



Scan to know paper details and
author's profile

Evaluating the Impact of Large Language Model AI on Acute Pancreatitis Management: A Chatgpt-based Investigation

Dylan Gracias, BBiomed, MD, Adrian Siu, BPharm (Hons), MD, MS, Ishith Seth, BBiomed (Hons), MD, MS, Dilshaad Dooreemeah, MBBS, FRACS & Angus Lee, MBBS, FRACS

ABSTRACT

Background: Evidence-based management of acute pancreatitis (AP) is important for patient outcomes. The present study evaluated suggestions by artificial intelligence (AI) chatbot system, ChatGPT, for the management of acute pancreatitis, its alignment with clinical guidelines, and assistance in clinical decision-making.

Methods: Six questions on pancreatitis management were curated by experienced RACS-qualified general surgeons and were put forth to ChatGPT. The chatbot was also asked to provide five high-level evidence references to support each of its responses. Each response was analyzed for its accuracy and comprehensiveness with respect to current internationally recognized guidelines and by two Board-Certified General Surgeons for acute pancreatitis management, as well as for its spelling, grammar, and reference quality. A five-point Likert Scale was utilized to analyze ChatGPT's responses, with scores ranging from 1 (strongly disagree) to 5 (strongly agree). Ten questions were designed to assess accuracy, consistency, informativeness, reliability, and coherence.

Keywords: chatGPT, acute pancreatitis, management, artificial intelligence, large language model.

Classification: NLM Code:W 26.5, WI 753, Q 171

Language: English

LJP Copyright ID: 392886



Great Britain
Journals Press

London Journal of Medical & Health Research

Volume 2024 | Issue 10 | Compilation 1.0



Evaluating the Impact of Large Language Model AI on Acute Pancreatitis Management: A Chatgpt-Based Investigation

Dylan Gracias, BBiomed, MD^α, Adrian Siu, BPharm (Hons), MD, MS^ο,
Ishith Seth, BBiomed (Hons), MD, MS^ρ, Dilshaad Dooreemeah, MBBS, FRACS^ω
& Angus Lee, MBBS, FRACS^{*}

ABSTRACT

Background: Evidence-based management of acute pancreatitis (AP) is important for patient outcomes. The present study evaluated suggestions by artificial intelligence (AI) chatbot system, ChatGPT, for the management of acute pancreatitis, its alignment with clinical guidelines, and assistance in clinical decision-making.

Methods: Six questions on pancreatitis management were curated by experienced RACS-qualified general surgeons and were put forth to ChatGPT. The chatbot was also asked to provide five high-level evidence references to support each of its responses. Each response was analyzed for its accuracy and comprehensiveness with respect to current internationally recognized guidelines and by two Board-Certified General Surgeons for acute pancreatitis management, as well as for its spelling, grammar, and reference quality. A five-point Likert Scale was utilized to analyze ChatGPT's responses, with scores ranging from 1 (strongly disagree) to 5 (strongly agree). Ten questions were designed to assess accuracy, consistency, informativeness, reliability, and coherence. These were independently rated by three junior doctors and two General Surgeons, with any scoring discrepancies resolved through consensus.

Results: ChatGPT successfully adhered to clinical guidelines when generating recommendations for the management of acute pancreatitis. The depth of information remained general and non-specific but was presented in an academic manner with appropriate grammar, spelling and

sentence structure. ChatGPT missed pertinent references, with some being totally fabricated or erroneous.

Conclusion: ChatGPT holds promise for delivering prompt and accessible medical information to non-experts, which may benefit in situations where medical professionals and resources may be scarce or patients are reluctant to seek such services. The inclusion of aberrant or fabricated references is a challenge for researchers and clinicians and breaches academic integrity. Ethically, it is imperative for researchers to exercise prudence when utilizing ChatGPT for research purposes.

Keywords: chatGPT; acute pancreatitis; management; artificial intelligence; large language model.

I. INTRODUCTION

Acute pancreatitis (AP) is an inflammatory disease state of the pancreas, with well-established etiology, diagnostic criteria, and management guidelines. Gallstones and excessive chronic alcohol consumption are the most common causes, contributing to 75%-80% of all cases [1,2]. The Revised Atlanta Classification is an established diagnostic tool for pancreatitis with an emphasis on management dependent on the underlying cause and severity of the disease [3,4].

Chat Generative Pre-trained Transformer (ChatGPT) is a large language model (LLM) which has gained significant popularity in the field of medicine [5]. LLMs are advanced machine learning platforms designed to understand and

generate human-like text based on the models that are trained on [6]. Launched by OpenAI in November 2022, this software has numerous potential applications in clinical environments and research, and the LLM has already shown promise in producing accurate responses to clinical questions on topics like neuropathic pain, cirrhosis, knee and hip osteoarthritis, scaphoid fractures, and radiology reporting [5]. While multiple studies demonstrate its ability to generate high quality scholarly outputs, issues with consistency, specificity, and accuracy in its answers are faults documented commonly, and concerns of sources of information question its ethical use in medicine and research [7,8].

As AI continues to evolve at a rapid pace, its applications in clinical environments have garnered increasing interest. This research article endeavored to evaluate the performance of the ChatGPT-4 model to provide comprehensive, accurate, and coherent responses to inquiries pertaining to acute pancreatitis. The authors posit that the amalgamation of AI and natural language processing models, such as ChatGPT, within the fields of medicine and research holds substantial potential for propelling advancements in these disciplines but clinicians should tread with caution.

II. METHODS

The study assessed ChatGPT-4's ability to provide accurate, extensive and well synthesized answers that were congruent with current evidence-based literature and international guidelines. Furthermore, the quality of references it provided was evaluated. There were no exclusion criteria to ChatGPT's generated answers. No institutional ethics were required for analyzing the artificial chatbot and for this type of study's design (observational case study).

