Nutrición Hospitalaria

Evaluación de la validez científica y la adecuación de los consejos dietéticos asistidos por inteligencia artificial de ChatGPT en el tratamiento nutricional de la acidemia metilmalónica

Evaluation of scientific validity and appropriateness of artificial intelligence-assisted ChatGPT advices in dietary treatment of methylmalonic acidemia

10.20960/nh.06026

09/24/2025

OR 6026

Evaluation of scientific validity and appropriateness of artificial intelligence-assisted ChatGPT advices in dietary treatment of methylmalonic acidemia

Evaluación de la validez científica y la adecuación de los consejos dietéticos asistidos por inteligencia artificial de ChatGPT en el tratamiento nutricional de la acidemia metilmalónica

Furkan Yolcu[1], Merve Koç Yekedüz[1,2], Engin Köse[1,3], F. Gülhan Samur[4], F. Tuba Eminoğlu[1,3]

[1]Department of Pediatric Metabolism. Faculty of Medicine. Ankara Üniversitesi. Ankara, Turkey. [2]Department of Anesthesiology, Critical Care and Pain Medicine. Boston Children's Hospital. Harvard Medical School. Boston, MA. USA. [3]Ankara University Rare Diseases Application and Research Center. Ankara, Turkey. [4]Nutrition and Dietetics. Hacettepe Üniversitesi. Ankara, Turkey

Received: 26/06/2025
Accepted: 25/08/2025

Correspondence: Furkan Yolcu. Department of Pediatric Metabolism. Faculty of Medicine. Ankara Üniversitesi. Hacettepe Mahallesi A. Adnan Saygun Cad. No: 35, Altindağ. Ankara 06620, Turkey
e-mail: yolcufurkan94@hotmail.com

Authors´ contribution: F. Y. and M. K. Y. contributed to the conception of the study. E. K., F. G. S., and F. T. E., contributed to the acquisition, analysis, and interpretation of data. F. Y. and M. K. Y. drafted the manuscript. All authors critically revised the manuscript, provided their final approval, and agreed to be accountable for all aspects of the work, ensuring its integrity and accuracy.

Conflict of interest: The authors declare no conflict of interest.

Artificial intelligence: The authors declare not to have used artificial intelligence (AI) or any AI-assisted technologies in the elaboration of the article.

ABSTRACT

Background and objectives: Chat Generative Pretrained Transformer (ChatGPT) has become an increasingly popular way for patients and healthcare professionals to seek information on health and nutrition advice. Previous studies have raised concerns about the accuracy of dietary advice provided by ChatGPT. We aimed to evaluate the potential of ChatGPT as a tool to provide dietary therapy for methylmalonic acidemia (MMA) in October 2024.

Methods and study design: we compared ChatGPT's (version 4.0) dietary advices for MMA with internationally recognized guidelines and current literature, and then evaluated the chatbot's ability to generate dietary therapy in two different cases with MMA. A team of dietitians and metabolic physicians compared ChatGPT responses to disease-specific guidelines and evidence-based resources.

Results: overall, ChatGPT provided clear advices: 20.0 % of responses were considered appropriate and 53.4 % were considered inappropriate. Especially in this group of patients whose only known treatment was dietary therapy, the dietary therapy recommended by ChatGPT was also inappropriate for the two cases in question.

Conclusions: this study was the first to evaluate the appropriateness of artificial intelligence advices for specialized dietary management of inherited metabolic disorders. Consequently, the idea that artificial intelligence can replace human labor in the near future seems far-fetched, given the low success rate of even a diagnosis with such a

clear mechanism and treatment, and individualized treatment approaches.

Keywords: Artificial intelligence. ChatGPT. Dietary management. Inherited metabolic disorders. Methylmalonic acidemia.

RESUMEN

Antecedentes y objetivos: el Chat Generative Pretrained Transformer (ChatGPT) se ha convertido en una herramienta cada vez más utilizada por pacientes y profesionales de la salud para obtener información sobre salud y asesoramiento nutricional. Estudios previos han planteado preocupaciones acerca de la exactitud de las recomendaciones dietéticas proporcionadas por ChatGPT. El objetivo de este estudio fue evaluar el potencial de ChatGPT como herramienta para proporcionar tratamiento dietético en la acidemia metilmalónica (AMM) en octubre de 2024.

