RESEARCH

The performance of ChatGPT in day surgery and pre-anesthesia risk assessment: a case-control study of 150 simulated patient presentations

Tingting Cheng[1†], Yu Li[2†], Jiaqiu Gu[3†], Yibo He[3], Guangbao He[3], Peipei Zhou[4], Shuyun Li[3], Hang Xu[3], Yang Bao[3*] and Xuejun Wang[5*]

Abstract

Background Day surgery has developed rapidly in China in recent years, although it still faces a shortage of anesthesiologists to handle pre-anesthesia routine before surgery. We hypothesized that ChatGPT may assist anesthesia practitioners in preoperative assessment and answer questions on the concerns of patients. The aims of this study were to examine the ability of ChatGPT to assess preoperative risk and determine its accuracy in answering questions regarding knowledge and management of day surgery anesthesia.

Methods One-hundred fifty patient profiles were generated to simulate day surgery patient presentations that involved complications of varying acuity and severity. The ChatGPT group and the expert group were both required to evaluate the profiles of 150 simulated patients to determine their ASA-PS classification and whether day surgery was recommended. ChatGPT was then asked to answer 131 questions about day surgery anesthesia that represented the most common issues encountered in clinical practice. The performance of ChatGPT was assessed and graded independently by two experienced anesthesiologists.

Results A total of 150 patient profiles were included in the study (75 males [50.0%] and 75 females [50.0%]). There was no difference between the ChatGPT group and the expert group for the ASA-PS classification and assessment of anesthesia risk in the patient profiles ($P > 0.05$). Regarding recommendation for day surgery in patients with certain comorbidities (ASA ≥ II), the expert group was inclined to require further examination or treatment. In addition, the proportion of conclusions made by ChatGPT was smaller than that of the experts (i.e., ChatGPT *n* (%) vs. expert *n* (%): day surgery can be performed, 67 (47.9) vs. 31 (25.4); needs further treatment and evaluation, 56 (37.3) vs. 66 (44.0); and day surgery is not recommended, 18 (12.9) vs. 29 (9.3), $P < 0.05$). We showed that ChatGPT had extensive knowledge related to day surgery anesthesia (94.0% correct), with most of the points (70%) considered comprehensive. The performance of ChatGPT was also better in the domains of peri-anesthesia concerns, lifestyle, and emotional support.

---

†Tingting Cheng, Yu Li, and Jiaqiu Gu share co-first authorship.

*Correspondence:
Yang Bao
jody_109123@126.com
Xuejun Wang
wangxuejunhsz9111@163.com
Full list of author information is available at the end of the article

Cheng *et al. Perioperative Medicine*   (2024) 13:111

Page 2 of 9

Conclusions ChatGPT can assist anesthesia practitioners and surgeons by alerting them to the ASA-PS classification and assessing perioperative risk in day surgery patients. ChatGPT can also be trusted to answer questions and concerns related to pre-anesthesia and therefore has the potential to provide important assistance in clinical work.

Keywords ChatGPT, Pre-anesthesia, Risk assessment, ASA-PS classification, Peri-anesthesia concerns, Day surgery, Day-case surgery

Introduction

Day-case surgery involves the management of a patient who is scheduled to be admitted and discharged on the same day (Bailey et al. 2019). This process involves admitting patients for investigation or an operation on a planned, nonresidential basis, with the provision of adequate facilities for recovery in a ward or separate unit (Goodwin et al. 1992). Historically, patients were typically hospitalized for surgical procedures and remained there until they regained independence, were able to walk on their own, and had their stitches removed. This process is attributed to a lack of comprehensive healthcare within the community, suboptimal home environments for patient care, a reduction in incomplete wound healing, and elevated rates of anesthetic and surgical complications (Ojo et al. 2010). The benefits of early mobilization after an operation are well recognized, while minimally invasive surgery is well established, resulting in more procedures being performed as day surgery (Albornoz. et al. 2011).

Several factors should be taken into account when considering the characteristics of the patient and selection of the procedure suitable for daycare surgery (Albornoz. et al. 2011). Firstly, the procedure carries a minimal risk of severe postoperative complications. Secondly, postoperative symptoms can typically be managed using oral medications or local anesthesia techniques. Thirdly, patients should be capable of independent movement prior to their discharge. Day surgery therefore facilitates a more rapid recovery, imposes less disruption on patients' lives, and also diminishes the risk of hospital-acquired infections.