We presented ChatGPT-4 with six questions focused on acute pancreatitis, predominantly based on its cause, diagnosis, and management. The questions were developed and evaluated in conjunction with two Royal Australasian College of Surgeons (RACS) certified General Surgeons (AL and DD) with enormous experience in the

field. Questions aimed to determine the precision and depth of generated answers. Concurrently, three authors (DG, AS, and IS), conducted independent scoping literature searches using PubMed, EMBASE, Scopus, Cochrane CENTRAL and Google Scholar databases to identify high quality and impactful research publications relevant to each question. The outputs the AI provided were analyzed for accuracy against well-established evidence-based research and international guidelines primarily derived from our initial database search, and secondly by two RACS-certified General Surgeons (AL and DD). This combination of evidence-based research and the expert opinions of two general surgeons served as our comparator in analysing ChatGPT's responses. Throughout the questioning, ChatGPT was asked to provide five high level references to support its answer, which was assessed for quality, appropriateness, and existence, and again compared to our initial scoping database searches on the subject. References that were not found through the aforementioned databases and Google, were deemed to be non-existent references.

To enhance the robustness of our study and validate our observations more comprehensively, ChatGPT's responses were subjected to a detailed analysis using a five-point Likert Scale. This scale ranged from "strongly disagree" (1) to "strongly agree" (5), and was meticulously designed to assess various dimensions of ChatGPT's output, including accuracy, consistency, informativeness, reliability, and coherence. To ensure a wide-ranging evaluation, ten meticulously crafted questions targeting these specific aspects were formulated. These questions were independently reviewed and scored by a diverse panel comprising three junior doctors (DG, AS, and IS) and two experienced General Surgeons (AL and DD). This composition of evaluators was strategically chosen to incorporate perspectives ranging from early-career medical professionals to seasoned experts, thereby enriching the assessment with varied levels of clinical expertise and experience. In addition to the Likert Scale evaluation, a qualitative component was integrated into the methodology. Panel members

were encouraged to provide narrative feedback and annotations on each of ChatGPT's responses, aiming to capture nuanced insights beyond the quantitative scores. This qualitative feedback focused on identifying specific strengths and weaknesses in the AI's responses, offering a more granular view of its performance in mimicking clinical reasoning and adhering to medical guidelines. In instances where discrepancies in scoring arose among the panel members, these were not merely noted but actively discussed in structured consensus meetings. These discussions aimed to explore the underlying reasons for differing opinions, fostering a deeper understanding of the evaluation criteria and potentially revealing any biases or interpretational variations.

IV. RESULTS

The first question posed was "In 200 words, what are the causes of acute pancreatitis? Provide 5 high level evidence references" (Figure 1). Overall ChatGPT-4 provided an extensive, factual and coherent answer which correctly identified gallstones and chronic alcohol use to be the two leading causes of acute pancreatitis with incidences similar to current literature [4,9]. Whilst large incidence ranges were given, it is well-known that incidence rates differ quite significantly between geographic regions and socioeconomic strata [10]. The brief pathophysiological description of gallstone pancreatitis was accurate, outlining that the obstructive pathology increases upstream pancreatic duct pressure [3]. In contrast, the pathophysiological description of alcohol-induced pancreatitis did not disclose the lack of a definitively known mechanism that exists and explored just one of several proposed mechanisms of how alcohol contributes to acute pancreatitis. Other common causes of AP were explored including hypertriglyceridemia, medication toxicity and infectious agents, with appropriate examples being given for the latter two [3,10]. A breadth of less common causes was correctly listed, providing a holistic answer within the proposed word limit. In saying this, idiopathic pancreatitis was not listed which makes up a

significant proportion of affected patients, as only 75-85% of etiologies are easily identified [9]. The answer was well synthesized and formatted making for easy viewing. All 5 references provided were high quality and relevant to the causes identified.

is

In 200 words, what are the causes of acute pancreatitis? Provide 5 high level evidence references.



Acute pancreatitis is an inflammatory condition of the pancreas characterized by abdominal pain and elevated pancreatic enzymes. The causes of acute pancreatitis can be divided into various categories:

1. Gallstones: Obstruction of the pancreatic duct by gallstones is one of the most common causes of acute pancreatitis, accounting for 30-60% of cases. Gallstones can lead to increased pressure in the pancreatic duct, causing inflammation and injury to the pancreas.

Reference: Tenner S, Baillie J, DeWitt J, Vege SS. American College of Gastroenterology guideline: management of acute pancreatitis. *Am J Gastroenterol*. 2013 Sep;108(9):1400-15.

2. Alcohol: Alcohol-induced pancreatitis is another major cause, responsible for 25-35% of cases. Chronic heavy alcohol consumption can cause direct toxic injury to pancreatic cells and increase the risk of pancreatitis.

Reference: Lankisch PG, Apte M, Banks PA. Acute pancreatitis. *Lancet*. 2015 Jul 4;386(9988):85-96.

3. Hypertriglyceridemia: High levels of triglycerides in the blood can lead to acute pancreatitis, typically when levels exceed 1000 mg/dL.

Reference: Scherer J, Singh VP, Pitchumoni CS, Yadav D. Issues in hypertriglyceridemic pancreatitis: an update. *J Clin Gastroenterol*. 2014 Mar;48(3):195-203.

4. Medications: Certain medications, such as diuretics, immunosuppressants, and some antiretroviral drugs, can induce pancreatitis as a side effect.

Reference: Nitsche CJ, Jamieson N, Lerch MM, Mayerle JV. Drug induced pancreatitis. *Best Pract Res Clin Gastroenterol*. 2010 Oct;24(5):143-55.

5. Infections: Viral infections (e.g., mumps, coxsackievirus), bacterial infections (e.g., Legionella, Mycoplasma), and parasitic infections (e.g., Ascaris lumbricoides) can cause acute pancreatitis.

Reference: Parenti DM, Steinberg W, Kang P. Infectious causes of acute pancreatitis. *Pancreas*. 1996 Nov;13(4):356-71.

Other less common causes include autoimmune conditions, trauma, metabolic disorders, and genetic factors. It is essential to identify and manage the underlying cause to effectively treat acute pancreatitis.

Figure 1: Prompt 'in 200 Words, What are the Causes of Acute Pancreatitis? Provide 5 High Level Evidence References'


The second question asked to the chatbot was "In 200 words, what is the diagnostic criteria for acute pancreatitis? Provide 5 high level evidence references" (Figure 2). The diagnostic criterion for acute pancreatitis is well-established and acknowledged internationally. ChatGPT was able

to correctly identify the Revised Atlanta Classification model and gave an in-depth explanation of what it entails to reach a diagnosis. Whilst accurate, there was no expansion of the answer to diagnose specific etiologies through investigations such as biliary tree imaging or lipid

profile laboratory investigations, which are important aspects of diagnosis that dictate specific and definitive management principles [10]. Of the 5 references supplementing its

answer, 4 were existent publications appropriate to the topic in question, however the final was a fabricated resource (Figure 2, reference #5, Van Geenen et al).

