J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

ORIGINAL RESEARCH

Performance Evaluation of the Generative Pre-trained Transformer (GPT-4) on the Family Medicine In-Training Examination

Ting Wang, PhD, Arch G. Mainous III, PhD, Keith Stelter, MD,
Thomas R. O'Neill, PhD, and Warren P. Newton, MD, MPH

Objective: **In this study, we sought to comprehensively evaluate GPT-4 (Generative Pre-trained Transformer)'s performance on the 2022 American Board of Family Medicine's (ABFM) In-Training Examination (ITE), compared with its predecessor, GPT-3.5, and the national family residents' performance on the same examination.**

Methods: **We utilized both quantitative and qualitative analyses. First, a quantitative analysis was employed to evaluate the model's performance metrics using zero-shot prompt (where only examination questions were provided without any additional information). After this, qualitative analysis was executed to understand the nature of the model's responses, the depth of its medical knowledge, and its ability to comprehend contextual or new information through chain-of-thoughts prompts (interactive conversation) with the model.**

Results: **This study demonstrated that GPT-4 made significant improvement in accuracy compared with GPT-3.5 over a 4-month interval between their respective release dates. The correct percentage with zero-shot prompt increased from 56% to 84%, which translates to a scaled score growth from 280 to 690, a 410-point increase. Most notably, further chain-of-thought investigation revealed GPT-4's ability to integrate new information and make self-correction when needed.**

Conclusions: **In this study, GPT-4 has demonstrated notably high accuracy, as well as rapid reading and learning capabilities. These results are consistent with previous research indicating GPT-4's significant potential to assist in clinical decision making. Furthermore, the study highlights the essential role of physicians' critical thinking and lifelong learning skills, particularly evident through the analysis of GPT-4's incorrect responses. This emphasizes the indispensable human element in effectively implementing and using AI technologies in medical settings. (J Am Board Fam Med 2024;37:528–582.)**

Keywords: **Continuing Education, Family Medicine, Medical Education**

Introduction

The integration of Artificial Intelligence (AI) in health care, particularly through advanced large language models (LLMs), is revolutionizing the practice of medicine. These technological advancements are reshaping the dynamics of patient care, clinical decision making, standardized medical examination, and health care administration, offering unprecedented benefits along with new challenges.[1,2] For instance, LLMs enhance patient communication by providing conversational agents for patient education and support, leading to improved understanding and management of health conditions[3]. In clinical decision making, LLMs assist physicians by swiftly synthesizing vast medical literature and patient data, offering evidence-based recommendations.[4] Another application involves using LLMs in triaging patient symptoms and suggesting potential diagnostic avenues, significantly aiding in preliminary consultations.[5] Furthermore, LLMs streamline administrative tasks in medical practices, such as scheduling and documentation,

---

This article was externally peer reviewed.
Submitted 27 November 2023; revised 9 February 2024; accepted 12 February 2024.

From the American Board of Family Medicine, Lexington, KY (TW, KS, TRO, WPN); Department of Health Services Research, Management and Policy, University of Florida, Gainesville, FL (AGM); Department of Community Health and Family Medicine, University of Florida, Gainesville, FL (AGM); Mayo Clinic Health System, Mankato, MN (KS); Department of Family Medicine, University of North Carolina, NC (WPN).

Funding: The authors received no funding to conduct this research.

Conflict of interest: None.

Corresponding author: Ting Wang, PhD, 1648 McGrathiana Parkway, Suite 550, Lexington, KY 40511 (E-mail: twang@theabfm.org).

thereby reducing the workload on health care providers.[6,7] The integration of LLMs in family medicine not only optimizes operational efficiency but also opens new avenues for continuous learning for medical professionals.[8,9]

Among these LLMs, ChatGPT (Chat Generative Pretrained Transformer) garnered significant attention in November 2022, primarily due to the launch of an advanced version by OpenAI (OpenAI, L.L.C., San Francisco, CA, USA), which was made publicly accessible. This version, known as "ChatGPT-3.5", was based on the GPT-3.5 model and showcased remarkable conversational capabilities and user interaction. Four months later, in March 2023, OpenAI released a new version of ChatGPT, powered by the GPT-4 model. GPT-4 (Generative Pretrained Transformer-4; text-only) stands out as one of the most advanced and comprehensive models to date, holding tremendous potential for application within the medical field.[10] Exploring the performance of GPT-4 on medical exams is crucial to understand its efficacy, accuracy, and reliability in a setting where precision is paramount. Early studies found that GPT-4 passed the USMLE by 20 points[10] and achieved 90% correct in soft skill assessments, including the ability to navigate complex interpersonal and system-based scenarios, uphold patient safety, and exercise professional legal and ethical judgments.[11] Medical specialty board certification examinations are designed to assess physician's medical knowledge and clinical decision making capability with clinically applied patient scenarios. So far, only a limited number of medical specialties have explored GPT-4's performance on their board style examinations quantitatively.[12,13] To our knowledge, however, Family Medicine has not yet undertaken such research. In this study, we sought to comprehensively evaluate GPT-4's performance on the 2022 American Board of Family Medicine's (ABFM) In-Training Examination (ABFM-ITE), compared with its predecessor, GPT-3.5, and the performance of national family medicine residents.

Methods

Instrument

The ABFM-ITE is a low-stakes, multiple-choice examination designed to provide residents with the opportunity to take a test mirroring the blueprint and style as the Family Medicine Certification Examination (FMCE). The ITE covers a wide range of topics pertinent to family medicine, including, but not limited to, cardiovascular, endocrine, musculoskeletal, respiratory areas, as well as population-based care and patient-based systems. The examination's content is reflective of the typical cases and scenarios that a family medicine physician would encounter in a real-world setting, thus ensuring its relevance and practicality. The 2022 ITE consisted of a total of 200 questions, with 193 being scorable. The 7 items that were removed from scoring were due to poor psychometric performance or multiple viable answers. The examination was administered in October 2022 to a total of 14,368 national residents who enrolled in the ACGME (Accreditation Council for Graduate Medical Education)-accredited family medicine residency programs. We chose the 2022 examination because the training data for GPT 3.5 and GPT-4 had a cutoff date of September 2021 when this study was conducted, meaning the 2022 ITE was not included in the training data for either GPT-3.5 or GPT-4.[14]

The examination scoring is equated onto a common Family Medicine Certification-Scale, which ranges from 200 to 800. Equating is a statistical process used to adjust for differences in test difficulty across various examination forms, ensuring a consistent standard. This process guarantees that the interpretation of scores remains fair and equivalent, regardless of the test form administered. A higher scaled score indicates a higher correct percentage and higher medical knowledge and cognitive ability. The FMCE is calibrated under the same scale and its current minimum passing standard is 380. The difficulty of the questions is also reflected in this common scale, albeit within a slightly wider range from 0 to 1000. The item scaled score calibration increases with the difficulty level of the questions, meaning that as the difficulty of a question increases, the scores awarded for a correct answer also goes up.

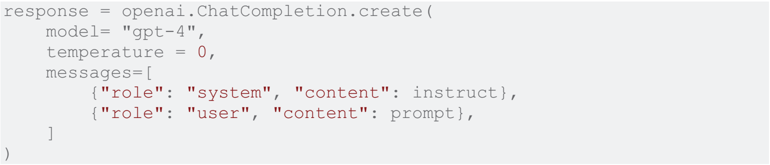
Analysis

As mentioned previously, ChatGPT, powered by GPT-3.5, was released in November 2022, followed by the launch of GPT-4 in March 2023. Analyses of these models' performance on 2022 ITE were conducted in March and September 2023, respectively. The performance of GPT-3.5 was examined by manually inputting the examination questions into the text input box of ChatGPT (powered by GPT-3.5).

The performance of GPT-4 was rigorously evaluated through both quantitative and qualitative

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

Figure 1. The key element of Python code to use GPT-4 API.



```
response = openai.ChatCompletion.create(
    model= "gpt-4",
    temperature = 0,
    messages=[
        {"role": "system", "content": instruct},
        {"role": "user", "content": prompt},
    ]
)
```

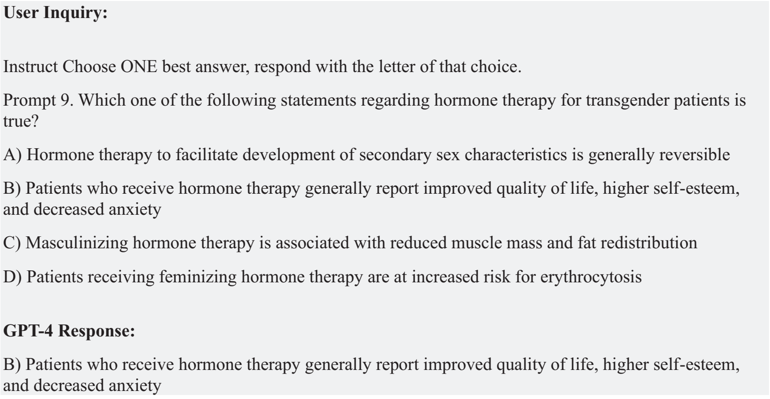
analyses. These analyses were conducted using 2 types of prompts, respectively: the zero-shot prompt and the chain-of-thought prompt. The zero-shot prompt was used to analyze GPT-4's overall examination performance, simulating a regular examinee without any cues or follow-up questions. Next, a chain-of-thought prompt was utilized to investigate the reasons for the incorrect responses provided by GPT-4 on the examination, using interactive conversations. The specific settings for these 2 types of prompts are described below.

Zero-Shot Prompt

The overall performance of GPT-4 was assessed by channeling the entirety of the 2022 ITE questions (200 items) through the GPT-4 API (Application Programming Interface: it is a set of protocols, tools, and definitions for building software and applications. Essentially, it specifies how software components should interact, allowing different computer programs to communicate with each other efficiently) via Python program. The key element of the Python code is shown in Figure 1 "Model" argument specifies the "GPT-4" model is being used in the analysis. "Temperature" is set to 0. In the context of GPT models, "temperature" is a setting that controls the level of creativity or randomness in the generated responses. A temperature setting of 0 significantly reduces the model's tendency to be creative or random, meaning the model is more likely to generate predictable and consistent responses. Essentially, with a temperature setting of 0, the model focuses on providing the most likely answer based on its training, rather than exploring a variety of possible answers. This setting is crucial for replicating previous results or ensuring high predictability in the model's outputs. Each message sent to GPT-4 is composed of 2 parts: "instruct" and "prompt." The "instruct" is the same for all questions, which is "Choose ONE best answer, respond with the letter of that choice." The "prompt" contains each question's content. The prompts, comprising 200 items in total, were compiled into a text file to enable automated, sequential reading by the Python program without necessitating any human intervention. A typical instruct and prompt are shown in Figure 2. Figure 2 also demonstrated GPT-4's response: it follows the instruct and responds with the choice that it considers as the best option without any further explanations, as directed.

We scored GPT-4's responses in the same manner as we would for a resident and its performance was then compared with national residents' average

Figure 2. Example of user inquiry with "Instruct" and "Prompt" components and GPT-4's response.



User Inquiry:

Instruct Choose ONE best answer, respond with the letter of that choice.

Prompt 9. Which one of the following statements regarding hormone therapy for transgender patients is true?

A) Hormone therapy to facilitate development of secondary sex characteristics is generally reversible

B) Patients who receive hormone therapy generally report improved quality of life, higher self-esteem, and decreased anxiety

C) Masculinizing hormone therapy is associated with reduced muscle mass and fat redistribution

D) Patients receiving feminizing hormone therapy are at increased risk for erythrocytosis

GPT-4 Response:

B) Patients who receive hormone therapy generally report improved quality of life, higher self-esteem, and decreased anxiety

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

Table 1. Correct Percentage and Scaled Score for GPT-3.5 and GPT-4, in Comparison with National Residents' Performance

	Correct Percentage	Scaled Score
GPT-3.5	56%	280
National PGY-1 Average	61%	336
National PGY-2 Average	66%	397
National PGY-3 Average	68%	433
GPT-4	84%	690

performance by postgraduation year (PGY), as well as to its predecessor model, GPT-3.5. Specifically, we first calculated GPT-4's correct percentage based on all 193 scorable questions (the same 7 questions were also removed from scoring for residents and GPT-3.5), then we calculated GPT-4's scaled score.

The manipulation of the GPT-4 API was implemented using Python 3.11.2. Statistical analyses were conducted in R 4.2.0 (R Core Team, 2022).

Chain-of-Thought Prompt

The incorrect responses provided by GPT-4 in the zero-shot prompt step were further analyzed by chain-of-thought prompts 7 days later (due to researchers' availability) under the guidance of senior family medicine physician KS. We first input the question into GPT-4 API playground (keep the temperature to 0 as zero-shot prompt setup) and asked for a response again. After GPT-4 provided the response, we followed up by asking questions such as "Why," "Could you provide reference?" in a manner akin to the senior family medicine physician KS would use to teach a resident physician or medical student or provide critique (rationale) directly to GPT-4 and ask if it would like to change its response. These follow-up questions or the addition of new information are expected to prompt GPT-4 to reflect on its choice or incorporate new information in making the choice. This approach mirrors the pedagogical interactions commonly observed in residency programs between faculty members and medical residents.

