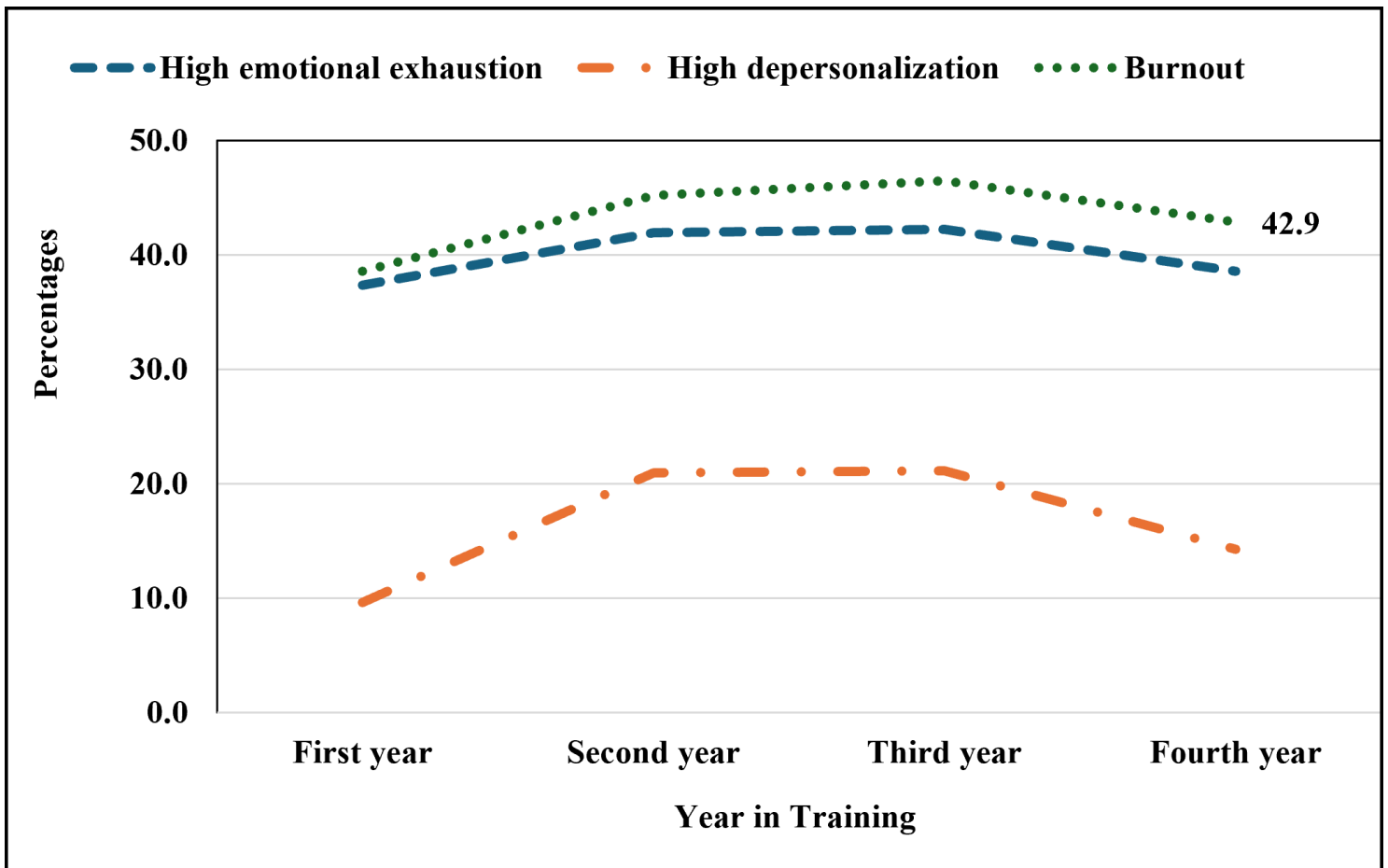


Figure 1. Manifestations of burnout, emotional exhaustion scores, and depersonalization scores by year in training among 2025 respondents.

depressive symptoms, fatigue, and suicidal ideation. Finally, the qualitative findings demonstrated that students perceive burnout to be driven less by a lack of wellness resources than by limited time, flexibility, and autonomy. Together, these findings suggest that institutional approaches focused primarily on wellness programming may be insufficient without concurrent efforts to address structural and contextual contributors to student well-being.

The prevalence of burnout observed in this study is consistent with reports from medical schools nationally and internationally. A systematic review by Frajerma, et al.²⁸ estimated that approximately 44% of medical students experience burnout, while other meta-analyses have reported prevalence estimates ranging from 37% to 52%, depending on the instruments used and study populations.²⁹ Our findings therefore suggest that burnout remains a persistent challenge despite increasing institutional attention to learner well-being. Although depersonalization declined

significantly compared with the 2019 survey, the persistence of emotional exhaustion and overall burnout indicates that meaningful improvements in student well-being remain difficult to achieve.

One notable finding was the significant increase in fatigue despite relatively stable rates of burnout and depression. Fatigue has increasingly been recognized as an important indicator of learner well-being that may precede or accompany burnout. The worsening fatigue observed in this study may reflect the cumulative effects of excessive workload, chronic sleep deprivation, and competing academic and clinical responsibilities, all of which have been associated with fatigue, burnout, and psychological distress among medical students and residents.³⁰⁻³³ These demands often intensify during the transition to clinical training, when students must simultaneously balance patient care responsibilities, licensing examination preparation, frequent evaluations, and uncertainty regarding residency ap-

Table 4. Themes and representative suggestions for reducing medical student burnout and promoting well-being from 2025 respondents.

Emerging themes	Key issues identified	Representative suggestions to promote wellness
Protected time and schedule flexibility	Lack of time for self-care, family responsibilities, appointments, and recovery; rigid attendance and absence policies	Wellness days, personal days, mental health days, designated non-clinical days during clerkships, increased vacation time, flexible absence policies, predictable schedules, advance notice of schedules.
Mental health services and counseling	Difficulty accessing services during clinical training; need for ongoing awareness and reduced stigma	Expanded counseling availability (evenings/weekends), periodic reminders of available resources, group counseling, increased support for mental health concerns, normalization of seeking counseling, improved access to psychiatric care.
Workload and curriculum burden	Excessive assignments, mandatory events, redundant content, and administrative tasks contributing to stress	Reduce nonessential assignments and modules, streamline curricular content, decrease mandatory attendance requirements, improve clarity regarding academic expectations and remediation processes.
Physical activity and healthy lifestyle promotion	Limited time and motivation for exercise and self-care	Free or discounted gym memberships, group exercise classes, yoga and meditation sessions, organized recreational activities, wellness programs incorporating physical movement.
Social support and community building	Feelings of isolation and need for stronger peer and mentor connections	Small-group activities, peer support opportunities, mentoring programs, social events, class gatherings, opportunities to engage with family and friends.
Financial support and resource accessibility	Financial burden from tuition, educational resources, examinations, and daily living expenses	Assistance with STEP examination fees, study resources (e.g., UWORLD, Amboss), grocery gift cards, food pantries, professional clothing assistance, scholarships, financial counseling.
Support during clinical training	Burnout associated with clerkship demands, limited autonomy, and emotionally challenging clinical experiences	Wellness days during clerkships, flexible leave policies, structured debriefing sessions, improved communication regarding schedules and evaluations, support for emotionally difficult patient encounters.
Student-centered wellness programming	Perception that mandatory wellness activities are less helpful than practical support	Therapy dog visits, massage chairs, free food, recreational events, outdoor activities, wellness activities integrated into existing curriculum time rather than added requirements.
Support for diverse student needs	Challenges faced by student parents, nontraditional students, students with disabilities, and students undergoing campus transitions	Family-friendly policies, support for dependent care responsibilities, accommodation for mental health conditions, resources for non-traditional students, assistance during campus transitions.
Institutional culture and communication	Desire for greater transparency, responsiveness, and demonstration of institutional support	Clear communication regarding academic expectations and residency preparation, reduced administrative messaging burden, increased transparency in decision-making, visible commitment to student well-being.

plications.^{9,30,32} Our qualitative findings support this interpretation, as students overwhelmingly attributed burnout to insufficient time, limited flexibility, and competing academic demands rather than a lack of wellness resources.

Although differences by class year were not statistically significant, a consistent pattern emerged in which burnout, depression, fatigue, and suicidal ideation peaked during the third year of medical school before improving during the fourth year. Similar findings have been reported in previous studies demonstrating that the transition into clinical clerkships represents one of the most stressful periods of undergraduate medical education.^{30,34} Third-year students must adapt to new clinical environments, frequent schedule changes, professional evaluations, and uncertainty regarding residency competitiveness, while also preparing for licensing examinations. Our multivariable analysis further supports this observation, with third-year students having more than twice the odds of screening positive for depression compared with first-year students. These findings suggest that wellness and mental health interventions may be particularly valuable during the transition into clinical training.

The finding that male students had lower odds of reporting high fatigue than female students also warrants consideration. Although gender remained significantly associated with fatigue in the adjusted model, the survey did not assess several personal and family factors that may contribute to differences in fatigue, including marital or partnership status, parenting, caregiving responsibilities, and other competing demands outside of medical school. These factors may represent important unmeasured confounders or contextual influences on student well-being. Importantly, the open-ended responses highlighted the challenges students face in balancing medical education with family and personal responsibilities, suggesting that well-being cannot be fully understood by examining academic demands in isolation. These competing priorities may become increasingly relevant as students progress through medical school and their personal circumstances evolve. In particular, the third year may represent a period when increasing academic and clinical demands coincide with changes in students' personal and family responsibilities. Students may marry, divorce, become parents, assume caregiving responsibilities, or take on other personal commitments that affect available time, stress, and overall well-being. Future studies should

therefore collect more comprehensive information about students' personal and family circumstances to better understand how these factors interact with medical education demands, fatigue, and burnout. Such information also would help clarify whether observed gender differences in fatigue reflect differences in the educational experience, differences in competing responsibilities, or a combination of these factors.

Consistent with previous literature, students experiencing burnout were significantly more likely to report depressive symptoms, fatigue, and suicidal ideation. These findings reinforce the well-established relationship between burnout and adverse mental health outcomes among medical students. Dyrbye and colleagues have repeatedly demonstrated that burnout is associated with depression, suicidal ideation, decreased empathy, and reduced professionalism among medical trainees.^{30,35,36} Although this study cannot establish causality because of its cross-sectional design, the strong associations observed underscore the importance of routinely assessing multiple dimensions of psychological distress rather than burnout alone.

Perhaps the most important contribution of this study comes from the qualitative findings. Students reported that the institution already offers numerous wellness resources, including counseling services, coaching, peer support, and wellness activities, but emphasized that they often lack sufficient time, flexibility, and autonomy to use them. Rather than requesting additional wellness lectures or programming, students consistently advocated for protected time off, flexible absence policies, fewer administrative requirements, predictable schedules, and reduced curricular burden. These findings align with growing evidence that organizational and educational factors contribute substantially to burnout and that interventions focused solely on individual resilience may be insufficient.^{37,38} The open-ended responses further suggest that the competing demands of medical education and students' personal lives should be considered when developing institutional approaches to well-being.

The qualitative findings also can be interpreted through the lens of Self-Determination Theory (SDT), which proposes that psychological well-being depends on satisfaction of three fundamental psychological needs: autonomy, competence, and relatedness.^{39,40} Students repeatedly described barriers that directly reflect frustration of these

needs. Requests for greater schedule flexibility, protected wellness time, and fewer mandatory activities reflect a desire for greater autonomy over learning and personal life. Calls for clearer expectations regarding residency preparation, remediation, and academic requirements suggest a need to strengthen students' sense of competence. Meanwhile, recommendations emphasizing mentoring, small-group learning communities, peer support, counseling, and opportunities to spend time with family and friends illustrate the importance of relatedness. The latter is particularly relevant given students' descriptions of competing family and personal responsibilities, which may influence their ability to maintain meaningful relationships and sources of support. According to SDT, environments that support these three needs foster intrinsic motivation, psychological well-being, and resilience, whereas environments that undermine them may contribute to emotional exhaustion, disengagement, and burnout.^{38,40}

Our findings therefore support a shift from primarily individual-focused wellness initiatives toward more system-level interventions. National organizations, including the National Academy of Medicine and the Association of American Medical Colleges, have increasingly emphasized that burnout largely is an organizational phenomenon requiring organizational solutions.^{41,42} Counseling services and wellness activities remain important components of student support, but these resources are unlikely to fully address burnout if students continue to experience excessive workload, inflexible scheduling, competing demands, and limited autonomy. Structural interventions such as protected wellness time, streamlined curricula, transparent communication, flexible leave policies, predictable scheduling, and enhanced support during clinical transitions may better address the underlying contributors identified by students. In addition, future institutional wellness assessments should consider students' personal and family circumstances so that interventions can account for the broader context in which students manage the demands of medical education.

Interestingly, despite relatively high rates of burnout, depression, fatigue, and suicidal ideation, most respondents continued to report good quality of life and satisfaction with their overall health. Similar findings have been observed elsewhere and may reflect the multidimensional nature of well-being.⁴³ Students may perceive their overall quality

of life positively because of supportive peer relationships, strong professional identity formation, meaningful patient interactions, or optimism regarding future careers despite experiencing substantial psychological distress.⁴⁴ This apparent paradox further emphasizes that a single global measure of quality of life may not adequately capture the complexity of learner well-being. A comprehensive assessment should instead consider multiple dimensions, including burnout, fatigue, depressive symptoms, social and family demands, and students' perceptions of autonomy, support, and overall quality of life.