IS In 200 words, what is the diagnostic criteria for acute pancreatitis? Provide 5 high level evidence references.

 The diagnosis of acute pancreatitis is based on a combination of clinical presentation, laboratory findings, and imaging studies. The Revised Atlanta Classification is the widely accepted diagnostic criteria for acute pancreatitis, which requires at least two of the following three features:

1. Abdominal pain: Acute onset of persistent, severe, and epigastric pain, often radiating to the back, is the most common presenting symptom.
2. Elevated pancreatic enzymes: Serum amylase and/or lipase levels at least three times the upper limit of normal, indicating pancreatic injury or inflammation.
3. Imaging findings: Characteristic findings of acute pancreatitis on abdominal imaging, such as computed tomography (CT), magnetic resonance imaging (MRI), or ultrasound.

References:

1. Banks PA, Bollen TL, Dervenis C, Gooszen HG, Johnson CD, Sarr MG, et al. Classification of acute pancreatitis—2012: revision of the Atlanta classification and definitions by international consensus. *Gut*. 2013 Jan;62(1):102-11.
2. Tenner S, Baillie J, DeWitt J, Vege SS. American College of Gastroenterology guideline: management of acute pancreatitis. *Am J Gastroenterol*. 2013 Sep;108(9):1400-15.
3. Working Group IAP/APA Acute Pancreatitis Guidelines. IAP/APA evidence-based guidelines for the management of acute pancreatitis. *Pancreatology*. 2013 Jul-Aug;13(4 Suppl 2):e1-15.
4. Forsmark CE, Vege SS, Wilcox CM. Acute Pancreatitis. *N Engl J Med*. 2016 Nov 17;375(20):1972-81.
5. Van Geenen EJ, van Santvoort HC, Besselink MG, van der Peet DL, van Erpecum KJ, Fockens P, et al. Lack of consensus on the diagnosis and treatment of acute pancreatitis: A survey amongst upper gastrointestinal surgeons, general surgeons and gastroenterologists. *Pancreatology*. 2012 May-Jun;12(3):266-70.

The diagnostic criteria for acute pancreatitis emphasize the importance of a comprehensive evaluation, including clinical presentation, biochemical markers, and imaging studies, for accurate diagnosis and appropriate management.

Figure 2: Prompt ‘in 200 words, what is the Diagnostic Criteria for Acute Pancreatitis? Provide 5 High Level Evidence References’

The following question was “In 300 words, how is acute pancreatitis managed? Provide 5 high level evidence references” (Figure 3). Given the array of

management strategies that can be employed based on disease severity and particular etiology, we encouraged a more comprehensive answer.

based on disease severity and particular etiology, we encouraged a more comprehensive answer. Once again, a detailed answer was provided with factual information primarily outlining the supportive care routinely applicable. This included fluid resuscitation, pain management and nutritional support which is the mainstay treatment for all cases, and often for those with mild disease [4,10]. Notably, early enteral feeding was mentioned but not stated as the standard of care which is not consistent with current guidelines [11]. The use of antibiotics was qualified in ChatGPT's response, as they are only indicated in infected pancreatic necrosis or extra pancreatic infections [12]. Intensive care involvement was correctly identified as an avenue in severe cases especially with organ failure [13].

Disappointingly, management directed at specific etiologies and complications were poorly answered, being vague and superficial. This aspect of the answer alluded to "endoscopic, radiological and surgical interventions," but no examples such as ERCP, percutaneous drainage, cholecystectomy nor their indications were given. Although five high-quality references were provided, some proved to be irrelevant or contradictory. For instance, one reference by Mounzer et al [31] was associated with the pain management point but ChatGPT lacked evidence directly related to pain management strategies. Another reference by Bakker et al [32] contradicted the nutritional support recommendation offered by ChatGPT, thereby undermining the reliability of the cited references.

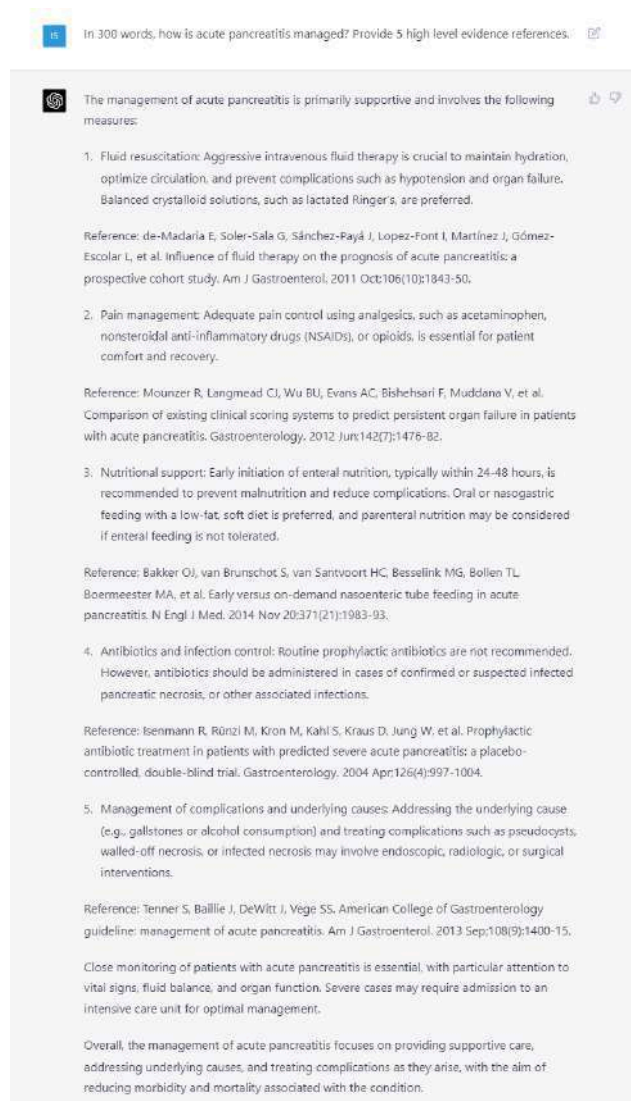


Figure 3: Prompt 'in 300 words: How is Acute Pancreatitis Managed? Provide 5 High Level Evidence References

The next question extended on the previous, asking more specifically “In 200 words, what’s the management for necrotizing pancreatitis? Provide 5 high level evidence references” (Figure 4).