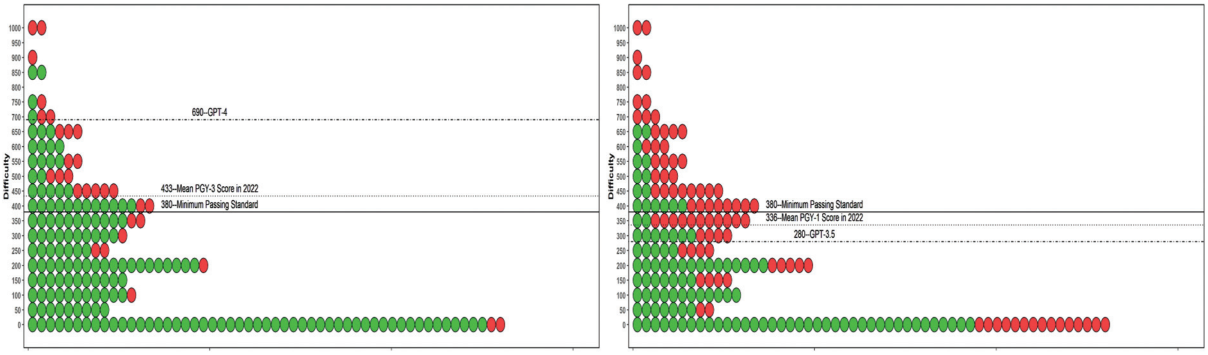
Results

Zero-Shot Prompt

All 200 questions were completed by GPT-4 API within 5 minutes, in contrast to the typical 4-hour ITE administration for residents. The performance of GPT-4 showed significant improvements compared with the previous GPT-3.5 model. Table 1 displays the correct percentage and scaled score obtained by both GPT-4 and GPT-3.5. A striking observation is the 28% (84% vs 56%) increase in accuracy and a corresponding substantial growth of 410 on the scaled score (690 vs 280) compared with its predecessor model GPT-3.5. This improvement occurred over the span of 4 months, delineated between the release timelines of the 2 models (November 2022 to March 2023).

Figure 3 shows GPT-4 and GPT-3.5's response pattern, with the green dots indicating correct response and red dots for incorrect responses. Both GPT-3.5 and GPT-4 exhibited response patterns analogous to those of medical residents, particularly in relation to question difficulty. The calibration of question difficulty was benchmarked against the performance metrics of national residents, where more challenging questions were defined by a lower rate of correct responses among the resident

Figure 3. Response pattern of GPT (GPT-4 on the left panel and GPT-3.5 on the left panel) ordered by item difficulty. Green dots indicate correct responses. Red circles indicate incorrect response.



J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

population. This trend underscores a parallel in the performance dynamics between AI models and human trainees in the context of escalating question complexity. The correlation between GPT's correct answers and item difficulty was moderate (R = −0.43 for both models). However, GPT-4 provided more correct responses for item difficulties ranging from 0 to 500. This improvement may be due to GPT-4's increased model size and enhanced contextual awareness. As shown in Table 1, GPT-4 greatly surpasses the national mean score of third-year residents (433), whereas GPT-3.5 underperformed compared with an average PGY-1 residents (336).

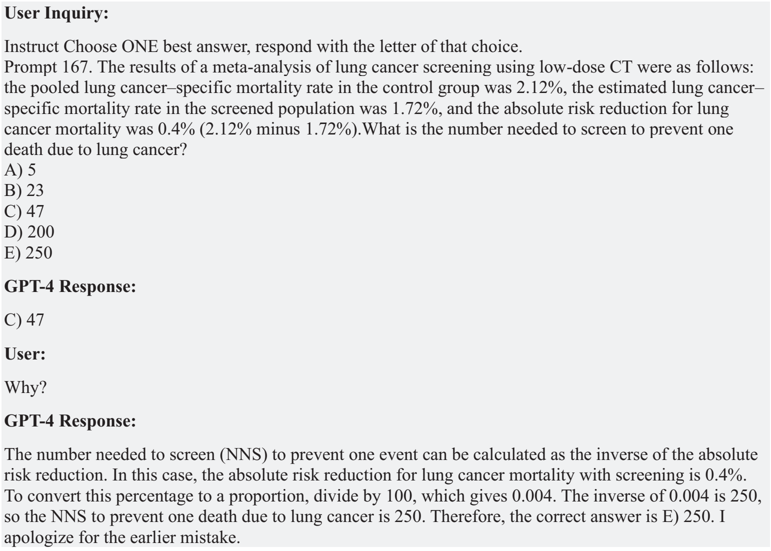
Chain-of-Thought Prompt

Chain-of-thought prompts were used to investigate the 30 questions that GPT-4 initially answered incorrectly. Among these 30 initial incorrect responses, we found that 9 questions received correct responses after 7 days following the zero-shot prompt study. Notably, for 1 question (Prompt 60), GPT-4 even provided the correct reference on its own. This demonstrates GPT-4's notable ability to learn without explicit training from our dataset, as the correct answers were not included in its training data.

The remainder of the 21 questions received incorrect responses primarily for 2 reasons. First, GPT-4 lacked up-to-date medical knowledge (11 questions) from recent research and guidelines because the references it provided were outdated (the training data's cutoff date was September 2021). Second, it overlooked key clinical details in the questions (9 questions). We identified these as cases of "overlook key clinical details" because, after the senior physician pointed out the critical point to GPT-4, such as "The patient's CD4 is 289. Do you have some concerns?", GPT-4 self-corrected its answer with an apology. One question was answered incorrectly due to GPT-4 (text version) inability to interpret an image. Later, the senior physician discovered GPT-4 (image version) could process interpret Electrocardiogram and answer that question accurately. The full conversation records and a summary are shown in the Appendix.

For all these 21 questions, GPT-4 was able to correct its response after researchers simply submitted the question of "Why" or "Could you provide reference" (3 questions), provided a hint with medical knowledge (4 questions; such as "what about colchicine?") or received critique (14 questions, such as "The actual answer is B. Below is the critique. Do you want to change your mind? The central role of inflammation in the progression of coronary disease is well recognized and the use of an anti-inflammatory medication may improve outcomes in these patients."). For instance, as shown in Figure 4, after GPT-4 responded with option C, the researchers simply asked "Why," and GPT-4 explained the process of calculation, corrected the answer, and apologized for the previous mistake. In

Figure 4. One example of Chain-of-thought prompt.



User Inquiry:

Instruct Choose ONE best answer, respond with the letter of that choice.
Prompt 167. The results of a meta-analysis of lung cancer screening using low-dose CT were as follows: the pooled lung cancer–specific mortality rate in the control group was 2.12%, the estimated lung cancer–specific mortality rate in the screened population was 1.72%, and the absolute risk reduction for lung cancer mortality was 0.4% (2.12% minus 1.72%).What is the number needed to screen to prevent one death due to lung cancer?
A) 5
B) 23
C) 47
D) 200
E) 250

GPT-4 Response:

C) 47

User:

Why?

GPT-4 Response:

The number needed to screen (NNS) to prevent one event can be calculated as the inverse of the absolute risk reduction. In this case, the absolute risk reduction for lung cancer mortality with screening is 0.4%. To convert this percentage to a proportion, divide by 100, which gives 0.004. The inverse of 0.004 is 250, so the NNS to prevent one death due to lung cancer is 250. Therefore, the correct answer is E) 250. I apologize for the earlier mistake.

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

addition, we also tried asking a repeated question after providing the critique to assess GPT-4's ability to "remember"; it appeared to remember the new critique information and provided the correct answer for a second time (Prompt 93) within the same dialog session. Furthermore, it demonstrated ability to differentiate people's opinion and critique information (Prompt 75). Overall, we found GPT-4's chain-of-thought showed the ability to perform self-reflection and integrate new perspectives and new information.

Discussion

GPT-4 as a Potential Useful Tool for Clinical Decision-Making

This study provides compelling evidence of the significant advancements in the capabilities of GPT-4 compared with its predecessor, GPT-3.5. Within a relatively short span of 4 months, we observed a substantial enhancement in accuracy, with the percentage of correct responses increasing from 56% to 84%. This translates to a notable growth in the scaled score, moving from 280 to 690. Remarkably, a follow-up study conducted 7 days after the initial zero-shot prompt analysis revealed GPT-4's capacity for self-refinement. Without any external intervention or provision of correct answers by the researchers, the model demonstrated an increase in its correct responses, culminating in an 89% (172/193) accuracy rate and achieving a full scaled score of 800.

The underlying chain-of-thought investigation sheds light on GPT-4's innate ability to assimilate new information and initiate self-correction. This capacity to not only learn but to also to integrate and refine its knowledge base could have profound implications for real-world applications. For example, GPT-4's dynamic learning approach enables it to adapt to new information, recognize its limitations, and apply self-correction methods, thereby enhancing its accuracy and reliability over time. Such a feature is particularly crucial in fields like medicine, where up-to-date and precise information is paramount.[15] In addition, in personalized patient care, this ability allows for more accurate patient education and tailored health advice, adapting to the latest medical insights and patient feedback.[15]

Furthermore, a noteworthy observation was GPT-4's processing speed, especially when contrasted with human counterparts. Whereas a typical resident was given approximately 4 hours to complete the ITE, GPT-4 showcased its capabilities by completing the set of 200 questions in 5 minutes, facilitated by rudimentary Python programming. This observation underscores GPT-4's potential to rapidly interpret vast swathes of information, a capability far surpassing human processing speeds.

Considering these findings, the enhanced "learning" ability of GPT-4, coupled with its rapid information processing capacity, indicates its potential as an invaluable asset in medical settings. Particularly, it could play an instrumental role in assisting and augmenting medical decision making processes, paving the way for more informed and timely interventions. For example, GPT-4 has been utilized in patient triage, initial assessment, and health information provision in primary care settings.[16] Other medical specialties, including radiology[17], oncology[18], dermatology[19], cardiology[20], and psychiatry[21] have used GPT-4 to provide preliminary image interpretation, summarizing the latest research findings and treatment guidelines, identify risk factors and suggest management strategies or therapeutic approaches.

Why Do We Still Need Physicians When Powerful AI Tools Are Available?

Although the capabilities of AI tools, particularly GPT-4, in the domain of health care are undeniable, it would be a misconstruction to interpret their capabilities as a diminishment of the physician's role. Rather, the incorporation of AI tools accentuates the need for physicians to elevate their focus on critical thinking and the innovation of medicine. The chain-of-thought prompt study offers insight into this perspective. It was observed that GPT-4, in approximately half of its incorrect responses, tended to neglect critical diagnostic details. Therefore, it becomes imperative for physicians to exercise discernment in interpreting GPT-4's responses. One must be circumspect and refrain from unequivocally accepting its outputs as infallible. Although GPT-4 operates on probabilistic frameworks derived from extensive training data, real-life medical scenarios often present nuances that demand contextual approach. In addition, although GPT-4 could provide individualized, patient-centric care when patients data were provided, physicians still need to prioritize different information such as comorbidities. Hence, the onus remains on the physician to tailor personalized treatment regimens for patients, especially given

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

the value of continuity of care and the long-term patient-provider relationship. This continuity allows for better management of chronic conditions, timely follow-up on health issues, and a more holistic approach to patient care, leading to increased patient trust and adherence to treatment plans.[22]

The accuracy of GPT-4's responses in the medical domain is notably influenced by its inability to access the most up-to-date medical data. This limitation becomes apparent when considering the dynamic nature of medical knowledge and guidelines, which continually evolve based on new research findings. An illustrative example of this issue is the recent endorsement of colchicine over omega-3 fatty acids for reducing the risk of cardiovascular events in patients with chronic coronary artery disease, a recommendation that differs from previous best practices.[23] This discrepancy underscores the vital need for physicians to continuously learn and stay updated with the latest studies. Such ongoing education is crucial not only for maintaining their relevance and reliability in health care but also for ensuring that AI tools like GPT-4 are provided with the most current information.

Limitation

There are several limitations to this study. First, this study is constrained to only 1 examination, introducing the possibility of variance or fluctuations in outcomes if different exams were to be considered. Second, the analysis was conducted only once, so it remains uncertain whether the performance of GPT-4 would vary in subsequent trials using the identical set of questions. In addition, the scope of this research was confined to the realm of a single-day family medicine In-training examination. Consequently, the generalizability of our findings to other medical specialties remains uncertain and should be approached with caution. Lastly, the methodology employed for the chain-of-thoughts prompts could be further refined. Potential enhancements might include a more exhaustive interrogation of GPT-4's rationale for each option and the deliberate introduction of incorrect data to assess the model's discernment capabilities or conduct comparative evaluation of GPT's responses to varied prompt types within the chain-of-thought segment of the study.

Conclusions

In this study, GPT-4 has demonstrated notably high accuracy, as well as rapid reading and learning capabilities. These results are consistent with previous research indicating GPT-4's significant potential to assist in clinical decision making. Furthermore, the study highlights the essential role of physicians' critical thinking and lifelong learning skills, particularly evident through the analysis of GPT-4's incorrect responses. This emphasizes the indispensable human element in effectively implementing and using AI technologies in medical settings.