Strengths and Limitations

This study has several strengths. It represents one of the few repeated institutional surveys directly comparing burnout and wellness outcomes over time using similar methodology. The inclusion of quantitative and qualitative data provides a more comprehensive understanding of student well-being than either approach alone. Additionally, incorporating suicidal ideation into the 2025 survey offers important contemporary data regarding mental health among medical students.

Several limitations should also be acknowledged. The cross-sectional design precludes causal inference, and self-reported measures may be subject to recall and social desirability biases. Although the response rate was comparable with many contemporary surveys of medical students, nonresponse bias remains possible. The study was conducted at a single medical school, which may limit generalizability to other institutions. The relatively small number of respondents from the Salina campus limited statistical power for campus-specific comparisons. Finally, the timing of data collection may have influenced the prevalence estimates. Data were collected between October 15 and December 31, 2025, during a period of varying academic and clinical demands. Because levels of burnout, depression, fatigue, and suicidal ideation may fluctuate throughout the academic year, the findings may not fully represent students' experiences at other time points. Longitudinal studies are needed to better characterize temporal changes in medical student well-being.

CONCLUSIONS

Burnout remains highly prevalent among KUSM medical

students and continues to be strongly associated with depression, fatigue, and suicidal ideation. Although prevalence of depersonalization was lower than observed in 2019, overall burnout remained common, and fatigue worsened. Third-year students consistently exhibited the numerically highest burden, highlighting an important opportunity for targeted intervention during the transition to clinical training. Importantly, students identified insufficient time, flexibility, and autonomy, not a lack of wellness resources, as the primary drivers of burnout. Viewed through the framework of SDT, these findings suggest that institutional efforts should move beyond expanding wellness programming toward creating learning environments that better support students' autonomy, competence, and relatedness. Such organizational approaches may provide a more sustainable strategy for improving medical student well-being.

REFERENCES

1. Santiago IS, de Castro E, Castro S, de Brito APA, et al. Stress and exhaustion among medical students: a prospective longitudinal study on the impact of the assessment period on medical education. *BMC Med Educ*. 2024 Jun 6;24(1):630. PMID: 38844948.
2. Ilić IM, Ilić MD. The relationship between the burnout syndrome and academic success of medical students: a cross-sectional study. *Arh Hig Rada Toksikol*. 2023 Jun 26;74(2):134-141. PMID: 37357875.
3. Kilic R, Nasello JA, Melchior V, Triffaux JM. Academic burnout among medical students: respective importance of risk and protective factors. *Public Health*. 2021 Sep;198:187-195. Epub 2021 Sep 1. PMID: 34478954.
4. Ofei-Dodoo S, Moser SE, Kellerman R, Wipperman J, Paolo A. Burnout and Other Types of Emotional Distress Among Medical Students. *Med Sci Educ*. 2019 Sep 16;29(4):1061-1069. PMID: 34457584.
5. Almutairi H, Alsubaiei A, Abduljawad S, et al. Prevalence of burnout in medical students: A systematic review and meta-analysis. *Int J Soc Psychiatry*. 2022 Sep;68(6):1157-1170. Epub 2022 Jul 1. PMID: 35775726.
6. Vidhukumar K, Hamza M. Prevalence and Correlates of Burnout among Undergraduate Medical Students - A Cross-sectional Survey. *Indian J Psychol Med*. 2020 Mar 9;42(2):122-127. PMID: 32346252.
7. Maslach C, Schaufeli WB, Leiter MP. Job burnout. *Annu Rev Psychol*. 2001;52:397-422. PMID: 11148311.
8. Dyrbye LN, Thomas MR, Massie FS, et al. Burnout and suicidal ideation among U.S. medical students. *Ann Intern Med*. 2008 Sep 2;149(5):334-41. PMID: 18765703.
9. Dyrbye LN, West CP, Satele D, et al. Burnout among U.S. medical students, residents, and early career physicians relative to the general U.S. population. *Acad Med*. 2014 Mar;89(3):443-51. PMID: 24448053.
10. Mohmand S, Monteiro S, Solomonian L. How are Medical Institutions Supporting the Well-being of Undergraduate Students? A Scoping Review. *Med Educ Online*. 2022 Dec;27(1):2133986. PMID: 36268575.
11. Shin P, Desai V, Conte AH, Qiu C. Time Out: The Impact of Physician Burnout on Patient Care Quality and Safety in Perioperative Medicine. *Perm J*. 2023 Jun 15;27(2):160-168. Epub 2023 Jun 6. Erratum in: *Perm J*. 2023 Sep 15;27(3):116. PMID: 37278062.
12. Cotobal Rodeles S, Martín Sánchez FJ, Martínez-Sellés M. Physician and Medical Student Burnout, a Narrative Literature Review: Challenges, Strategies, and a Call to Action. *J Clin Med*. 2025 Mar 26;14(7):2263. PMID: 40217713.
13. Liaison Committee on Medical Education. Functions and structure of a medical school: standards for accreditation of medical education programs leading to the MD degree. 2023. <https://lcme.org/standards-and-rules/>. Accessed June 16, 2026.
14. Kim HY. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restor Dent Endod*. 2017 May;42(2):152-155. Epub 2017 Mar 30. PMID: 28503482.
15. The University of Kansas School of Medicine. Campuses. <https://www.kumc.edu/school-of-medicine/campuses.html>. Accessed June 18, 2026.
16. West CP, Dyrbye LN, Sloan JA, Shanafelt TD. Single item measures of emotional exhaustion and depersonalization are useful for assessing burnout in medical professionals. *J Gen Intern Med*. 2009 Dec;24(12):1318-21. Epub 2009 Oct 3. PMID: 19802645.
17. Levis B, Sun Y, He C, et al. Accuracy of the PHQ-2 Alone and in Combination With the PHQ-9 for Screening to Detect Major Depression: Systematic Review and Meta-analysis. *JAMA*. 2020 Jun 9;323(22):2290-2300. PMID: 32515813.
18. Manea L, Gilbody S, Hewitt C, et al. Identifying de-

- pression with the PHQ-2: A diagnostic meta-analysis. *J Affect Disord.* 2016 Oct;203:382-395. Epub 2016 Jun 6. PMID: 27371907.
19. Millner AJ, Lee MD, Nock MK. Single-Item Measurement of Suicidal Behaviors: Validity and Consequences of Misclassification. *PLoS One.* 2015 Oct 23;10(10):e0141606. PMID: 26496707.
 20. West CP, Tan AD, Habermann TM, Sloan JA, Shanafelt TD. Association of resident fatigue and distress with perceived medical errors. *JAMA.* 2009 Sep 23;302(12):1294-300. PMID: 19773564.
 21. World Health Organization (1996) WHOQOL-BREF. introduction, administration, scoring and generic version of the assessment. field trial version. Program on mental health. WHO, Geneva. http://www.who.int/mental_health/media/en/76.pdf. Accessed June 26, 2026.
 22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009 Apr;42(2):377-81 Epub 2008 Sep 30. PMID: 18929686.
 23. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform.* 2019 Jul;95:103208. Epub 2019 May 9. PMID: 31078660.
 24. Creswell JW. Mapping the developing landscape of mixed methods research. In: Tashakkori A, Teddlie C. (Eds.) *Sage Handbook of Mixed Methods in Social & Behavioral Research*. 2nd Edition. Thousand Oaks, CA: Sage, 2010, pp. 45-68. ISBN: 9781412972666.
 25. Cuschieri S. The STROBE guidelines. *Saudi J Anaesth.* 2019 ;13(Suppl 1):S31-S34. PMID: 30930717.
 26. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *J Med Internet Res.* 2004 Sep 29;6(3):e34. Erratum in: doi:10.2196/jmir.2042. PMID: 15471760.
 27. Greenbook. What Every Researcher Should Know About Statistical Significance. <https://www.greenbook.org/marketing-research/statistical-significance-03377>. Accessed June 26, 2026.
 28. Frajerman A, Morvan Y, Krebs MO, Gorwood P, Chauvette B. Burnout in medical students before residency: A systematic review and meta-analysis. *Eur Psychiatry.* 2019 Jan;55:36-42. Epub 2018 Oct 29. PMID: 30384110.
 29. Rotenstein LS, Ramos MA, Torre M, et al. Prevalence of Depression, Depressive Symptoms, and Suicidal Ideation Among Medical Students: A Systematic Review and Meta-Analysis. *JAMA.* 2016 Dec 6;316(21):2214-2236. PMID: 27923088.
 30. Dyrbye LN, Thomas MR, Shanafelt TD. Medical student distress: causes, consequences, and proposed solutions. *Mayo Clin Proc.* 2005 Dec;80(12):1613-22. PMID: 16342655.
 31. Wolf MR, Rosenstock JB. Inadequate Sleep and Exercise Associated with Burnout and Depression Among Medical Students. *Acad Psychiatry.* 2017 Apr;41(2):174-179. Epub 2016 Mar 14. PMID: 26976402.
 32. Ishak W, Nikraves R, Lederer S, Perry R, Ogunyemi D, Bernstein C. Burnout in medical students: a systematic review. *Clin Teach.* 2013 Aug;10(4):242-5. PMID: 23834570.
 33. Hill MR, Goicochea S, Merlo LJ. In their own words: stressors facing medical students in the millennial generation. *Med Educ Online.* 2018 Dec;23(1):1530558. PMID: 30286698.
 34. Radcliffe C, Lester H. Perceived stress during undergraduate medical training: a qualitative study. *Med Educ.* 2003 Jan;37(1):32-8. PMID: 12535113.
 35. Dyrbye LN, Thomas MR, Power DV, et al. Burnout and serious thoughts of dropping out of medical school: a multi-institutional study. *Acad Med.* 2010 Jan;85(1):94-102. PMID: 20042833.
 36. West CP, Dyrbye LN, Shanafelt TD. Physician burnout: contributors, consequences and solutions. *J Intern Med.* 2018 Jun;283(6):516-529. Epub 2018 Mar 24. PMID: 29505159.
 37. National Academies of Sciences, Engineering, and Medicine; National Academy of Medicine; Committee on Systems Approaches to Improve Patient Care by Supporting Clinician Well-Being. *Taking Action Against Clinician Burnout: A Systems Approach to Professional Well-Being*. Washington (DC): National Academies Press (US); 2019 Oct 23. PMID: 31940160.
 38. Deci EL, Ryan RM. Self-determination theory: a macrotheory of human motivation, development, and health. *Can Psychol.* 2008;49(3):182-185. doi:10.1037/a0012801.
 39. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development,

- and well-being. *Am Psychol.* 2000 Jan;55(1):68-78. PMID: 11392867.
40. Kusurkar RA, Croiset G, Ten Cate TJ. Twelve tips to stimulate intrinsic motivation in students through autonomy-supportive classroom teaching derived from self-determination theory. *Med Teach.* 2011;33(12):978-82. PMID: 22225435.
41. National Academy of Medicine. National plan for health workforce well-being. Washington (DC): National Academy of Medicine; 2022. <https://nam.edu/publications/national-plan-for-health-workforce-well-being/>. Accessed July 15, 2026.
42. Association of American Medical Colleges. Commitment to clinician well-being and resilience [Internet]. Washington (DC): Association of American Medical Colleges; 2018. <https://www.aamc.org/media/23696/download>. Accessed July 15, 2026.
43. Tempiski P, Santos IS, Mayer FB, et al. Relationship among Medical Student Resilience, Educational Environment and Quality of Life. *PLoS One.* 2015 Jun 29;10(6):e0131535. PMID: 26121357.
44. Ofei-Dodoo S, Irwin G, Wright B, et al. Burnout, Depression, Anxiety, and Stress Among Resident Physicians 18 Months Into the COVID-19 Pandemic: A Cross-Sectional Study. *Kans J Med.* 2022 Nov 28;15(3):403-411. PMID: 36467446.

Keywords: *professional burnout; medical students; depression; fatigue; mental health*

Training and Assessment of Non-Technical Human Factor Skills Among Anesthesiology Residents: An Exploratory Educational Quality Improvement Initiative

Sukruta Pradhan, M.D.

William Krogman, M.S.

Dakota Rapp, M.D.

Sawyer Schmidt, B.S.

James Walker, M.D.

Brian Mahoney, D.O.

Blake Gorman, D.O.

Department of Anesthesiology, The University of Kansas School of Medicine-Wichita, Wichita, Kansas



Received June 25, 2026; Accepted for publication October 1, 2026; Published online October 16, 2026, Kans J Med 2026 Sept-Oct; 19:148-159. <https://doi.org/10.17161/kjm.vol19.25684>.

Conflict of Interest: None.

Corresponding Author:

William Krogman, M.S.

Department of Anesthesiology,
The University of Kansas School of
Medicine-Wichita
929 N St Francis St., Room 8081,
Wichita, KS 67214
wkrogman@kumc.edu

ABSTRACT

Introduction. Safe anesthesia requires technical and non-technical skills to prevent adverse events. Non-technical skills training is not routinely incorporated into formal anesthesia education in the United States. This quality educational improvement initiative aimed to provide clinical anesthesia (CA) residents with non-technical skills training.

Methods. Three anesthesiologists completed virtual Anaesthetists' Non-Technical Skills (ANTS) training in May 2023. An anesthesiologist developed two high-stress simulation scenarios. Fourteen residents were randomized into dyads for the first simulation/post-simulation survey, followed by a two-part didactic seminar on human factors based on the ANTS framework and a post-didactic survey. One hundred twenty days later, residents were re-randomized for a second simulation and another post-simulation survey. Two ANTS-trained anesthesiologists reviewed video recordings of both simulations.