Pre-anesthetic assessment is carried out to determine whether the planned procedure is appropriate for the patient and has an acceptably low risk for perioperative complications. The assessment of anesthesia risk prior to surgery is essential to identify and exclude patients with medium- to high-risk profiles who are considered unsuitable for day-case procedures in outpatient clinics, emergency departments, or primary care settings. Of these evaluations, the American Society of Anesthesiologists-Physical Status (ASA-PS) classification is the most basic and effective (Voney et al. 2007). The ASA-PS classification is a suitable index for assessing the physical status of surgical patients and predicting adverse events during surgical anesthesia, such as the length of stay in the operating room, duration of anesthesia and surgery, and blood loss and infections, with these factors resulting in higher ASA-PS classification (Ansell et al. 2004). The increasing number of patients requiring day-case surgery and anesthesia poses a growing challenge of ensuring adequate preoperative assessment. The large population of China has resulted in a substantial demand for surgeries, although this is hindered by a shortage of anesthesiologists.

Large language models (LLM) are sophisticated artificial intelligence (AI) systems that utilize deep neural networks to learn and generate human-like natural language through training on extensive text datasets (Thirunavukarasu et al. 2023). The Chat Generative Pre-trained Transformer (ChatGPT) was developed in recent years and is considered to have the ability to help doctors relieve the pressure of preoperative assessments. ChatGPT is an advanced natural language processing (NLP) model pioneered by OpenAI (Ali et al. 2023). The model exhibits a human-like capacity for generating text, especially in the domain of chatbot dialogues. Many medical professionals are already investigating the possibility of using ChatGPT in clinical work (Grünebaum et al. 2023; Lahat et al. 2023). However, to date there have been no extensive studies that have evaluated the effectiveness of ChatGPT in answering questions accurately and holistically about day-case surgery based on preoperative assessments.

The day surgery procedures performed at our medical center are typically straightforward and uncomplicated, although patients with more complex conditions require a greater allocation of labor costs. The aim of the current study was to use ChatGPT to assess the physical status of simulated patients prior to surgery and to examine the dependability and precision of ChatGPT in the preoperative evaluation of anesthesia. The precision, comprehensiveness, and consistency of ChatGPT's responses to frequently asked questions about treatment and patient care prior to anesthesia were also assessed.

Methods

OpenAI has released ChatGPT-4. The current study used the ChatGPT-4 online interface (ChatGPT Version 4.0), a model trained by OpenAI. ChatGPT was questioned on

Cheng *et al. Perioperative Medicine*   (2024) 13:111

Page 3 of 9

the patient's profiles, including medical history, physical examination, current vital signs, and the results of diagnostic tests. A panel of experienced anesthesiologists then evaluated the correctness and accuracy of ChatGPT's responses, with the panel consisting of three anesthesiologists, all of whom had over 10 years of clinical experience.

Ability of ChatGPT for preoperative risk assessment

Hypothetical and standardized patient profiles were generated in order to simulate day surgery patient presentations for each of the defined complications according to the checklist components of the process-oriented score (PRO-score) for risk evaluation of adults (Vogelsang et al. 2020). The score is based on national and international guidelines for preoperative risk assessment of adult patients and contains current guidelines and the assessment of vital signs (Vogelsang et al. 2020; Kristensen et al. 2014). The contents are expressed as 234 single parameters, with these items sorted by organ systems and following the airway (A), breathing (B), circulation (C), disability (D), and endocrinology (E) scheme. Overall, a total of 150 patient profiles and presentations were generated for the purpose of this study (Supplementary Table 1).

Based on the patient's information such as history, symptoms, and test results we provided, ChatGPT drew conclusions about the patient's ASA-PS classification (Doyle et al. 2024) (Table 1) and also determined whether they could undergo the day surgery procedure. For the expert group, two anesthesiologists were invited to analyze the ASA-PS classification of the patients and conclude whether day surgery could be performed, with a third experienced expert presiding over controversial results.

In order to ensure the uniformity of the patient cohort, those with complex conditions were not simulated in our selection: (1) patients with difficult airways caused by anatomical problems in the upper respiratory tract, (2) comatose patients, and (3) patients with heart murmurs.

The following were the reasons why these three types of factors were not selected:

(1) Due to the invention of video laryngoscopy, common difficult airways (i.e., small mandibles or short spacing) are no longer challenging problems in airway management. However, surgery and tumor space-occupying lesions remain difficult to treat. Since the identification of difficult airways in patients requires professional anesthesiologists, it was not conducive to providing clinical treatment for the problems described in this experiment.
(2) Patients with a high Glasgow score or coma were not admitted to the day surgery clinic.
(3) Heart murmurs are more subjective and could not be evaluated scientifically and so were assessed by more objective echocardiography.