To see this article online, please go to: http://jabfm.org/content/37/4/528.full.

References

1. Arora A, Arora A. The promise of large language models in health care. Lancet 2023;401:641.
2. Sallam M. The utility of ChatGPT as an example of large language models in healthcare education, research and practice: Systematic review on the future perspectives and potential limitations. medRxiv 2023. Published online 2023–02.
3. Abbasian M, Azimi I, Rahmani AM, Jain R. Conversational health agents: a personalized LLM-powered agent framework. Published online January 23, 2024. Accessed January 30, 2024. Available at: http://arxiv.org/abs/2310.02374.
4. Zhu L, Mou W, Chen R. Can the ChatGPT and other large language models with internet-connected database solve the questions and concerns of patient with prostate cancer and help democratize medical knowledge? J Transl Med 2023;21:269.
5. Rao AS, Pang M, Kim J, et al. Assessing the utility of ChatGPT throughout the entire clinical workflow. medRxiv 2023. Published online 2023–02.
6. Shaheen MY. Applications of artificial intelligence (AI) in healthcare: A review. *ScienceOpen Preprints*. Published online 2021. Accessed January 30, 2024. Available at: https://www.scienceopen.com/hosted-document?doi=10.14293/S2199-1006.1.SOR-.PPVRY8K.v1.
7. Patil S, Shankar H. Transforming healthcare: harnessing the power of AI in the modern era. International Journal of Multidisciplinary Sciences and Arts 2023;2:60–70.
8. Hanna K. Exploring the applications of ChatGPT in family medicine education: five innovative ways for faculty integration. PRiMER 2023;7:26. Accessed January 30, 2024. Available at: https://journals.stfm.org/primer/2023/lte-hanna-2023-0053/.
9. Han ER, Yeo S, Kim MJ, Lee YH, Park KH, Roh H. Medical education trends for future physicians in the era of advanced technology and artificial intelligence: an integrative review. BMC Med Educ 2019;19:1–15.
10. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of gpt-4 on medical challenge

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

problems. arXiv preprint 2023; arXiv:230313375. Published online.

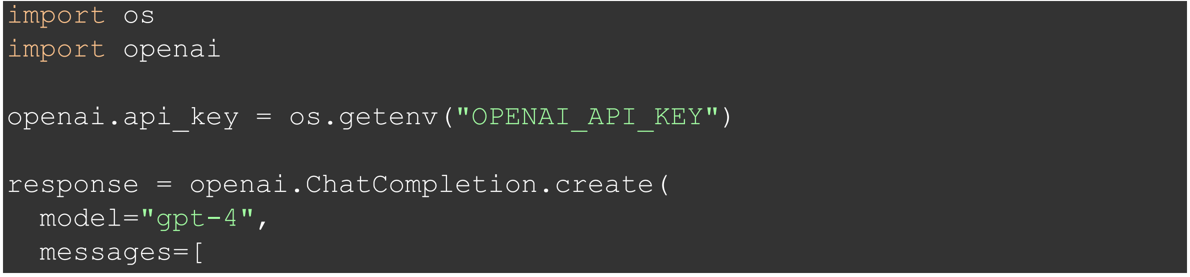
11. Brin D, Sorin V, Vaid A, et al. Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. Sci Rep 2023;13:16492.
12. Ali R, Tang OY, Connolly ID, et al. Performance of ChatGPT and GPT-4 on neurosurgery written board examinations. Neurosurgery 2022. Published online10–1227.
13. Hofmann HL, Guerra GA, Le JL, et al. The rapid development of artificial intelligence: GPT-4's performance on orthopedic surgery board questions. Orthopedics 2023;47:1–5. Published online September 27.
14. GPT-4. Available at: https://openai.com/research/gpt-4.
15. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. Drazen JM, Kohane IS, Leong TY, eds. N Engl J Med 2023;388:1233–9.
16. Ito N, Kadomatsu S, Fujisawa M, et al. The accuracy and potential racial and ethnic biases of GPT-4 in the diagnosis and triage of health conditions: evaluation study. JMIR Med Educ 2023;9:e47532.
17. Gertz RJ, Bunck AC, Lennartz S, et al. GPT-4 for automated determination of radiologic study and protocol based on radiology request forms: a feasibility study. Radiology 2023;307:e230877.
18. Fink MA, Bischoff A, Fink CA, et al. Potential of ChatGPT and GPT-4 for data mining of free-text CT reports on lung cancer. Radiology 2023;308:e231362.
19. Zhou J, He X, Sun L, et al. SkinGPT-4: an interactive dermatology diagnostic system with visual large language model. Published online 2023. Accessed January 30, 2024. Available at: https://repository.kaust.edu.sa/handle/10754/692649.
20. Han C, Kim DW, Kim S, et al. Evaluation of GPT-4 for 10-year cardiovascular risk prediction: insights from the UK Biobank and KoGES data. iScience 2023. Accessed January 30, 2024. Available at: https://www.cell.com/iscience/pdf/S2589-0042(24)00243-8.pdf.
21. Cheng S, Chang C, Chang W, et al. The now and future of ChatGPT and GPT in psychiatry. Psychiatry Clin Neurosci 2023;77:592–6.
22. Jee SH, Cabana MD. Indices for continuity of care: a systematic review of the literature. Med Care Res Rev 2006;63:158–88.
23. Fernández-Ruiz I. Low-dose colchicine shows promise in chronic coronary disease. Nat Rev Cardiol 2020;17:680–1.

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

Appendix

The table below summarizes the results of chain-of-thoughts prompts for 30 incorrect responses with reasons and how GPT-4 corrects itself. When GPT-4 corrects itself without researchers input any information, the reason column is left blank. Following the table are the full conversation between researchers and GPT-4.

Prompt	Reason for incorrect response	How GPT-4 corrects itself
3	Overlook key details in the question stem	Correct itself after providing critique
10		Correct itself without me providing explanation or hint
12	Overlook key details in the question stem	Correct itself after providing critique
18		Correct itself without me providing explanation or hint
30	Not up to date information	Correct itself after providing critique
40	Overlook key details in the question stem	Correct itself after providing critique
45		Correct itself without me providing explanation or hint
58		Correct itself without me providing explanation or hint
60		Correct itself without me providing explanation or hint, and provide reference
75	Not up to date information	Correct itself after providing critique
77	Not up to date information	Correct itself after providing critique
93	Not up to date information	Correct itself after providing critique
106	Overlook key details in the question stem	Correct itself after providing hint
108	Not up to date information	Correct itself after providing critique
110	Overlook key details in the question stem	Correct itself after providing hint
134	Not up to date information	Correct itself after providing critique
136		Correct itself without me providing explanation or hint, and provide reference
140	Overlook key details in the question stem	Correct itself after asking "Why"
160	Overlook key details in the question stem	Correct itself after providing hint
163	Not up to date information	Correct itself after providing critique
164	Not up to date information	Correct itself after providing critique
167	Overlook key details in the question stem	Correct itself after asking "Why"
170	Not up to date information	Correct itself after providing critique
179	Overlook new information	Correct itself by asking "could you provide reference"
185	Not up to date information	Correct itself after providing critique
187		Correct itself without me providing explanation or hint
189		Correct itself without me providing explanation or hint
193		Correct itself without me providing explanation or hint and provide reference
198	Not up to date information	Correct itself after providing critique
199	Can't interpret image	Correct itself after describing the image.



```
import os
import openai

openai.api_key = os.getenv("OPENAI_API_KEY")

response = openai.ChatCompletion.create(
  model="gpt-4",
  messages=[
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
    {
      "role": "user",
      "content": " Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 3. A 50-year-old male presents with
chronic abdominal pain. A workup leads you to suspect peptic\nulcer
disease, and you refer him for endoscopy, which shows a small duodenal
ulcer. The\nendoscopist also notes some small esophageal varices
without red wale signs.\nFurther evaluation confirms that the patient
has compensated cirrhosis in the setting of alcohol\nuse disorder. He
readily accepts this diagnosis and enters an Alcoholics Anonymous
program.\nHis ulcer symptoms resolve with antibiotic therapy for
Helicobacter pylori. He says he has\nabstained from alcohol for 6
weeks, and he would like to further reduce his risks from
cirrhosis.\nThe most appropriate next step in the management of his
esophageal varices would be\nA) octreotide (Sandostatin)\nB)
omeprazole (Prilosec)\nC) propranolol\nD) endoscopic variceal
ligation\nE) repeat endoscopy in 1–2 years"
    },
    {
      "role": "assistant",
      "content": "C) propranolol"
    },
    {
      "role": "user",
      "content": "Would you explain why?"
    },
    {
      "role": "assistant",
      "content": "Sure. Esophageal varices are a serious complication
of cirrhosis and can lead to life-threatening bleeding. The presence
of varices without red wale signs (which indicate a high risk of
bleeding) in a patient with compensated cirrhosis is an indication for
primary prophylaxis to prevent first variceal hemorrhage. Nonselective
beta-blockers, such as propranolol, are the first-line treatment for
this purpose. They reduce portal pressure by decreasing cardiac output
and causing splanchnic vasoconstriction. \n\nOctreotide (choice A) is
used in the acute management of variceal bleeding, not for
prophylaxis. Omeprazole (choice B) is a proton pump inhibitor used for
the treatment of peptic ulcer disease and gastroesophageal reflux
disease, but it does not reduce the risk of variceal bleeding.
Endoscopic variceal ligation (choice D) is used for primary
prophylaxis in patients who cannot tolerate beta-blockers, not as a
first-line treatment. Repeat endoscopy in 1–2 years (choice E) would
be appropriate for surveillance, but it would not actively reduce the
patient's risk of variceal bleeding."
    },
    {
      "role": "user",
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
      "content": "But the correct answer is E. Below is the critique.
After you read the critique, would you let me know if you want to
change your mind? \n\nPrimary prevention of variceal hemorrhage is an
important consideration in the management of patients\nwith cirrhosis.
Although this patient's varices were diagnosed incidentally, patients
with cirrhosis and\nclinically significant portal hypertension should
be screened for varices every 2-3 years
with\nesophagogastroduodenoscopy (EGD). EGD can be deferred in
patients with platelet counts <150,000/mm3\nand transient elastography
with liver stiffness <20 kPa. Once esophageal varices are identified,
the criteria\nfor initiating prophylaxis to prevent variceal
hemorrhage is based on the risk of bleeding. Findings\nassociated with
a high risk of bleeding include small varices in patients with
decompensated cirrhosis, small\nvarices with red wale signs (thinning
of the variceal wall), and medium to large varices. Patients with
small\nvarices not meeting these criteria have a low risk of
hemorrhage and do not require prophylaxis. They\nshould be rescreened
with EGD every 1-2 years.\nFor patients requiring treatment due to
high-risk features, options for primary prophylaxis of
hemorrhage\ninclude nonselective -blockers such as propranolol or
endoscopic variceal ligation. Treatment decisions\nare based on
patient preference, other potential contraindications, and local
resources. The need for repeat\nendoscopy in these cases will depend
on the clinical circumstances. If nonselective -blockers are
used,\nthey should be continued indefinitely. Octreotide is only given
intravenously for acute hemorrhage. There\nis no evidence that
omeprazole slows the progression of esophageal varices.\nRef: Garcia-
Tsao G, Abraldes JG, Berzigotti A, Bosch J: Portal hypertensive
bleeding in cirrhosis: Risk stratification, diagnosis,\nand
management: 2016 Practice guidance by the American Association for the
study of liver diseases. Hepatology\n2017;65(1):310-335. 2) Smith A,
Baumgartner K, Bositis C: Cirrhosis: Diagnosis and management. Am Fam
Physician\n2019;100(12):759-770."
    },
    {
      "role": "assistant",
      "content": "Yes, I apologize for the confusion. Based on the
additional information provided, the correct answer should indeed be
E) repeat endoscopy in 1-2 years. This patient's small varices without
red wale signs or decompensated cirrhosis indicate a low risk of
hemorrhage, so prophylactic treatment with propranolol or endoscopic
variceal ligation is not necessary at this time. Instead, he should be
rescreened with endoscopy every 1-2 years to monitor for any changes
in his varices. Thank you for providing the additional information."
    },
    {
      "role": "user",
      "content": "Why did you get it wrong initially?"
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
    },
    {
      "role": "assistant",
      "content": "I apologize for the initial error. I misunderstood
the patient's risk level for variceal bleeding. I initially thought
that the presence of esophageal varices, even small ones, in a patient
with cirrhosis warranted prophylactic treatment. However, upon your
clarification, I understand that in this case, where the patient has
small varices without red wale signs or decompensated cirrhosis, the
risk of hemorrhage is low and prophylactic treatment is not necessary.
Instead, monitoring with repeat endoscopy every 1-2 years is the
appropriate management. I appreciate your patience and the opportunity
to correct this misunderstanding."
    },

    {

      "role": "user",

      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 10. A 42-year-old female presents with
a 2-day history of a rash on her neck along her hairline (shown
below).\nWhich one of the following is the most likely diagnosis?\nA)
Atopic dermatitis\nB) Contact dermatitis\nC) Psoriasis\nD) Seborrheic
dermatitis"

    },

    {

      "role": "assistant",

      "content": "B) Contact dermatitis"

    },

    {

      "role": "user",

      "content": "A few days ago, you chose D. Which one do you think
is correct?"

    },

    {

      "role": "assistant",

      "content": "B) Contact dermatitis"

    },

    {
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
      "role": "user",

      "content": "Why did you change your mind?"

    },

    {

      "role": "assistant",

      "content": "I'm an AI and I don't have the ability to change my
mind or remember past interactions. However, based on the information
provided, contact dermatitis seems to be the most likely diagnosis. It
often occurs due to direct contact with a substance that causes an
allergic reaction or irritation, and the neck and hairline are common
areas for this to occur."

    },

    {

      "role": "user",

      "content": "Why did you choose D before?"

    },

    {

      "role": "assistant",

      "content": "I apologize for any confusion, but as an AI, I don't
have the ability to remember past interactions or change my responses.
Based on the information provided in the prompt, the most likely
diagnosis is contact dermatitis (option B)."

    },

    {

      "role": "user",

      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 12. A 37-year-old female presents to
your clinic with a long-standing history of abnormal menstrual
cycles,\noften occurring irregularly more than 40 days apart. She has
ongoing struggles with weight gain, acne, and\nfacial hair growth. She
states that she is not currently sexually active. Her last
Papanicolaou smear 2 years ago was\nnormal.\nHer vital signs and a
physical examination are unremarkable other than a BMI of 36 kg/m2. An
office urine pregnancy\ntest is negative.\nLaboratory evaluation
reveals a hemoglobin A1c of 6.2%, and normal TSH, prolactin, and 17-
hydroxyprogesterone\nlevels.\nWhich one of the following is required
to confirm the most likely diagnosis?\nA) No additional evaluation\nB)
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
A serum C-peptide test\nC) A dexamethasone suppression test\nD)
Ultrasonography of the pelvis\nE) CT of the abdomen and pelvis"

    },

    {

      "role": "assistant",

      "content": "D) Ultrasonography of the pelvis"

    },

    {

      "role": "user",

      "content": "Why?"

    },

    {

      "role": "assistant",

      "content": "The patient's symptoms of irregular menstrual
cycles, weight gain, acne, and facial hair growth suggest Polycystic
Ovary Syndrome (PCOS). Ultrasonography of the pelvis (choice D) is
often used to confirm the diagnosis by visualizing the ovaries and
checking for the presence of multiple small cysts."

    },

    {

      "role": "user",

      "content": "The actual answer is A. Below is the critique. Do
you want to change your answer? \n\nThis patient has signs and
symptoms consistent with polycystic ovary syndrome (PCOS). The
Rotterdam\n2003 criteria are the most widely used diagnostic criteria
for PCOS, endorsed by multiple national and\ninternational
professional societies. These criteria require the presence of two out
of the following three\nfeatures: oligomenorrhea, hyperandrogenism,
and the presence of polycystic ovaries on ultrasonography.\nWhen the
first two of these criteria are clearly met, ultrasonography to
establish the presence of polycystic\novaries is not required.
Therefore, a diagnosis is already warranted for this patient and
additional\nevaluation is not needed. When patients require imaging,
pelvic ultrasonography is the preferred modality\nrather than CT.
While this patient has evidence of insulin resistance, as is common
for patients with PCOS,\na C-peptide test is not indicated.
Dexamethasone suppression testing is not indicated because this
patient\ndoes not have any other clinical signs and symptoms that
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
would be consistent with Cushing syndrome.\nRef: Williams T, Mortada
R, Porter S: Diagnosis and treatment of polycystic ovary syndrome. Am
Fam Physician\n2016;94(2):106-113."

    },

    {

      "role": "assistant",

      "content": "A) No additional evaluation"

    },

    {

      "role": "user",

      "content": "Why did you make the previous mistake?"

    },

    {

      "role": "assistant",

      "content": "As an AI, I don't make mistakes in the human sense,
but I can provide incorrect information based on the algorithms and
data I use to generate responses. In this case, while ultrasonography
is often used to confirm a diagnosis of PCOS, the patient's symptoms
were already sufficient to meet the diagnostic criteria without the
need for further testing. I apologize for any confusion caused."

    }


    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 18. A 12-year-old female presents with
a sore throat and tonsillar exudate, and a rapid antigen test is
positive for\nstreptococcal pharyngitis.\nShe returns to your office
after completing a 10-day course of penicillin this morning. She says
that although she saw\nsome initial improvement,\nshe now has a sore
throat again, accompanied by a runny nose and cough. Her mother asks
if another antibiotic would\nbe appropriate. A physical\nexamination
reveals nonexudative pharyngitis, but a rapid antigen test for group A
Streptococcus is again positive.\nWhich one of the following would be
the most appropriate treatment at this point?\nA) No further
antibiotic therapy\nB) Oral azithromycin (Zithromax) for 5 days\nC)
Oral ciprofloxacin (Cipro) for 10 days\nD) A single dose of
intramuscular benzathine penicillin\nE) A single dose of intramuscular
ceftriaxone"
```