Results. Cohen's kappa demonstrated moderate agreement ($\kappa = 0.43$; 95% CI, 0.36-0.49; $p < 0.001$) between evaluators. Wilcoxon signed-rank tests indicated improved non-technical skill use among CA1 residents, while senior residents generally performed better during the first simulation. Fisher's exact tests showed no significant associations in resident knowledge or confidence following training.

Conclusions. Non-technical skills training may benefit residents early in CA training, while senior residents may have different learning needs. Simulation-based training appears feasible; however, formal, recurring, and consistent education may be needed to reinforce these skills throughout residency.

INTRODUCTION

Non-technical skills are important in health care but often lack structured, dedicated training during graduate medical education in the United States. One exception is Stanford University's immersive sim-

ulation-based learning program through the Center for Immersive and Simulation-based Learning, which offers training in holistic technical/non-technical skills through immersive simulation learning.¹ Establishing a structured training regimen that integrates technical and non-technical skills during residency may therefore facilitate the application of human factor principles in clinical practice.

Safe and efficient anesthesia care requires both technical and non-technical skills which work together to support effective task management and prevent adverse events.² Although technical skills routinely are taught and practiced, non-technical skills receive less structured attention.³ One study has demonstrated that non-technical skills can be taught through cognitive aids, repeated scenario-based training, and systematic approaches to patient care, with systematic approaches associated with improved problem-solving skills and confidence in high-stress situations.⁴ Non-technical skills encompass a range of interpersonal, cognitive, and behavioral skills, whereas human factors refers to the effects of communication, teamwork, task management, situational awareness, culture, and organizational factors on human behavior and performance, as well as the application of this knowledge in clinical settings.⁵

Although human factors are addressed throughout anesthesiology residency, they are not always taught through structured, dedicated curricula.⁶ This is important because human factors and non-technical skills substantially contribute to patient safety, particularly during high-stress situations. Studies have reported that 82% (294/359) of incidents resulting from preventable adverse events were associated with human error,⁷ with communication breakdowns between health care professionals at different levels contributing substantially to these errors.⁸ Another study found that deficits in nontechnical skills contributed to 47.7% (2,677/5,610) of reported incidents, with task management identified as the most prevalent contributing factor.⁹ Deficiencies in non-technical skills may therefore increase the risk of errors and adverse events.¹⁰ Structured training in these skills is becoming increasingly common during residency, particularly in European and Asian countries.^{6,11-13}

Despite these findings, relatively few anesthesiology studies have examined the effects of non-technical skills training in clinical settings, particularly in the United States.^{14,15} In contrast, research in other specialties is more

extensive.^{9,13-20} The Royal College of Anaesthetists in the United Kingdom developed the Anaesthetists' Non-Technical Skills (ANTS) framework and associated training and evaluation program.¹² ANTS evaluates four categories of non-technical skills: Situation Awareness, Team Working, Task Management, and Decision Making across 15 behavioral elements. Each element is rated from poor (1) to good (4), providing a structured assessment of non-technical skill performance (Table 1). Research has demonstrated that ANTS is a reliable and valid tool for evaluating anesthesiologists' non-technical skills in simulated and clinical settings.^{21,22} Its versatility also allows for implementation across different study designs and educational interventions.²³

Table 1: Summary of the elements included in the Anaesthetists' Non-Technical Skills protocol and evaluation.

Skill Elements	
Task Management	Planning and preparing
	Prioritizing
	Providing and maintaining standards
	Identifying and utilizing resources
Team Working	Coordinating activities with team
	Exchanging information
	Using authority and assertiveness
	Assessing capabilities
Situation Awareness	Supporting others
	Gathering information
	Recognizing and understanding
	Anticipating
Decision Making	Identifying options
	Balancing risk and selecting options
	Re-evaluating

Note: There are 15 skill elements categorized into four main groups.

The aim of this exploratory educational quality improvement initiative was to educate anesthesiology residents about human factors and evaluate their use of non-technical skills during one-off simulated high-stress scenarios. Using the ANTS framework, we assessed residents' non-technical skills during an initial simulation, followed by two didactic sessions on human factors, and then a second simulation. We hypothesized that residents would demonstrate higher ANTS scores during the second simulation after receiving human factors education and targeted training in non-technical skills.

METHODS

Quality Initiative Design and Protocol

Three anesthesiology faculty members virtually attended the ANTS training course in May 2023 to learn about human factors and standardize future evaluations. Following the training, one faculty member independently developed two high-stress simulations for use at a simulation center (Supplemental S1: Simulation Scenarios; only available online at journals.ku.edu/kjm). Residents in their clinical anesthesia (CA) years (CA1 = 4, CA2 = 5, CA3 = 5; n = 14) were randomly assigned to dyads and participated in a baseline simulation with less than 24 hours' notice in September 2023 (Supplemental ST1: Group Combinations). Randomization was used because resident pairings vary during routine clinical duties and may involve different levels of experience.

Each dyad was debriefed immediately after the simulation by the faculty member who developed the scenario. Debriefing was conducted at the faculty member's discretion without standardized questions but followed the PAUSE format (Provide summary, Analyze initial reactions, Understand the sentiment, Silence for introspection, and Education).²⁴ Within 1 month of the baseline simulation, the same faculty member presented a two-part, 3-hour didactic seminar on non-technical skills based on the ANTS training course. Each session lasted approximately 1.5 hours and was delivered 3 weeks apart because of faculty scheduling. Pre- and post-didactic surveys assessed residents' perspectives on non-technical skills training (Supplemental S2: Didactic Survey).

Approximately 120 days after the baseline simulation, residents were re-randomized into dyads and participated in a second, distinct simulation with less than 24 hours' notice in January 2024 (Supplemental ST1: Group Combinations). No dyad was repeated between simulations. The 120-day interval was selected based on several factors: (a) evidence that long-term skill retention is assessed beyond 3 months,²⁵ (b) availability of the University of Kansas School of Medicine-Wichita Simulation Center, (c) allowing residents time to practice skills in the clinical setting following the didactic intervention, (d) faculty availability, and (e) facilitating program planning for resident absences from clinical duties. Residents were immediately debriefed after the second simulation using the same approach as the

baseline simulation. Post-simulation surveys were collected after both simulations to assess the usefulness of the scenarios (Supplemental S3: Simulation Survey).

Simulation Development

Two distinct simulations were developed in conjunction with the Simulation Center to evaluate non-technical skills after training. New simulations were developed to prevent residents from anticipating the scenarios, ensure that cases reflected anesthesiology-specific situations, and minimize familiarity with previously used simulations. The Simulation Center facilitated the simulations and provided real-time changes in vital signs, manikins capable of simulating breath sounds and heart rates, and video recording for subsequent evaluation. An experienced operating room nurse was available as a resource during each simulation.

The simulated operating room was arranged to reflect the layout familiar to residents, except that an anesthesia machine was unavailable at the Simulation Center and therefore was not included. Before entering the simulation room, residents received a case description indicating that the patient was prepped and draped and that the surgeon was scrubbing in but was not yet in the room. Residents were given several minutes to familiarize themselves with the room and equipment before the scenario began (Supplemental S1: Simulation Scenarios). The baseline simulation involved anaphylaxis, whereas the follow-up simulation involved local anesthetic systemic toxicity. Each simulation lasted approximately 40 minutes, including debriefing.

Following both simulations, video recordings from three camera angles were obtained from the Simulation Center. Videos were edited to remove extraneous downtime, such as room setup, and resident debriefings. Edited recordings were provided to the two remaining ANTS-trained anesthesiology faculty members for evaluation using the ANTS framework.¹² Evaluators were blinded to the simulation sequence and were unaware of which recording represented the baseline or follow-up simulation. However, they were aware of group composition because they knew the residents and could identify the dyads during video review. They had no prior knowledge of the order in which the simulations were conducted. Resident debriefings were excluded from the recordings provided to evaluators to minimize potential assessment bias. Evaluations were conducted from February through March 2024.

Statistical Analysis

Quadratic weighted Cohen's kappa was calculated to assess inter-rater reliability between evaluators. Baseline and follow-up assessments were compared using the Wilcoxon signed-rank test for individual residents' ANTS element scores. Residents were analyzed individually rather than as dyads. Statistical significance was set at $p < 0.05$.

Given the small sample size ($n = 14$) and the estimated effect size (Cohen's $d = 0.09$), the study had limited statistical power and a corresponding high probability of Type II error (94.5%). Cohen's d was used to quantify the magnitude of the observed effect independent of sample size. Statistical power and the corresponding probability of Type II error were estimated using a power analysis.

Survey results were used to provide additional information regarding residents' knowledge and confidence in nontechnical skills. The surveys were developed specifically for this study and were not formally validated; however, three ANTS-trained faculty members reviewed the surveys for clarity and content relevance.

Fisher's exact tests and Cramer's V were used to evaluate associations among survey responses. Fisher's exact test was used to assess associations involving the Likert-type response categories, which ranged from Very Low to Very High. Cramer's V was used to estimate the effect size for each Fisher's exact test.

Thematic analysis was used to evaluate responses to the simulation surveys. Three authors independently identified keywords and concepts within the responses and subsequently discussed and categorized them into themes using predefined characteristics. All statistical analyses were performed using R[®] statistical software (version 4.1.1; R Foundation for Statistical Computing, Vienna, Austria),²⁶ and figures were created using the R package ggplot2 version 3.4.0.²⁷

IRB Statement

This study was reviewed and approved as a quality improvement initiative by the Ascension Health Institutional Review Board (RKS20230030). Written informed consent was obtained from participants for publication of the findings. The manuscript adheres to the Standards for Quality Improvement Reporting Excellence (*SQUIRE* 2.0) guidelines for quality improvement reporting.²⁸

RESULTS

Assessment of Non-Technical Skill Use in Simulations

Upon completion of the resident evaluations, quadratic weighted Cohen's kappa was used to assess inter-rater reliability between the two evaluators. The analysis demonstrated moderate agreement ($\kappa = 0.43$; 95% CI, 0.36-0.49; $p < 0.001$). A Wilcoxon signed-rank test was then used to compare ANTS scores between the baseline and follow-up simulations for each individual resident within each evaluator (Table 2).

According to Evaluator 1, 75% of CA1 residents ($n = 3$) demonstrated a significant improvement in the use of non-technical skills between the baseline and follow-up simulations. Among CA2 and CA3 residents, 40% ($n = 4$) demonstrated a significant decline in non-technical skill use, while the remaining 60% ($n = 6$) showed no significant difference between simulations.

According to Evaluator 2, 100% of CA1 residents ($n = 4$) demonstrated a significant improvement in non-technical skill use between simulations. Among CA2 and CA3 residents, 10% ($n = 1$) demonstrated a significant improvement, 50% ($n = 5$) demonstrated a significant decline, and 40% ($n = 4$) showed no significant difference between simulations (Table 2).

Simulation and Didactic Survey Evaluation

Residents completed anonymous surveys following each simulation (two surveys per resident, one for each simulation) and the didactic seminar series. Qualitative analysis of the post-simulation survey identified six themes that residents recommended be retained in future simulations and six themes for improvement (Table 3; $n = 23$; 82.1% response rate). The most frequently identified theme for retention was the uncommon nature of the cases, with residents noting that simulation provided an opportunity to practice rare clinical scenarios without consequences to patients. The most frequently identified area for improvement was the design of scenarios to facilitate one-on-one training.

The post-didactic survey assessed residents' perceptions of the seminar's usefulness in developing nontechnical skills (Figure 1) and their confidence across nontechnical skill areas (Figure 2; $n = 13$; 92.9% response rate). Overall, residents were receptive to the seminar, with 84.6% ($n = 11$) reporting confidence in their nontechnical skills, although

Table 2: Wilcoxon Signed-Rank tests between baseline and follow-up simulations for each resident. ANTS scores were averaged between both evaluators for each element.