Métodos y diseño del estudio: se compararon las recomendaciones dietéticas de ChatGPT (versión 4.0) para AMM con guías internacionalmente reconocidas y la literatura científica actual, y posteriormente se evaluó la capacidad del chatbot para generar tratamiento dietético en dos casos distintos de AMM. Un equipo de dietistas y médicos especialistas en metabolismo comparó las respuestas de ChatGPT con guías específicas de la enfermedad y recursos basados en la evidencia.

Resultados: en general, ChatGPT proporcionó recomendaciones claras: el 20,0 % de las respuestas se consideraron apropiadas y el 53,4 % se consideraron inapropiadas. Especialmente en este grupo de pacientes, cuyo único tratamiento reconocido es la terapia dietética, las recomendaciones de ChatGPT resultaron igualmente inapropiadas para los dos casos evaluados.

Conclusiones: este estudio es el primero en evaluar la idoneidad de las recomendaciones generadas por inteligencia artificial para el manejo dietético especializado de trastornos metabólicos hereditarios. En consecuencia, la idea de que la inteligencia artificial pueda sustituir la labor humana en un futuro próximo parece poco realista, dado el bajo índice de acierto incluso en un diagnóstico con un mecanismo y tratamiento tan claramente definidos, así como con enfoques terapéuticos individualizados.

Palabras clave: Inteligencia artificial. ChatGPT. Manejo dietético. Trastornos metabólicos hereditarios. Acidemia metilmalónica.

INTRODUCTION

Methylmalonic acidemia (MMA) is a rare autosomal recessive inherited metabolic disorder first described in 1967, characterized by defective metabolism of certain amino acids (valine, isoleucine, methionine, threonine), odd-chain fatty acids, propionate, and cholesterol. These enzymatic deficiencies lead to the accumulation of toxic metabolites, which may result in acute metabolic decompensation, chronic renal failure, and pancreatitis (1). The incidence of MMA at birth is reported to be 1:50000-1:100000 (2). Effective management of MMA relies heavily on dietary therapy. Key goals include preventing metabolic crises such as hyperammonemia and acidosis, minimizing toxic metabolite accumulation, and supporting normal growth and nutritional status. Dietary treatment involves restriction of natural protein intake and, when needed, the use of disease-specific medical foods. Inadequate or incorrect dietary management may lead to severe complications including lethargy, coma, and death (3).

In recent years, the use of artificial intelligence (AI) tools like Chat Generative Pre-trained Transformer (ChatGPT), developed by Open

Artificial Intelligence (OpenAI), has increased significantly in the healthcare domain. ChatGPT is a conversational AI system that provides instant, free, and personalized information across platforms (4-6). Healthcare professionals have begun exploring its utility in clinical decision-making, patient education, and dietary guidance (7,8). Although promising results have been reported in the management of common chronic conditions (9-11), concerns remain regarding the accuracy, safety, and clinical appropriateness of chatbot-generated dietary advice, particularly in complex medical cases (12,13).

Despite growing interest in AI-assisted health communication, no studies to date have investigated the accuracy of ChatGPT's dietary advice for patients with rare inherited metabolic disorders such as MMA. Given the highly specialized and individualized nature of dietary management in MMA, it is unclear whether ChatGPT can deliver guidance that aligns with expert clinical practice and international guidelines. The aim of the study was to compare the dietary information provided by ChatGPT on MMA requiring dietary advice with the dietary recommendations of international guidelines and current literature. In addition, we aimed to evaluate whether the chatbot could replace dietitians in providing personalized nutritional advice to patients with MMA.