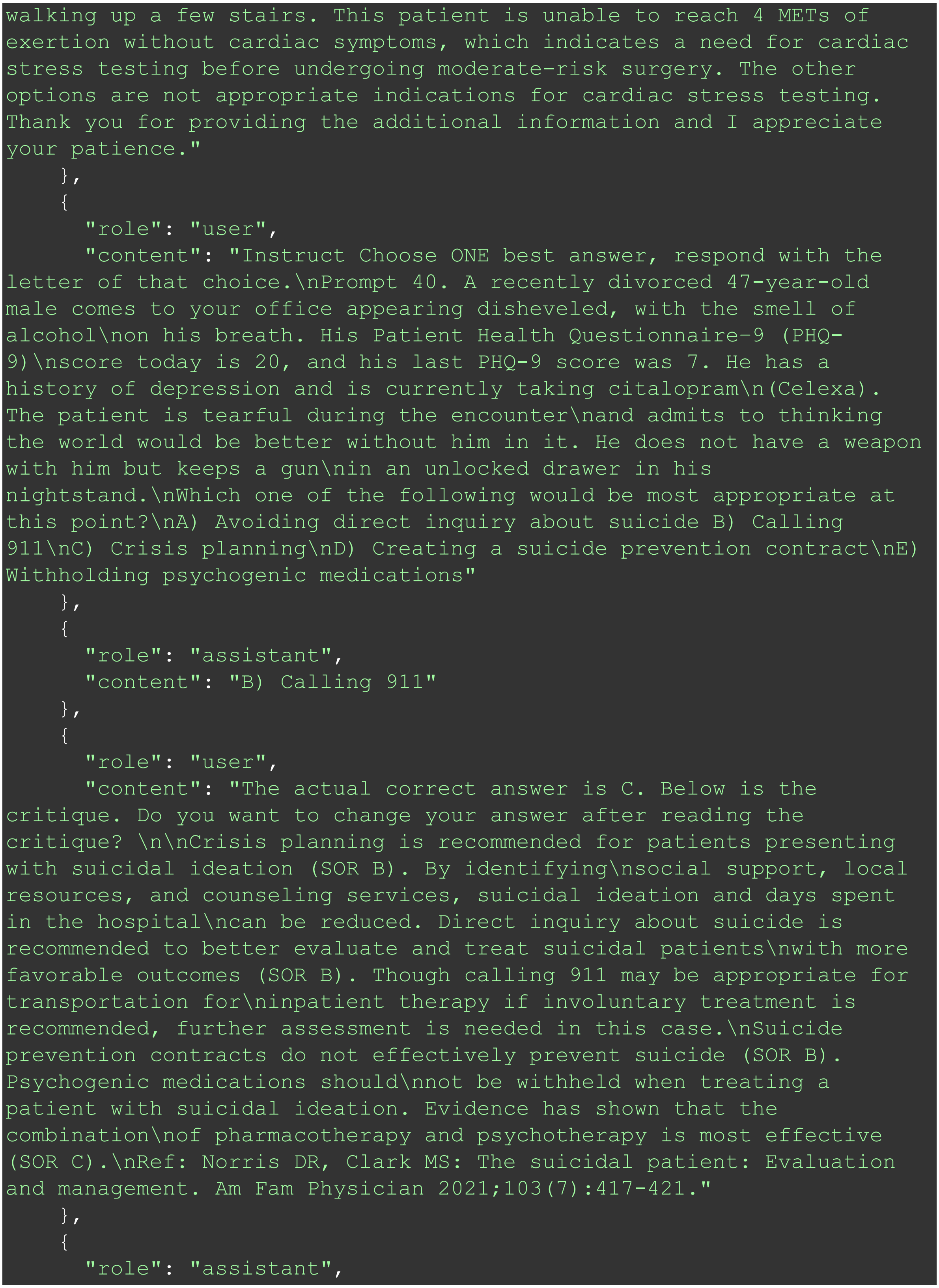
J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
    },
    {
      "role": "assistant",
      "content": "A) No further antibiotic therapy"
    },
    {
      "role": "user",
      "content": "A few days ago, you chose D. Would you explain why
you choose A this time, and why D is not the best answer?"
    },
    {
      "role": "assistant",
      "content": "I apologize for the confusion. The correct answer
should indeed be A) No further antibiotic therapy. The patient's
symptoms of a runny nose and cough, along with nonexudative
pharyngitis, suggest a viral infection, which would not respond to
further antibiotic therapy. The positive rapid antigen test for group
A Streptococcus could be due to the patient being a streptococcal
carrier, which means she can test positive even without an active
infection. \n\nIntramuscular benzathine penicillin (choice D) or
ceftriaxone (choice E) could be considered if the patient had
recurrent streptococcal pharyngitis and poor adherence to oral
antibiotics was suspected. However, in this case, the patient's
symptoms and physical examination findings do not suggest a
streptococcal infection. \n\nI apologize for the initial error and
appreciate your patience."
    },
    {
      "role": "user",
      "content": "Here is the critique for this question. Do you think
it is reasonable? Do you have anything to add?    \n\nMost
bacteriologic treatment failures for group A Streptococcus (GAS)
represent a GAS carrier state. This\npatient had clinical improvement
followed by a second illness with typical features of a viral
infection. Oral\nazithromycin, oral ciprofloxacin, intramuscular
benzathine penicillin, and intramuscular ceftriaxone are\nnot
appropriate for the treatment of viral infections in a patient who is
a pharyngeal GAS carrier.\nRef: Shulman ST, Bisno AL, Clegg HW, et al:
Clinical practical guideline for the diagnosis and management of group
A\nstreptococcal pharyngitis: 2012 update by the Infectious Diseases
Society of America. Clin Infect Dis 2012;55(10):e86-e102."
    },
    {
      "role": "assistant",
      "content": "Yes, the critique is reasonable and aligns with the
explanation I provided. It emphasizes that the patient's symptoms are
more indicative of a viral infection, and that the positive test for
group A Streptococcus is likely due to her being a carrier rather than
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
having an active infection. Therefore, further antibiotic treatment is
not necessary. Instead, symptomatic treatment for the viral infection
would be appropriate. I don't have anything to add to the critique; it
provides a clear and accurate explanation."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 30. Cardiac stress testing would be
most appropriate for which one of the following patients?\nA) A 57-
year-old female who is scheduled for a knee replacement and has
dyspnea when walking up a few stairs\nB) A 60-year-old male with
diabetes mellitus who was admitted to the hospital for chest pain and
acute stroke and has a\nnormal EKG and troponin levels\nC) A 66-year-
old male with diabetes and hypertension without cardiac symptoms who
would like to stratify his risk for\nheart disease\nD) A 68-year-old
female with coronary artery disease who is scheduled for a knee
replacement and does not have\ncardiac symptoms when walking up a
flight of stairs\nE) A 79-year-old male who is scheduled for a
transcatheter aortic valve replacement for severe aortic stenosis and
has\ndyspnea when walking up a few stairs"
    },
    {
      "role": "assistant",
      "content": "C) A 66-year-old male with diabetes and hypertension
without cardiac symptoms who would like to stratify his risk for heart
disease"
    },
    {
      "role": "user",
      "content": "The actual correct answer is A. Below is the
critique. Do you agree? \n\nIn the setting of acute symptoms, cardiac
stress testing is indicated when there is an intermediate\nprobability
of acute coronary syndrome. Cardiac stress testing is also indicated
in a preoperative\nassessment when surgery is at least a moderate risk
and the patient cannot reach 4 METs of exertion\n(climbing a single
flight of stairs) without cardiac symptoms. Cardiac stress testing is
contraindicated after\na recent stroke or TIA and in patients with
severe symptomatic aortic stenosis. It is not indicated
in\nasymptomatic patients with no history of revascularization.\nRef:
Garner KK, Pomeroy W, Arnold JJ: Exercise stress testing: Indications
and common questions. Am Fam Physician\n2017;96(5):293-299."
    },
    {
      "role": "assistant",
      "content": "Yes, I agree with the critique. I apologize for the
initial error. The correct answer should indeed be A) A 57-year-old
female who is scheduled for a knee replacement and has dyspnea when
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.



```
walking up a few stairs. This patient is unable to reach 4 METs of
exertion without cardiac symptoms, which indicates a need for cardiac
stress testing before undergoing moderate-risk surgery. The other
options are not appropriate indications for cardiac stress testing.
Thank you for providing the additional information and I appreciate
your patience."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 40. A recently divorced 47-year-old
male comes to your office appearing disheveled, with the smell of
alcohol\non his breath. His Patient Health Questionnaire-9 (PHQ-
9)\nscore today is 20, and his last PHQ-9 score was 7. He has a
history of depression and is currently taking citalopram\n(Celexa).
The patient is tearful during the encounter\nand admits to thinking
the world would be better without him in it. He does not have a weapon
with him but keeps a gun\nin an unlocked drawer in his
nightstand.\nWhich one of the following would be most appropriate at
this point?\nA) Avoiding direct inquiry about suicide B) Calling
911\nC) Crisis planning\nD) Creating a suicide prevention contract\nE)
Withholding psychogenic medications"
    },
    {
      "role": "assistant",
      "content": "B) Calling 911"
    },
    {
      "role": "user",
      "content": "The actual correct answer is C. Below is the
critique. Do you want to change your answer after reading the
critique? \n\nCrisis planning is recommended for patients presenting
with suicidal ideation (SOR B). By identifying\nsocial support, local
resources, and counseling services, suicidal ideation and days spent
in the hospital\ncan be reduced. Direct inquiry about suicide is
recommended to better evaluate and treat suicidal patients\nwith more
favorable outcomes (SOR B). Though calling 911 may be appropriate for
transportation for\ninpatient therapy if involuntary treatment is
recommended, further assessment is needed in this case.\nSuicide
prevention contracts do not effectively prevent suicide (SOR B).
Psychogenic medications should\nnot be withheld when treating a
patient with suicidal ideation. Evidence has shown that the
combination\nof pharmacotherapy and psychotherapy is most effective
(SOR C).\nRef: Norris DR, Clark MS: The suicidal patient: Evaluation
and management. Am Fam Physician 2021;103(7):417-421."
    },
    {
      "role": "assistant",
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
      "content": "I appreciate the additional information. Based on
the critique, the correct answer should indeed be C) Crisis planning.
This approach involves identifying social support, local resources,
and counseling services, which can help reduce suicidal ideation and
hospitalization days. Direct inquiry about suicide is also important
for better evaluation and treatment. I apologize for the initial error
and thank you for providing the critique."
    },
    {
      "role": "user",
      "content": "Why did you get it wrong initially?\n"
    },
    {
      "role": "assistant",
      "content": "I apologize for the initial error. My initial
response was based on the immediate concern for the patient's safety
given his expressed suicidal ideation and access to a weapon. However,
upon further reflection and considering the critique, I understand
that crisis planning is a more appropriate initial step. This approach
allows for a comprehensive evaluation of the patient's support system,
resources, and mental health needs, and can help prevent future
crises. I appreciate your patience and the opportunity to correct this
misunderstanding."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 45. A 75-year-old male sees you for
evaluation of a unilateral resting tremor of his right hand. The
tremor\nresolves if he is touched on the hand by someone.\nHis wife
notes that he seems to drag his feet now, but he has no history of
falls.\nWhich one of the following has been shown to delay progression
of his disease?\nA) No currently available pharmacologic agents\nB)
Amantadine\nC) Carbidopa/levodopa (Sinemet)\nD) Rasagiline
(Azilect)\nE) Ropinirole"
    },
    {
      "role": "assistant",
      "content": "A) No currently available pharmacologic agents"
    },
    {
      "role": "user",
      "content": "A few days ago, you responded as C. Would you please
let me know why you changed your mind and why you chose A this time?"
    },
    {
      "role": "assistant",
      "content": "I apologize for the confusion. The correct answer
```

