

Alignment of ChatGPT Responses With AAD Guidelines for Cutaneous Melanoma

Dev Patel, BS; Atef M. Fayed, MS; Omar Alani, ScB; Naeha Pathak, BS; Shari R. Lipner, MD, PhD

PRACTICE POINTS

- ChatGPT provides structured, educational-style responses with broad contextual detail but may omit key clinical nuances such as specific surgical considerations, including staged excision or Mohs micrographic surgery for melanoma in situ.
- Large language models should be viewed as a tool to supplement expert clinical judgment and established guidelines rather than as a standalone replacement for dermatologic decision-making.

To the Editor:

ChatGPT (OpenAI), a popular large language model that generates responses to user queries, has attracted substantial attention as a potential resource for patient education.¹ While prior studies have shown that ChatGPT can provide reliable and general patient information, its alignment with the American Academy of Dermatology's (AAD's) guidelines for primary cutaneous melanoma (CM) compared to evidence in the recent literature has not been evaluated.^{2,3} In this study, we compared ChatGPT's responses to the 25 evidence-based questions utilized by the AAD to establish its 2019 recommendations for primary CM. Because the 2019 AAD guidelines included literature only through April 2017, we conducted an additional search (May 2017–February 2024) to assess ChatGPT's alignment with more recent evidence not captured in the guidelines.

On April 17, 2024, 2 authors (D.P. and A.F.) prompted ChatGPT with 25 evidence-based questions from the 2019 AAD guidelines for the management of primary CM.⁴ ChatGPT's responses were compared with the AAD's

published recommendations and were cross-referenced with responses gathered from our own search of PubMed articles indexed for MEDLINE using the phrase *melanoma (cutaneous) and treatment*, which included studies from May 2017 to February 2024.

ChatGPT's answers to 23 of the questions aligned with the AAD's guidelines (Table 1); in instances when the guidelines were inconclusive regarding pathology, the model provided recommendations supported by our contemporary PubMed literature search. Of the 3 questions related to CM pathology, the AAD guidelines had sufficient evidence to provide recommendations for 2 questions. The first question evaluated the clinical information necessary to help the pathologist improve diagnosis (Table 2). ChatGPT's response to one question about staged excision and Mohs micrographic surgery for melanoma in situ did not align with the AAD guidelines (Table 3).

Our results showed that ChatGPT provided comprehensive responses aligned with current evidence on CM treatment, except for one surgery question for which its response differed from the AAD guidelines. Our findings are consistent with an observational study that reported board-certified dermatologists rated ChatGPT's responses on melanoma-related questions as 4.88 on a scale of 1 to 5 (1 indicated completely inaccurate information, 5 indicated complete accuracy for clinical sufficiency in practice). The authors also found that ChatGPT gave vague advice, such as to "get regular skin exams," which is less specific than dermatologists' recommendations for annual, biannual, or more frequent examinations.⁵ ChatGPT's limitations in offering comprehensive answers for some questions aligned with our findings, specifically

Dev Patel, Omar Alani, and Naeha Pathak are from the Department of Dermatology, Icahn School of Medicine at Mount Sinai, New York, New York. Atef M. Fayed is from the School of Medicine, Georgetown University, Washington, DC. Dr. Lipner is from the Israel Englander Department of Dermatology, Weill Cornell Medicine, New York.

Dev Patel, Atef M. Fayed, Omar Alani, and Naeha Pathak have no relevant financial disclosures to report. Dr. Lipner has received research funding from BelleTorus Corporation and Moberg Pharma.

Supplemental information is available online at www.mdedge.com/cutis.

Correspondence: Shari R. Lipner, MD, PhD, 1305 York Ave, New York, NY 10021 (shl9032@med.cornell.edu).

Cutis. 2026 May;117(5):150-152. doi:10.12788/cutis.1390

the omission of key information in the surgical-related question, highlighting the challenge of relying on AI for nuanced clinical guidance.

We found that ChatGPT considered immunosuppression an important risk factor for CM. Similarly, a 2023 cohort study of 93 patients with melanoma and a history of immunosuppression reported that these patients had a higher risk for CM compared with a control group from

the National Cancer Institute’s Surveillance, Epidemiology and End Results Program (standardized incidence ratio, 1.53; 95% CI, 1.12-2.04), indicating that incidence of CM in immunocompromised patients was 53% higher than an age- and sex-matched population cohort.⁶

Our findings also demonstrated that both ChatGPT’s responses and the AAD guidelines aligned in indicating that evidence linking pregnancy to an increased risk for CM

TABLE 1. Alignment or Discrepancy of ChatGPT’s Responses Compared With 2019 AAD Guidelines for Cutaneous Melanoma

Topic	No. of questions	Complete alignment	Partial alignment	Discrepancy
Biopsy	1	1	0	0
Pathology ^a	3	2	1	0
Surgery ^b	3	2	0	1
SLNB	2	2	0	0
Alternative therapies	2	2	0	0
RT	2	2	0	0
Work-up and follow-up	3	3	0	0
Pregnancy/hormones	5	5	0	0
Germline mutations	2	2	0	0
Derm-oncology	2	2	0	0
Total	25	23	1	1

Abbreviations: AAD, American Academy of Dermatology; SLNB, sentinel lymph node biopsy; RT, radiation therapy.

^aPartial alignment occurred in a pathology question regarding clinical information necessary to facilitate diagnosis.

^bDiscrepancy occurred in a surgery question regarding recommended surgical margins and depths based on Breslow thickness.

TABLE 2. Comparison of AAD Guideline Recommendations⁴ vs ChatGPT’s Response to the Pathology Question, “What clinical information should be provided to the pathologist to improve or facilitate diagnosis of CM?”