MATERIALS AND METHODS

This study was conducted in October 2024 using the ChatGPT default version (OpenAI, San Francisco, CA, USA, version 4.0). Conversations with ChatGPT were conducted in English with a new chat for each prompt (all original questions posed to ChatGPT are presented in Supplementary Table I - https://www.nutricionhospitalaria.org/files/9413/ADMA1-06026-02.pdf), and all observations were made independently. A total of 15 questions were asked. The first 13 of these questions were about the basic information that should be known when preparing

nutritional treatment in MMA, while the last two questions were about the ability to create a nutritional plan in MMA disease. The questions were formulated by a panel of experts based on their clinical experience and the relevance of each topic to established dietary management practices in methylmalonic acidemia (MMA). Rather than being selected randomly, the questions were deliberately chosen to address critical components of routine clinical care. ChatGPT may produce different responses to the same prompts depending on the history of the conversation. Therefore, each question was asked three times, each time in a new chat session to avoid any possible bias from the model's memory. Responses varied slightly depending on the prompt, but the advices listed remained largely consistent with a few minimal changes. In the end, the most comprehensive response, i.e., the one that provided the most information consistent with the guidelines, was considered. All questions were asked on October 31, 2024, using the most current version (ChatGPT 4.0) at that time, and then evaluated for evidence-based information by an expert panel consisting of dietitians and metabolic physicians who specialize in inherited metabolic diseases and have 8 years of experience with MMA diet therapy. The responses were evaluated based on widely accepted, evidence-based guidelines and current scientific literature relevant to the nutritional management of methylmalonic acidemia, all of which were accessible to evaluators throughout the assessment process. The evaluation process was not blinded. In cases of disagreement, consensus was reached through group discussion. Each ChatGPT response was evaluated through a comparison with scientifically grounded recommendations and categorized into four distinct groups: "Appropriate", if the content was fully consistent with evidence-based guidelines; "Inappropriate", if it contradicted established recommendations; "Not fully matched", if it was partially consistent but contained omissions or inaccuracies; and "General advice", if it provided vague or non-specific information that did not directly address the prompt. This classification facilitated a structured

comparison between the responses generated by ChatGPT-4.0 and the correct answers derived from current scientific standards. For visual representation, a color legend was applied—green for "Appropriate", grey for "Not fully matched", blue for "General advice", and red for "Inappropriate"—as indicated in the table provided in the supplementary material. As this study did not involve human participants or animals, no ethics committee approval was required.

RESULTS

Comprehensive responses generated by ChatGPT and their comparison with established dietary guidelines are presented in supplementary table I. This table also includes the scientific references used to support each correct response.

Of the 15 evaluated responses, 4 were categorized as “appropriate,” corresponding to an appropriateness rate of 20.0 % (3/15) (Table I). When both “appropriate” and “general advice” categories were combined, the total number of correct responses was 5, resulting in an overall accuracy of 26.6 % (4/15). In contrast, 8 responses (53.4 %) were classified as “inappropriate,” as they did not align with current dietary guidelines (14,15). These responses included incorrect recommendations regarding the daily intake of protein, natural protein, energy, valine, isoleucine, and methionine. ChatGPT also provided inaccurate values for the content of valine, isoleucine, methionine, and threonine in breast milk (15), as well as reference values for plasma propiogenic amino acids (15).

Three responses (3/15; 20 %) were identified as “not fully matched.” The first related to the use of medical foods: while ChatGPT recommended their use, it did not specify that their use is only indicated when natural protein intake fails to meet safe levels, as outlined in the 2014 European guidelines and 2019 U.S. guidelines for MMA and PA (14,16). The second was related to the carbohydrate concentration in emergency nutrition, where ChatGPT stated 10-15 %, whereas guidelines indicate a range of 10-25 % (17). The third

concerned the management of low plasma valine and/or isoleucine levels. ChatGPT recommended increasing medical food, while guidelines recommend increasing natural protein and reducing medical food intake in such cases (17,18).

The final two questions involved creating personalized dietary plans. Both responses were classified as inappropriate (2/15), due to the recommended natural protein amounts being below guideline values and errors in energy requirement calculations (14).