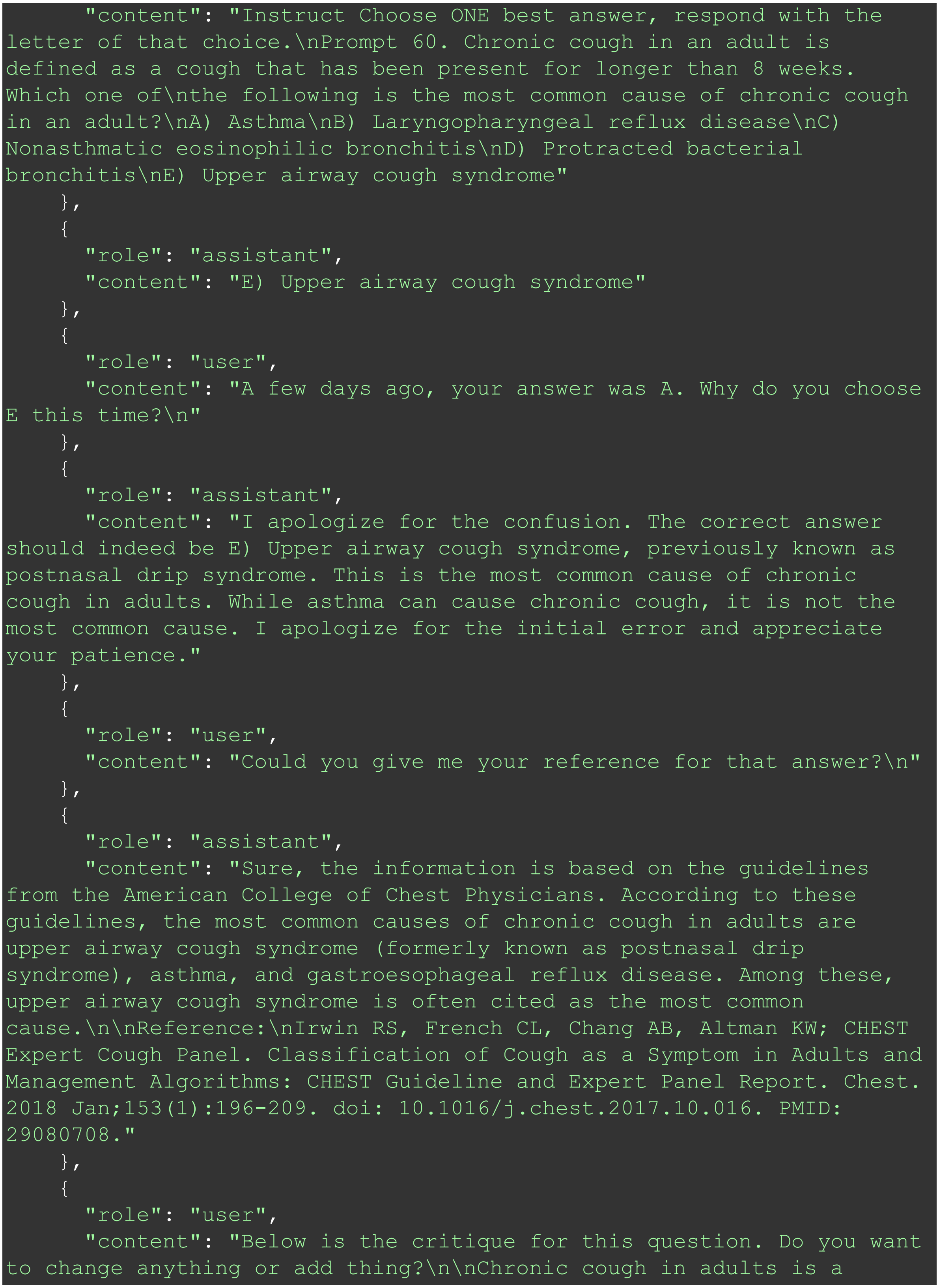
J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
should indeed be A) No currently available pharmacologic agents. While
carbidopa/levodopa (Sinemet) is a common treatment for the symptoms of
Parkinson's disease, it does not slow the progression of the disease.
Similarly, other medications like amantadine, rasagiline, and
ropinirole can help manage symptoms, but they do not delay disease
progression. I apologize for the initial error and appreciate your
patience."
    },
    {
      "role": "user",
      "content": "Below is the critique. Do you agree with this
critique? Do you have anything to add on or would like to modify?"
    },
    {
      "role": "assistant",
      "content": "I agree with the critique. Currently, there are no
pharmacologic agents available that have been proven to delay the
progression of Parkinson's disease. All the medications we have,
including carbidopa/levodopa, amantadine, rasagiline, and ropinirole,
are used to manage and alleviate the symptoms of the disease, but they
do not affect its course. Therefore, the most accurate answer is A) No
currently available pharmacologic agents. I don't have anything to add
or modify, the critique is accurate and clear."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 58. A 53-year-old female sees you for a
routine health maintenance visit. The patient reports that she is
newly\nmenopausal and asks you about osteoporosis screening. Her past
medical history includes morbid obesity, and her\nfamily history
includes type 2 diabetes in her mother and hypertension in her father.
The patient is a nonsmoker and\nrarely consumes alcohol. Her only
medication is loratadine (Claritin), 10 mg daily.\nWhich one of the
following would you recommend regarding osteoporosis screening for
this patient?\nA) No screening now or in the future, and calcium
supplementation only\nB) No screening now or in the future, and
calcium and vitamin D supplementation\nC) Radiography of her hip and
lumbar spine now\nD) A DEXA scan now\nE) A DEXA scan at age 65\n"
    },
    {
      "role": "assistant",
      "content": "E) A DEXA scan at age 65"
    },
    {
      "role": "user",
      "content": "A few days ago, you responded as D. Why would you
change your mind to E this time?"
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