ChatGPT displayed its answer in a similar format to the previous and in doing so was largely repetitive in the information it provided. It failed to clearly discriminate between sterile versus infected pancreatitis, which have different management approaches as well as complication and mortality rates (5-10% versus 20-30%, respectively) [10,14]. Fluid resuscitation, nutritional support and intensive care strategies once again were cornerstones in their answers; however, pain management was excluded.

Prophylactic antibiotic therapy and radiological/endoscopic/surgical drainage was correctly established for infected necrotizing pancreatitis management [14]. However, this was not accompanied by specific management principles of sterile pancreatitis, which generally consists of supportive treatment and delayed necrosectomy in symptomatic cases. Its last point was additionally futile given acute pancreatitis, regardless of necrosis, is often managed with multidisciplinary collaboration between general surgeons, gastroenterologists, intensivists, pathologists, radiologists and dieticians [15]. As such, this doesn’t necessarily represent a key management principle of necrotizing pancreatitis. Two of the five references were shared with the previous question. One reference was fabricated (Figure 4, reference #1, Mayerle et al), whilst the remaining four were high quality publications related to the question.



In 200 words, what's the management for necrotising pancreatitis? Provide 5 high level evidence references.



The management of necrotizing pancreatitis focuses on supportive care, infection control, and intervention for complications when necessary. Key aspects include:



1. Fluid resuscitation and organ support: Aggressive intravenous fluid therapy is crucial, and organ support, such as mechanical ventilation or renal replacement therapy, may be required in severe cases.

Reference: Mayerle J, Dummer A, Sessler M, Malla SR, van den Brandt C, Teller S, et al. Intra-abdominal pressure during fluid resuscitation is a sensitive end point for fluid management in acute pancreatitis. *Ann Surg*. 2020 May;271(5):942-51.

2. Nutritional support: Enteral nutrition, preferably via a nasogastric or nasoenteric route, should be initiated early to prevent malnutrition and reduce complications.

Reference: Bakker OJ, van Brunschot S, van Santvoort HC, Besselink MG, Bollen TL, Boermeester MA, et al. Early versus on-demand nasoenteric tube feeding in acute pancreatitis. *N Engl J Med*. 2014 Nov 20;371(21):1983-93.

3. Antibiotics and infection control: Prophylactic antibiotics are not routinely recommended; however, they should be administered if infected pancreatic necrosis or other associated infections are suspected or confirmed.

Reference: Isenmann R, Rünzi M, Kron M, Kahl S, Kraus D, Jung W, et al. Prophylactic antibiotic treatment in patients with predicted severe acute pancreatitis: a placebo-controlled, double-blind trial. *Gastroenterology*. 2004 Apr;126(4):997-1004.

4. Minimally invasive intervention: In cases of infected pancreatic necrosis, a step-up approach, including percutaneous or endoscopic drainage followed by minimally invasive necrosectomy if needed, is preferred over open necrosectomy.

Reference: van Santvoort HC, Besselink MG, Bakker OJ, Hofker HS, Boermeester MA, Dejong CH, et al. A step-up approach or open necrosectomy for necrotizing pancreatitis. *N Engl J Med*. 2010 Apr 22;362(16):1491-502.

5. Multidisciplinary management: Close collaboration with a multidisciplinary team, including gastroenterologists, surgeons, interventional radiologists, and critical care specialists, is essential for optimal patient outcomes.

Reference: Vege SS, DiMaggio MJ, Forsmark CE, Martel M, Barkun AN. Initial medical treatment of acute pancreatitis: American Gastroenterological Association Institute technical review. *Gastroenterology*. 2018 Apr;154(4):1103-39.

Figure 4: Prompt 'in 200 words, What's the Management for Necrotizing Pancreatitis? Provide 5 High Level Evidence References'

The fifth question was designed with more what is the optimal timing of laparoscopic surgical focus, asking ChatGPT, "In 200 words, cholecystectomy for gallstone pancreatitis?"

Provide 5 high level evidence references" (Figure 5). ChatGPT correctly identified the decision around surgery timing depends on the severity of pancreatitis. While a complex decision, ChatGPT correctly indicated early laparoscopic cholecystectomy (ELC) in mild disease, versus delayed laparoscopic cholecystectomy (DLC) in moderate to severe disease and provided timelines for management. Although no consensus exists on the definition of "early," ELC within 72 hours as

recommended by ChatGPT, is evidenced by numerous studies such as the PONCHO trial [16]. ChatGPT also synthesized the underlying reasons behind these guidelines, including reducing recurrence, complication, morbidity and mortality rates [17]. Five true and high-quality publications were referenced, including a key systematic review by van Baal et al, and a publication on the PONCHO trial by da Costa et al.

In 200 words, what is the optimal timing of laparoscopic cholecystectomy for gallstone pancreatitis? Provide 5 high level evidence references.