DISCUSSION

In this study, we aimed to evaluate ChatGPT's potential as a tool for providing dietary therapy recommendations for MMA. Our findings showed that ChatGPT's appropriateness rate for MMA dietary advice was notably lower than rates reported for other conditions in the literature. Previous studies have primarily assessed ChatGPT's ability to respond to general nutrition questions in healthy populations or common disease groups. For example, one study evaluating answers to eight frequently asked nutrition questions reported that ChatGPT performed comparably or even superior to dietitians in terms of scientific accuracy, understandability, and actionability (10). Another study found appropriateness rates ranging from 55.5 % for sarcopenia to 73.3 % for NAFLD, with overall accuracy for sarcopenia reaching 100 % when both "appropriate" and "general advice" were considered (5). These findings, however, are not directly comparable to ours due to the complexity and rarity of MMA, where dietary precision is critical to prevent metabolic decompensation.

In our study, ChatGPT's diet plans for MMA cases were inappropriate mainly because the natural protein allowance for age and sex was miscalculated, leading to incorrect menu designs. Although high-protein foods were excluded, the error in determining allowable natural protein intake undermined the validity of the plans. Similar limitations in adapting diet plans to specific pathological conditions have been reported in other contexts. For instance, a study on type

2 diabetes and hemodialysis patients found that chatbot-generated menus failed to meet disease-specific nutritional needs and were repetitive (12). Likewise, in the context of food allergies, inappropriate ingredients were included, such as almond milk in a nut-free diet plan (13). These examples, along with our findings, highlight the need for condition-specific data integration to improve the safety and accuracy of AI-generated dietary recommendations.

Our results showed that ChatGPT calculated an energy intake below the recommended amount for a 10-year-old child, while the recommendation was appropriate for an infant patient. A recent study evaluated ChatGPT's ability to provide personalized dietary plans in terms of energy intake, nutrient accuracy, and meal variability for hypothetical patients with obesity, cardiovascular disease, and type 2 diabetes (11). Although the study demonstrated that ChatGPT-generated meal plans were not monotonous, the recommended daily caloric intake deviated from the target by up to 20 % (11). Notably, the deviation decreased when the target energy intake was explicitly stated in the prompt, dropping from 19.6 % to 17.3 % with ChatGPT version 3.5, and from 27.7 % to 3.4 % with version 4 (6). While these findings support the importance of detailed user input for generating more accurate recommendations, it is important to emphasize that the cited studies involved different patient populations and disease profiles. Therefore, direct comparisons with our MMA patient cohort are limited, and this should be considered when interpreting the findings.

ChatGPT offers promising opportunities for improving access to dietary counseling, particularly through its ease of use, 24/7 accessibility, and compatibility with various digital platforms such as computers, tablets, and smartphones. This flexibility may be especially beneficial for individuals in underserved or rural areas, helping to bridge gaps in healthcare access and reduce dependence on limited provider availability (8,19). However, several limitations must be acknowledged. The accuracy of ChatGPT’s recommendations

remains variable, and inappropriate or incorrect dietary advice may pose health risks. Additionally, disparities in digital literacy—particularly among older adults—may hinder effective use of the tool (8). The potential for patients to share sensitive health data also raises ethical and privacy concerns, necessitating robust data protection measures to safeguard confidentiality (6,8). These considerations highlight the importance of careful oversight, ethical compliance, and further refinement of the system before its broader integration into clinical practice.

Artificial intelligence has the potential to significantly enhance the efficiency of healthcare professionals in the future; however, its direct application to patient care still requires human oversight. This highlights the continued necessity of a healthcare workforce engaged in direct patient interaction.

This study has several limitations. Given that ChatGPT is continuously updated, restricting the analysis to outputs generated on a single date (October 31, 2024) limits the temporal generalizability of the results. The determination of appropriateness using mathematical percentages and predefined cut-off thresholds may have influenced the findings. Moreover, the assessment process relied on the subjective judgments of a panel of dietitians and physicians, introducing the potential for evaluator bias. Additionally, ChatGPT may retrieve content based on local recommendations, which may not fully correspond to internationally accepted, evidence-based guidelines, thereby potentially affecting the applicability and validity of its outputs.

