International Surgery Journal
Kreilach A et al. Int Surg J. 2026 May;13(5):724-728
http://www.ijsurgery.com
pISSN 2349-3305 | eISSN 2349-2902

DOI: https://dx.doi.org/10.18203/2349-2902.isj20261160

Original Research Article

Accuracy and efficiency of utilizing Microsoft Excel search functions and artificial intelligence for retrospective data analysis

Alec Kreilach[1], Marlee Franden[2], Elise Paranto[1], Gopal Kowdley[1]*

[1]Department of Surgery, TidalHealth, Salisbury, MD, USA
[2]Rowan-Virtua School of Osteopathic Medicine, Stratford, NJ, USA

Received: 12 January 2026
Revised: 23 March 2026
Accepted: 20 April 2026

***Correspondence:**
Dr. Gopal Kowdley,
E-mail: gopal.kowdley@tidalhealth.org

ABSTRACT

Background: Artificial intelligence (AI) has become integrated into many aspects of modern medicine, including medical research. Manual retrospective chart review has been regarded as an effective yet time-consuming and taxing process.

Methods: To identify an accurate yet faster process, a retrospective electronic chart analysis was performed comparing three modalities: Microsoft Excel search function and Microsoft CoPilot were compared against manual review in identifying appendiceal neoplasms in patients undergoing laparoscopic or open appendectomy from an extracted database. The time taken to identify key words establishing neoplastic processes and the number of words found were recorded for manual search, Microsoft Excel search function, and Microsoft CoPilot.

Results: The Microsoft Excel search function identified more terms than Microsoft CoPilot; however, it did not find all the instances found in manual search. The Microsoft Excel search function process took the least amount of time (2 minutes), and manual search took the most (2 hours 22 minutes). Although Microsoft CoPilot was the most dynamic tool, it was not as fast as the Microsoft Excel search function, nor as accurate as the other two modalities.

Conclusions: Although the Microsoft Excel search function had the best speed, it was a static tool that did not find every key word and could not extrapolate data. Based on this paper and other literature, AI has the potential to assist in retrospective chart analysis. However, the search function remains faster and much more accurate than AI at this time.

Keywords: Artificial intelligence, Retrospective, Database review

INTRODUCTION

Artificial intelligence (AI) was first described in 1950 by Alan Turing as a basic series of 'if, then rules' and has exploded in complexity to the current day iteration that mimics the human brain with highly sophisticated neural networks. In modern medicine, AI has a plethora of uses such as reading radiographic images, creating differential diagnoses, and generation of medical documentation.[1] AI has percolated many aspects of healthcare and medical research, but there is a paucity of studies that have used AI as an assistive technology in retrospective chart analysis.

Retrospective analysis of electronic medical records (EMR) is a well-known and widely used tool by researchers to answer specific questions in medicine.[2] Traditional retrospective chart reviews analyze data within medical records that were originally intended for documentation, not research purposes.[3] Retrospective chart reviews allow for inexpensive yet time consuming analysis of things such as rare pathologies and diseases that take many years to develop.[3] Accuracy of manual chart review can vary between individuals depending on the individual reviewer. A recent study has shown that when reviewing a chart manually, it takes a reviewer around 7.56 minutes per chart.[4] This same study also found that accuracy in finding specific end points via manual review varies widely between 74.7% to 98.6%.[4] Another inherent challenge with analyzing data from a chart review is that many of the data are considered ‘free text’; making it difficult to reliably abstract from the source as well as analyze with many traditional statistical tools.[2] In attempt to combat these challenges, recent authors used the macro function in Microsoft Excel to be able to quickly and precisely analyze an unstructured data set to answer specific questions.[2] We have set out to determine the accuracy and speed using the commercially available resources of Microsoft Excel search function and Microsoft CoPilot compared to manual review for data analysis as part of a retrospective chart review process.