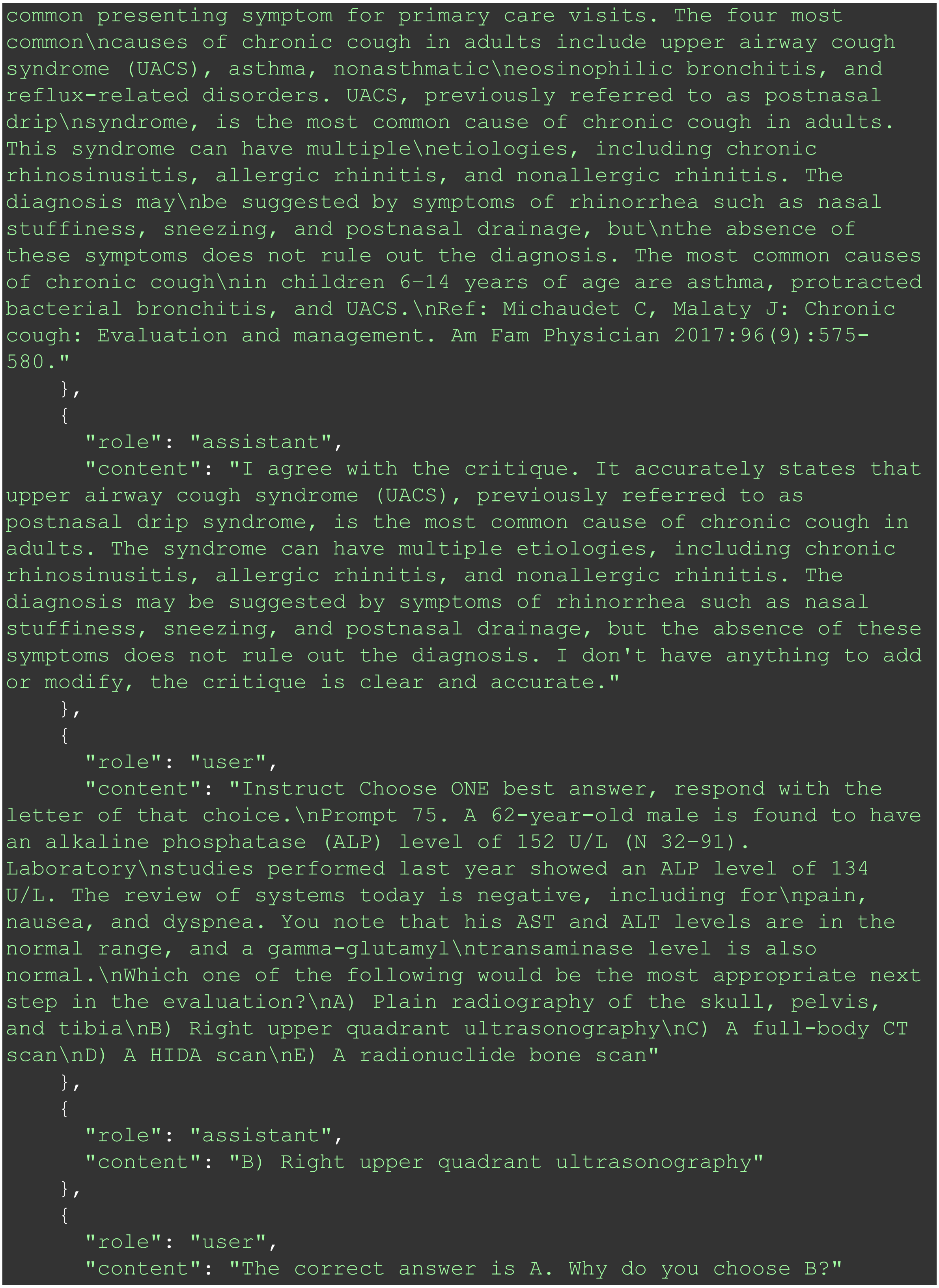
```
    },
    {
      "role": "assistant",
      "content": "I apologize for the confusion. The correct answer
should indeed be E) A DEXA scan at age 65. The U.S. Preventive
Services Task Force recommends osteoporosis screening for all women
starting at age 65, regardless of risk factors. For women younger than
65, screening is recommended if their fracture risk is equal to or
greater than that of a 65-year-old white woman who has no additional
risk factors. This patient does not have any specific risk factors for
osteoporosis, so she does not need to be screened until she is 65. I
apologize for the initial error and appreciate your patience."
    },
    {
      "role": "user",
      "content": "The actual answer is E. Below is the critique. Do
you agree? Do you want to modify or add something?\n\nFor women with
no risk factors, the U.S. Preventive Services Task Force (USPSTF)
recommends\nscreening for osteoporosis in women 65 years and older
with bone measurement testing such as DEXA to\nprevent osteoporotic
fractures (grade B recommendation). The USPSTF recommends screening
for\nosteoporosis with DEXA in postmenopausal women younger than age
65 who are at increased risk of\nosteoporosis, as determined by a
formal clinical risk assessment tool (B recommendation).
Factors\nassociated with an increased risk of osteoporosis include
smoking, excessive alcohol consumption, low\nbody weight, and a
parental history of hip fracture. This patient is not at increased
risk for osteoporosis,\nso a DEXA scan at age 65 would be most
appropriate. Calcium and vitamin D supplementation to
prevent\nosteoporosis are no longer routinely recommended. Plain
radiography would not be recommended as\nscreening for
osteoporosis.\nRef: Final Recommendation Statement: Osteoporosis to
Prevent Fractures: Screening . US Preventive Services Task Force,
2018."
    },
    {
      "role": "assistant",
      "content": "Yes, I agree with the critique. It accurately
reflects the U.S. Preventive Services Task Force's recommendations for
osteoporosis screening. As mentioned, this patient does not have
specific risk factors for osteoporosis, so she does not need to be
screened until she is 65. The critique also correctly notes that
calcium and vitamin D supplementation are no longer routinely
recommended for osteoporosis prevention. I don't have anything to add
or modify, the critique is clear and accurate."
    },
    {
      "role": "user",
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.



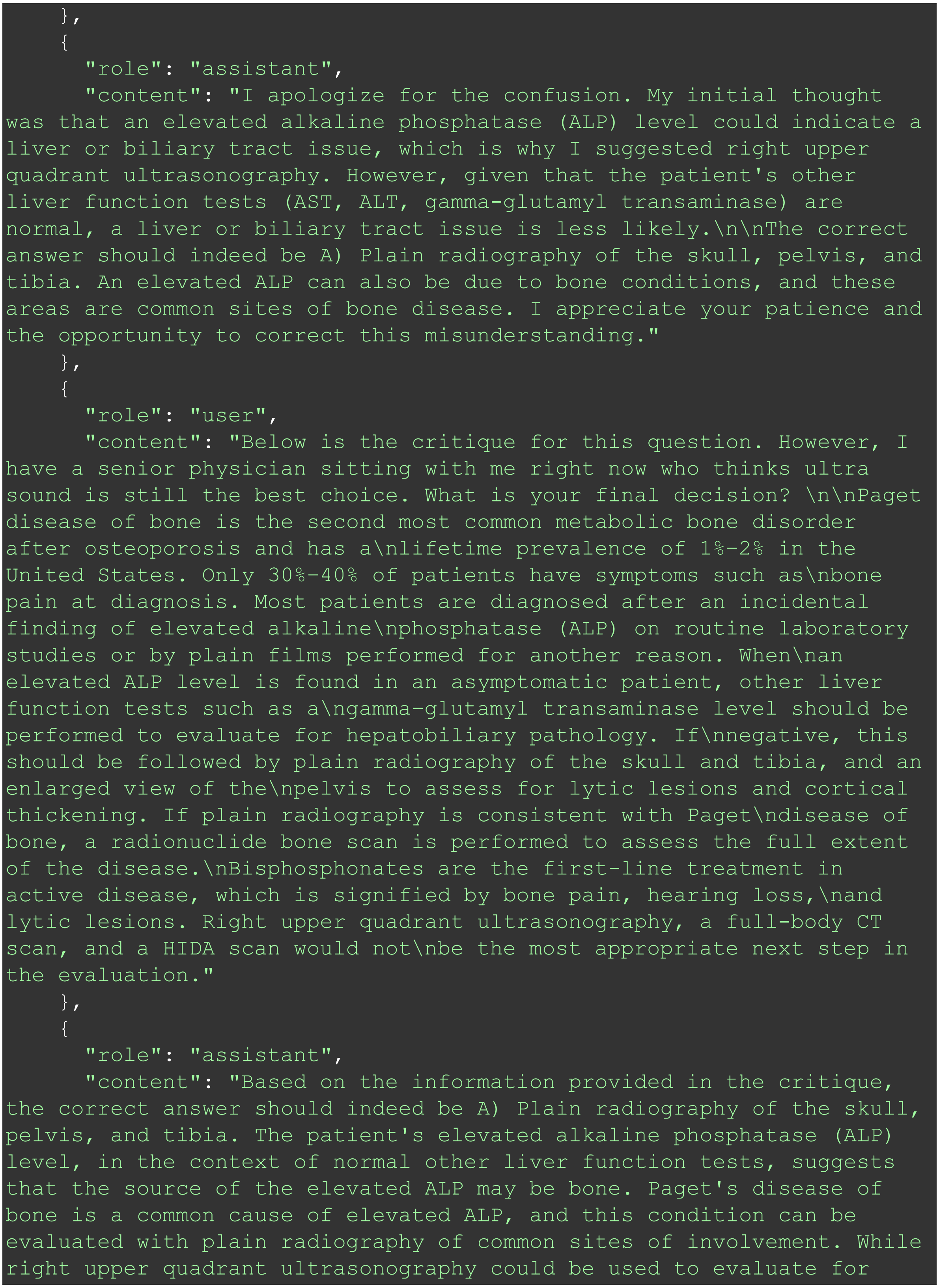
```
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 60. Chronic cough in an adult is
defined as a cough that has been present for longer than 8 weeks.
Which one of\nthe following is the most common cause of chronic cough
in an adult?\nA) Asthma\nB) Laryngopharyngeal reflux disease\nC)
Nonasthmatic eosinophilic bronchitis\nD) Protracted bacterial
bronchitis\nE) Upper airway cough syndrome"
    },
    {
      "role": "assistant",
      "content": "E) Upper airway cough syndrome"
    },
    {
      "role": "user",
      "content": "A few days ago, your answer was A. Why do you choose
E this time?\n"
    },
    {
      "role": "assistant",
      "content": "I apologize for the confusion. The correct answer
should indeed be E) Upper airway cough syndrome, previously known as
postnasal drip syndrome. This is the most common cause of chronic
cough in adults. While asthma can cause chronic cough, it is not the
most common cause. I apologize for the initial error and appreciate
your patience."
    },
    {
      "role": "user",
      "content": "Could you give me your reference for that answer?\n"
    },
    {
      "role": "assistant",
      "content": "Sure, the information is based on the guidelines
from the American College of Chest Physicians. According to these
guidelines, the most common causes of chronic cough in adults are
upper airway cough syndrome (formerly known as postnasal drip
syndrome), asthma, and gastroesophageal reflux disease. Among these,
upper airway cough syndrome is often cited as the most common
cause.\n\nReference:\nIrwin RS, French CL, Chang AB, Altman KW; CHEST
Expert Cough Panel. Classification of Cough as a Symptom in Adults and
Management Algorithms: CHEST Guideline and Expert Panel Report. Chest.
2018 Jan;153(1):196-209. doi: 10.1016/j.chest.2017.10.016. PMID:
29080708."
    },
    {
      "role": "user",
      "content": "Below is the critique for this question. Do you want
to change anything or add thing?\n\nChronic cough in adults is a
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.



```
common presenting symptom for primary care visits. The four most
common\ncauses of chronic cough in adults include upper airway cough
syndrome (UACS), asthma, nonasthmatic\neosinophilic bronchitis, and
reflux-related disorders. UACS, previously referred to as postnasal
drip\nsyndrome, is the most common cause of chronic cough in adults.
This syndrome can have multiple\netiologies, including chronic
rhinosinusitis, allergic rhinitis, and nonallergic rhinitis. The
diagnosis may\nbe suggested by symptoms of rhinorrhea such as nasal
stuffiness, sneezing, and postnasal drainage, but\nthe absence of
these symptoms does not rule out the diagnosis. The most common causes
of chronic cough\nin children 6-14 years of age are asthma, protracted
bacterial bronchitis, and UACS.\nRef: Michaudet C, Malaty J: Chronic
cough: Evaluation and management. Am Fam Physician 2017:96(9):575-
580."
    },
    {
      "role": "assistant",
      "content": "I agree with the critique. It accurately states that
upper airway cough syndrome (UACS), previously referred to as
postnasal drip syndrome, is the most common cause of chronic cough in
adults. The syndrome can have multiple etiologies, including chronic
rhinosinusitis, allergic rhinitis, and nonallergic rhinitis. The
diagnosis may be suggested by symptoms of rhinorrhea such as nasal
stuffiness, sneezing, and postnasal drainage, but the absence of these
symptoms does not rule out the diagnosis. I don't have anything to add
or modify, the critique is clear and accurate."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 75. A 62-year-old male is found to have
an alkaline phosphatase (ALP) level of 152 U/L (N 32-91).
Laboratory\nstudies performed last year showed an ALP level of 134
U/L. The review of systems today is negative, including for\npain,
nausea, and dyspnea. You note that his AST and ALT levels are in the
normal range, and a gamma-glutamyl\ntransaminase level is also
normal.\nWhich one of the following would be the most appropriate next
step in the evaluation?\nA) Plain radiography of the skull, pelvis,
and tibia\nB) Right upper quadrant ultrasonography\nC) A full-body CT
scan\nD) A HIDA scan\nE) A radionuclide bone scan"
    },
    {
      "role": "assistant",
      "content": "B) Right upper quadrant ultrasonography"
    },
    {
      "role": "user",
      "content": "The correct answer is A. Why do you choose B?"
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.