Baseline	CA1-1	CA1-2	CA1-3	CA1-4	CA2-1	CA2-2	CA2-3	CA2-4	CA2-5	CA3-1	CA3-2	CA3-3	CA3-4	CA3-5
Task Management	3	2	3	2	3	3	3	3	3	3	4	3	3	4
	2	3	3	2.5	4	3	3	3.5	3.5	3	3.5	4	3	3.5
	2.5	2.5	3	2.5	4	3	3.5	3.5	3.5	3	3.5	4	3	3.5
	2.5	3	3	2.5	3	2.5	3	3.5	3.5	2.5	3.5	3.5	2.5	3
Team Working	3	2.5	3	2.5	4	3	3	3.5	3.5	3	2.5	3.5	3	2.5
	2.5	3	3.5	2	3.5	2.5	3	3.5	3.5	2.5	3	3.5	3	3.5
	2.5	2.5	2.5	2	4	2.5	3.5	3.5	4	2	3.5	4	3	3
	2.5	2	2.5	2.5	3.5	3	3	4	4	3	3.5	3	3	3
Situation Awareness	3	3	3	3	3.5	3	3	3.5	3.5	3	4	3.5	3.5	4
	2.5	2.5	3	2.5	3.5	3	3	3.5	3.5	3	3.5	4	3.5	3
	1.5	2	3	1.5	4	3	2.5	4	4	3	3.5	4	3	3.5
	1.5	2	3	1.5	4	3	3	4	4	3	3.5	4	3	3.5
Decision Making	2.5	2	3	2	4	2.5	2.5	4	4	3	3	3.5	3	3
	2	2.5	2.5	2	4	3	2.5	4	4	3	3.5	4	3	3.5
	2.5	2.5	3	2	4	3	3	4	4	3	4	4	3.5	4
Follow-up	CA1-1	CA1-2	CA1-3	CA1-4	CA2-1	CA2-2	CA2-3	CA2-4	CA2-5	CA3-1	CA3-2	CA3-3	CA3-4	CA3-5
Task Management	4	3	3	2	3	3	3	2	4	3	3	3	3	4
	3	3	3	2.5	3	3.5	3	3	3.5	2.5	2.5	4	2.5	3
	3.5	2.5	3	2.5	3	3.5	3	3	3.5	3	3	3.5	3	3.5
	3.5	3	3.5	2.5	3	3	3	2.5	3.5	2.5	3.5	3.5	2	3.5
Team Working	4	3.5	4	2.5	3.5	3	3	3	4	2.5	3	4	2.5	3.5
	4	3.5	3.5	2.5	3.5	3.5	3	3.5	3.5	2.5	3	4	2	3.5
	3.5	3.5	3.5	2.5	3.5	3	2.5	3.5	4	2	2.5	3.5	2	3.5
	3	3	3.5	2	2.5	3.5	2.5	2	3.5	2.5	2.5	3.5	2.5	3
Situation Awareness	4	3.5	4	3.5	3.5	3.5	3	3.5	3	3	3	3.5	3	3.5
	3.5	3	3.5	3	3.5	3	2.5	3.5	3.5	2.5	2.5	3.5	2	3.5
	2.5	3.5	3.5	2.5	3	3.5	2	3	3.5	2	2	3.5	2	2.5
	3	3	3.5	2.5	2.5	3	2	2.5	3	2	2	3.5	2	3
Decision Making	3.5	3.5	3.5	2.5	3.5	3.5	2	2.5	3.5	2	2	4	2	3.5
	3.5	3	4	2.5	3	3.5	2	2.5	3.5	2.5	2	4	2.5	3.5
	3.5	3.5	3.5	3	3.5	3.5	2.5	3	3.5	2.5	2.5	3.5	2.5	3.5
p-value	<0.001	<0.001	<0.001	0.006	<0.001	<0.001	0.001	<0.001	0.207	0.001	<0.001	0.546	<0.001	1
Direction	▲	▲	▲	▲	▼	▲	▼	▼	—	▼	▼	—	▼	—

Note: ANTS elements are rated on a scale of 1 (poor) to 4 (good). ▲ indicates a significant difference in evaluation between simulations, where the resident performed better during the follow-up simulation. ▼ indicates a significant difference in evaluation between simulations, where the resident performed worse during the follow-up simulation. No ▲ or ▼ indicates no significant differences in performance between simulations. ANTS = Anaesthetists' non-technical skills; CA = clinical anesthesia year.

Table 3: Thematic analysis of the post-simulation surveys.

Question	Theme	Responses	Definition of Theme
Positive points in the simulation session that should be repeated in the future?	Scenario	18	A good balance of uncommon but potential case that could and should be experienced during residency.
	Critical Thinking	2	Fast pace and high stress situation, coupled with unknown scenario, resulted in deeper levels of thinking.
	Peer Partner	8	Approaching stressful situations with a peer enhanced the experience.
	Nursing Staff	4	Having an experienced OR nurse participate in the simulation made the experience more realistic and immersible.
	Simulation Environment	9	The setup of the simulation (draped/prepared manikin, connected IV, available syringes, etc.) made the situation more believable and realistic while allowing a safe environment to make mistakes.
	Debrief	6	Immediate debriefing by the attending following the simulation allowed for highlighting of strengths and deficiencies.
Some areas for improvement for future sessions?	Lab Familiarity	4	Let residents familiarize themselves with the simulation lab before the beginning of the scenario.
	Real-time Feedback	4	As the monitors can be delayed when making changes, provide real-time changes to numbers as the case progresses, if residents are moving too quickly.
	Equipment Availability	4	Have more equipment available for the simulations (i.e., ultrasound, ECHO).
	Simulations	6	Have more simulations: one-on-one, more complex, and/or cases pertaining to the resident's current level of training.
	Handoffs	4	Present the case details in the form of a handoff to make the simulation more realistic.
	Chaos	3	Introduce more chaos to the simulation to make it more high stress (i.e., have music on in the background, provide more ambiguity to the case details) to help highlight non-technical skills a bit better.

Note: The specific response categories to each question are displayed as themes, with the number of similar responses falling into that theme. The theme definitions provide a further explanation of the responses.

OR = operating room; IV = intravenous; ECHO = echocardiogram.

fewer residents expressed interest in additional didactic training (Figure 1). Fisher's exact test showed no statistically significant association between the seminar and residents' knowledge of nontechnical skills ($p = 0.743$). Cramer's V was 0.337.

Regarding resident confidence, responses indicated some uncertainty across each of the four major nontechnical skill categories (Figure 2). Fisher's exact test showed no statistically significant association between the seminar and residents' confidence in their nontechnical skills ($p = 0.949$). Cramer's V was 0.223.

DISCUSSION

The authors of this educational quality improvement initiative hypothesized that residents would have higher ANTS scores during the follow-up simulation after receiving didactic education. The observed improvement in non-tech-

nical skill behaviors was evident among residents early in their clinical anesthesia training but was not observed among CA2 and CA3 residents (Table 2). However, given the small sample size and high likelihood of Type II error, these findings should be interpreted cautiously.

The lack of observed improvement among CA2 and CA3 residents may have been influenced by opportunities to incorporate non-technical skills into their developing clinical practice. Residents were expected to actively practice non-technical skills between simulations. Although senior residents may have had more opportunities to incorporate these skills into clinical practice, they may also have already developed established approaches to managing high-acuity situations. In contrast, junior residents generally are more closely supervised by attending physicians, providing greater opportunities for faculty instruction, even when faculty have not received formal training in non-technical skills. Additionally, the improved scores among CA1 res-

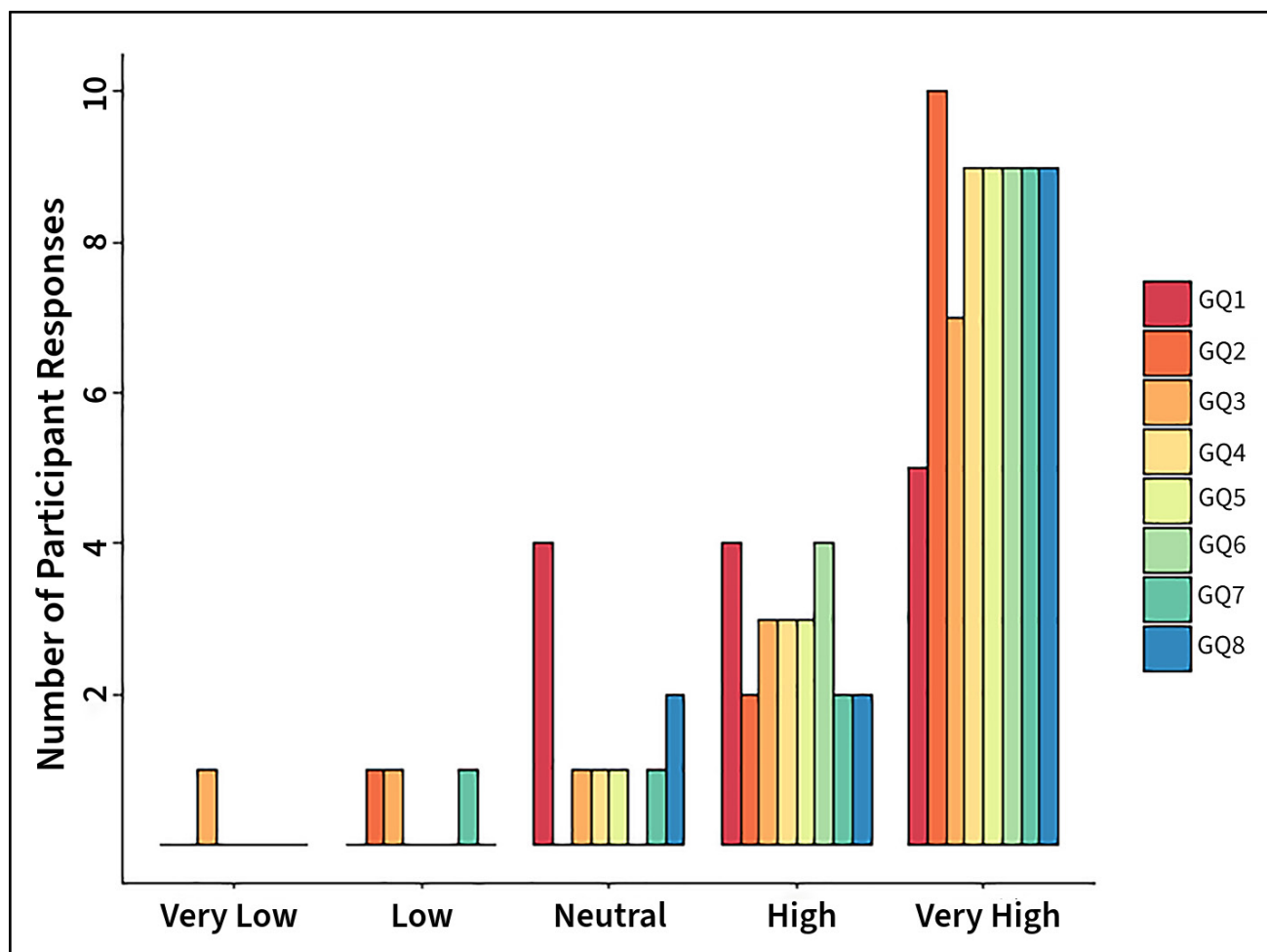


Figure 1: Survey responses from the residents about the general helpfulness of the didactic seminar in the development of non-technical skills.

Notes: GQ1: You were aware that non-technical skills could influence patient outcomes before the didactic presentation. GQ2: Did the didactic presentations further your understanding of non-technical skills? GQ3: Would you like more didactic presentations about non-technical skills in the future? GQ4: Are you likely to practice your non-technical skills more than before the didactic presentation? GQ5: You will practice your non-technical skills often. GQ6: How likely are you to consciously incorporate non-technical skills into your practice? GQ7: New skills/strategies were gained that you can apply to your practice. GQ8: You feel more confident in handling high-risk, high-stress situations because of your non-technical skill trainings. GQ = general questions.

idents may have been partly attributable to greater confidence during the follow-up simulation and/or an additional 120 days of clinical anesthesia training compared with the baseline simulation. There also is evidence that junior residents may be more receptive to introspection and reflection, whereas senior residents may favor more consistent cause-and-effect reasoning.²⁹ When considered alongside the post-didactic survey findings, which demonstrated a moderate association between the seminar and resident knowledge (Figure 1), and weak-to-moderate association between seminar and resident confidence (Figure 2), these results suggest that senior residents may require more structured and longitudinal non-technical skills education.

However, definitive conclusions cannot be drawn because longitudinal changes in non-technical skills during clinical practice were not assessed.

Previous studies have demonstrated that non-technical skills can be successfully taught⁴ and that training may lead to improvement over time.^{6,11-13} Simulation also has been identified as an effective approach for training non-technical skills.^{4,16,30} Despite evidence supporting effective training methods, evaluation and assessment of these skills have been limited, particularly within trainee programs.^{31,32} However, this area is receiving increasing attention.³³ Our findings support the feasibility of using simulation to train non-technical skills and suggest that developing knowledge

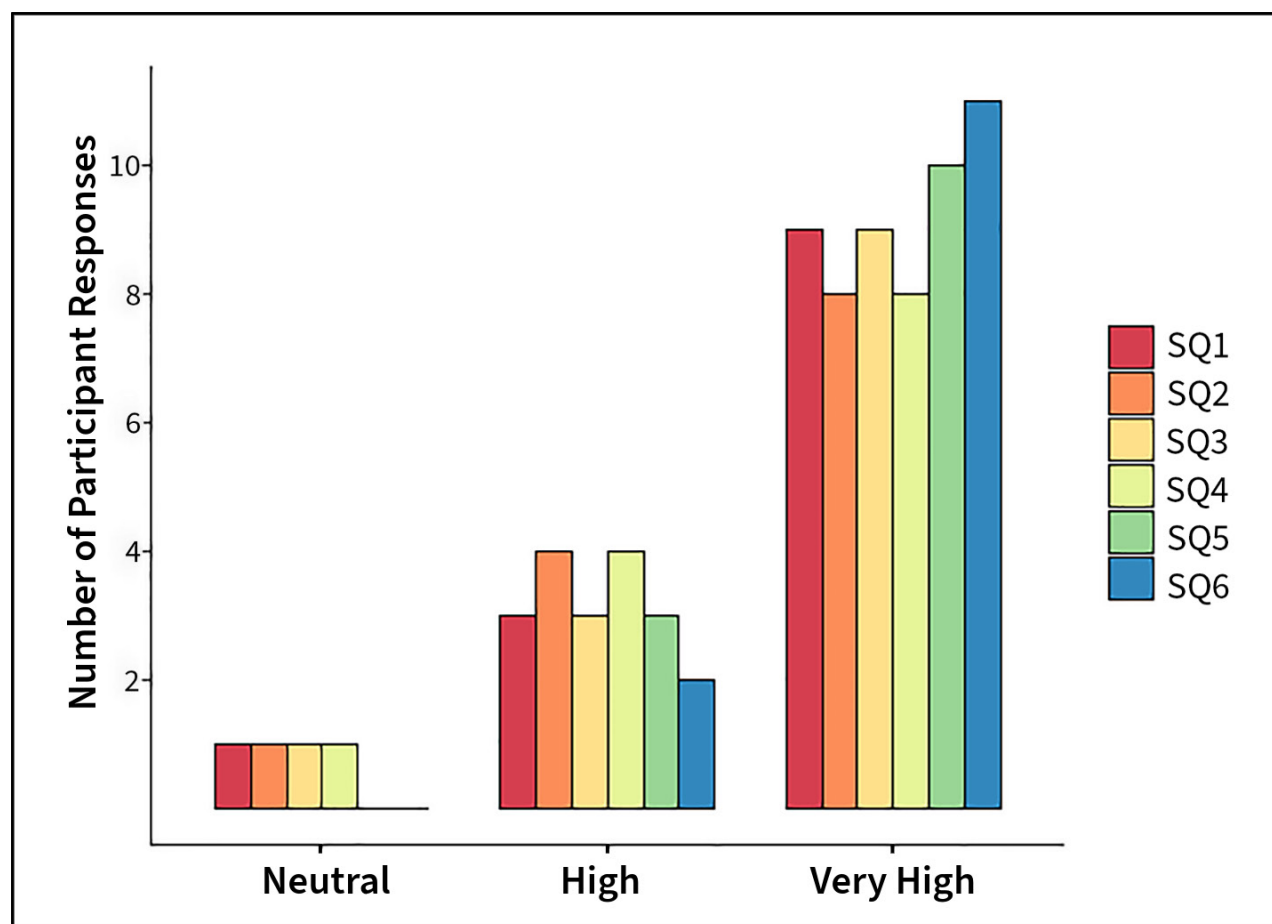


Figure 2: Survey responses from the residents about their confidence in the different non-technical skill areas.