METHODS

Institutional IRB approval for the chart review was obtained from the home institution, TidalHealth Regional Medical Center in Salisbury, MD. Our study took place between February 2025 and January 2026. A retrospective electronic chart review of our institution’s Epic charts was undertaken to obtain pathological reports of all patients undergoing laparoscopic or open appendectomy for acute appendicitis between 01 January 2019 and 31 December 2022. Patients were excluded from the study if they were under the age of 18 or if they received an appendectomy for reasons other than acute appendicitis. An epic analyst at our institution pulled all charts within the given date range by current procedural terminology (CPT) code for open and laparoscopic appendectomy; 44960 and 44970, respectively. The analyst placed the pathology reports for each instance of appendectomy into a Microsoft Excel sheet with corresponding medical record numbers (MRN). We chose this specific database to attempt to understand the incidence of mucinous appendiceal neoplasms at our institution. During this data set review, we thought it to be prudent to try finding a method to reduce time spent on data review with similar if not better accuracy than gold standard manual data review. The data was de-identified, and the charts were then searched for key words: ‘mucinous, carcinoma, neoplasm, tumor’ using the search function in Microsoft Excel. These terms were selected to account for variation in pathologist wording during dictations.

Microsoft CoPilot version 1.25054.80.0 was then queried with de-identified data of the Excel sheet, in multiple runs to learn with AI. We were not attempting to identify the full capabilities of CoPilot, rather, we were looking at its use by the researcher without significant knowledge of AI. A total of four runs were completed, timed, and the correct entries/errors were recorded. The first run of CoPilot was queried with the phrase "identify all instances of neoplastic pathology”. The second run query was “add instances of carcinoma and highlight all cells containing carcinoma in the report”. Third query stated, “highlight all instances of tumor and hyperplasia and polyps”. Query number four, “highlight in orange all instances of neoplastic processes such as cancer, carcinoma, polyp, hyperplastic, and mucinous”.

Microsoft Excel using the search function was interrogated first, followed by manual chart review and finally Microsoft CoPilot was queried. Each of the three search modalities was timed. Findings and differences in the Excel and CoPilot arms were recorded against the gold standard of manual data review. The manual review was independently performed by two second year surgical residents with similar novice expertise in data review experience. The two reviewers identified the same incident cases. To analyze the data, two by two tables were created comparing both Microsoft Excel to manual review as well as each Microsoft CoPilot run to manual review. Sensitivity, specificity, positive predictive value, negative predictive value and Chi-square analysis were all calculated with the assistance of online tools.[5,6] Data reporting was not aligned with either STARD – AI or DECIDE – AI reporting guidelines, as the aim of the study was to assess use of CoPilot by a novel user and not to evaluate the full range of capabilities as an AI system.

RESULTS

There was a total of 631 records reviewed for patients in our institution who underwent appendectomy between 01 January 2019 and 31 December 2022. A manual search was performed by the two reviewers, and this yielded 14 instances of neoplasia within the pathology report, establishing our gold standard for comparison. Instances found in manual search included the terms ‘neoplasm, tumor, carcinoma, mucinous, sessile serrated adenoma’. Manual search run time was 2 hours 37 minutes for reviewer 1. Reviewer two search time was 1 hour 34 minutes. Group discussion of findings between reviewers took 15 minutes and 56 seconds. The Microsoft Excel search uncovered 10 out of 14 instances of neoplasia, missing the sessile serrated adenoma/polyp (Table 1). One false positive during Microsoft Excel word search came from a report of ‘Mucin consistent with perforation'. The run time for Microsoft Excel word search was 2 minutes.

The first CoPilot run yielded two results, both of which were low-grade mucinous appendiceal neoplasms. The first run time was 2 min 37 seconds. The second run added one more unique appendiceal neoplasm which was an adenocarcinoma of the appendix. One of the previously found mucinous neoplasms was also found in the second run.

Lastly, for the second run, a false positive was elucidated which was a tubular adenoma of the cecum. The run time for the second search was 2 min and 21 seconds. For run number three, there were no neoplasms found on this search. The third run time was 1 min and 21 seconds. The last CoPilot search identified three appendiceal neoplasms, including an adenocarcinoma, villous adenoma, sessile serrated polyp, as well as 10 non-appendiceal neoplastic processes, 9 of which were lymphoid hyperplasia and one instance of a hyperplastic polyp (Table 1). Run four time was 2 min 28 seconds.

Table 1: Demonstration of the number of correct terms identified and missed for each search modality.