The optimal timing of laparoscopic cholecystectomy for gallstone pancreatitis depends on the severity of the pancreatitis. Current evidence suggests the following:

1. Mild gallstone pancreatitis: Early laparoscopic cholecystectomy, preferably within 72 hours of symptom onset or during the index hospitalization, is recommended to reduce the risk of recurrent pancreatitis and other gallstone-related complications.
Reference: Aboulian A, Chan T, Yaghoobian A, Kaji AH, Putnam B, Neville A, et al. Early cholecystectomy safely decreases hospital stay in patients with mild gallstone pancreatitis: a randomized prospective study. *Ann Surg.* 2010 Apr;251(4):615-9.
2. Moderate to severe gallstone pancreatitis: In more severe cases, cholecystectomy is typically delayed until clinical and laboratory parameters have improved, and local complications, such as pancreatic necrosis, have stabilized. This often occurs within 2 to 4 weeks of symptom onset.
Reference: van Baal MC, Besselink MG, Bakker OJ, van Santvoort HC, Schaapherder AF, Nieuwenhuijs VB, et al. Timing of cholecystectomy after mild biliary pancreatitis: a systematic review. *Ann Surg.* 2012 May;255(5):860-6.
Early cholecystectomy in mild cases and a more conservative approach in moderate to severe cases have been shown to reduce morbidity, mortality, and the risk of recurrent pancreatitis.
Reference: Tenner S, Baillie J, DeWitt J, Vege SS. American College of Gastroenterology guideline: management of acute pancreatitis. *Am J Gastroenterol.* 2013 Sep;108(9):1400-15.
Reference: Ito K, Ito H, Whang EE. Timing of cholecystectomy for biliary pancreatitis: do the data support current guidelines? *J Gastrointest Surg.* 2008 Oct;12(12):2164-70.
Reference: Da Costa DW, Bouwense SA, Schepers NJ, Besselink MG, van Santvoort HC, van Brunschot S, et al. Same-admission versus interval cholecystectomy for mild gallstone pancreatitis (PONCHO): a multicentre randomised controlled trial. *Lancet.* 2015 Oct 17;386(10000):1261-8.

Figure 5: Prompt 'in 200 words, what is the Optimal timing of Laparoscopic Cholecystectomy for Gallstone Pancreatitis? Provide 5 high Level Evidence References'

The final question requested was “In 200 words, provide future directions of research for the management of acute pancreatitis? Provide 5 high level evidence references” (Figure 6). ChatGPT accurately identified the significance of biomarkers related to inflammatory pathophysiology and targeted therapies [18] but responses were not new ideas or very timely. While nutritional support was discussed, the cited reference was outdated [44], and more recent

findings on parenteral nutrition and immuno-enhanced nutrients or probiotics were omitted [19]. The suggestion to investigate genetics in acute pancreatitis was pertinent, but the accompanying reference was not only outdated, but also more tailored to chronic pancreatitis [43]. Overall, ChatGPT's response was adequate, touching on several key research areas, but only three of five references were high-quality and relevant.

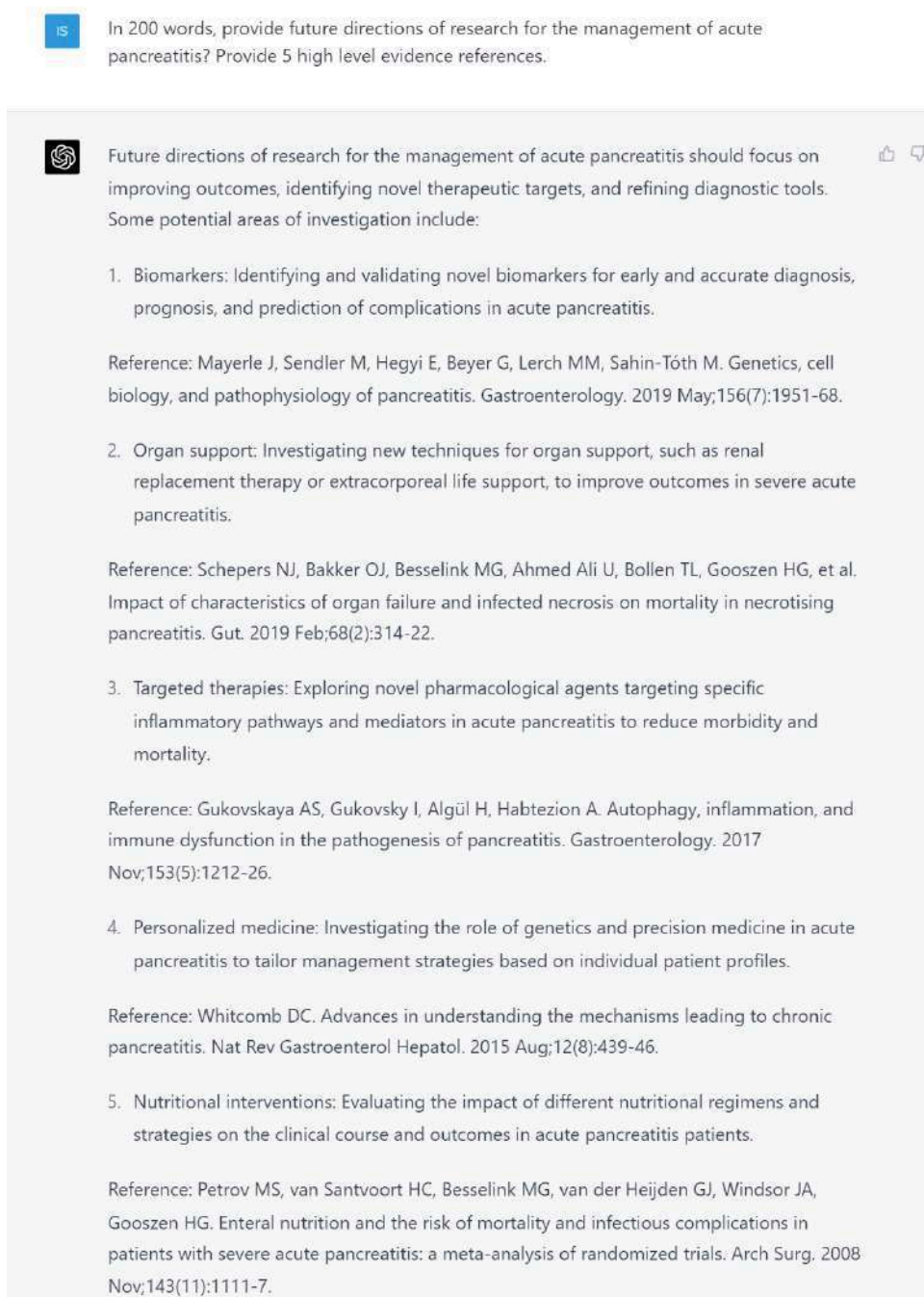


Figure 6: Prompt ‘in 200 words, Provide Future Directions of Research for the Management of Acute Pancreatitis? Provide 5 High Level Evidence References

In performing a subjective qualitative assessment of ChatGPT's abilities using a Likert Scale (Table 1) we were able to unify its attributes and summarize the overall performance of ChatGPT in this study. Our findings demonstrated scattered scores for desirable characteristics of the Chatbot. ChatGPT's ability to comprehend complex questions and provide accurate answers received scores of 4, and its ability to synthesize its

answers coherently and utilise academic terminology received scores of 5. We were indifferent to its ability to provide in-depth and wholistic information, hence these attributes were given scores of 3. Concerningly, questions directed to its referencing abilities received scores of 2. A high score of 5 was given for its general medical knowledge and 4 for its specific medical knowledge.