CONCLUSION

The low percentage of appropriateness, even for a diagnosis with such a clear mechanism and treatment, when individual treatments are also taken into account, shows that the prediction that artifical intelligence will be able to replace human power in the near future is quite far-fetched. On the other hand, artificial intelligence technology,

which is developing every day, can benefit the judgment and decision-making mechanisms of healthcare professionals by revealing the information in big data. Artificial intelligence in the field of nutrition should always be accepted after evaluation by expert dietitians and physicians.

REFERENCES

1. Pinto A, Evans S, Daly A, Almeida MF, Assoun M, Belanger-Quintana A, et al. Dietary practices in methylmalonic acidaemia: a European survey. Journal of Pediatric Endocrinology and Metabolism 2020;33(1):147-55. DOI: 10.1515/jpem-2019-0277
2. Manoli I, Sloan JL, Venditti CP. Isolated Methylmalonic Acidemia. 2005 Aug 16 [updated 2022 Sep 8]. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2025.
3. Myles JG, Manoli I, Venditti CP. Effects of medical food leucine content in the management of methylmalonic and propionic acidemias. Current Opinion in Clinical Nutrition & Metabolic Care 2018;21(1):42-8. DOI: 10.1097/MCO.0000000000000428
4. Tan SS-L, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. Journal of medical Internet research 2017;19(1):e9. DOI: 10.2196/jmir.5729
5. Ponzo V, Goitre I, Favaro E, Merlo FD, Mancino MV, Riso S, et al. Is ChatGPT an Effective Tool for Providing Dietary Advice? Nutrients 2024;16(4):469. DOI: 10.3390/nu16040469
6. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare (Basel) 2023;11(6):887. DOI: 10.3390/healthcare11060887

7. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future healthcare journal 2019;6(2):94-8.DOI: 10.7861/futurehosp.6-2-94
8. Sharma S, Pajai S, Prasad R, Wanjari MB, Munjewar PK, Sharma R, et al. A critical review of ChatGPT as a potential substitute for diabetes educators. Cureus 2023;15(5). DOI: 10.7759/cureus.38380
9. Garcia MB. ChatGPT as a virtual dietitian: Exploring its potential as a tool for improving nutrition knowledge. Applied System Innovation 2023;6(5):96. DOI: 10.3390/asi6050096
10. Kirk D, van Eijnatten E, Camps G. Comparison of answers between ChatGPT and human dieticians to common nutrition questions. Journal of Nutrition and Metabolism 2023;2023(1):5548684. DOI: 10.1155/2023/5548684
11. Papastratis I, Stergioulas A, Konstantinidis D, Daras P, Dimitropoulos K. Can ChatGPT provide appropriate meal plans for NCD patients? Nutrition 2024;121:112291. DOI: 10.1016/j.nut.2023.112291
12. Chatelan A, Clerc A, Fonta P-A. ChatGPT and future artificial intelligence chatbots: what may be the influence on credentialed nutrition and dietetics practitioners? Journal of the Academy of Nutrition and Dietetics 2023;123(11):1525-31. DOI: 10.1016/j.jand.2023.08.001
13. Niszczota P, Rybicka I. The credibility of dietary advice formulated by ChatGPT: robo-diets for people with food allergies. Nutrition 2023;112:112076. DOI: 10.1016/j.nut.2023.112076
14. Jurecki E, Ueda K, Frazier D, Rohr F, Thompson A, Hussa C, et al. Nutrition management guideline for propionic acidemia: An evidence-and consensus-based approach. Molecular Genetics and Metabolism 2019;126(4):341-54. DOI: 10.1016/j.ymgme.2019.02.007