Test		True neoplasm	Negative neoplasm	
Microsoft Excel search function accuracy	Neoplasm identified	10	1	Positive predictive value: 90.91%
	Negative neoplasm by test	4	616	Negative predictive value: 99.35%
		Sensitivity: 71.43%	Specificity: 99.84%	χ^2=405.9093, p≤0.001
Microsoft CoPilot search run 1	Neoplasm identified	2	0	Positive predictive value: 100%
	Negative neoplasm by test	12	617	Negative predictive value: 98.09%
		Sensitivity: 14.29%	Specificity: 100%	χ^2=88.4231, p≤0.001
Microsoft CoPilot search run 2	Neoplasm identified	2	1	Positive predictive value: 66.67%
	Negative neoplasm by test	12	616	Negative predictive value: 98.09%
		Sensitivity: 14.29%	Specificity: 99.84%	χ^2=57.7105, p≤0.001
Microsoft CoPilot search run 3	Neoplasm identified	0	0	Positive predictive value: 97.78%
	Negative neoplasm by test	14	617	Negative predictive value: 97.78%
		Sensitivity: 0%	Specificity: 100%	χ^2=57.7105, p≤0.001
Microsoft CoPilot search run 4	Neoplasm identified	3	10	Positive predictive value: 23.08%
	Negative neoplasm by test	11	607	Negative predictive value: 98.22%
		Sensitivity: 21.43%	Specificity: 98.38%	χ^2=26.6185, p≤0.001

DISCUSSION

This study evaluated the accuracy and speed of three modalities that may be used in retrospective electronic medical record review. The Microsoft Excel search function and Microsoft CoPilot were compared against manual search in an attempt to identify a more accurate and time-efficient method for medical data review. With regards to accuracy, we found that the Microsoft Excel search function was superior to Microsoft CoPilot; however, it did not find all terms identified in manual search. The Microsoft Excel search function took the least amount of time out of the three methods.

Manual chart review has been widely used in retrospective analysis, as search tools and AI have not always been available. Accuracy and speed are variable based on the individual conducting the review, though this process is widely regarded as demanding and time-consuming. A recent study that evaluated speed and accuracy of manual chart review for sacral wound complications compared to a medical center's internal ChatGPT found that AI took 1.03 minutes per chart with 95.7% accuracy whereas manual review took 7.56 minutes per chart with accuracy varying between 74.7-98.6%.[4] In our study, when compared to the two other modalities, manual search of our database took a considerably longer time. This is especially important in the current clinical setting, in which clinicians are placed under considerable pressure to perform more tasks in less time. However, neither the search function nor the AI tool, in our study, matched the accuracy of the manual search. Therefore, it raises the question of whether a less time-consuming process is worth the reduced accuracy. It is understood that other statistical software packages with different AI tools are available on the market; however, these were not used in this study, and these may have yielded other results.

The Microsoft CoPilot function overall was faster than manual search, but it had the lowest accuracy in identifying neoplastic processes. This tool was dynamic, and the specific neoplasms identified changed with each run. However, it still was not as accurate as the search tool, even after four runs. It also took time to learn how to use this tool, and we found it to be difficult to use. It is also understood, that the purpose of this tool was not to assess the complete capabilities of the AI tool, rather to understand its utility in a pragmatic way, without formal AI training.

Recent literature has described AI-assisted chart review utilizing various forms of artificial intelligence. For example, Laursen et al describe the use of artificial intelligence in identifying hemorrhage during chart review. In this study, participants were evaluated on their ability to identify sentences relevant to hemorrhage both with and without AI assistance in a timed setting. The study found that the physicians participating in the study missed up to 46% of relevant sentences when AI assistance was not used compared to when it was used.[7] Additionally, the study reported that its participants generally regarded AI assistance as an advantage and could see its potential in saving time and resources. Although AI seemed to underperform based on the parameters of our study (understanding that we used the tools to analyze a database with extracted elements, as opposed to using the tool in the actual chart review step), this is not to say it cannot assist with retrospective chart review in the future. With adequate training of the AI model, it could potentially outperform both of the other modalities. An AI model was used to create Gleason grades for prostate cancer biopsies and was trained on a set of 550 biopsy samples and the AI model outperformed 10 of the 15 pathologists that it was compared against.[8] This shows that AI can be very accurate when an appropriately trained model is applied.

The Microsoft Excel search function was found to be easy to use and was considerably faster than manual search. Compared to Microsoft CoPilot, however, this tool is static. It only has the capacity to find specific words and cannot extrapolate to find information outside of these given words. It did, however, have better accuracy than the AI tool in our study. The tool is as helpful as the terms used in the search, so these should be carefully chosen.