Notes: SQ1: You are confident in your situational awareness skills (gathering information, recognizing and understanding, anticipating) after the didactic presentation. SQ2: You are confident in your decision-making skills (identifying options, balancing risk and selecting options, re-evaluating) after the didactic presentation. SQ3: You are confident in your teamwork skills (coordinating activities, exchanging information, using authority and assertiveness, assigning capabilities, supporting others) after the didactic presentation. SQ4: You are confident in your task management skills (planning and preparing, prioritizing, providing and maintaining standards, identifying and utilizing resources) after the didactic presentation. SQ5: The coping strategies discussed throughout the presentation will enhance your clinical skills and expertise, leading to less mistakes and anticipation of problems. SQ6: You will practice and use the coping strategies in your practice. SQ = skill-based questions.

and awareness of these skills may represent an important initial step toward behavioral skill development. More consistent and longitudinal education may be needed to reinforce these skills. One-time educational experiences may not provide sufficient opportunities for residents to consolidate skill development,^{17,18} which is consistent with findings from this exploratory initiative. Structured curricular training also has been associated with significant improvements in non-technical skills.¹⁹ One study demonstrated significant improvements in three ANTS categories: team working, situation awareness, and decision making; and five ANTS elements following simulation-based team training.³⁴ Another study examined the relationship between

team training delivered through two simulations and team culture and communication using paired-sample *t* tests.²⁰ Participants demonstrated movement away from autonomous cultural ideals, stronger teamwork, and a more favorable communication climate.²⁰ These findings are particularly relevant given the documented contribution of human error, including non-technical skill deficits and communication failures, to adverse events.

In contrast, a meta-analysis by Leuschner, et al.¹⁵ demonstrated substantial variation in non-technical skill levels following training, with no significant improvement among individual surgeons or anesthesiologists but improvement when assessed at the whole-theatre team level. These find-

ings differ from those of a more recent meta-analysis, which found that non-technical skills training was effective among emergency and critical care physicians, although substantial heterogeneity was present.³⁵ Together, these findings suggest that the effectiveness of non-technical skills training may depend on the learners, training approach, clinical context, and level at which skills are evaluated.

Limitations

This exploratory educational quality improvement initiative had several limitations. First, using surveys to assess resident perceptions of the simulations and didactic seminar may have introduced response bias. For the post-didactic assessment, a pre- and post-test of non-technical skills knowledge may have provided a more direct assessment of residents' understanding and knowledge acquisition. However, repeated testing may introduce recall or testing effects. To assess knowledge retention while minimizing the influence of immediate recall, future evaluations could incorporate a delayed retention assessment, such as repeating the assessment at least 90 days after the seminar.²⁵

Second, there was a 7-month interval between the faculty ANTS training and the follow-up simulation. This delay resulted from logistical factors, including the availability of the annual online ANTS course, which has since been expanded, finalization of the simulation scenarios, coordination with the Simulation Center, preparation of the didactic seminars, and scheduling time for all residents to be released from clinical duties. Although the evaluators completed a brief ANTS refresher training before conducting the evaluations, the extended interval may have contributed to the modest inter-rater reliability observed in this study. In addition, the faculty received only one formal training experience in non-technical skills through the ANTS course. This level of training may have been insufficient to ensure consistent evaluation, as suggested by the inter-rater reliability findings. This limitation is particularly relevant given the limited availability of formal human factors expertise within U.S. residency programs.¹⁵

Third, we did not evaluate the clinical impact of non-technical skills training on patient outcomes. Previous research has suggested that improvements in non-technical skills may be associated with reductions in adverse clinical events, including cardiac arrests.¹⁸ Although residents were expected to practice non-technical skills during their clinical

duties following the didactic intervention, adherence to this expectation was not assessed.

Fourth, simulations were conducted with dyads rather than individual residents, preventing assessment of individual performance in isolation. This approach may have influenced skill performance based on residents' learning preferences^{36,37} and level of training. Future studies could evaluate individual and team-based performance separately to better understand these effects. Additionally, more CA-1 residents were paired with senior residents during the follow-up simulation, which may have impacted both residents' ANTS assessments, as the senior resident could have taken on more of a teaching role, compared with the baseline simulation (Supplemental ST1: Group Combinations). Although residents performed the simulation as a dyad, the evaluators were evaluating on an individual basis, to try and mitigate this type of bias.

Fifth, the evaluators could have possibly diagnosed the simulation scenarios due to monitor readings and resident conversations throughout the recordings. However, even if the evaluators deduced the case (i.e., that one was anaphylaxis), they still would not know if that was the first or second simulation.

Finally, separating technical and non-technical skills during evaluation may have been challenging because evaluators were familiar with the participating residents, potentially introducing observer or confirmation bias. The ANTS framework was intended to reduce this potential bias by providing a structured evaluation platform based on 15 defined skill elements.¹² Nevertheless, future studies could further minimize this limitation by using independent evaluators who are blinded to resident identity and training level when feasible.

Overall, the small sample size, single-institution design, use of resident dyads, and limited duration of formal non-technical skills education restrict the generalizability of these findings. However, the methodology could be adapted to other specialties to evaluate the feasibility of developing competency-based curricula for non-technical skills.

One important barrier to the regular incorporation of non-technical skills training is the limited number of faculty formally trained to teach and evaluate these skills. Developing faculty expertise may increase recognition of human factors in the operating room and provide residents with

greater exposure to appropriate behaviors, coaching, and feedback. Similarly, formal and structured resident education can establish a foundation for continued development and application of non-technical skills throughout residency and into independent practice.

As a next step, we aim to expand this experience into a competency-based medical education curriculum by integrating ANTS training into the existing simulation curriculum for the anesthesiology residency. Residents will participate in repeated simulation scenarios with deliberate practice, longitudinal observation by ANTS-trained faculty, and structured faculty evaluation and feedback. This curriculum, combined with didactic instruction in non-technical skills, may provide continuous and structured education throughout residency training.

CONCLUSIONS

This exploratory educational quality improvement initiative evaluated the feasibility and resident acceptability of an ANTS-based approach to non-technical skills education and explored changes in observed non-technical skill performance following a multifaceted educational intervention. The findings support the feasibility of incorporating structured simulation and didactic education into anesthesiology training. Further research with larger samples, longitudinal assessment, trained evaluators, and repeated educational exposure is needed to determine how non-technical skills can be effectively developed, assessed, and sustained throughout residency training.

REFERENCES

- Center for Immersive and Simulation-based Learning. Leaders in Simulation-based Medical Education. 2026. <https://cisl.stanford.edu/>. Accessed Sept 15, 2026.
- Radhakrishnan B, Katikar MD, Myatra SN, Gautam PL, Vinayagam S, Saroa R. Importance of non-technical skills in anaesthesia education. *Indian J Anaesth*. 2022 Feb; 66:64-69. PMID: 35309030.
- Bromiley M. The husband's story: from tragedy to learning and action. *BMJ Qual Saf*. 2015 Jul; 24:425-427. Epub 2015 May 13. PMID: 25972221.
- Mossenson AI, Tuyishime E, Rawson D, et al. Promoting anaesthesia providers' non-technical skills through the Vital Anaesthesia Simulation Training (VAST) course in a low-resource setting. *Br J Anaesth*. 2020 Feb; 124:206-213. Epub 2019 Dec 12. PMID: 31839255.
- Reason J. Human error: models and management. *BMJ*. 2000 Mar; 320(7237):768-70. PMID: 10720363.
- Matos FM, Martins MR, Martins I. Non-technical skills progression during anesthesiology residency in Portugal: the impact of a National Pedagogical Plan. *Med Educ Online*. 2020 Dec; 25:1800980. PMID: 32815792.
- Cooper JB, Newbower RS, Long CD, McPeck B. Preventable anesthesia mishaps: a study of human factors. *Anesthesiology*. 1978 Dec; 49(6):399-406. PMID: 727541.
- Francis T, Davidson M, Senese L, et al. Exploring the use of social network analysis methods in process improvement within healthcare organizations: a scoping review. *BMC Health Serv Res*. 2024 Sep 5; 24(1):1030. PMID: 39237937.
- Reader T, Flin R, Lauche K, Cuthbertson BH. Non-technical skills in the intensive care unit. *Br J Anaesth*. 2006 May; 96(5):551-559. Epub 2006 Mar 27. PMID: 16567346.
- Flin R, Patey R. Non-technical skills for anaesthetists: developing and applying ANTS. *Best Pract Res Clin Anaesthesiol*. 2011 Jun; 25(2):215-227. PMID: 21550546.
- Oofuvong M, Pattaravit N, Kanjanawanichkul O, Siripruekpong S, Nuanjun K, Suwannarat B. Are technical skills assessed using medical knowledge associated with non-technical skill knowledge in anesthesia resident training? *Siriraj Med J*. 2022; 74:844-856. DOI: 10.33192/Smj.2022.99.
- Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. *Br J Anaesth*. 2003 May; 90(5):580-588. PMID: 12697584.
- O'Keeffe DA, Losty M, Traynor O, Doherty EM. Objective assessment of surgical trainees' non-technical skills: Improved performance following a two-year program of instruction. *Am J Surg*. 2020 Dec; 220:1566-1571. Epub 2020 May 8. PMID: 32444063.
- Oeppen RS, Davidson M, Scrimgeour DS, Rahimi S, Brennen PA. Human factor awareness and recognition during multidisciplinary team meetings. *J Oral Pathol Med*. 2019 Sep; 48(8):656-661. Epub 2019 Apr 9. PMID: 30908725.
- Leuschner S, Leuschner M, Kropf S, Niederbichler AD.

- Non-technical skills training in the operating theater: A meta-analysis of patient outcomes. *Surgeon*. 2019 Aug; 17(4):233-243. Epub 2018 Aug 7. PMID: 30093229.
16. Magill JC, Tolley N. Non-technical skills simulation training. *Curr Otorhinolaryngol Rep*. 2020; 8:106-110. DOI: 1007/s40136-020-00269-8.
17. Griffin C, Aydin A, Brunckhorst O, et al. Non-technical skills: a review of training and evaluation in urology. *World J Urol*. 2020 Jul; 38(7):1653-1661. Epub 2019 Sep 17. PMID: 31529246.
18. Coggins A, Desai M, Nguyen K, Moore N. Early acquisition of non-technical skills using a blended approach to simulation-based medical education. *Adv Simul (Lond)*. 2017 Aug; 2:12. PMID: 29450013.
19. Dedy NJ, Bonrath EM, Ahmed N, Grantcharov TP. Structured training to improve nontechnical performance of junior surgical residents in the operating room: A randomized controlled trial. *Ann Surg*. 2016 Jan; 263(1):43-49. PMID: 25775073.
20. Kirschbaum KA, Rask JP, Brennan M, Phelan S, Fortner SA. Improved climate, culture, and communication through multidisciplinary training and instruction. *Am J Obstet Gynecol*. 2012 Sep; 207(3):200.e1-7. Epub 2012 Jun 20. PMID: 22840971.
21. Boet S, Larrigan S, Martin L, Liu H, Sullivan KJ, Etherington C. Measuring non-technical skills of anaesthesiologists in the operating room: a systematic review of assessment tools and their measurement properties. *Br J Anaesth*. 2018 Dec; 121(6):1218-1226. Epub 2018 Sep 6. PMID: 30442248.
22. Zwaan L, Tjon Soei Len L, Wagner C, van Groenigen D, Kolenbrander M, Krage R. The reliability and usability of the Anesthesiologists' Non-Technical Skills (ANTS) system in simulation research. *Adv Simul (Lond)*. 2016 Jun; 1:18. PMID: 29449987.
23. Kang J, Hu J, Yan C, Xing X, Tu S, Zhou F. Development and applications of the Anaesthetists' Non-Technical Skills behavioural marker system: a systematic review. *BMJ Open*. 2024 Mar; 14(3):e075019. Epub 2024 Mar 20. PMID: 38508635.
24. Clapper TC, Leighton K. Incorporating the Reflective Pause in Simulation: A Practical Guide. *J Contin Educ Nurs*. 2020 Jan 1; 51(1):32-38. PMID: 31895468.
25. Kovacs E, Jenei ZM, Csordas K, et al. The timing of testing influences skill retention after basic life support training: a prospective quasi-experimental study. *BMC Med Educ*. Dec 2019; 19(1): 452. PMID: 31801502.
26. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>. Accessed September 5, 2021.
27. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. Cham, Switzerland: Springer International Publishing. 2016.
28. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for Quality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf*. 2015 Dec; 25(12):986-992. Epub 2015 Sep 14. PMID: 26369893.
29. Choi H, Oh J, Kim CK, Ryu H, Ryu Y. Residents need competence not confidence: A retrospective evaluation of the new competency education program for Korean neurology residents. *PLoS One*. Oct 2023; 18(10): e0290503. PMID: 37796843.
30. Kelly FE, Frerk C, Bailey CR, et al. Human factors in anaesthesia: a narrative review. *Anaesthesia*. 2023 Apr; 78:479-490. Epub 2023 Jan 11. PMID: 36630729.
31. Gaba DM, DeAnda A. The response of anesthesia trainees to simulated critical incidents. *Anesth Analg*. 1989; 68(4):444-451. PMID: 2929977.
32. Fletcher G, McGeorge P, Flin RH, Glavin RJ, Maran NJ. The role of non-technical skills in anaesthesia: a review of current literature. *Br J Anaesth*. 2002 Mar; 88(3):418-429. PMID: 11990277.
33. Moll-Khosrawi P, Kamphausen A, Hampe W, Schulte-Uentrop L, Zimmermann S, Kubitz JC. Anaesthesiology Students' Non-Technical Skills: Development and evaluation of a behavioural marker system for students (AS-NTS). *BMC Med Educ*. 2019 Jun; 19:205. Epub 2019 Jun 13. PMID: 31196070.
34. Finstad AS, Bjørshol CA, Aase I, Røislien J, Ballangrud R. Assessment of Anaesthesia Teams' Non-Technical Skills in Clinical Practice before and after Simulation-Based Team Training: A Quasiexperimental Study. *Anesthesiol Res Pract*. 2024 Jul; 2024:2021671. PMID: 39006532.
35. Sanchez-Marco M, Escribano S, Rubio-Aparicio M, Julia-Sanchis R, Cabanero-Martinez MJ. Effectiveness of nontechnical skills educational intervention in the