Performance of ChatGPT for answering questions regarding anesthesia

ChatGPT was instructed to answer frequently asked questions about knowledge or concerns related to peri-anesthesia were collected from questions on the networks and from professional associations and institutions, and were also the most commonly encountered conditions in clinical practice. Each question was entered as a separate, independent prompt using the "New Chat" function. There were a total of 145 questions about day surgery anesthesia. Of these questions, 11 similar or duplicate questions were deleted, while 134 questions were answered by ChatGPT. Two questions with no clear answer and one question unrelated to the topic were excluded, leaving a total of 131 questions in the final analysis (Fig. 1, Supplementary Table 2).

We considered the 131 questions to be unique and pertinent and encompassed a wide array of physical status evaluations. We therefore categorized the 131 questions into 5 distinct groups: (1) Basic knowledge: the procedures to anesthesia, including issues related to anesthesia education (e.g., whether to continue taking high blood pressure medication), and postoperative rehabilitation

Table 1 ASA-PS classification: physical status of the patients as classified by the American Society of Anesthesiologists

ASA-PS classification	Patient's physical status
I	Normal and healthy patient
II	Patient with mild systemic disease without a limitation of activity
III	Patient with severe systemic disease that limits activity but is not incapacitating
IV	Patient with systemic disease that is incapacitating and a constant threat to life
V	Moribund patient who is not expected to survive without the operation
VI	A declared brain-dead patient whose organs are being removed for donor purposes

E is appended to any of the classifications if the patient requires an emergency operation

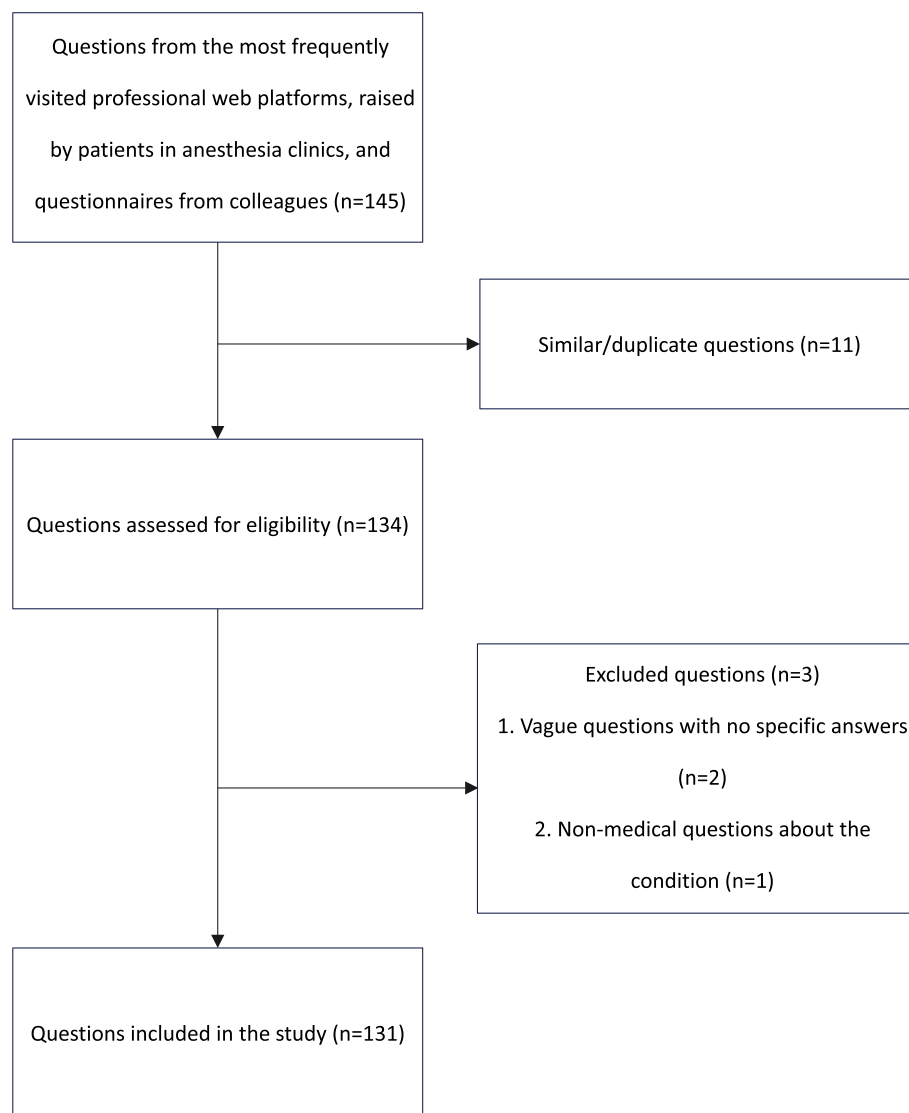Cheng *et al. Perioperative Medicine*   (2024) 13:111

Page 4 of 9

Fig. 1 Flow chart of question selection for pre-anesthesia assessment. Frequently asked questions about knowledge or concerns related to peri-anesthesia were collected from questions on networks, professional associations and institutions, and experiences in clinical practice

(e.g., how long will I be awake after anesthesia?); (2) peri-anesthesia concerns and preoperative preparation (e.g., Why do I need to fast before surgery?); (3) emotional support including problems about emotional support related to surgery or anesthesia (e.g., how does anesthesia affect my body?); (4) lifestyle (e.g., I have insomnia, does it matter after anesthesia?); and (5) others.

