

Assessment of the information provided by ChatGPT regarding exercise for patients with type 2 diabetes: a pilot study

Seung Min Chung ,¹ Min Cheol Chang ²

To cite: Chung SM, Chang MC. Assessment of the information provided by ChatGPT regarding exercise for patients with type 2 diabetes: a pilot study. *BMJ Health Care Inform* 2024;**31**:e101006. doi:10.1136/bmjhci-2023-101006

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/bmjhci-2023-101006>).

Received 20 December 2023
Accepted 21 June 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Division of Endocrinology and Metabolism, Department of Internal Medicine, College of Medicine, Yeungnam University, Daegu, The Republic of Korea
²Department of Physical Medicine and Rehabilitation, College of Medicine, Yeungnam University, Daegu, The Republic of Korea

Correspondence to

Professor Min Cheol Chang; wheel633@ynu.ac.kr

ABSTRACT

Objectives We assessed the feasibility of ChatGPT for patients with type 2 diabetes seeking information about exercise.

Methods In this pilot study, two physicians with expertise in diabetes care and rehabilitative treatment in Republic of Korea discussed and determined the 14 most asked questions on exercise for managing type 2 diabetes by patients in clinical practice. Each question was inputted into ChatGPT (V.4.0), and the answers from ChatGPT were assessed. The Likert scale was calculated for each category of validity (1–4), safety (1–4) and utility (1–4) based on position statements of the American Diabetes Association and American College of Sports Medicine.

Results Regarding validity, 4 of 14 ChatGPT (28.6%) responses were scored as 3, indicating accurate but incomplete information. The other 10 responses (71.4%) were scored as 4, indicating complete accuracy with complete information. Safety and utility scored 4 (no danger and completely useful) for all 14 ChatGPT responses.

Conclusion ChatGPT can be used as supplementary educational material for diabetic exercise. However, users should be aware that ChatGPT may provide incomplete answers to some questions on exercise for type 2 diabetes.

INTRODUCTION

Diabetes has become a widespread epidemic, primarily due to the increase in prevalence and incidence of type 2 diabetes (T2D).¹ According to the latest report from the International Diabetes Federation, the global prevalence of T2D in adults was 536.6 million people (10.5%) in 2021.² The number of individuals with T2D is expected to increase to 783.2 million (12.2%) by 2045.² T2D is widely known to increase the risk of cardiovascular disease, chronic renal disease, blindness and amputation.³ To manage T2D and prevent its complications, regular exercise is one of the key therapeutic factors, together with medication and diet.⁴

Regular exercise improves glucose tolerance, increases peripheral and hepatic insulin

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Opinions on using large language model in clinical practice vary considerably.

WHAT THIS STUDY ADDS

⇒ ChatGPT provided relatively valid, safe and useful information about exercise for type 2 diabetes.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ After receiving diabetes self-management and education from medical professionals, ChatGPT can be used as supplementary educational material for diabetic exercise.

sensitivity, reduces glycosylated haemoglobin and promotes the uptake and utilisation of glucose by muscles.⁴ To exercise using proper methods and to be mindful of precautions during exercise are crucial for patients with T2D. In the past, patients with T2D had no choice but to visit a hospital or clinic to receive explanations from physicians about exercise that can be used to manage T2D and prevent complications. However, ChatGPT may be fruitful in helping physicians manage time efficiently while validating information during patients' visits.

With the recent development of the internet, patients can obtain medical information on specific disorders or conditions online.^{5 6} However, the internet provides abundant information beyond what patients specifically seek to know. Therefore, it can be challenging for patients to read, select and acquire personally relevant information.

Recently, large language models (LLMs), which are sophisticated artificial intelligence (AI) models that excel in natural language processing tasks, were developed.⁷ These models are trained using deep learning techniques on massive amounts of internet text data, allowing them to understand and respond to a wide

range of topics.⁸ ChatGPT is the most popular LLM and was developed by OpenAI based on the generative pretrained transformer (GPT) architecture.^{9–11} The primary function of ChatGPT is to provide human-like answers to natural language questions in real time.^{9–11} It is anticipated to be available for application in the medical field.^{9–11} ChatGPT is expected to be a useful search engine for patients with T2D. However, the usefulness and accuracy of the provided information have not been evaluated. Our hypothesis is that ChatGPT can be used as educational material for diabetic exercise.

Therefore, in the current study, we assessed the validity, safety and utility of ChatGPT for patients with T2D seeking information about exercise.

METHODS

This was a pilot, cross-sectional study. Similar to the systematic reviews and meta-analyses, two physicians (MCC and SMC) who work in Republic of Korea, each with approximately 15 and 7 years of experience in rehabilitative treatment and diabetes care, respectively, used the modified Delphi technique to assess the information provided by ChatGPT: premeeting question development, face-to-face consensus meeting and postmeeting feedback.¹² Two authors discussed and determined the questions regarding diabetic exercise in clinical practice. Each question was keyed into ChatGPT, and the answers from ChatGPT were assessed by the two authors. Any discrepancies in the assessment were discussed until a consensus was reached.