Table 1: Evaluation of ChatGPT's Responses using a Likert Scale

	Strongly Disagree-1	Disagree-2	Neither Agree or Disagree	Agree -4	Strongly Agree - 5
ChatGPT provides accurate answers consistently when answering questions				X	
Chat CPT can Comprehend complex questions and provide appropriately directed answers accordingly				X	
ChatCPT Provides in-depth information when answering questions			X		
ChatGPT provides a wholistic breadth information when answering question			X		
ChatGPT provides true citations consistently when prompted		X			
ChatGPT provides high quality citations consistently when prompted		X			
ChatCPT produces coherent and well synthesized answers					X
ChatCPT uses academic medical terminology appropriately when answering questions					X
ChatCPT is a valuable source of general medical knowledge					X
ChatCPT is a valuable source of specific medical knowledge				X	

V. DISCUSSION

ChatGPT consistently performed excellently when generating responses to questions on the management of acute pancreatitis. As other studies have noted, the chatbot has the potential to assist within clinical environments as an adjunctive management tool, or for patients to utilize at home for their self-education. Further to its application here, there is convincing potential to adopt its use more confidently and safely in information gathering for both patients and clinicians, as well as in the research domain. Specifically, the chatbot consistently constructed accurate, easily understandable sound answers, with minimal pitfalls, which would rival the communication of many doctor-patient conversations where medical jargon is a barrier to

patient understanding. In addition, ChatGPT often derived its answers from high quality medical publications on request, which may enhance doctor confidence when utilising AI. Currently, the use of LLMs may improve patient outcomes in remote and rural regions or emergencies where prompt diagnosis and management in the context of resource scarcity would optimize patient safety.

We hypothesize that multiple factors contribute to the quality disparity between this study and prior ChatGPT research. Our study used the latest ChatGPT version with enhancements by OpenAI, potentially improving learning, data collection, and answer quality. Perhaps, with acute pancreatitis having well-studied and established guidelines in general surgery, the likelihood of

ChatGPT providing aberrant information is minimized, compared with more nuanced and evolving topics. The relative consistency in the literature for pancreatitis would presumably allow for greater accuracy, consistency in responses and basis for choosing references. Despite this, investigations into ChatGPT-4 for more poorly understood pathologies are yet to be undertaken, and similar improvements could be observed in the advent of the new version.

Although the upsides are clear, ChatGPT has previously shown inadequate consistency and accuracy in information gathering depicted across an array of topics that preclude its use in clinical medicine. A recent systematic review explored ChatGPT's strengths and limitations for its use in healthcare education, research, and practice, finding numerous strengths which were also accompanied by concerns in 58/60 (96.7%) of its analysed records. This encompassed ethical, copyright, legal, risk of bias, plagiarism, inaccuracy with risk of hallucination, limited knowledge and incorrect citation concerns [20]. Herein, the generation of irrelevant and non-existent citations was infrequent but none-the-less present in this study, as has been problematic previously [21]. So far, ChatGPT has been restricted to certain areas of medicine and this is the first to investigate its use with acute pancreatitis being the clinical presentation.

Despite promising results in our study, it would be negligent not to identify the pitfalls of ChatGPT which could lead to disastrous consequences in clinical practice, research and education. For example, the occasional provision of outdated, inaccurate, or superficial information highlights the need for further model refinement and updated medical literature training before such a tool is confidently used for information gathering and synthesis. Specifically, the generation of fabricated or irrelevant references, despite infrequent, raises great concern and can be fatal in the real clinical setting. This proves to be a drastic issue yet to be resolved by the developers, which questions the reliability, consistency and evidence base of answers produced by the Chatbot. In the field of medicine, where clinical decisions are derived from foundational evidence,

this is very problematic, and a major contributor to inaccurate information and knowledge gaps that ChatGPT may produce. To maintain scientific integrity, a necessary improvement of ChatGPT lies in enhancing its citation detection capabilities through vigorous cross checking with reputable peer-reviewed databases such as what we have done through our methodology. Particularly, artificial intelligence systems need to be created to avoid the common "plausible combination" type references whereby a combination of vaguely relevant sources will be summated to produce a false reference. This remains the most worrying aspect of ChatGPT, with additional academic integrity, plagiarism and ethical concerns stemming from this consistent downfall. It should also be noted that ChatGPT can only gather data prior to September 2021, and thereby may miss emerging research or guidelines. ChatGPT's current knowledge limitations and inability to consider individual patient factors also flies in the face of modern medicine's best attempt at attending to each patient as individuals, and the multidisciplinary team who take center stage in many walks of medicine. How ChatGPT will fit into these paradigms of clinical practice will be interesting in the coming years. If these concerns can be addressed, the application of large language models in clinical practice could drive transformative changes in healthcare.

For its use in clinical medicine, ChatGPT needs to be able to reliably deliver sound information based on high quality evidence-based literature. Currently, concerns surrounding its data collection and reliability of answers raise hesitation to the medical community to confidently adopt it into their practice. As a result, it is of vital importance that all users, especially medical practitioners, exercise caution when using these LLMs and understand their limitations.

Lateral to ChatGPT's potential in clinical practice, is its use in medical education. The promising but not completely refined results that this study and others have shown, opens the door for using ChatGPT as a tool to educate medical students and doctors especially on well-established and well understood common pathologies. Harnessing

this as a modern education technology can help overview, summarise and consolidate key clinical learning points for different pathologies as well as the basic scientific understanding behind clinical practices. Once again, limitations would be shared with its use in clinical practice, and therefore not only should be used cautiously, but may have the potential to be abused through plagiarism.