The responses of ChatGPT to the 131 questions were graded independently by 2 anesthesiologists, and if there were controversial results, these were resolved by a third reviewer. The performance of ChatGPT was graded and divided into four types: (1) Comprehensive, (2) correct but inadequate, (3) mixed with correct and incorrect/outdated data, and (4) completely incorrect.

Statistical analysis

The data collected were analyzed using standard statistical methods. All the calculations were performed using IBM SPSS Statistical Package version 28. Descriptive statistics were calculated to describe the baseline characteristics of the simulated patients. Categorical variables were expressed as frequencies and percentages.

The differences in ASA assessment results between the ChatGPT group and the expert group were analyzed using the Mann-Whitney U-test. The results of the conclusion of recommendations for day surgery between the two groups were analyzed by the chi-square test. All patients who were labelled ASA $\geq$ II during the initial risk assessment were later re-evaluated, with differences in

Cheng *et al. Perioperative Medicine*   (2024) 13:111

Page 5 of 9

the recommendations between the two groups analyzed by the chi-square test. A *P*-value < 0.05 between the two groups was considered to indicate a statistically significant difference.

The proportions of each aforementioned grading for responses of each pre-anesthesia domain were calculated and reported as percentages.

Results

Assessments of the 150 simulated patients (75 males [50.0%] and 75 females [50.0%]) were performed during the study. The characteristics of the patients including sex, age, and BMI are shown in Table 2. The comorbidities (e.g., airway-related diseases, pulmonary inhalation risk, respiratory diseases, circulation-related diseases, neurological disorders, endocrine/blood system diseases, allergic history, and operation history) of the simulated patients are listed in Table 3.

The two experts provided distinct ASA-PS scores in three cases (3/150, 2%) but reached different conclusions regarding the feasibility of day surgery in seven cases (7/150, 4.6%). A third expert was required to reach a consensus for these disputed conclusions. There were no significant differences between the ChatGPT and expert responses in the majority of cases ($P=0.064, >0.05$) (Table 4). However, there were some differences between ChatGPT and the experts regarding the conclusion as to whether day surgery could be performed after comprehensive consideration of the patients' conditions (ASA ≥ II).

For patients with certain comorbidities (ASA ≥ II), the expert group was more inclined to evaluate whether the patient was suitable for day surgery after further

Table 2 Characteristics of the simulated patients

Demographics	N (%)
Sex, M/F	75/75 (50.0/50.0)
Age, years	57.77 ± 21.56
10 to 29	18 (12.0)
30 to 39	19 (12.7)
40 to 49	18 (12.0)
50 to 59	21 (14.0)
60 to 69	25 (16.7)
70 to 79	31 (20.7)
> 80	18 (12.0)
BMI, kg/m²	23.12 ± 3.58
< 18.5	13 (8.7)
18.5 to 23.9	77 (51.3)
24 to 27.9	44 (29.3)
> 28	16 (10.7)

Table 3 Comorbidities of the simulated patients

Comorbidities	N (%)
Airway-related diseases	23 (15.3)
OSAS	2 (1.3)
Airway inflammation	1 (0.7)
Tumors affect the airway	20 (13.3)
Pulmonary inhalation risk	39 (26.0)
Respiratory diseases	28 (18.7)
COPD	8 (5.3)
Asthma	5 (3.3)
Pulmonary infection	8 (5.3)
History of pulmonary bullae	3 (2.0)
Hydrothorax	1 (0.7)
Lung tumor	3 (2.0)
Circulation-related diseases	35 (23.3)
Hypertension	12 (8.0)
CAD	5 (3.3)
Arrhythmia	17 (11.3)
CHD	1 (0.7)
Neurological disorders	21 (14.0)
Parkinson's disease	3 (2.0)
Myasthenia gravis	1 (0.7)
Mitochondrial myopathy	1 (0.7)
History of cerebral infarction	2 (1.3)
History of cerebral hemorrhage	2 (1.3)
History of brain tumors	2 (1.3)
Alzheimer's disease	2 (1.3)
History of seizures	2 (1.3)
Mental disorder	6 (4.0)
Endocrine/blood system diseases	28 (18.7)
Diabetes	12 (8.0)
Thyroid disease	6 (4.0)
Abnormal renal function	2 (1.3)
Abnormal liver function	2 (1.3)
Anemia	3 (2.0)
History of allergy	7 (4.7)
Antibiotic allergy	1 (0.7)
Blood transfusion allergy	1 (0.7)
Iodine allergy	1 (0.7)
Propofol allergy	2 (1.3)
Soybean allergy	1 (0.7)
Milk allergy	1 (0.7)
History of surgery	39 (26.0)
Abdominal surgery	9 (6.0)
Thoracic surgery	13 (8.7)
Orthopedic surgery	3 (2.0)
Head and neck surgery	14 (9.3)