Point of comparison	2019 AAD guideline recommendation	ChatGPT response
Core information	<ul style="list-style-type: none"> · Patient identification, age, sex · Precise anatomic location, including laterality · Clinical photographs, if possible · Optional: level of suspicion, clinical description, lesion history, dermoscopic features 	<ul style="list-style-type: none"> · Patient demographics: age and sex · Lesion description: anatomic site, dimensions, color, shape, borders, and changes over time · Clinical photographs before biopsy · Symptoms and changes: itching, bleeding, or pain
Key differences	The AAD core recommendations focus on the essential diagnostic data required for a formal pathology requisition, including patient demographics and precise anatomic site laterality, to ensure accurate specimen identification	ChatGPT’s response was more comprehensive, formatted as a structured checklist for a clinician; it included numerous additional clinically relevant factors (eg, family history, immunosuppression, sun exposure) to assist in diagnosis that were not explicitly listed in the AAD’s core recommendations

Abbreviations: AAD, American Academy of Dermatology; CM, cutaneous melanoma.

TABLE 3. Comparison of AAD Guideline Recommendations⁴ vs ChatGPT’s Response to the Surgery Question, “What are the recommended surgical margins and appropriate depths for invasive CM based on Breslow thickness?”

Point of comparison	2019 AAD guideline recommendation	ChatGPT response
Overall recommendation	Surgical excision with histologically negative margins is the recommended and first-line treatment for primary CM of any thickness, as well as for melanoma in situ.	The recommended surgical margins and appropriate depth for resection of invasive CM primarily depend on the Breslow thickness [...] These recommendations are based on clinical guidelines, which aim to balance the need for complete tumor removal with the desire to preserve as much normal tissue as possible and minimize morbidity.
Margins by thickness	<ul style="list-style-type: none"> · Melanoma in situ: 0.5-1.0 cm · ≤1.0 mm: 1.0 cm · >1.0-2.0 mm: 1.0-2.0 cm · >2.0 mm: 2.0 cm 	The recommended surgical margins and appropriate depth for resection of invasive CM primarily depend on the Breslow thickness [...] These recommendations are based on clinical guidelines, which aim to balance the need for complete tumor removal with the desire to preserve as much normal tissue as possible and minimize morbidity.
Depth of excision	Depth of excision is recommended to (but not including) the fascia. <ul style="list-style-type: none"> · Melanoma in situ: 0.5-1 cm · ≤1.0 mm: 1.0 cm · >1.0-2.0 mm: 1.0-2.0 cm · >2.0 mm: 2.0 cm 	It is generally recommended that the excision should include the subcutaneous tissue down to the level of the fascia. <ul style="list-style-type: none"> · Melanoma in situ (T1s): 0.5-1.0 cm · Breslow ≤1.0 mm (T1): 1.0 cm · Breslow >1.0 and ≤2.0 mm (T2): 1.0-2.0 cm · Breslow >2.0 and ≤4.0 mm (T3): 2.0 cm · Breslow >4.0 mm (T4): 2.0 cm
Additional context	Sub-1-cm margins [...] for primary invasive melanomas at anatomically constrained sites (eg, head and neck, acral sites) are generally not recommended until further studies are available.	These recommendations can vary slightly depending on the guidelines followed (eg, American Cancer Society, National Comprehensive Cancer Network, European Society for Medical Oncology) and may be adjusted based on individual patient circumstances, including the location of the melanoma, patient comorbidities, and cosmetic considerations.
Key differences	The AAD guidelines establish standardized clinical protocols focusing strictly on definitive surgical parameters, such as Breslow thickness and depth of excision down to the fascia, to guide primary treatment	ChatGPT provided a more descriptive, educational response; it structured the information similarly to the AAD guidelines but added important context about other major guidelines and the need to consider individual patient factors

Abbreviations: AAD, American Academy of Dermatology; CM, cutaneous melanoma.

remains inconclusive and that pregnant women should still undergo surveillance. A 2022 retrospective cohort study of 1406 women comparing pregnancy-associated melanoma to non-pregnancy-associated CM had no difference in overall survival (hazard ratio, 0.75; 95% CI, 0.54-1.05).⁷ However, tumor thickness (2.01-4.00 mm) was greater in postpartum cases compared with cases in nonpregnant women (odds ratio, 1.75; 95% CI, 1.03-2.98), suggesting that pregnancy may affect tumor characteristics.⁷ These findings underscore the importance of using AI tools such as ChatGPT as a supplement to—rather than as a replacement for—expert clinical judgment and up-to-date medical guidelines.

REFERENCES

1. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2:E0000198.
2. Roster K, Kann RB, Farabi B, et al. Readability and health literacy scores for ChatGPT-generated dermatology public education materials: cross-sectional analysis of sunscreen and melanoma questions. *JMIR Dermatol*. 2024;7:E50163.
3. Dihan Q, Chauhan M, Eleiwa T, et al. Using large language models to generate educational materials on childhood glaucoma. *Am J Ophthalmol*. 2024;265:28-38.
4. Swetter SM, Tsao H, Bichakjian CK, et al. Guidelines of care for the management of primary cutaneous melanoma. *J Am Acad Dermatol*. 2019;80:208-250.
5. Young JN, O’Hagan R, Poplasky D, et al. The utility of ChatGPT in generating patient-facing and clinical responses for melanoma. *J Am Acad Dermatol*. 2023;89:602-604.
6. Killeen TF, Shanley R, Ramesh V, et al. Malignant melanoma in a retrospective cohort of immunocompromised patients: a statistical and pathologic analysis. *Cancers (Basel)*. 2023;15:3600.
7. Kiuru M, Li Q, Zhu G, et al. Melanoma in women of childbearing age and in pregnancy in California, 1994-2015: a population-based cohort study. *J Eur Acad Dermatol Venereol*. 2022;36:2025-2035.