Our study had multiple limitations that should be acknowledged to guide future research in this field. Firstly, our study design did not analyse any temporal data in the form of re-asking the same question to the Chatbot at a different point in time to compare generated answers over time. This would be valuable as the Chatbot continues to develop. Additionally, word limits were generally kept to 200 words which may restrict the quality and depth of answer provided by ChatGPT. This decision was made to encourage a response that was precise with enough room to briefly expand on its answer, which we believe is reflective of how doctors would primarily utilise Chatbots.

Finally, questions were generally open-ended which may impact the Chatbot's interpretation and therefore answer. These limitations encourage the need for further evaluation of LLMs as they continue to improve prior to integration to healthcare. Thus, future research should encompass the evaluation of specific and targeted questions and do so at different time points as the technology develops. Strategies like cross-verification with current medical databases to mitigate poor quality referencing would be insightful to ensure AI's safe integration into clinical decision-making. We focused on evaluating AI responses qualitatively against clinical guidelines, which begs the need for complimentary quantitative analysis.

VI. CONCLUSION

ChatGPT had excellent responses to clinical questions regarding acute pancreatitis management. While ChatGPT shows promise in supporting clinical decision-making, it is essential to ensure its responsible integration into clinical practice through continuous refinement, up-to-date training, and collaboration with

healthcare professionals. Future research should continue to evaluate different aspects of LLMs potential and applications in medicine as they continue to be refined and perform statistical analysis of their utility and limitations. Furthermore, wholistic future research should focus on evaluating the long-term effects of AI chatbots on patient outcomes, satisfaction, and healthcare resource utilization in various clinical settings.

Statements and Declarations:

ACKNOWLEDGEMENTS

None

Conflict of Interest

The authors declare no Conflict of Interest for this article.

Financial Disclosure and Products

No authors have received any funding or support

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

1. Cappell MS: Acute pancreatitis: etiology, clinical presentation, diagnosis, and therapy. *Medical Clinics of North America*. 2008, 92: 889-923. DOI: 10.1016/j.mcna.2008.04.013
2. Nesvaderani M, Eslick GD, Cox MR: Acute pancreatitis: update on management. *The Medical Journal of Australia*. 2015, 202: 420-423. DOI: 10.5694/mja14.01333.
3. Mederos MA, Reber HA, Girgis MD: Acute pancreatitis: a review. *JAMA*. 2021, 325: 382-390. DOI: 10.1001/jama.2020.20317.
4. Van Dijk SM, Hallensleben ND, van Santvoort HC, et al.: Acute pancreatitis: recent advances through randomised trials. *Gut*. 2017, 66:2024-2032. DOI:10.1136/gutjnl-2016-313595
5. Seth I, Rodwell A, Tso R, Valles J, Bulloch G, Seth N: A Conversation with an Open Artificial Intelligence Platform on Osteoarthritis of the

- Hip and Treatment. *Journal of Orthopedics and Sports Medicine*. 2023, 5:112-120.
6. Wu Y, Jiang AQ, Li W, et al.: Autoformalization with large language models. *Advances in Neural Information Processing Systems*. 2022, 35:32353-32368.
 7. Lund BD, Wang T: Chatting about ChatGPT: how may AI and GPT impact academia and libraries? *Library Hi Tech News*. 2023.
 8. Liebrez M, Schleifer R, Buadze A, Bhugra D, Smith A: Generating scholarly content with ChatGPT: ethical challenges for medical publishing. *The Lancet Digital Health*. 2023, 5:e105-e106. Doi:10.1016/S2589-7500(23)0019-5
 9. Wang G-J, Gao C-F, Wei D, Wang C, Ding S-Q: Acute pancreatitis: etiology and common pathogenesis. *World journal of gastroenterology: WJG*. 2009, 15:1427. DOI: 10.3748/wjg.15.1427.
 10. Gapp J, Chandra S: Continuing Education Activity. StatPearls Publishing; 2023.
 11. Oláh A, Romics Jr L: Enteral nutrition in acute pancreatitis: a review of the current evidence. *World Journal of Gastroenterology: WJG*. 2014, 20:16123. DOI:10.3748/wjg.v20.i43.16123
 12. Tenner S, Baillie J, DeWitt J, Vege SS: American College of Gastroenterology guideline: management of acute pancreatitis. *Official journal of the American College of Gastroenterology| ACG*. 2013, 108:1400-1415. DOI: 10.1038/ajg.2013.218
 13. Banks PA, Freeman ML, Gastroenterology PPCotACo: Practice guidelines in acute pancreatitis. *Official journal of the American College of Gastroenterology| ACG*. 2006, 101:2379-2400. DOI:10.1111/j.1572-0241.2006.00856.x
 14. Boumitri C, Brown E, Kahaleh M: Necrotizing Pancreatitis: Current Management and Therapies. *Clin Endosc*. 2017, 50:357-365. DOI: 10.5946/ce.2016.152
 15. Haj-Mirzaian A, Patel BN, Fishman EK, Zaheer A: Value of multidisciplinary collaboration in acute and chronic pancreatitis. *Abdominal Radiology*. 2020, 45:1458-1467. DOI: 10.1007/s00261-019-02320-9
 16. Da Costa DW, Bouwense SA, Schepers NJ, et al.: Same-admission versus interval cholecystectomy for mild gallstone pancreatitis (PONCHO): a multicentre randomised controlled trial. *The Lancet*. 2015, 386:1261-1268. DOI:10.1016/S0140-6736(15)00274-3
 17. Nealon WH, Bawduniak J, Walser EM: Appropriate timing of cholecystectomy in patients who present with moderate to severe gallstone-associated acute pancreatitis with peripancreatic fluid collections. *Annals of surgery*. 2004, 239:741. DOI: 10.1097/01.sla.0000128688.97556.94
 18. Silva-Vaz P, Abrantes AM, Castelo-Branco M, Gouveia A, Botelho MF, Tralhão JG: Multifactorial scores and biomarkers of prognosis of acute pancreatitis: applications to research and practice. *International journal of molecular sciences*. 2020, 21:338. DOI: 10.3390/ijms21010338
 19. Liu Y, Wan Z, Liao D: Efficacy of enteral nutrition for patients with acute pancreatitis: A systematic review and meta-analysis of 17 studies. *Experimental and Therapeutic Medicine*. 2023, 25:1-12. Doi:10.3892/etm.2023.11883
 20. Sallam M: ChatGPT Utility in Health Care Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. In *Healthcare*. Volume 11. MDPI; 2023:887. <https://doi.org/10.3390/healthcare11060887>
 21. Chen T-J: ChatGPT and other artificial intelligence applications speed up scientific writing. *Journal of the Chinese Medical Association*. 2023:10.1097. DOI: 10.1097/JCMA.0000000000000900
 22. Tenner S, Baillie J, DeWitt J, Vege SS: American College of Gastroenterology guideline: management of acute pancreatitis. *Am J Gastroenterol*. 2013, 108:1400-1415; 1416. 10.1038/ajg.2013.218
 23. Lankisch PG, Apte M, Banks PA: Acute pancreatitis. *Lancet*. 2015, 386:85-96. 10.1016/s0140-6736(14)60649-8
 24. Scherer J, Singh VP, Pitchumoni CS, Yadav D: Issues in hypertriglyceridemic pancreatitis: an update. *J Clin Gastroenterol*. 2014, 48:

- 195-203. 10.1097/01.mcg.0000436438.60145.5a.
25. Nitsche CJ, Jamieson N, Lerch MM, Mayerle JV: Drug induced pancreatitis. *Best Pract Res Clin Gastroenterol.* 2010, 24:143-155. 10.1016/j.bpg.2010.02.002
 26. Parenti DM, Steinberg W, Kang P: Infectious causes of acute pancreatitis. *Pancreas.* 1996, 13:356-371. 10.1097/00006676-199611000-00005.
 27. Banks PA, Bollen TL, Dervenis C, Gooszen HG, Johnson CD, Sarr MG, et al. Classification of acute pancreatitis-2012: revision of the Atlanta classification and definitions by international consensus. *Gut.* 2013 Jan; 62(1): 102-11.
 28. IAP/APA evidence-based guidelines for the management of acute pancreatitis. *Pancreatology.* 2013, 13:e1-15. 10.1016/j.pan.2013.07.063
 29. Forsmark CE, Vege SS, Wilcox CM: Acute Pancreatitis. *N Engl J Med.* 2016, 375: 1972-1981. 10.1056/NEJMra1505202.
 30. de-Madaria E, Soler-Sala G, Sánchez-Payá J, et al.: Influence of fluid therapy on the prognosis of acute pancreatitis: a prospective cohort study. *Am J Gastroenterol.* 2011, 106:1843-1850. 10.1038/ajg.2011.236
 31. Mounzer R, Langmead CJ, Wu BU, et al.: Comparison of existing clinical scoring systems to predict persistent organ failure in patients with acute pancreatitis. *Gastroenterology.* 2012, 142:1476-1482; quiz e1415-1476. 10.1053/j.gastro.2012.03.005
 32. Bakker OJ, van Brunschot S, van Santvoort HC, et al.: Early versus on-demand nasoenteric tube feeding in acute pancreatitis. *N Engl J Med.* 2014, 371:1983-1993. 10.1056/NEJMoa1404393
 33. Isenmann R, Rünzi M, Kron M, Kahl S, Kraus D, Jung W, et al. Prophylactic antibiotic treatment in patients with predicted severe acute pancreatitis: a placebo-controlled, double-blind trial. *Gastroenterology.* 2004 Apr; 126(4):997-1004.
 34. van Santvoort HC, Besselink MG, Bakker OJ, et al.: A step-up approach or open necrosectomy for necrotizing pancreatitis. *N Engl J Med.* 2010, 362:1491-1502. 10.1056/NEJMoa0908821.
 35. Vege SS, DiMagno MJ, Forsmark CE, Martel M, Barkun AN: Initial Medical Treatment of Acute Pancreatitis: American Gastroenterological Association Institute Technical Review. *Gastroenterology.* 2018, 154:1103-1139. 10.1053/j.gastro.2018.01.031
 36. Aboulian A, Chan T, Yaghoobian A, et al.: Early cholecystectomy safely decreases hospital stay in patients with mild gallstone pancreatitis: a randomized prospective study. *Ann Surg.* 2010, 251:615-619. 10.1097/SLA.ob013e3181c38f1f
 37. Van Baal MC, Besselink MG, Bakker OJ, et al.: Timing of cholecystectomy after mild biliary pancreatitis: a systematic review. *Ann Surg.* 2012, 255:860-866. 10.1097/SLA.ob013e3182507646.
 38. Ito K, Ito H, Whang EE: Timing of cholecystectomy for biliary pancreatitis: do the data support current guidelines? *J Gastrointest Surg.* 2008, 12:2164-2170. 10.1007/s11605-008-0603-y
 39. Da Costa DW, Bouwense SA, Schepers NJ, et al.: Same-admission versus interval cholecystectomy for mild gallstone pancreatitis (PONCHO): a multicentre randomised controlled trial. *The Lancet.* 2015, 386:1261-1268.
 40. Mayerle J, Sendler M, Hegyi E, Beyer G, Lerch MM, Sahin-Tóth M: Genetics, Cell Biology, and Pathophysiology of Pancreatitis. *Gastroenterology.* 2019, 156:1951-1968.e1951. 10.1053/j.gastro.2018.11.081
 41. Schepers NJ, Bakker OJ, Besselink MG, et al.: Impact of characteristics of organ failure and infected necrosis on mortality in necrotising pancreatitis. *Gut.* 2019, 68:1044-1051. 10.1136/gutjnl-2017-314657
 42. Gukovskaya AS, Gukovsky I, Algül H, Habtezion A: Autophagy, inflammation, and immune dysfunction in the pathogenesis of pancreatitis. *Gastroenterology.* 2017, 153:1212-1226.
 43. Whitcomb DC: Mechanisms of disease: Advances in understanding the mechanisms leading to chronic pancreatitis. *Nat Clin Pract Gastroenterol Hepatol.* 2004, 1:46-52. 10.1038/ncpgasthep0025.

44. Petrov MS, Loveday BP, Pylypchuk RD, McIlroy K, Phillips AR, Windsor JA: Systematic review and meta-analysis of enteral nutrition formulations in acute pancreatitis. *Br J Surg.* 2009, 96:1243-1252. 10.1002/bjs.6862.