In our study, the order of operations was word search, followed by the manual search, followed by Microsoft CoPilot. This may have caused biasing in the results of the study. If the manual search function had been performed first, this would have influenced the terms that were searched within Microsoft Excel (and may have altered the accuracy more in favor of the search function). This approach, however, allowed us to test the validity of the search function without the biasing of the manual review. This order of operations did set us up well for the CoPilot search as it allowed us to more completely train the model for all instances of neoplasia and not just malignancies of the appendix. While CoPilot failed to perform well in both the time and accuracy categories, it may become superior in the near future. With enough training of the model, CoPilot theoretically will be able to search charts rapidly and effectively but needs to be trained on what constitutes neoplasia to be able to appropriately apply to the dataset. We understand that we did not use CoPilot to search the entire charts, rather, the extracted Excel database in our study.

Limitations

There are several possible limitations to this study. For example, manual search time may be considerably impacted by individual factors, including user familiarity with the EMR and user experience with manual search. Manual search time is variable. Similarly, AI "learns" from the user, and its performance may vary based on user familiarity and the feedback given. To account for this "learning process", four runs were completed in this part of our project. Lastly, another limitation of the study was the use of an epic analyst to pull data from the charts. While many institutions do not have this capability, individuals looking to reproduce this study can use tools within epic such as the slicer dicer program to retrieve similar data sets and export them into an Excel spreadsheet for analysis.

It has been described that Microsoft Excel 'macros' can be used to make large data sets more manageable and in a fraction of the time it takes to 'clean' the data manually.[9] With this idea in mind, an area of future study may be to evaluate how accurately and rapidly AI can 'clean' datasets and compare to the already described 'macros' in Excel.

CONCLUSION

Overall, the Microsoft Excel search function and Microsoft CoPilot were less time-consuming than manual search but did not prove as accurate. AI has become a widely accessible tool, and more research should be conducted to determine its role in retrospective chart review. Its adaptability and ability to "learn" from its users is valuable in these processes. However, it requires the user to dedicate time to learning how to use it and may not be as easy to use as other tools. The search tool, on the other hand, was relatively easy to use and found more instances of neoplasia than Microsoft CoPilot. *Funding: No funding sources*

Conflict of interest: None declared
Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. Gastrointest Endosc. 2020;92(4):807-12.
2. Bauzon J, Romero-Velez G, Sehnem L, Shin J, Siperstein A, Jin J. Comparative Analysis of the Accuracy of Microsoft Excel Macros in Retrospective Chart Review Studies. J Surg Res. 2025;311:92-7.
3. Gearing RE, Mian IA, Barber J, Ickowicz A. A methodology for conducting retrospective chart review research in child and adolescent psychiatry. J Can Acad Child Adolesc Psychiatry. 2006;15(3):126-34.
4. Friedman R, Lisk R, Cordero-Bermudez K, Singh S, Ghani S, Gillette BM, et al. Optimizing Chart Review Efficiency in Pressure Injury Evaluation Using ChatGPT. Ann Plast Surg. 2025;94(4S Suppl 2):S327-33.
5. MedCalc Software Ltd. Diagnostic Test Evaluation. Available at: https://www.medcalc.org/en/calc/diagnosic_test.php. Accessed on 04 April 2026.
6. Social Science Statistics. Chi-Square Test Calculator. Available at: https://www.socscistatistics.com/tests/chisquare/calculator/. Accessed on 04 April 2026.

7. Laursen MS, Pedersen JS, Hansen RS, Savarimuthu TR, Lynggaard RB, Vinholt PJ. Doctors Identify Hemorrhage Better during Chart Review when Assisted by Artificial Intelligence. Appl Clin Inform. 2023;14(4):743-51.
8. Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. Lancet Oncol. 2020;21(2):233-41.
9. Bauzon J, Murphy C, Wahi-Gururaj S. Using macros in microsoft excel to facilitate cleaning of research data. J Community Hosp Intern Med Perspect. 2021;11(5):653-7.

Cite this article as: Kreilach A, Franden M, Paranto E, Kowdley G. Accuracy and efficiency of utilizing Microsoft Excel search functions and artificial intelligence for retrospective data analysis. Int Surg J 2026;13:724-8.

\