OSAS obstructive sleep apnea syndrome, *CAD* coronary artery disease, *CHD* congenital heart disease

Cheng *et al. Perioperative Medicine*   (2024) 13:111

Page 6 of 9

Table 4 Concordance between ChatGPT and expert preoperative assessment

ASA	ChatGPT *n* (%)	Expert *n* (%)	
I	10 (6.7)	28 (18.7)	Mann-Whitney *U*-test/*P*
II	66 (44.0)	58 (38.7)	12,548.00/*P* = 0.064, > 0.05
III	63 (42.0)	52 (34.7)	
IV	10 (6.7)	12 (8.0)	
V	1 (0.7)	0 (0.0)	
Conclusion			
Day surgery can be performed	76 (50.7)	55 (36.7)	χ^2/*P* 6.761/*P* = 0.034, < 0.05
Needs further treatment and evaluation	56 (37.3)	66 (44.0)	
Day surgery is not recommended	18 (12.0)	29 (9.3)	

examination or treatment or considered that the patient's current physical condition was not suitable for day surgery. In contrast, this proportion of conclusions made by ChatGPT was smaller (ChatGPT *n* (%) vs. expert *n* (%): day surgery can be performed, 67 (47.9) vs. 31 (25.4); needs further treatment and evaluation, 56 (37.3) vs. 66 (44.0); day surgery is not recommended, 18 (12.9) vs. 29 (9.3), *P* = 0.001, < 0.05) (Table 5).

ChatGPT answered 131 questions that were then evaluated by three anesthesiologists. For the answers of ChatGPT on basic knowledge related to anesthesia, the anesthesiologists considered that 70% were comprehensive, 24% were correct but inadequate, and 6% were mixed with correct and incorrect/outdated data. For peri-anesthesia concerns, 95.3% were comprehensive, and 4.7% were correct but inadequate. The responses of ChatGPT to questions regarding emotional support were all 100% and have been all praised by experts as comprehensive. Although there was no standard answer to questions on emotional support, it was possible to judge ChatGPT's potential in this area. In terms of answers to lifestyle and other questions, ChatGPT was 91.7% comprehensive. The expert panel did not consider any of the answers to be completely incorrect. Taken together, these results indicated that although the evaluation of basic knowledge was relatively low, ChatGPT still had a certain reference value for solving problems related to anesthesia (Fig. 2).

Discussion

It is widely recognized that China has a vast population, yet the nation continues to grapple with significant medical challenges, particularly due to a shortage of anesthesiologists. At present, the number of day surgery procedures is increasing, and therefore, preoperative evaluation is a very important part of anesthesia (Ojo et al. 2010). Currently, ChatGPT's exploration in the medical field is still focused mainly on medical education and scientific writing, and there is relatively little use of it in clinical and research scenarios (Kung et al. 2023; Shay et al. 2023). One of the key benefits of ChatGPT is its ability to provide instant, accurate, and personalized responses to a wide range of questions related to health care (Cascella et al. 2023; Liu et al. 2023; Odom-Forren 2023). A study by Gupta et al. (2024) searched the database to determine how ChatGPT could be helpful to anesthesia providers, including preoperative management, ICU management, pain management, and palliative care. The results of Gupta's study showed that ChatGPT can be extremely useful for anesthesiologists, especially for determining the dose of anesthetics, assisting in retrieving research materials, or providing guidance on how to perform certain procedures.

ASA is an important index for preoperative evaluation of both anesthesia and surgical risk and has been used widely, resulting in the index being recognized throughout the world (Riley et al. 2014; Mayhew et al. 2019).