The 14 questions most frequently asked by patients with T2D were developed based on personal perspective: (1) What is the benefit of exercise for type 2 diabetes patients?, (2) Which type of exercise training should type 2 diabetes patients do? (3) How much intensity should be exercised in patients with type 2 diabetes? (4) How often should type 2 diabetes patients exercise? (5) How long should type 2 diabetes patients exercise? (6) How much weight should type 2 diabetes patients lose to achieve metabolic benefits? (7) Do type 2 diabetes patients require exercise stress testing before starting exercise? (8) How should type 2 diabetes patients prevent hypoglycaemia during exercise? (9) How should type 2 diabetes patients prevent hyperglycaemia during exercise? (10) Which kind of exercise should diabetic neuropathy patients do? (11) Which kind of exercise should diabetic retinopathy patients do? (12) What kind of exercise should diabetic kidney patients do? (13) When should type 2 diabetes patients exercise, before or after meals? (14) Which time of the day should type 2 diabetes patients exercise?

We used ChatGPT (V.4.0) to ask questions related to exercise for patients with T2D in November 2023. A Likert scale was used to evaluate the validity, safety and utility of the answers generated by ChatGPT. The Likert scale is an ordinal scale frequently used in medical education research.¹³ The Likert scale typically ranges from 1 to 5: completely disagree, disagree, neutral, agree and completely agree. In this study, each score for the validity, safety and utility was divided into 4 points, and a score of 4 means the most highly valid, safe

and useful answers, and 1 point denotes the incomplete or incorrect answers. The Likert scale for evaluating the validity, safety and utility of the answers generated by ChatGPT was categorised as follows:

- ▶ Validity:
 - Completely erroneous information (all the information that ChatGPT answered cannot be found in medical sources or is inaccurate or incomplete).
 - Partially erroneous information (some of the information that ChatGPT answered cannot be found in medical sources or contains inaccuracies or incompleteness).
 - Reliable but incomplete information (all the information that ChatGPT answered is found in medical sources and accurate but with some incomplete elements).
 - Completely reliable and complete information (all the information that ChatGPT answered is found in medical sources and complete).
- ▶ Safety:
 - Significant and certain danger to the patient's condition.
 - Moderate potential danger to the patient's condition.
 - Minimal potential danger to the patient's condition.
 - No danger.
- ▶ Utility:
 - Not useful for the patient (no useful information).
 - Partially useful for the patient (more than 0% and less than 50% of the information provided is useful).
 - Moderately useful for the patient ($\geq 50\%$ of the information provided is useful, but not 100%).
 - Completely useful (100% of the information provided is useful).

RESULTS

The answers generated by ChatGPT and the Likert scales for each answer are presented in online supplemental data and [table 1](#). ChatGPT generally provided a well-organised list of instructions using technical terminology. The contents were consistent with the position statements of the American Diabetes Association and American College of Sports Medicine^{14 15} and the practice guidelines of the Korean Diabetes Association.¹⁶

The validity of each question ranged from 3 to 4, suggesting that the answers of ChatGPT were accurate. However, four answers (questions 4, 5, 7 and 11) (28.6%) provided incomplete information. In question 4 (frequency of exercise), ChatGPT recommended at least 150 min of aerobic activities per week, which can be broken down into 30 minutes a day, 5 days a week. However, there was no information about exercising at least 3 days per week and not resting for two consecutive days.^{14 15} In question 5 (duration of exercise), ChatGPT recommended that flexibility training be performed for 10–30 min. However, there were no instructions to maintain the stretch for 10–30 s per stretch.¹⁵ In question 7 (pre-exercise evaluation), ChatGPT recommended that patients