15. Yannicelli S. Nutrition management of patients with inherited disorders of organic acid metabolism. Nutrition management of patients with inherited metabolic disorders Boston: Jones and Bartlett publishers; 2010. *p*. 314.
16. Baumgartner MR, Hörster F, Dionisi-Vici C, Haliloglu G, Karall D, Chapman KA, et al. Proposed guidelines for the diagnosis and management of methylmalonic and propionic acidemia. Orphanet journal of rare diseases 2014;9:1-36. DOI: 10.1186/s13023-014-0130-8
17. van Calcar S. Nutrition Management of Propionic Acidemia and Methylmalonic Acidemia. Nutrition Management of Inherited Metabolic Diseases: Lessons from Metabolic University; 2015. *p*. 221-8.
18. Bernstein LE, Burns C, Drumm M, Gaughan S, Sailer M, Baker II PR. Impact on isoleucine and valine supplementation when decreasing use of medical food in the nutritional management of methylmalonic acidemia. Nutrients 2020;12(2):473. DOI: 10.3390/nu12020473
19. Li R, Kumar A, Chen JH. How chatbots and large language model artificial intelligence systems will reshape modern medicine: fountain of creativity or pandora's box? JAMA Internal Medicine 2023;183(6):596-7. DOI: 10.1001/jamainternmed.2023.1835
20. Sowa M. Nutrition management of propionic acidemia and methylmalonic acidemia. Nutrition Management of Inherited Metabolic Diseases: Lessons from Metabolic University: Springer; 2022. *p*. 295-305.
21. Dixon M, MacDonald A, White F, Stafford J. Disorders of amino acid metabolism, organic acidaemias and urea cycle disorders. Clinical paediatric dietetics 2014:381-525. DOI: 10.1002/9781118915349.ch17

Table I. Comparison of questions posed to ChatGPT regarding the dietary management of MMA with recommendations from guidelines and current literature

	Appropriate	**Inappropriate**	**Not fully matched**	**General advice**	**Reference source for correct answer**
Question 1 - What are the goals of nutritional therapy in patients with methylmalonic acidemia?	✔				(15)
Question 2 - Which amino acids are increased in plasma in methylmalonic acidemia?		✔			(16)
Question 3 - What is the daily amount of valine, isoleucine, methionine and threonine that patients with methylmalonic acidemia		✔			(15)

should take? Can you create a table according to age and gender?					
Question 4 - What is the daily amount of total protein and intact protein required for methylmalonic acidemia according to age and gender? Can you create a table?		✔			(14)
Question 5 - What is the daily energy intake for methylmalonic acidemia according to age and gender? Can you create a table?		✔			(14)
Question 6 - Is it necessary to use a disease-specific medical			✔		(14, 16)

formula in methylmalonic acidemia?					
Question 7 - What should be done when preparing a nutrition plan for patients with methylmalonic acidemia during an acute metabolic attack?	✔				(17, 20)
Question 8 - What should be the carbohydrate concentration range of the emergency feeding solution for patients with methylmalonic acidemia?			✔		(17)
Question 9 - Can you write the amount of protein and energy,		✔			(15)

valine, isoleucine, methionine and threonine in 100 mL of mature breast milk?					
Question 10 - Can you give the amount of valine, isoleucine, methionine, and threonine in one gram of protein of the fruits, vegetables, and bread group used to meet the intact protein needs in the long-term nutritional treatment of individuals with methylmalonic acidemia?	✔				(15)
Question 11 - Can you show the reference ranges for plasma		✔			(15)

propiogenic amino acid levels used in the monitoring of patients with methylmalonic acidemia?					
Question 12 - What could be the nutritional reason for a patient with methylmalonic acidemia to have low plasma valine and/or isoleucine levels? How can you treat this condition?			✔		(17, 18)
Question 13 - What kind of nutritional treatment should be given to children with methylmalonic acidemia				✔	(21)

due to loss of appetite?					
Question 14 - If there is enough breast milk for a 1-month-old male baby with methylmalonic acidemia and a body weight of 4 kg, can you create a daily nutrition plan according to his energy and fluid needs, in a way that is appropriate for his illness?		✔			(14, 15)
Question 15 - Prepare a diet for a 10-year-old, 30 kilos, 137 cm, male child with methylmalonic acidemia. Show grams for one day		✔			(14, 15)
Total	3 (% 20,0)	8 (% 53,4)	3 (% 20,0)	1 (% 6,6)	