Table 5 Comparison of assessment results for patients with comorbidities

	ChatGPT *n* (%)	Expert *n* (%)	
ASA ≥ 2	140 (93.3)	122 (81.3)	
Conclusion			
Day surgery can be performed	67 (47.9)	31 (25.4)	χ^2/*P*
Needs further treatment and evaluation	55 (39.3)	62 (50.8)	15.052/*P* = 0.001, < 0.05
Day surgery is not recommended	18 (12.9)	29 (23.8)	

B: Basic knowledge; P: Peri-anesthesia concern; E: Emotional support; L: Lifestyle; O:Others
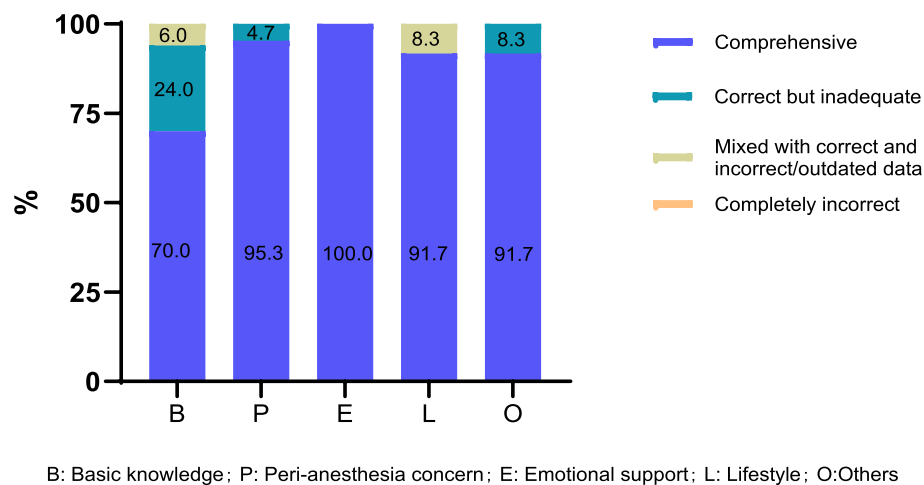
Fig. 2 Grading of responses by ChatGPT to questions related to peri-anesthesia. The percentage of responses graded as comprehensive, correct but inadequate, mixed with correct and incorrect/outdated data, and completely incorrect are provided

Lim et al. (2023) used ChatGPT to evaluate 10 standardized hypothetical patient scenarios and suggested that ChatGPT was able to classify ASA-PS consistently and correctly in multiple simulated patient scripts with appropriate justification and had similar performance to that of human anesthesiologists in the majority of cases. Our current study expanded the size of the patient cohort, broadened the spectrum of diseases under investigation, and showed that ChatGPT had a significant degree of utility for assessing the physical status of patients according to the ASA classification system, with its evaluations aligning largely with those of an expert panel.

This study used ChatGPT to analyze data of the patient's medical history, examination outcomes, surgical procedures, and anesthesia techniques. Leveraging this data, surgeons and anesthesiologists can acquire suitable risk assessment indicators, thereby saving a significant amount of energy and time. There were no significant differences between ChatGPT and the responses of the experts in the majority of cases. This underscores Chat-GPT's proficiency for evaluating the physical condition of the simulated patients and achieving the correct ASA ratings. This ability of ChatGPT makes it an exceptionally efficient method for preoperative evaluation. As far as we are aware, this is the first study to assess the ability of ChatGPT to make ASA grading and preoperative evaluation of patients, a function that would have major clinical value.

To guarantee that all the anesthesiologists use the same criteria as ChatGPT for considering suitability for daycare surgery, experts need to make their judgments based on the standardized guidelines for day-case surgery 2019 (Bailey et al. 2019). In our current study, ChatGPT and the clinical experts may had different views as to whether patients were eligible for day surgery procedures in some condition. ChatGPT mostly recommended patients for day surgery directly after assessing their physical condition, surgical method, and anesthesia risk. For patients with an ASA ≥ 2, the panel preferred to recommend further examination and treatment before considering the suitability for day surgery. Even for more seriously ill patients, the panel recommended canceling day surgery at a higher rate.

Do these results indicate that ChatGPT is more liberal when considering the risks of anesthesia surgery while the panel is more conservative? We considered the reasons for this difference may be as follows:

1. Possibly related to the working habits of each expert group, with some groups having a stricter level for indication of day surgery, while others do not. Ansell and his colleges carried out a retrospective case-controlled review of 896 ASA III patients who had undergone day case procedures and concluded that with good pre-assessment and adequate preparation, these patients could be treated safely in the day surgery setting (Ansell et al. 2004). Alternatively, Rasmussen considered that fitness for a procedure should relate to the patient's functional status as determined by a pre-anesthetic assessment, and not by ASA physical status, age, or body mass index (Rasmussen et al. 2015).

2. ChatGPT analyzes a patient's objective indicators with the conclusion made after synthesizing all these indicators, thereby defining the reference value.