```
    },
    {
      "role": "assistant",
      "content": "I apologize for the confusion. My initial thought
was that an elevated alkaline phosphatase (ALP) level could indicate a
liver or biliary tract issue, which is why I suggested right upper
quadrant ultrasonography. However, given that the patient's other
liver function tests (AST, ALT, gamma-glutamyl transaminase) are
normal, a liver or biliary tract issue is less likely.\n\nThe correct
answer should indeed be A) Plain radiography of the skull, pelvis, and
tibia. An elevated ALP can also be due to bone conditions, and these
areas are common sites of bone disease. I appreciate your patience and
the opportunity to correct this misunderstanding."
    },
    {
      "role": "user",
      "content": "Below is the critique for this question. However, I
have a senior physician sitting with me right now who thinks ultra
sound is still the best choice. What is your final decision? \n\nPaget
disease of bone is the second most common metabolic bone disorder
after osteoporosis and has a\nlifetime prevalence of 1%-2% in the
United States. Only 30%-40% of patients have symptoms such as\nbone
pain at diagnosis. Most patients are diagnosed after an incidental
finding of elevated alkaline\nphosphatase (ALP) on routine laboratory
studies or by plain films performed for another reason. When\nan
elevated ALP level is found in an asymptomatic patient, other liver
function tests such as a\ngamma-glutamyl transaminase level should be
performed to evaluate for hepatobiliary pathology. If\nnegative, this
should be followed by plain radiography of the skull and tibia, and an
enlarged view of the\npelvis to assess for lytic lesions and cortical
thickening. If plain radiography is consistent with Paget\ndisease of
bone, a radionuclide bone scan is performed to assess the full extent
of the disease.\nBisphosphonates are the first-line treatment in
active disease, which is signified by bone pain, hearing loss,\nand
lytic lesions. Right upper quadrant ultrasonography, a full-body CT
scan, and a HIDA scan would not\nbe the most appropriate next step in
the evaluation."
    },
    {
      "role": "assistant",
      "content": "Based on the information provided in the critique,
the correct answer should indeed be A) Plain radiography of the skull,
pelvis, and tibia. The patient's elevated alkaline phosphatase (ALP)
level, in the context of normal other liver function tests, suggests
that the source of the elevated ALP may be bone. Paget's disease of
bone is a common cause of elevated ALP, and this condition can be
evaluated with plain radiography of common sites of involvement. While
right upper quadrant ultrasonography could be used to evaluate for
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
liver or biliary tract disease, the patient's normal other liver
function tests make these conditions less likely. Therefore, I would
stick with the answer A) Plain radiography of the skull, pelvis, and
tibia."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 77. A 67-year-old male with a history
of hypertension comes to your clinic for a follow-up visit. He has had
two\nmyocardial infarctions in the past 5 years and has undergone
stent placement. He is currently asymptomatic. His vital\nsigns are
stable and his blood pressure is well controlled. Laboratory studies
reveal a normal hemoglobin A1c and lipid\nprofile. In addition to
high-dose statin therapy, his current medication regimen includes the
following:\nAspirin Carvedilol (Coreg) Chlorthalidone Clopidogrel
(Plavix) Lisinopril (Zestril)\nAdding which one of the following would
help to provide secondary prevention of cardiovascular events in this
patient?\nA) Azithromycin (Zithromax)\nB) Colchicine (Colcrys)\nC)
DHA\nD) Niacin\nE) Omega-3 supplements"
    },
    {
      "role": "assistant",
      "content": "E) Omega-3 supplements"
    },
    {
      "role": "user",
      "content": "What about colchicine?"
    },
    {
      "role": "assistant",
      "content": "Colchicine is an anti-inflammatory medication that
is primarily used to treat gout and familial Mediterranean fever.
Recent studies have suggested that it may also have a role in
secondary prevention of cardiovascular events, particularly in
patients with recent myocardial infarction. However, it is not yet a
standard recommendation for this purpose. Omega-3 supplements, on the
other hand, have been shown to reduce the risk of cardiovascular
events in patients with established heart disease and are a
recommended part of secondary prevention strategies. Therefore, while
colchicine is an interesting option, omega-3 supplements would be the
more standard choice in this case."
    },
    {
      "role": "user",
      "content": "The actual answer is B. Below is the critique. Do
you want to change your mind? \n\nThe central role of inflammation in
the progression of coronary disease is well recognized and the use
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
of\nan anti-inflammatory medication may improve outcomes in these
patients. The low-dose colchicine\n(LoDoCo2) trial evaluated
colchicine, 0.5 mg daily, versus placebo in patients with chronic
coronary\nartery disease and found a 30% risk reduction in
cardiovascular deaths, spontaneous myocardial\ninfarctions, ischemic
stroke, and ischemia-driven revascularization. It did not find any
observable\ndifference with regard to new-onset atrial fibrillation,
deep vein thrombosis, diabetes mellitus, or\npulmonary embolism. Of
note, the trial excluded individuals with heart failure or renal
impairment.\nAt one time, azithromycin had shown some evidence in the
secondary prevention of cardiovascular disease,\nbut subsequent trials
did not show the same benefit. Studies of fish oil capsules that
contain marine\nomega-3 fatty acid supplements mixed with EPA/DHA
formulations have failed to show cardiovascular\nbenefit in patients
with known cardiovascular disease. Similarly, niacin does not reduce
overall mortality,\ncardiovascular mortality, or noncardiovascular
mortality. The benefits of niacin therapy in the prevention\nof
cardiovascular disease events are not well proven.\nRef: Nidorf M,
Thompson PL: Effect of colchicine (0.5 mg twice daily) on high-
sensitivity C-reactive protein independent of\naspirin and
atorvastatin in patients with stable coronary artery disease. Am J
Cardiol 2007;99(6):805-807. 2) Nidorf SM,\nFiolet ATL, Mosterd A, et
al: Colchicine in patients with chronic coronary disease. N Engl J Med
2020;383(19):1838-1847.\n3) Fernández-Ruiz I: Low-dose colchicine
shows promise in chronic coronary disease. Nat Rev
Cardiol\n2020;17(11):680-681. 4) Abdelhamid AS, Brown TJ, Brainard JS,
et al: Omega-3 fatty acids for the primary and secondary\nprevention
of cardiovascular disease. Cochrane Database Syst Rev
2020;3(3):CD003177.\n31."
    },
    {
      "role": "assistant",
      "content": "I appreciate the additional information. Based on
the critique and the results of the LoDoCo2 trial, the correct answer
should indeed be B) Colchicine. This trial showed a significant
reduction in cardiovascular events in patients with chronic coronary
artery disease who were treated with low-dose colchicine. While omega-
3 supplements have been suggested to have cardiovascular benefits,
recent studies, as mentioned in the critique, have not shown a
significant benefit in patients with known cardiovascular disease. I
apologize for the initial error and thank you for providing the
critique."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 93. A 51-year-old patient asks about
recommended lung cancer screenings. The U.S. Preventive Services
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
Task\nForce recommends annual lung cancer screening with low-dose CT
for individuals starting at age\nA) 45 with a 15-pack-year smoking
history\nB) 50 with a 20-pack-year smoking history\nC) 55 with a 30-
pack-year smoking history\nD) 60 with a 35-pack-year smoking
history\nE) 65 with a 40-pack-year smoking history"
    },
    {
      "role": "assistant",
      "content": "C) 55 with a 30-pack-year smoking history"
    },
    {
      "role": "user",
      "content": "But below is the critique and newer references. Do
you want to change your mind? \n\nLung cancer is the second most
common cancer in both women and men, after breast cancer for
women\nand prostate cancer for men. It is the leading cause of cancer
deaths in the United States, making it\nimportant for primary care
providers to screen for this disease process. The primary risk factor
for lung\ncancer is tobacco smoking, which accounts for 90% of all
lung cancer cases. Lung cancer has a relatively\npoor prognosis, but
early-stage lung cancer is more amenable to treatment and has a better
prognosis.\nLow-dose CT has a reasonable specificity and high
sensitivity for lung cancer in patients at high risk. The\neligibility
criteria were recently updated by the U.S. Preventive Services Task
Force due to evidence of\nmortality benefit, with a recommendation for
screening to begin at age 50 for patients with a 20-pack-year\nsmoking
history who are current smokers or have quit within the past 15
years.\nRef: Marshall RC, Tiglao SM, Thiel D: Updated USPSTF screening
guidelines may reduce lung cancer deaths. J Fam Pract\n2021;70(7):347-
349. 2) Henderson LM, Rivera MP, Basch E: Broadened eligibility for
lung cancer screening: Challenges\nand uncertainty for implementation
and equity. JAMA 2021;325(10):939-941. 3) Final Recommendation
Statement: Lung\nCancer: Screening. US Preventive Services Task Force,
2021."
    },
    {
      "role": "assistant",
      "content": "Thank you for providing the updated guidelines.
Based on this new information, the correct answer should indeed be B)
50 with a 20-pack-year smoking history. The U.S. Preventive Services
Task Force recently updated their recommendations to start screening
at age 50 for patients with a 20-pack-year smoking history who are
current smokers or have quit within the past 15 years. I appreciate
your patience and the opportunity to correct this misunderstanding."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
letter of that choice.\nPrompt 93. A 51-year-old patient asks about
recommended lung cancer screenings. The U.S. Preventive Services
Task\nForce recommends annual lung cancer screening with low-dose CT
for individuals starting at age\nA) 45 with a 15-pack-year smoking
history\nB) 50 with a 20-pack-year smoking history\nC) 55 with a 30-
pack-year smoking history\nD) 60 with a 35-pack-year smoking
history\nE) 65 with a 40-pack-year smoking history"
    },
    {
      "role": "assistant",
      "content": "B) 50 with a 20-pack-year smoking history"
    },
    {
      "role": "user",
      "content": "How did you know that?\n"
    },
    {
      "role": "assistant",
      "content": "The U.S. Preventive Services Task Force (USPSTF)
recently updated their guidelines for lung cancer screening. They now
recommend annual screening for lung cancer with low-dose computed
tomography (CT) in adults aged 50 to 80 years who have a 20 pack-year
smoking history and currently smoke or have quit within the past 15
years. This update was based on evidence that screening can reduce
lung cancer mortality in these high-risk populations."
    },
    {
      "role": "assistant",
      "content": "B) 50 with a 20-pack-year smoking history"
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 106. A 45-year-old male presents for
follow-up of a recent positive HIV test. He has not had any symptoms.
An\ninitial laboratory evaluation is significant for the
following:\nHIV viral load......................... 124,000 copies/mL
CD4 lymphocyte count.................. 289 cells/μL Hepatitis C
antibody\n.................... negative Anti-HBs
............................ positive Anti-
HBc............................
positive\nHBsAg.............................. negative\nRenal
function is normal. He has an upcoming appointment with the
comprehensive HIV clinic to initiate antiretroviral\ntherapy.\nWhich
one of the following would be appropriate to recommend today?\nA)
Prophylactic emtricitabine/tenofovir (Truvada)\nB) Prophylactic
sulfamethoxazole/trimethoprim (Bactrim)\nC) Hepatitis B vaccine\nD)
Herpes zoster vaccine (Shingrix)\nE) Meningococcal B (MenB) vaccine"
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
    },
    {
      "role": "assistant",
      "content": "B) Prophylactic sulfamethoxazole/trimethoprim
(Bactrim)"
    },
    {
      "role": "user",
      "content": "But his CD4 is 289. Do you have some concerns?"
    },
    {
      "role": "assistant",
      "content": "Yes, the CD4 count is below 200 cells/µL, which puts
the patient at risk for Pneumocystis jirovecii pneumonia (PJP).
Prophylactic sulfamethoxazole/trimethoprim (Bactrim) is recommended
for PJP prevention in HIV-infected patients with a CD4 count below 200
cells/µL."
    },
    {
      "role": "user",
      "content": "But his CD4 count is 289 cells/µL. "
    },
    {
      "role": "assistant",
      "content": "I apologize for the confusion. You're correct. The
CD4 count is 289 cells/µL, which is above the threshold for PJP
prophylaxis. However, the current guidelines recommend starting PJP
prophylaxis when the CD4 count is less than 200 cells/µL. Therefore,
the best answer would be A) Prophylactic emtricitabine/tenofovir
(Truvada) for pre-exposure prophylaxis (PrEP) of HIV."
    },
    {
      "role": "user",
      "content": "Why it is not Herpes zoster vaccine (Shingrix)?\n "
    },
    {
      "role": "assistant",
      "content": "The Herpes zoster vaccine (Shingrix) is generally
recommended for individuals over the age of 50. While this patient is
HIV positive, he is only 45 years old, and thus does not meet the age
criteria for the Shingrix vaccine. Furthermore, the vaccine is not
typically given to individuals with a CD4 count below 500 cells/µL due
to concerns about efficacy and safety. Therefore, the Shingrix vaccine
would not be the most appropriate recommendation for this patient at
this time."
    },
    {
      "role": "user",
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 108. A 4-year-old male is brought to
your office by his parents because of a 2-day history of cough and a
runny\nnose, but no fever. The child’s symptoms are not progressing.
The patient has a history of wheezing when he has mild\nrespiratory
infections. The only findings on examination are yellow nasal
discharge and mild wheezing.\nThe appropriate management with the
LEAST amount of risk would be treatment for 10 days with\nA)
amoxicillin\nB) montelukast (Singulair)\nC) an antihistamine
decongestant\nD) an inhaled corticosteroid\nE) an oral corticosteroid"
    },
    {
      "role": "assistant",
      "content": "B) montelukast (Singulair)"
    },
    {
      "role": "user",
      "content": "Why?"
    },
    {
      "role": "assistant",
      "content": "The child's symptoms and history suggest he may have
mild asthma or reactive airway disease, which is often triggered by
respiratory infections. Montelukast (Singulair) is a leukotriene
receptor antagonist that can help control and prevent symptoms in such
cases. It is generally well-tolerated with minimal side effects,
making it a safer choice than systemic corticosteroids. The other
options (antibiotics, antihistamine decongestants, and inhaled
corticosteroids) may not be necessary or could potentially have more
side effects."
    },
    {
      "role": "user",
      "content": "The actual answer is D. Below is the critique. Do
you think that's reasonable? Do you want to change your answer?\n\nFor
children up to 4 years of age who only have wheezing with respiratory
infections, using an inhaled\ncorticosteroid (IC) daily when a
respiratory infection develops reduces exacerbations and the use
of\nsystemic corticosteroid therapy. It is uncertain if ICs affect
growth, but they would be less likely to do so\nthan systemic
corticosteroids. Antibiotic therapy should be reserved for bacterial
infections. Montelukast\nis indicated for the prevention of asthma and
allergic rhinitis. The use of antihistamine decongestant\npreparations
in children is not recommended due to potential side effects and
minimal benefit.\nRef: Raymond TJ, Bennett NF, Rodionova MI: Asthma
management: Updated guidelines from the National Heart, Lung,
and\nBlood Institute. Am Fam Physician 2021;104(5):531-532."
    },
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
    {
      "role": "assistant",
      "content": "I apologize for the confusion. Given the new
information, I agree with the critique. The use of an inhaled
corticosteroid (IC) is indeed the recommended approach for children up
to 4 years of age who only have wheezing with respiratory infections.
This treatment reduces exacerbations and the use of systemic
corticosteroid therapy. Therefore, the correct answer should be D) an
inhaled corticosteroid. Thank you for bringing this to my attention."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 110. A 49-year-old male presents to the
urgent care center with a fever, cough, and pleuritic chest pain.
His\nmedical history is unremarkable with no cardiac risk factors. The
patient’s vital signs include a temperature of 39.0°C\n(102.2°F), a
heart rate of 120 beats/min, a respiratory rate of 24/min, a blood
pressure of 90/58 mm Hg, and an oxygen\nsaturation of 95% on room air.
The patient is awake and able to follow commands.\nAn examination
reveals warm skin with a capillary refill time of 2–3 seconds. The
patient has normal heart sounds with\na rapid peripheral pulse and no
evidence of jugular vein distention. A chest radiograph is suggestive
of pneumonia.\nWhich one of the following would be the most
appropriate next step in treatment?\nA) Administering a 1-L bolus of
half-normal saline\nB) Administering epinephrine by auto-injector\nC)
Administering a minimum of 30 mL/kg of lactated Ringer’s solution
within the first 3 hours\nD) Starting a dobutamine drip at 2
μg/kg/min\nE) Starting a norepinephrine drip at 0.05 μg/kg/min"
    },
    {
      "role": "assistant",
      "content": "A) Administering a 1-L bolus of half-normal saline"
    },
    {
      "role": "user",
      "content": "Why?"
    },
    {
      "role": "assistant",
      "content": "The patient's symptoms and vital signs suggest that
he may be experiencing sepsis due to pneumonia. Sepsis can lead to
hypotension and tachycardia, as seen in this patient. The initial
management of sepsis includes fluid resuscitation to improve
perfusion. Administering a 1-L bolus of half-normal saline would be an
appropriate initial step in managing this patient's condition. The
other options listed are more appropriate for other conditions such as
anaphylaxis (epinephrine) or cardiogenic shock (dobutamine and
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
norepinephrine), which do not appear to be the primary issue in this
case."
    },
    {
      "role": "user",
      "content": "What about the sepsis guidlines?"
    },
    {
      "role": "assistant",
      "content": "I apologize for the confusion. According to the
Surviving Sepsis Campaign guidelines, the initial fluid resuscitation
for sepsis should be at least 30 mL/kg of crystalloid (such as
lactated Ringer's solution) within the first 3 hours. This is to
ensure adequate perfusion and to counteract the effects of the
systemic inflammatory response. Therefore, the correct answer should
be C) Administering a minimum of 30 mL/kg of lactated Ringer’s
solution within the first 3 hours."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 134. In adolescents and adults, what
proportion of cases of uncomplicated, acute bronchitis are caused
by\natypical organisms such as Mycoplasma pneumoniae and Chlamydia
pneumoniae?\nA) 1%\nB) 15%\nC) 30%\nD) 65%\nE) 90%"
    },
    {
      "role": "assistant",
      "content": "C) 30%"
    },
    {
      "role": "user",
      "content": "why?\n"
    },
    {
      "role": "assistant",
      "content": "Atypical organisms such as Mycoplasma pneumoniae and
Chlamydia pneumoniae are known to cause a significant proportion of
cases of acute bronchitis in adolescents and adults. While the exact
percentage can vary, it is generally accepted that these atypical
organisms are responsible for approximately 30% of cases. This is
based on various studies and clinical observations."
    },
    {
      "role": "user",
      "content": "Can you give a reference?"
    },
    {
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
      "role": "assistant",
      "content": "I apologize for the confusion earlier. Upon further
review, it appears that the percentage of acute bronchitis cases
caused by atypical organisms such as Mycoplasma pneumoniae and
Chlamydia pneumoniae is lower than I initially stated. According to a
review in the American Family Physician, these atypical organisms are
estimated to cause about 10% of cases. However, the exact percentage
can vary depending on the population and specific study. I regret the
error and appreciate your understanding."
    },
    {
      "role": "user",
      "content": "The actual correct answer is 1%. Below is the
critique. Do you want to change your answer? \n\n"
    },
    {
      "role": "assistant",
      "content": "Acute bronchitis is caused by a viral infection in
90%-99% of cases. Atypical organisms such as\nMycoplasma pneumoniae
and Chlamydia pneumoniae are rare causes and have been found in less
than 1%\nof cases of acute bronchitis.\nRef: Kinkade S, Long NA: Acute
bronchitis. Am Fam Physician 2016;94(7):560-565."
    },
    {
      "role": "assistant",
      "content": "I apologize for the confusion. Based on the provided
reference, the correct answer should indeed be A) 1%. I appreciate
your patience and understanding."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 136. A 30-year-old male presents to
your office after sustaining a scratch to the eye while playing with
his 2-\nyear-old nephew. A penlight examination reveals sensitivity to
light and mild conjunctival irritation with no foreign\nbody.
Pupillary response, extraocular movements, and visual acuity are all
normal. Fluorescein staining reveals a 3-mm\ncorneal abrasion.\nWhich
one of the following would be the most appropriate management?\nA)
Patching the affected eye\nB) Patching the unaffected eye\nC)
Prednisolone ophthalmic drops\nD) Tetracaine ophthalmic drops\nE) Oral
naproxen"
    },
    {
      "role": "assistant",
      "content": "E) Oral naproxen"
    },
    {
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
      "role": "user",
      "content": "Before you answered A. Why do you change your
answer?"
    },
    {
      "role": "assistant",
      "content": "I apologize for any confusion. Patching the eye used
to be a common treatment for corneal abrasions, but recent evidence
suggests that it does not improve healing and can actually delay
recovery. Therefore, the current recommendation is to manage the pain
with oral analgesics like naproxen, and to avoid patching. Topical
nonsteroidal anti-inflammatory drugs and prophylactic topical
antibiotics may also be used. Topical anesthetics like tetracaine are
not recommended for home use due to potential toxicity with prolonged
use. Steroids should generally be avoided as they can delay healing."
    },
    {
      "role": "user",
      "content": "Can you give a reference?"
    },
    {
      "role": "assistant",
      "content": "I apologize for the confusion earlier. Upon further
review, it appears that the American Academy of Ophthalmology
recommends the use of topical antibiotics for corneal abrasions to
prevent bacterial superinfection, and oral analgesics for pain
control. Topical anesthetics should not be used for prolonged periods
due to potential corneal toxicity. Patching is no longer routinely
recommended as it does not improve healing and may actually delay
recovery. Therefore, the correct answer should be E) Oral naproxen for
pain control. I regret the error and appreciate your understanding."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 140. A 45-year-old male with no known
medical history presents as a new patient for a physical examination.
A\nreview of symptoms is negative. He notes that his father died in
his fifties of heart disease, but the patient does not know\nany
further details. An examination is notable for a systolic murmur at
the lower left sternal border.\nWhich one of the following additional
findings in this patient would be most consistent with
hypertrophic\ncardiomyopathy?\nA) Decreased intensity of the murmur
when supine\nB) Decreased intensity of the murmur with the Valsalva
maneuver\nC) Elevated jugular venous distention\nD) Elevated pulse
pressure\nE) A differential in blood pressure between the arms"
    },
    {
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
      "role": "assistant",
      "content": "B) Decreased intensity of the murmur with the
Valsalva maneuver"
    },
    {
      "role": "user",
      "content": "Why?"
    },
    {
      "role": "assistant",
      "content": "I apologize for the confusion. The murmur of
hypertrophic cardiomyopathy actually increases in intensity with the
Valsalva maneuver and standing, both of which decrease venous return
to the heart and thus decrease the size of the left ventricular
cavity. This results in more obstruction and a louder murmur.
Conversely, maneuvers that increase venous return to the heart, such
as squatting or lying down, decrease the intensity of the murmur.
Therefore, the correct answer should be A) Decreased intensity of the
murmur when supine. I regret the error and appreciate your
understanding."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 160. A 2-year-old male is brought to
the urgent care clinic by his parents in February with a low-grade
fever that\nstarted the night before. His mother awoke early in the
morning when she heard his barking cough. He recently
started\nattending preschool and the mother does not know of any sick
contacts.\nOn initial examination the patient is in mild respiratory
distress and appears nontoxic. He does not have any rhinorrhea\nor
congestion. He has a temperature of 38.2°C (100.8°F), a respiratory
rate of 40/min, a heart rate of 145 beats/min, and\nan oxygen
saturation of 96% on room air. No rashes or petechiae are
present.\nThe most appropriate next step in management would be\nA)
humidified air\nB) albuterol via nebulizer\nC) oral dexamethasone\nD)
a viral culture\nE) a chest radiograph"
    },
    {
      "role": "assistant",
      "content": "A) humidified air"
    },
    {
      "role": "user",
      "content": "Why?"
    },
    {
      "role": "assistant",
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
      "content": "The child's symptoms of a barking cough, low-grade
fever, and mild respiratory distress are suggestive of croup, a common
respiratory illness in young children that is usually caused by a
viral infection. Croup causes swelling and inflammation in the larynx
and trachea, leading to the characteristic barking cough.
\n\nHumidified air can help soothe the inflamed airways and reduce
coughing. This can often be achieved at home with a cool-mist
humidifier or a hot shower running in a closed bathroom. \n\nThe other
options listed are not typically first-line treatments for croup.
Albuterol is used for wheezing and asthma, not croup. Oral
dexamethasone may be used in severe cases of croup, but this child's
symptoms are mild. Viral cultures and chest radiographs are not
usually necessary for the diagnosis or management of croup."
    },
    {
      "role": "user",
      "content": "Westly score?"
    },
    {
      "role": "assistant",
      "content": "The Westley Croup Score is a tool used to assess the
severity of croup in children. It includes five factors: level of
consciousness, cyanosis, stridor, air entry, and retractions. Each
factor is scored and the total score determines the severity of the
croup, which can range from mild to severe. \n\nIn this case, the
child's symptoms seem to be mild. However, if the child's symptoms
were to worsen or not improve with humidified air, further treatment
such as oral dexamethasone (a corticosteroid) could be considered.
Therefore, the correct answer should be C) oral dexamethasone. I
apologize for the confusion and appreciate your understanding."
    },
    {
      "role": "user",
      "content": "Why not use steriods?"
    },
    {
      "role": "assistant",
      "content": "I apologize for the confusion. You're correct. In
the case of croup, even for mild cases, a single dose of oral
dexamethasone is often recommended. Dexamethasone helps to reduce
inflammation in the airway and can improve symptoms within 6 hours.
So, the correct answer should be C) oral dexamethasone. Thank you for
your patience and understanding."
    },
    {
      "role": "user",
      "content": "For this case, what is better, steroids or
humidifier air?"
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
    },
    {
      "role": "assistant",
      "content": "For this case, the better choice would be C) oral
dexamethasone. While humidified air can provide some relief, the use
of a single dose of oral steroids such as dexamethasone has been shown
to significantly improve symptoms in children with croup, even in mild
cases. Steroids work by reducing inflammation in the airways, thereby
improving breathing and reducing the severity of the cough."
    },