context of emergencies: a systematic review and meta-analysis. *Australian Crit Care*. 2023 Nov; 36(6):1159-1171. Epub 2023 Feb 28. PMID: 36858860.

36. Lou Y, Abrami PC, d'Apollonia S, et al. Small group and individual learning with technology: a meta-analysis. *Rev Educ Res*. 2001; 71(3):449-521. DOI: 10.3102/00346543071003449.
37. Kirschner F, Paas F, Kirschner PA. Individual and group-based learning from complex cognitive tasks: Effects on retention and transfer efficiency. *Comput Human Behav*. 2009; 25(2):306-314. DOI: 10.1016/j.chb.2008.12.008.

Keywords: *simulation training; clinical competence; quality improvement; medical residency; coping skills*

Clinical Workflow Associated with Spanish-Speaking Patient Encounters at JayDoc Community Clinic-Wichita: A Quality Improvement Project

Leah C. Duncan, MS-3

Wyl Halimeh, M.D., M.S.

Dania Jumpa, M.D.

Carlos Schwindt, MS-4

Evie Schmitz, MS-3

Dulcinea Rakestraw, Ph.D., MPH

The University of Kansas School of
Medicine-Wichita, Wichita, Kansas

Received June 22, 2026; Accepted for
publication October 15, 2026; Pub-
lished online October 16, 2026, Kans J
Med 2026 Sept-Oct; 19:160-166. <https://doi.org/10.17161/kjm.vol19.25667>.

Acknowledgements: The authors
would like to thank Hayrettin Okut,
Ph.D. for his review and advice on the
data analysis.

Conflict of Interest: None.

Corresponding Author:

Leah C. Duncan, MS-3

The University of Kansas School of
Medicine-Wichita

1010 N. Kansas St., Wichita, KS 67214

lduncan@kumc.edu

ABSTRACT

Introduction. JayDoc Community Clinic-Wichita is a student-run free clinic serving a predominantly Spanish-speaking patient population. Authors of this study evaluated whether interpreter staffing represents a potentially modifiable systems-level factor associated with clinic workflow.

Methods. This retrospective observational quality improvement evaluation included 715 patient encounters nested within 128 clinic sessions at JayDoc from November 2023 to July 2025. Encounter-level workflow differences by Spanish-speaking status were evaluated using linear mixed-effects models accounting for clustering within clinic sessions and adjusting for clinic day. Clinic-session-level multivariable linear regression evaluated associations among Spanish-speaking patient volume, in-person interpreter staffing, and clinic workflow.

Results. After accounting for clustering and clinic day, Spanish-speaking encounters were 5.3 minutes longer than non-Spanish-speaking encounters, although the difference was not statistically significant ($p = 0.065$). Spanish-speaking encounters had significantly longer medical preparation time (adjusted mean difference, 2.57 minutes; 95% CI, 0.83 to 4.31; $p = 0.004$) and wait-to-present time (adjusted mean difference, 3.46 minutes; 95% CI, 0.47 to 6.45; $p = 0.024$). At the clinic-session level, each additional Spanish-speaking patient was associated with a 6.7-minute increase in clinic-session duration after controlling for total patient volume and interpreter staffing ($p < 0.001$). Each 1-unit increase in the Spanish-speaking patient-to-interpreter ratio was associated with a 3.72-minute increase in clinic-session duration ($p = 0.052$).

Conclusions. Spanish-speaking encounters had longer medical preparation and wait-to-present times, while greater Spanish-speaking patient volume was associated with longer clinic-session duration. In-person interpreter staffing was not independently associated with session duration. These findings support further evaluation of strategies to better align interpreter resources with patient needs.

INTRODUCTION

In the United States, more than one in five individuals aged 5 years or older speak a language other than English, and 61.0% of these individuals speak Spanish.¹ In Wichita, Kansas, the Hispanic/Latino population increased from 58,352 in 2010 to 72,786 in 2020,² and 18.4% of city residents report speaking a language other than English at home.³

JayDoc Clinic, a student-run free clinic at the University of Kansas School of Medicine-Wichita, serves this region and a high proportion of Spanish-speaking patients. Language differences can create barriers to effective health care when patients and clinicians do not share a common language, potentially affecting communication, patient understanding, and participation in clinical decision-making. Limited access to trained medical interpreters may further compound these challenges, particularly when clinics rely on family members, students, or clinical staff without formal interpreter training. Ad hoc interpretation may introduce inaccuracies or omissions, limit discussion of sensitive information, and place additional communication demands on patients and clinicians.

Before 2023, access to formal interpreter services at JayDoc was limited, and family members, volunteer physicians, medical students, or nursing staff frequently assisted with interpretation. In 2023, JayDoc implemented a formal interpreter service that increased access to trained interpreters, including multiple in-person interpreters during clinic sessions and telephone interpretation when in-person interpreters were unavailable.

This change provided an opportunity to examine the relationship between interpreter staffing and clinic workflow. Student-run free clinics commonly experience long patient visits and slower clinic flow, which may contribute to patient dissatisfaction and underscore the need to monitor clinic efficiency.⁴ Insufficient interpreter coverage may contribute to downstream workflow delays.⁵

Improved access to in-person interpreters may facilitate more efficient communication and timely care while addressing language-related disparities in health care. The objective of this study was to examine associations among Spanish-speaking patient volume, interpreter staffing, and clinic workflow in a student-run free clinic. Specifically, we evaluated differences in the duration of various phases of patient encounters between Spanish-speaking

and non-Spanish-speaking patients and examined whether Spanish-speaking patient volume and interpreter staffing were associated with overall clinic-session duration.

METHODS

Setting and Context

This retrospective observational quality improvement evaluation involved review of data collected from November 2023 to July 2025 at JayDoc Community Clinic-Wichita. JayDoc is a student-run free clinic that operates twice weekly, on Monday evenings and Saturday mornings, year-round. The clinic primarily is staffed by a multidisciplinary team of volunteer interpreters, medical students, pharmacy students, premedical students, medical residents, pharmacists, and attending physicians. Approximately 70% of patient encounters involve Spanish-speaking patients, making interpreter services an integral component of routine clinic operations.

Staffing models differ by clinic day. Monday clinics are staffed by an Internal Medicine resident, whereas Saturday clinics include a Family Medicine resident and attending physician, in addition to varying numbers of volunteer interpreters and student learners. Typically, six patients are scheduled for Monday clinics and eight patients for Saturday clinics. Each clinic session generally lasts approximately 3 hours. Clinic-session duration was defined as the time elapsed between the start-of-clinic huddle, led by the student director, and the discharge of the final patient. These times were recorded on a data collection form maintained by the JayDoc Board of Directors.

Intervention

Review of clinic tracking-board data by the JayDoc board consistently identified in-person interpreter staffing as a potential operational bottleneck associated with prolonged clinic visits and reduced workflow efficiency. In response, the clinic initiated an ongoing quality improvement effort to improve access to interpreter services and better align interpreter staffing with patient demand.

As part of this initiative, the clinic partnered with a new interpreter service, expanded in-person interpreter staffing during clinic sessions, increased the use of telephone interpretation, and adjusted interpreter staffing over time based on patient demand, workflow observations, and op-

erational needs. These changes were introduced incrementally rather than at a single predefined intervention time point and were refined throughout the study period as part of the clinic's continuous quality improvement process. The intervention was intended to improve communication efficiency, reduce workflow bottlenecks, and better align interpreter resources with the linguistic needs of the clinic population.

Measures and Data Collection

Encounter-level and clinic-session-level data were extracted from the clinic's electronic time-stamped tracking board. Variables included Spanish-speaking status, total patient volume, Spanish-speaking patient volume, number of in-person interpreters available, and workflow intervals.

Spanish-speaking status was recorded at the patient-encounter level based on the clinic's tracking data. In-person interpreter staffing was defined as the number of in-person interpreters available during each clinic session. Telephone interpretation was available when in-person interpreters were unavailable; however, telephone interpreter use was not consistently captured in the tracking dataset and therefore could not be quantified. Accordingly, the interpreter staffing measure reflects in-person staffing rather than total interpreter capacity.

The tracking board routinely records timestamps for each phase of the patient encounter, including wait-to-be-roomed time, premedical student intake time, preparation time for medicine and pharmacy, pharmacy student encounter, medical student encounter, wait time to presentation to the supervising attending, attending encounter, discharge time, and overall clinic-session duration. These workflow intervals represent the standard sequence of care at JayDoc and were selected because they could reasonably be influenced by interpreter availability and communication-related workflow delays.

Encounter-level outcomes included total patient visit duration and individual workflow intervals. At the clinic-session level, overall clinic-session duration was the primary outcome used to evaluate associations with patient volume, Spanish-speaking patient volume, and in-person interpreter staffing.

Workflow data routinely were collected during each clinic session. As volunteers advanced patients through each stage of care, they entered workflow milestones into the

tracking board, which automatically recorded timestamps in Microsoft® Excel® (Microsoft Corp., Redmond, WA) and calculated workflow intervals and summary metrics. Formal validation of timestamp accuracy was not performed; however, clinic leadership emphasized contemporaneous data entry and monitored data completeness during clinic sessions to maximize data accuracy. Some degree of human error, including delayed or missed entries, was considered unavoidable.

Before analysis, clinic sessions were excluded if they had substantial missing workflow documentation (approximately >30% incomplete tracking-board data), missing clinic-session duration, or operational characteristics that were not representative of routine clinic workflow. The latter included sessions staffed primarily by board members rather than the usual complement of student volunteers, residents, and attending physicians. In total, 27 clinic sessions were excluded.

Analysis

Analyses were conducted at two levels: patient encounter and clinic session. Encounter-level analyses included 715 individual patient encounters and evaluated differences in workflow intervals according to Spanish-speaking status. Clinic-session-level analyses included 128 clinic sessions and evaluated associations between clinic characteristics; including total patient volume, Spanish-speaking patient volume, in-person interpreter staffing, and Spanish-speaking patient-to-interpreter ratio; and overall clinic-session duration.

Descriptive statistics were used to summarize encounter-level workflow intervals and clinic-session-level characteristics. Continuous variables were summarized using means and standard deviations and medians, quartiles, ranges, minimums, and maximums, as appropriate. Encounter-level workflow intervals were compared between Spanish-speaking and non-Spanish-speaking patient encounters using linear mixed-effects regression models. Spanish-speaking status and clinic day (Monday vs. Saturday) were included as fixed effects, and clinic session was included as a random intercept to account for correlation among patient encounters occurring within the same clinic session. Separate models were estimated for each workflow interval using restricted maximum likelihood estimation. Model estimates are reported as adjusted mean differences

with 95% confidence intervals (CIs) and two-sided *P* values. No adjustments for multiplicity were applied.

Clinic-session-level analyses used the clinic session as the unit of analysis because interpreter staffing, patient volume, and overall clinic-session duration were measured at the session level. Each clinic session constituted one observation (*n* = 128). Separate multivariable linear regression models were constructed to evaluate the associations of language-service demand, in-person interpreter staffing, and interpreter workload with clinic-session duration while adjusting for relevant covariates.

Spanish-speaking patient volume was used as a proxy for language-service demand, and the number of in-person interpreters represented interpreter staffing capacity. Interpreter workload was operationalized as the Spanish-speaking patient-to-interpreter ratio, calculated as the number of Spanish-speaking patients divided by the number of in-person interpreters available during the clinic session. Because the Spanish-speaking patient-to-interpreter ratio was mathematically derived from Spanish-speaking patient volume and in-person interpreter staffing, the ratio and its component variables were not included simultaneously in the same regression model. Variance inflation factors (VIFs) and tolerance statistics were examined to assess multicollinearity among predictors included in each model.