Table 1 Likert scores of each answer generated by ChatGPT

| Question | Validity | Safety | Utility | Total |
|---|----------|--------|---------|-------|
| 1 What is the benefit of exercise for type 2 diabetes patients? | 4 | 4 | 4 | 12 |
| 2 Which type of exercise training should type 2 diabetes patients do? (Type) | 4 | 4 | 4 | 12 |
| 3 How much intensity should be exercised in patients with type 2 diabetes? (Intensity) | 4 | 4 | 4 | 12 |
| 4 How often should type 2 diabetes patients exercise? (Frequency) | 3 | 4 | 4 | 11 |
| 5 How long should type 2 diabetes patients exercise? (Duration) | 3 | 4 | 4 | 11 |
| 6 How much weight should type 2 diabetes patients lose to achieve metabolic benefits? (Weight loss) | 4 | 4 | 4 | 12 |
| 7 Do type 2 diabetes patients require exercise stress testing before starting exercise? (Pre-exercise evaluation) | 3 | 4 | 4 | 11 |
| 8 How should type 2 diabetes patients prevent hypoglycaemia during exercise? (Management of hypoglycaemia) | 4 | 4 | 4 | 12 |
| 9 How should type 2 diabetes patients prevent hyperglycaemia during exercise? (Management of hyperglycaemia) | 4 | 4 | 4 | 12 |
| 10 What kind of exercise should diabetic neuropathy patients do? (Precaution against health complications) | 4 | 4 | 4 | 12 |
| 11 Which kind of exercise should diabetic retinopathy patients do? (Precaution against health complications) | 3 | 4 | 4 | 11 |
| 12 Which kind of exercise should diabetic kidney patients do? (Precaution against health complications) | 4 | 4 | 4 | 12 |
| 13 When should type 2 diabetes patients exercise before or after meals? (Exercise timing) | 4 | 4 | 4 | 12 |
| 14 Which time of the day should type 2 diabetes patients exercise? (Exercise timing) | 4 | 4 | 4 | 12 |

with T2D with known cardiovascular disease or related symptoms, those aged over 40, and those who have ≥ 1 risk factor for heart disease (smoking, hypertension, dyslipidaemia, family history of heart disease or overweight) undergo exercise stress testing. However, stress testing is also recommended for patients with T2D aged over 30 with >10 years of diabetes.^{15 17} In question 11 (precaution against health complications—diabetic retinopathy), ChatGPT recommended high-intensity exercise and activities that lower the head, which could increase intraocular pressure. ChatGPT highly recommended that patients consult an eye specialist before starting or modifying their exercise routine. However, ChatGPT did not mention that exercise is contraindicated for anyone with unstable or untreated proliferative retinopathy, recent pan-retinal photocoagulation or other recent surgical eye treatment.¹⁵

ChatGPT scored 4 for the safety of every generated answer. It always emphasised an individualised exercise strategy, mentioned safety tips and recommended consulting with health professionals. It also scored 4 for the utility of every question, suggesting that the provided information benefited patients.

DISCUSSION

We hypothesised that ChatGPT can be used as educational material for diabetic exercise. 14 questions regarding exercise

for patients with T2D were posed to ChatGPT. The Likert scores of each question ranged from 11 to 12. The answers were systematic with easy readability. Four (28.6%) out of 14 answers had incomplete elements, but the presented information was accurate, safe and useful. ChatGPT always emphasised an individualised exercise approach and recommended consulting a health professional. Our hypothesis has been proven to some extent. However, since ChatGPT's answers are accurate but sometimes incomplete, it cannot replace face-to-face education and should be used as supplementary material.

Diabetes self-management and education (DSME) is a process that promotes the acquisition of knowledge and skills to improve glycaemic control and quality of life and reduce acute and chronic diabetes complications.¹⁸ While health-care professionals typically provide initial DSME, ongoing support may be provided through various community-based resources. Recently, the American Diabetes Association also recommended using telehealth and other digital health solutions to deliver DSME.¹⁹ As interest in AI-based LLMs is rapidly increasing, ChatGPT could be another source of DSME.

Opinions on using ChatGPT in healthcare vary considerably.²⁰ Argentine dermatologists adopted an intermediate stance towards ChatGPT and suggest that the reliability of ChatGPT should be currently questioned.²¹ It is reported

that ChatGPT can be used as a search engine and a source for obtaining medical information. However, there are limitations, such as the possibility of generating inaccurate or even erroneous responses.²² When the four domains of DSME (diet and exercise; hypoglycaemia and hyperglycaemia education; insulin storage; insulin administration) were posed to GPT3, it revealed incomplete knowledge of various insulin types and regimens, which might induce potential safety issues.²³ In this study, ChatGPT lacked some indications for pre-exercise evaluation or contraindications for exercise among diabetic retinopathy patients. Since incomplete information may be provided, we recommend patients use it as a reference after receiving their initial education from medical staff. In addition, physicians should always be aware of both the strengths and weaknesses of ChatGPT and use them as a DSME tool with critical thinking. Appropriate enhancements of ChatGPT may help physicians manage their time efficiently during patients' visits, which needs to be further validated through further studies.

This study had certain limitations. First, the Likert scale was not validated for evaluating the answers from ChatGPT. However, it is a tool frequently used in medical education research; therefore, it was considered an acceptable tool for this pilot study. Second, the evaluation of the validity, safety and utility of the answers was relatively subjective. However, we attempted to conduct an objective evaluation based on international statements and consensus between the two physicians. Lastly, we did not assess patient satisfaction with the received ChatGPT responses. In the future, further studies compensating for these limitations are warranted.