3. The medical staff could not consider which decision was right or wrong, with the actual decision based on

Cheng *et al. Perioperative Medicine*   (2024) 13:111

Page 8 of 9

either the surgeon or anesthesiologist's understanding of the guidelines or the patient's condition.

Although our study showed the benefits of ChatGPT as a tool, there remain problems that need to be considered. Firstly, the correctness and validity of the content must be considered, as incorrect content may mislead the patient. While ChatGPT can provide numerous information and helpful assistance, at present, it cannot completely replace human healthcare workers in all situations. Compared to search engines, it is not possible to find the source of ChatGPT's information. Taken together, these findings indicate that ChatGPT sometimes answers questions incorrectly, with the information it uses currently not updated since 2021. In addition, ChatGPT cannot access the Internet in real time. To overcome these shortcomings, manual auditing can be used to screen the content generated and allow its accuracy to be judged (Lee et al. 2023).

Secondly, attention should be paid to ethical and privacy issues. During the process of communicating with ChatGPT, patients provide their personal basic information and medical conditions, sometimes including sensitive pictures of their private parts. Although ChatGPT claims that it does not save conversations with users, it needs to be understood that sensitive health information may be damaged or abused during transmission and browsing. It is therefore necessary to implement sound data protection measures, including encryption of sensitive information and secure data transmission. In addition, because ChatGPT has extremely rich emotional value, it is necessary to be careful that anxious patients do not become psychologically dependent on this "friend". Strict ethical and privacy regulations therefore need to be established to limit the scale of information input and emotional output of ChatGPT.

ChatGPT has the potential to revolutionize the way programs are evaluated for patients by providing accurate and effective clinical help. With any new technology, there are shortcomings that need to be addressed, although the potential benefits of ChatGPT in the field of ambulatory surgical evaluation are enormous. Development is the name of the day, and therefore, healthcare workers need to keep up with this trend and explore this promising area of technology. In this regard, it has been proposed that LLM represented by ChatGPT has the potential to add a new dimension to solving clinical problems. It is also important to realize that ChatGPT is just a machine and cannot replace the humanity and compassion that are so essential to our profession (Odom-Forren 2023).

ChatGPT can therefore be regarded as impartial and potentially offers a sense of reassurance. In a profit-oriented healthcare system, such as that of the USA, it is evident that financial incentives can influence the guidance provided to patients. Therefore, having an independent assessor to oversee medical assessments and decisions would be highly beneficial. As we continue to explore the possibilities of AI in health care, it is important to embrace these new technologies and use them to augment, rather than replace, important clinical work.

While our research offers valuable insights, it is important to acknowledge its limitations. We used simulated patient data, which although closely mirroring real patients inevitably differs in certain aspects. This discrepancy may introduce some degree of error that future studies using real patient data could address. In addition, the limited number of simulated cases restricted the generalizability of our findings. Expanding the case pool would therefore enhance the robustness of our conclusions. Furthermore, our study focused solely on day surgery patients. Further research is needed to assess the applicability of ChatGPT in pre-anesthetic evaluation of major surgery or emergency procedures.

Conclusions

ChatGPT can assist anesthesia practitioners and surgeons by alerting them to the ASA-PS classification and assessing perioperative risk in patients. ChatGPT can also be trusted to answer questions and concerns related to pre-anesthesia, thereby providing important assistance in clinical work.

Cheng *et al. Perioperative Medicine*   (2024) 13:111

Page 9 of 9

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Anesthesiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China. [2]School of Clinical Medicine, Qinghai University, Xining, Qinghai 810000, China. [3]Department of Anesthesiology, Jiading District Central Hospital, Shanghai University of Medicine & Health Sciences, Shanghai 201800, China. [4]Department of Anesthesiology, Children's Hospital of Shanghai, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China. [5]Department of Anesthesiology, Qinghai Red Cross Hospital, Xining, Qinghai 810000, China.

References
Albornoz SCd, Cobb J, Diaz T, et al. Day case and short stay surgery: 2. Anaesthesia. 2011;66:417–34. https://doi.org/10.1111/j.1365-2044.2011.06651.x.

Ali H, Qadir J, Alam T, et al. Revolutionizing Healthcare with foundation AI models. Stud Health Technol Inform. 2023;305:469–70. https://doi.org/10.3233/shti230533.

Ansell GL, Montgomery JE. Outcome of ASA III patients undergoing day case surgery. Br J Anaesth. 2004;92:71–4. https://doi.org/10.1093/bja/aeh012.

Bailey CR, Ahuja M, Bartholomew K, et al. Guidelines for day-case surgery 2019: guidelines from the Association of Anaesthetists and the British Association of Day Surgery. Anaesthesia. 2019;74:778–92. https://doi.org/10.1111/anae.14639.