Regression coefficients (β), standard errors, 95% CIs, *P* values, and adjusted *R*² values were reported, as appropriate. Statistical significance was defined as a two-sided *p* < 0.05.

This project was reviewed by the University of Kansas Medical Center Institutional Review Board (IRB) and determined to constitute a quality improvement project. This manuscript was prepared in accordance with the Standards for Quality Improvement Reporting Excellence (*SQUIRE*) 2.0 guidelines.⁶ All analyses were conducted using SAS® version 9.4 (SAS Institute Inc., Cary, NC).

RESULTS

Among the 715 patient encounters, 422 occurred on Saturday and 293 occurred on Monday. Spanish-speaking patients accounted for 286 (67.8%) Saturday encounters, 208 (71.0%) Monday encounters, and 494 (69.1%) encounters overall. Saturday clinics had a mean of 2.43 interpreters per session (SD = 0.88; range, 1-5), whereas Monday clinics had a mean of 1.95 interpreters per session (SD = 0.38; range, 1-3).

In encounter-level linear mixed-effects models accounting for clustering of patient encounters within clinic sessions and adjusting for clinic day, Spanish-speaking

Table 1: Comparison between Spanish-speaking and non-Spanish-speaking encounters.

Outcome (min)	Non-Spanish-speaking patients Mean $\pm$ SD	Spanish-speaking patients Mean $\pm$ SD	Mean difference (95% CI)*	<i>P</i> value
Total patient visit duration	98.7 $\pm$ 37.1	107.9 $\pm$ 38.1	5.30 (−0.33 to 10.92)	0.065
Wait to be roomed	10.4 $\pm$ 17.7	11.8 $\pm$ 17.7	0.46 (−2.29 to 3.21)	0.743
Premed time	7.2 $\pm$ 6.6	7.7 $\pm$ 6.4	0.46 (−0.59 to 1.51)	0.393
Prep time (medical)	7.8 $\pm$ 9.1	10.7 $\pm$ 11.2	2.57 (0.83 to 4.31)	0.004
Prep time (pharmacy)	18.9 $\pm$ 27.1	19.3 $\pm$ 30.7	−3.12 (−12.82 to 6.59)	0.529
Medicine encounter time	24.4 $\pm$ 9.7	24.3 $\pm$ 10.2	−0.23 (−2.17 to 1.72)	0.818
Pharmacy encounter time	19.3 $\pm$ 13.1	18.7 $\pm$ 15.9	−0.63 (−5.53 to 4.27)	0.801
Wait to present (medical)	19.2 $\pm$ 11.6	22.7 $\pm$ 12.3	3.46 (0.47 to 6.45)	0.024
Attending time	9.8 $\pm$ 6.1	11.6 $\pm$ 6.7	1.59 (−0.24 to 3.42)	0.089
Discharge time	28.4 $\pm$ 18.5	26.6 $\pm$ 17.9	−0.33 (−3.80 to 3.15)	0.855

*Adjusted mean differences and *p* values were estimated using linear mixed-effects models with Spanish-speaking status and clinic day (Monday vs. Saturday) included as fixed effects and clinic session included as a random intercept to account for clustering of patient encounters within clinic sessions. Positive estimates indicate longer workflow times among Spanish-speaking encounters. Sample sizes varied by outcome because of missing workflow timestamps. Abbreviations: SD, standard deviation; CI, confidence interval.

encounters had an adjusted mean total patient visit duration that was 5.30 minutes longer than that of non-Spanish-speaking encounters; however, this difference was not statistically significant (95% CI, -0.33 to 10.92; $p = 0.065$). Spanish-speaking encounters were associated with significantly longer medical preparation time (adjusted mean difference, 2.57 minutes; 95% CI, 0.83-4.31; $p = 0.004$) and wait-to-present time (adjusted mean difference, 3.46 minutes; 95% CI, 0.47-6.45; $p = 0.024$). No statistically significant differences were observed in wait-to-be-roomed time, premedical student intake time, medical student encounter time, attending encounter time, or discharge time (Table 1).

At the clinic-session level, after adjusting for total patient volume and in-person interpreter staffing, each additional Spanish-speaking patient was associated with a 6.70-minute increase in clinic-session duration (95% CI, 3.18-10.22; $p < 0.001$). In a separate model adjusting for total patient volume, each 1-unit increase in the Spanish-speaking patient-to-interpreter ratio was associated with a 3.72-minute increase in clinic-session duration; however, this association was not statistically significant (95% CI, -0.03 to 7.47; $p = 0.052$). In another model adjusting for total patient volume, each additional in-person interpreter was associated with a 2.21-minute increase in clinic-session duration, although this association was not statistically significant (95% CI, -3.03 to 7.45; $p = 0.406$). No substantial multicollinearity was identified among predictors included in the multivariable models, with VIFs ranging from 1.04 to 2.16 and tolerance values ranging from 0.46 to 0.96 (Table 2).

DISCUSSION

In this quality improvement evaluation, differences in workflow associated with Spanish-speaking patient encounters were concentrated in specific phases of care rather than in overall patient visit duration. After accounting for clustering of encounters within clinic sessions and adjusting for clinic day, Spanish-speaking encounters had an estimated 5.3-minute longer patient visit duration than non-Spanish-speaking encounters; however, this difference was not statistically significant. Significant differences were observed in medical preparation and wait-to-present times, two workflow phases involving preparation for and communication of patient information. At the clinic-session level, greater Spanish-speaking patient volume was associated with longer clinic-session duration. Given that approximately 70% of patient encounters involved Spanish-speaking patients, these findings highlight the operational importance of aligning interpreter resources with language-service demand.

The longer workflow intervals were concentrated in phases of clinic flow that are communication intensive. Specifically, medical preparation time was approximately 2.6 minutes longer and wait-to-present time approximately 3.5 minutes longer among Spanish-speaking encounters after accounting for clinic-session clustering and clinic day. Medical preparation time corresponds to medical student chart review and preparation for the patient encounter, whereas wait-to-present time corresponds to the period during which learners wait to present patient cases

Table 2: Multivariable regression analysis of Spanish-speaking patient volume and clinic workflow.

Model and Predictor	β (min)	SE	95% CI	<i>P</i> value	Adjusted R^2	<i>n</i>	Covariates
Model 1: Spanish-speaking patient volume Each additional Spanish-speaking patient	6.70	1.8	3.18 to 10.22	<0.001	0.265	128	Total patient volume; in-person interpreters
Model 2: Interpreter workload Each 1-unit increase in Spanish-speaking patients per interpreter	3.72	1.9	-0.03 to 7.47	0.052	0.207	128	Total patient volume
Model 3: In-person interpreter staffing Each additional in-person interpreter	2.21	2.6	-3.03 to 7.45	0.406	0.187	128	Total patient volume

Abbreviations: β , regression coefficients; SE, standard error; CI, confidence interval; Adjusted R^2 , adjusted R-squared; *n*, sample size.

to the resident or attending. These phases may be particularly sensitive to language-related workflow demands, including time spent waiting for an in-person interpreter or coordinating telephone interpretation services. Because interpreter modality was not consistently captured at the encounter level, however, the observed differences cannot be attributed specifically to limited in-person interpreter availability.

At the clinic-session level, each 1-unit increase in the Spanish-speaking patient-to-interpreter ratio was associated with an estimated 3.72-minute increase in clinic-session duration; however, the confidence interval included the null value (95% CI, -0.03 to 7.47), and the association was not statistically significant ($p = 0.052$). This uncertainty may reflect the relatively small number of clinic sessions available for the clinic-session-level analyses, which may have limited the precision of the estimates. These findings indicate that the relationship between language-service demand and in-person interpreter capacity warrants further prospective evaluation rather than establishing that increased interpreter staffing improves clinic efficiency.

These findings can be considered in the context of prior studies examining interpreter services and clinical workflow. Johnston, et al.⁷ reported that total and Spanish-speaking patient visit times were significantly shorter when either video or in-person interpreter services were used in an otolaryngology clinic. Fagan, et al.⁸ found that hospital-based interpreters were more time-efficient than telephone-based interpretation and estimated that providing hospital-based interpreters for Spanish-speaking patients could allow additional visits in an outpatient teaching clinic. Although these studies support the potential operational importance of interpreter modality and availability, their findings should be interpreted cautiously when compared with the present study because the settings, study designs, and outcomes differed.

Unlike many prior studies that evaluated individual patient visits, this quality improvement initiative examined both encounter-level workflow and clinic-session-level interpreter staffing from a systems perspective as part of an ongoing quality improvement initiative. Rather than focusing solely on encounter-level differences, we evaluated how language-service demand, relative to in-person interpreter staffing, was associated with the operation of the clinic as a whole. This system-level approach provides practical in-

formation for clinic leadership regarding the relationship between language-service demand, interpreter resources, and clinic workflow. Student-run free clinics and other resource-limited ambulatory settings may benefit from using routinely collected workflow data to monitor these relationships and inform interpreter scheduling. However, because the number of in-person interpreters was not independently associated with clinic-session duration in the present study, these findings should not be interpreted as demonstrating that increasing interpreter staffing alone improves clinic efficiency.

Interpreter services are necessary for effective communication with patients who have limited English proficiency, while also representing an operational resource that requires planning and coordination. In this setting, the findings suggest that interpreter demand should be considered when planning clinic staffing and workflow. Monitoring Spanish-speaking patient volume, interpreter availability, and communication-intensive workflow intervals may help clinic leadership identify periods in which interpreter resources and patient demand are misaligned. Such monitoring may provide a basis for testing different interpreter staffing models in future quality improvement initiatives. Because the present study did not demonstrate an independent association between the number of in-person interpreters and clinic-session duration, prospective evaluation is needed to determine whether specific staffing strategies improve workflow efficiency without compromising the quality or accessibility of interpreter services.

Limitations

Several limitations should be considered when interpreting these findings. First, this study was conducted at a single student-run free clinic, which may limit generalizability to other practice settings. Second, because this was an observational quality improvement project and the intervention was introduced incrementally and refined over 20 months, causal relationships between interpreter staffing and workflow efficiency cannot be established. Although mixed-effects models accounted for clustering of patient encounters within clinic sessions and clinic-session-level regression models adjusted for measured staffing and patient-volume characteristics, residual confounding by factors such as patient complexity, learner experience, interpreter experience, and staffing composition may remain.

Third, in-person interpreter staffing was measured at the clinic-session level and does not represent total interpreter access or actual interpreter utilization by individual patients. Telephone interpretation was available when in-person interpreters were unavailable, but its use was not consistently captured. Therefore, associations involving in-person interpreter staffing should not be interpreted as estimates of the effects of overall interpreter access or receipt of interpreter services during individual encounters. Fourth, the relatively small number of clinic sessions may have limited the precision of clinic-session-level estimates, particularly for the association between the Spanish-speaking patient-to-interpreter ratio and clinic-session duration. Fifth, clinic sessions occurring on Mondays and Saturdays could not be differentiated in the data analysis and differences in interpreter staffing could not be accounted for.

Finally, workflow measures were derived from routinely collected operational data and were not formally validated, making delayed or missed timestamp entry a potential source of measurement error. The quality improvement initiative also focused on operational efficiency and did not evaluate patient satisfaction, communication quality, clinical outcomes, or the costs associated with interpreter services. Future quality improvement initiatives should prospectively evaluate interpreter staffing models while incorporating patient-centered outcomes, including satisfaction, communication quality, and access to care. Evaluation of the sustainability and cost-effectiveness of different interpreter staffing strategies would further inform efforts to optimize clinic workflow while maintaining equitable care for linguistically diverse patient populations.

REFERENCES

1. United States Census Bureau. New Data on Detailed Languages Spoken at Home and the Ability to Speak English. 2025. <https://www.census.gov/newsroom/press-releases/2025/2017-2021-acs-language-use-tables.html>. Accessed July 2, 2026.
2. City of Wichita. Wichita Census 2020. City of Wichita GIS. 2020. <https://wichita-census-2020-cityofwichita.hub.arcgis.com/>. Accessed July 2, 2026.
3. United States Census Bureau. QuickFacts Wichita city, Kansas. 2024. <https://www.census.gov/quickfacts/fact/table/wichitacitykansas/POP815224>. Accessed July 2, 2026.
4. Hu T, Leung FH. An evaluation of wait-times at an interprofessional student-run free clinic. *J Interprof Care* 2016 Jul; 30(4):532-535. PMID: 27332500.
5. Clark NM, Hirsch JL, Wallace CR, et al. Assessing community needs among Spanish-speaking patients in a free clinic setting. *J Prim Care Community Health* 2024; 15:21501319241276810. PMID: 39305068.
6. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf*. 2016 Dec;25(12):986-992. Epub 2015 Sep 14. PMID: 26369893.
7. Johnston DR, Lavin JM, Hammer AR, et al. Effect of dedicated in-person interpreter on satisfaction and efficiency in otolaryngology ambulatory clinic. *Otolaryngol Head Neck Surg* 2021 May; 164(5):944-951. PMID: 32957819.
8. Fagan MJ, Diaz JA, Reinert SE, et al. Impact of interpretation method on clinic visit length. *J Gen Intern Med* 2003 Aug; 18(8):634-638. PMID: 12911645.

Keywords: *medically uninsured; Hispanic or Latino; communication barriers; limited English proficiency; student run clinic*

Delayed Eyelid Masses: When to Suspect Infiltrative Mimickers

Emmanuel Vazquez, M.D.

Universidad Autónoma de
Bucaramanga (UNAB), Bucaramanga,
Santander, Colombia
Fundación Oftalmológica de Santander
(FOSCAL), Floridablanca, Santander,
Colombia

Received September 6, 2026; Accepted
for publication September 24, 2026;
Published online October 16, 2026,
Kans J Med 2026 Sept-Oct; 19:167-
169. [https://doi.org/10.17161/kjm.
vol19.26103](https://doi.org/10.17161/kjm.vol19.26103).