In conclusion, although some responses were incomplete, ChatGPT can be considered an educational material for achieving information regarding diabetic exercise. However, it should be kept in mind that ChatGPT has some limitations in information acquisition and, therefore, ChatGPT can be used as supplementary educational material for diabetic exercise after receiving DSME from medical professionals.

Contributors Conceptualisation: MCC, Investigation: SMC, MCC, Writing—original draft: SMC, MCC, Writing—review and editing: MCC.

Funding This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NO.00219725).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Ethical committee approval was not required due to the absence of patients and identifiable data.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which

permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Seung Min Chung <http://orcid.org/0000-0003-3336-7557>

Min Cheol Chang <http://orcid.org/0000-0002-7629-7213>

REFERENCES

- Colberg SR, Sigal RJ, Fernhall B, *et al*. Exercise and type 2 diabetes: the American college of sports medicine and the American diabetes Association: joint position statement executive summary. *Diabetes Care* 2010;33:2692–6.
- Sun H, Saeedi P, Karuranga S, *et al*. IDF diabetes atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res Clin Pract* 2022;183:109119.
- Goyal RS, Jialal I. Type 2 diabetes: statpearls. 2013. Available: <https://www.ncbi.nlm.nih.gov/books/NBK513253/>
- Kirwan JP, Sacks J, Nieuwoudt S. The essential role of exercise in the management of type 2 diabetes. *Cleve Clin J Med* 2017;84:S15–21.
- Chang MC, Park D. Youtube as a source of information on epidural steroid injection. *J Pain Res* 2021;14:1353–7.
- Lee H, Chang MC. Youtube as a source of information regarding the effect of vitamin C on coronavirus disease. *Complement Ther Med* 2022;67:102827.
- Miao H, Ahn H. Impact of ChatGpt on interdisciplinary nursing education and research. *Asian Pac Isl Nurs J* 2023;7:e48136.
- Clusmann J, Kolbinger FR, Muti HS, *et al*. The future landscape of large language models in medicine. *Commun Med (Lond)* 2023;3:141.
- Nedbal C, Naik N, Castellani D, *et al*. Chatgpt in urology practice: revolutionizing efficiency and patient care with generative artificial intelligence. *Curr Opin Urol* 2024;34:98–104.
- Ramamurthi A, Are C, Kothari AN. From ChatGpt to treatment: the future of AI and large language models in surgical oncology. *Indian J Surg Oncol* 2023;14:537–9.
- Yu P, Xu H, Hu X, *et al*. n.d. Leveraging generative AI and large language models: a comprehensive roadmap for Healthcare integration. *Healthcare* 11:2776.
- Gagnier JJ, Morgenstern H, Altman DG, *et al*. Consensus-based recommendations for investigating clinical heterogeneity in systematic reviews. *BMC Med Res Methodol* 2013;13:106.
- Sullivan GM, Artino AR. Analyzing and interpreting data from likert-type scales. *J Grad Med Educ* 2013;5:541–2.
- Colberg SR, Sigal RJ, Yardley JE, *et al*. Physical activity/exercise and diabetes: a position statement of the American diabetes association. *Diabetes Care* 2016;39:2065–79.
- Kanaley JA, Colberg SR, Corcoran MH, *et al*. Exercise/physical activity in individuals with type 2 diabetes: a consensus statement from the American college of sports medicine. *Med Sci Sports Exerc* 2022;54:353–68.
- Hur KY, Moon MK, Park JS, *et al*. Clinical practice guidelines for diabetes mellitus of the Korean diabetes association. *Diabetes Metab J* 2021;45:461–81.
- Colberg SR, Sigal RJ, Fernhall B, *et al*. Exercise and type 2 diabetes: the American college of sports medicine and the American diabetes association: joint position statement. *Diabetes Care* 2010;33:e147–67.
- Powers MA, Bardsley J, Cypress M, *et al*. Diabetes self-management education and support in type 2 diabetes: a joint position statement of the American diabetes association, the American Association of diabetes educators, and the academy of nutrition and dietetics. *Diabetes Care* 2015;38:1372–82.
- ElSayed NA, Aleppo G, Aroda VR, *et al*. 5. Facilitating positive health behaviors and well-being to improve health outcomes: standards of care in diabetes-2023. *Diabetes Care* 2023;46:S68–96.
- 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.
- Ko EA, Torre AC, Hernandez B, *et al*. Argentine dermatology and chat-GPT: infrequent use and intermediate stance. *Clin Exp Dermatol* 2024;7:734–6.
- Dave T, Athaluri SA, Singh S. ChatGpt in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell* 2023;6.
- Sng GGR, Tung JYM, Lim DY, *et al*. Potential and pitfalls of ChatGpt and natural-language artificial intelligence models for diabetes education. *Diabetes Care* 2023;46:e103–5.