Cascella M, Montomoli J, Bellini V, et al. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst. 2023;47:33. https://doi.org/10.1007/s10916-023-01925-4.

Doyle DJ, Hendrix JM, Garmon EH. American Society of Anesthesiologists Classification. StatPearls. Treasure Island (FL) ineligible companies. Disclosure: Joseph Maxwell Hendrix declares no relevant financial relationships with ineligible companies. Disclosure: Emily Garmon declares no relevant financial relationships with ineligible companies.: StatPearls Publishing. Copyright © 2024, StatPearls Publishing LLC.; 2024.

Goodwin AP, Ogg TW. Preoperative preparation for day surgery. Br J Hosp Med. 1992;47:197–201.

Grünebaum A, Chervenak J, Pollet SL, et al. The exciting potential for ChatGPT in obstetrics and gynecology. Am J Obstet Gynecol. 2023;228:696–705. https://doi.org/10.1016/j.ajog.2023.03.009.

Gupta B, Ahluwalia P, Gupta A, et al. ChatGPT in anesthesiology practice - a friend or a foe. Saudi J Anaesth. 2024;18:150–3. https://doi.org/10.4103/sja.sja_336_23.

Kristensen SD, Knuuti J, Saraste A, et al. 2014 ESC/ESA Guidelines on non-cardiac surgery: cardiovascular assessment and management: the Joint Task Force on non-cardiac surgery: cardiovascular assessment and management of the European Society of Cardiology (ESC) and the European Society of Anaesthesiology (ESA). Eur Heart J. 2014;35:2383–431. https://doi.org/10.1093/eurheartj/ehu282.

Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLiS Digit Health. 2023;2:e0000198. https://doi.org/10.1371/journal.pdig.0000198.

Lahat A, Shachar E, Avidan B, et al. Evaluating the use of large language model in identifying top research questions in gastroenterology. Sci Rep. 2023;13:4164. https://doi.org/10.1038/s41598-023-31412-2.

Lee SW, Choi WJ. Utilizing ChatGPT in clinical research related to anesthesiology: a comprehensive review of opportunities and limitations. Anesth Pain Med. 2023;18:244–51. https://doi.org/10.17085/apm.23056.

Lim DYZ, Ke YH, Sng GGR, et al. Large language models in anaesthesiology: use of ChatGPT for American Society of Anesthesiologists physical status classification. Br J Anaesth. 2023;131:e73–5. https://doi.org/10.1016/j.bja.2023.06.052.

Liu J, Wang C, Liu S. Utility of ChatGPT in clinical practice. J Med Internet Res. 2023;25:e48568. https://doi.org/10.2196/48568.

Mayhew D, Mendonca V, Murthy BVS. A review of ASA physical status - historical perspectives and modern developments. Anaesthesia. 2019;74:373–9. https://doi.org/10.1111/anae.14569.

Odom-Forren J. The role of ChatGPT in perianesthesia nursing. J Perianesth Nurs. 2023;38:176–7. https://doi.org/10.1016/j.jopan.2023.02.006.

Ojo EO, Sharpin C. Day case surgery and developing countries: a review. Niger J Clin Pract. 2010;13:459–66.

Rasmussen LS, Steinmetz J. Ambulatory anaesthesia and cognitive dysfunction. Curr Opin Anaesthesiol. 2015;28:631–5. https://doi.org/10.1097/aco.0000000000000247.

Riley R, Holman C, Fletcher D. Inter-rater reliability of the ASA physical status classification in a sample of anaesthetists in Western Australia. Anaesth Intensive Care. 2014;42:614–8. https://doi.org/10.1177/0310057x1404200511.

Shay D, Kumar B, Bellamy D, et al. Assessment of ChatGPT success with specialty medical knowledge using anaesthesiology board examination practice questions. Br J Anaesth. 2023;131:e31–4. https://doi.org/10.1016/j.bja.2023.04.017.

Thirunavukarasu AJ, Ting DSJ, Elangovan K, et al. Large language models in medicine. Nat Med. 2023;29:1930–40. https://doi.org/10.1038/s41591-023-02448-8.

Vogelsang H, Herzog-Niescery J, Botteck NM, et al. Improvement in pre-operative risk assessment in adults undergoing noncardiac surgery by a process-oriented score: a prospective single-centre study. Eur J Anaesthesiol. 2020;37:629–35. https://doi.org/10.1097/eja.0000000000001190.

Voney G, Biro P, Roos M, et al. Interrelation of peri-operative morbidity and ASA class assignment in patients undergoing gynaecological surgery. Eur J Obstet Gynecol Reprod Biol. 2007;132:220–5. https://doi.org/10.1016/j.ejogrb.2006.04.028.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.