Conflict of Interest: *None.*

Corresponding Author:

Emmanuel Vazquez, M.D.
Calle 157 #19-55, Campus el Bosque,
Universidad Autónoma de Bucaraman-
ga, Floridablanca, Santander, Colombia
emmanuelvazquez1404@gmail.com

To the Editor: I read with great interest the report by Le and Burgoyne of a 64-year-old woman with painless unilateral lower eyelid fullness that developed three years after malar hyaluronic acid filler injection. The authors correctly identified a delayed filler-related complication, supported by the intraoperative gelatinous material, acellular foreign material on histopathology, the subsequently disclosed history, and improvement after hyaluronidase.¹ I wish to complement their discussion by highlighting four relevant infiltrative mimics.

First, sarcoidosis may involve the eyelids, lacrimal gland, extraocular muscles, and orbital tissues. Eyelid granulomas, dacryoadenitis, uveitis, or proptosis may provide diagnostic clues. A compatible biopsy supports definite ocular sarcoidosis, whereas bilateral hilar adenopathy may support a presumed diagnosis.² In this case, the normal ocular examination and absence of granulomas or an inflammatory infiltrate made sarcoid involvement unlikely.

Second, the diagnostic criteria for IgG4-related ophthalmic disease integrate imaging, histopathology, and serum IgG4 concentration. Imaging may show enlargement of the lacrimal gland, trigeminal nerve, or extraocular muscles, as well as masses involving other ophthalmic tissues. Histopathology typically demonstrates marked lymphocytic and plasmacytic infiltration, sometimes with fibrosis, and abundant IgG4-positive plasma cells; serology alone does not establish a definite diagnosis.³ Isolated lower eyelid involvement, normal ocular motility, and the absence of an inflammatory infiltrate did not favor this diagnosis.

Third, ocular adnexal lymphoma may involve the orbit, conjunctiva, lacrimal apparatus, or eyelid. In the series by Boltezar et al.,⁴ the median age was 68 years, and mucosa-associated lymphoid tissue lymphoma accounted for 71.6% of cases. Ocular adnexal lymphoma may manifest as an eyelid mass, ptosis, or painless proptosis; on imaging, an orbital mass typically is poorly defined and molds to adjacent structures without bony erosion. Diagnosis requires biopsy with histopathologic, immunophenotypic, and molecular studies. The absence of malignancy and the presence of acellular material did not support this diagnosis.

Finally, idiopathic orbital inflammation may cause conjunctival injection, periorbital edema, proptosis, pain, diplopia, or reduced ocular motility. Evaluation integrates clinical findings, magnetic resonance imaging, and blood tests, with biopsy reserved for cases in which the diagnosis remains uncertain.⁵ In this case, chronic painless edema without proptosis or motility impairment, together with the absence of an inflammatory infiltrate, made this possibility less likely.

Together, these comparisons complement the authors' sound assessment. This case reinforces the importance of specifically asking about remote cosmetic procedures and integrating the chronology, anatomic location, imaging when uncertainty persists, and tissue analysis before attributing an eyelid mass to a benign cause.

REFERENCES

1. Le DN, Burgoyne JK. Isolated, Unilateral Lower Eyelid Mass Presenting as a Delayed Ocular Complication of Dermal Filler. *Kans J Med.* 2026;19(4):119-121. PMID: 42639601.
2. Rosenbaum JT, Pasadhika S. Ocular Sarcoidosis. *Clin Chest Med.* 2024;45(1):59-70. PMID: 38245371.
3. Takahira M, Goto H, Azumi A. The 2023 revised diagnostic criteria for IgG4-related ophthalmic disease. *Jpn J Ophthalmol.* 2024;68(4):293-301. PMID: 39037515.
4. Boltezar L, Strbac D, Pizem J, Hawlina G. Ocular adnexal lymphoma - a retrospective study and review of the literature. *Radiol Oncol.* 2024;58(3):416-424. PMID: 39287167.
5. Otsuka M, Yunoki T, Ozaki H, Hayashi A. Clinical Characteristics of Idiopathic Orbital Inflammation Syndrome in Relation to Intraocular Pressure. *Clin Ophthalmol.* 2022;16:1467-1473. PMID: 35592671.

In Reply to Emmanuel Vazquez, M.D.: We thank Dr. Vazquez for their comments on our article. We agree with the author's discussion of four relevant infiltrative mimics and the clinical features that should raise suspicion for these conditions. Although this patient's presentation was not consistent with any of these more serious conditions, we agree that clinicians should consider and recognize these entities when evaluating eyelid masses.

Ocular sarcoidosis can involve any ocular structure, and

its heterogeneous clinical presentation can pose a significant diagnostic challenge. The revised International Workshop on Ocular Sarcoidosis (IWOS-2019) and Standardization of Uveitis Nomenclature (SUN) criteria can help guide diagnosis; however, definitive diagnosis relies on biopsy and histopathologic confirmation.¹⁻³ Although granulomatous uveitis remains the predominant manifestation of ocular sarcoidosis, we have encountered a case in which unilateral eyelid mass/swelling was the sole manifestation and was confirmed by open biopsy and histopathologic examination.

IgG4-related ophthalmic disease can present as masses, enlargement, or hypertrophic lesions involving various ophthalmic tissues.⁴ We also have encountered a case presenting with eyelid swelling and a conjunctival mass.

We agree that ocular adnexal lymphoma and idiopathic orbital inflammation may both manifest as eyelid masses and should therefore be included in the differential diagnosis. Another relevant mimic is thyroid orbitopathy, which may present with asymmetric eyelid edema, usually accompanied by mild proptosis, in its early stages.⁵

Overall, we believe Dr. Vazquez's comments complement our discussion and further highlight the importance of maintaining a broad differential diagnosis that includes infiltrative mimics. The widespread use of dermal fillers in the periorbital region can result in long-term complications that often are bilateral and recognizable; however, as illustrated by this case,⁶ patients may not recall prior filler treatments or may not consider them relevant to their current presentation.

REFERENCES

1. Sandu CA, Donica VC, Ganea CV, Bîlha MI, Alexa AI, Bogdănici CM. Ocular sarcoidosis: Contemporary insights and future directions. *Rom J Ophthalmol* 2025 Oct-Dec; 69(4):488-494. PMID: 41971214.
2. Mochizuki M, Smith JR, Takase H, Kaburaki T, Acharya NR, Rao NA; International Workshop on Ocular Sarcoidosis Study Group. Revised criteria of International Workshop on Ocular Sarcoidosis (IWOS) for the diagnosis of ocular sarcoidosis. *Br J Ophthalmol* 2019 Oct; 103(10):1418-1422. PMID: 30798264.
3. Standardization of Uveitis Nomenclature (SUN) Working Group. Development of classification criteria for

the uveitides. Am J Ophthalmol 2021 Aug; 228:96-105. PMID: 33848532.

4. Takahira M, Goto H, Azumi A. The 2023 revised diagnostic criteria for IgG4-related ophthalmic disease. Jpn J Ophthalmol 2024 Jul; 68(4):293-301. PMID: 39037515.
5. Wiersinga WM, Eckstein AK, Žarković M. Thyroid eye disease (Graves' orbitopathy): clinical presentation, epidemiology, pathogenesis, and management. Lancet Diabetes Endocrinol 2025 Jul; 13(7):600-614. PMID: 40324443.
6. Le DN, Burgoyne JK. Isolated, Unilateral Lower Eyelid Mass Presenting as a Delayed Ocular Complication of Dermal Filler. Kans J Med. 2026;19(4):119-121. PMID: 42639601.

In Reply to Vazquez

Diana N. Le, MS-4¹

Jennifer K. Burgoyne, M.D.^{1,2,3}

¹The University of Kansas School of Medicine-Wichita, Wichita, Kansas

²Department of Surgery

³ICT Eye LLC, Wichita, Kansas

Published in Final Edited Form as:

Kans J Med 2026 Sept-Oct; 19:170-171. <https://doi.org/10.17161/kjm.vol19.26197>.

Conflict of Interest: None.

Corresponding Author:

Diana Le
2002 South Beech Street
Wichita, KS 67207
dle4@kumc.edu

We thank Dr. Vazquez for their comments on our article. We agree with the author's discussion of four relevant infiltrative mimics and the clinical features that should raise suspicion for these conditions. Although this patient's presentation was not consistent with any of these more serious conditions, we agree that clinicians should consider and recognize these entities when evaluating eyelid masses.

Ocular sarcoidosis can involve any ocular structure, and its heterogeneous clinical presentation can pose a significant diagnostic challenge. The revised International Workshop on Ocular Sarcoidosis (IWOS-2019) and Standardization of Uveitis Nomenclature (SUN) criteria can help guide diagnosis; however, definitive diagnosis relies on biopsy and histopathologic confirmation.¹⁻³ Although granulomatous uveitis remains the predominant manifestation of ocular sarcoidosis, we have encountered a case in which unilateral eyelid mass/swelling was the sole manifestation and was confirmed by open biopsy and histopathologic examination.

IgG4-related ophthalmic disease can present as masses, enlargement, or hypertrophic lesions involving various ophthalmic tissues.⁴ We also have encountered a case presenting with eyelid swelling and a conjunctival mass.

We agree that ocular adnexal lymphoma and idiopathic orbital inflammation may both manifest as eyelid masses and should therefore be included in the differential diagnosis. Another relevant mimic is thyroid orbitopathy, which may present with asymmetric eyelid edema, usually accompanied by mild proptosis, in its early stages.⁵

Overall, we believe Dr. Vazquez's comments complement our discussion and further highlight the importance of maintaining a broad differential diagnosis that includes infiltrative mimics. The widespread use of dermal fillers in the periorbital region can result in long-term complications that often are bilateral and recognizable; however, as illustrated by this case,⁶ patients may not recall prior filler treatments or may not consider them relevant to their current presentation.

REFERENCES

1. Sandu CA, Donica VC, Ganea CV, Bilha MI, Alexa AI, Bogdănici CM.

- Ocular sarcoidosis: Contemporary insights and future directions. *Rev J Ophthalmol* 2025 Oct-Dec; 69(4):488-494. PMID: 41971214.
2. Mochizuki M, Smith JR, Takase H, Kaburaki T, Acharya NR, Rao NA; International Workshop on Ocular Sarcoidosis Study Group. Revised criteria of International Workshop on Ocular Sarcoidosis (IWOS) for the diagnosis of ocular sarcoidosis. *Br J Ophthalmol* 2019 Oct; 103(10):1418-1422. PMID: 30798264.
 3. Standardization of Uveitis Nomenclature (SUN) Working Group. Development of classification criteria for the uveitides. *Am J Ophthalmol* 2021 Aug; 228:96-105. PMID: 33848532.
 4. Takahira M, Goto H, Azumi A. The 2023 revised diagnostic criteria for IgG4-related ophthalmic disease. *Jpn J Ophthalmol* 2024 Jul; 68(4):293-301. PMID: 39037515.
 5. Wiersinga WM, Eckstein AK, Žarković M. Thyroid eye disease (Graves' orbitopathy): clinical presentation, epidemiology, pathogenesis, and management. *Lancet Diabetes Endocrinol* 2025 Jul; 13(7):600-614. PMID: 40324443.
 6. Le DN, Burgoyne JK. Isolated, Unilateral Lower Eyelid Mass Presenting as a Delayed Ocular Complication of Dermal Filler. *Kans J Med.* 2026; 19(4):119-121. PMID: 42639601.

Have a manuscript ready to publish?

Visit our website for instructions on submitting a manuscript. journals.ku.edu/kjm

Publication Staff

Samuel Ofei-Dodoo, Ph.D., MPA, M.A., CPH
Editor-in-Chief/Managing Editor

Justin Moore, M.D., FACP
Medical Editor

Jon P. Schrage, M.D.
Editor-in-Chief Emeritus

K. James Kallail, Ph.D.
Associate Editor Emeritus

Leslie Hunt, B.A.
Publication Manager

Abigail Tatpati
Intern

Editorial Board Members

Nisha Agasthya, M.D.
Kamran Ali, M.D.
Darren Farley, M.D.
Mark E. Harrison, M.D.
Bernard F. Hearon, M.D.
Justin T. Helberg, M.D.
Samuel Joseph Hund, M.D.
Cyrus Munguti, MBChB
Missy Norton, Pharm.D.
Hayrettin Okut, Ph.D.
Tiffany Schwasinger-Schmidt, M.D., Ph.D.
Wade T. Swenson II, M.D.
Rosey Zacula, M.A., Statistician



**Kansas Journal of Medicine is partially supported by
Wichita Medical Research & Education Foundation**
wichitamedicalresearch.org



Medical Science
1010 N. Kansas • Wichita, KS 67214 • 316-293-3811 • kjm@kumc.edu

The University of Kansas prohibits discrimination on the basis of race, color, ethnicity, religion, sex, national origin, age, ancestry, disability, status as a veteran, sexual orientation, marital status, parental status, gender identity, gender expression, and genetic information in the university's programs and activities. Retaliation is also prohibited by university policy and procedures. The following person has been designated to handle inquiries regarding the nondiscrimination policies and is the Title IX coordinator for all KU and KUMC campuses: Associate Vice Chancellor for the Office of Civil Rights and Title IX, civilrights@ku.edu, Room 1082, Dole Human Development Center, 1000 Sunnyside Avenue, Lawrence, KS, 66045, 785-864-6414, 711 TTY. Reports can be submitted by contacting the Title IX coordinator as provided herein or using the Title IX online report form.