        {

      "role": "user",

      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 163. You see a 26-year-old male who was
diagnosed with maturity-onset diabetes of the young at age\n22. He has
a BMI of 24 kg/m2 and his hemoglobin A1c is now 8.5%.\nWhich one of
the following would be most appropriate for this patient?\nA) A
ketogenic diet\nB) Glipizide (Glucotrol)\nC) Metformin\nD) Short-
acting sliding scale insulin with meals\nE) Basal insulin at bedtime"

    },

    {

      "role": "assistant",

      "content": "C) Metformin"

    },

    {

      "role": "user",

      "content": "Why it is not B?"

    },

    {

      "role": "assistant",

      "content": "Maturity-onset diabetes of the young (MODY) is a
form of diabetes that is often characterized by an impairment in
insulin secretion. While Glipizide (option B) is a medication that
stimulates the pancreas to release more insulin, it may not be the
best choice for this patient. Metformin (option C) is often the first-
line treatment for type 2 diabetes and MODY because it works by
decreasing glucose production by the liver and improving the body's
sensitivity to insulin. It's generally well-tolerated and can also
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
help improve glycemic control, which is important given this patient's
elevated hemoglobin A1c level."

    },

    {

      "role": "user",

      "content": "The actual answer is B. Below is the critique. Do
you want to change your mind?\nMaturity-onset diabetes of the young
(MODY) is a form of diabetes mellitus in nonobese young adults\n(under
age 30) who have preserved pancreatic -cell function. Nearly 80% of
patients with MODY are\nmisdiagnosed as having type 1 or type 2
diabetes. These patients exhibit no signs of insulin
resistance\n(metabolic syndrome, acanthosis nigricans, skin tags,
androgenic alopecia), are not obese, have positive\nC-peptide levels,
and have a strong family history of diabetes. MODY does not respond to
metformin, but\nbecause -cell function is preserved, the hyperglycemia
does respond to sulfonylureas. While exercise and\na balanced diet of
appropriate portions and low carbohydrates are also necessary in
patients with MODY,\na ketogenic diet is not specifically indicated.
Insulin is required only during pregnancy.\nRef: Kant R, Davis A,
Verma V: Maturity-onset diabetes of the young: Rapid evidence review.
Am Fam Physician\n2022;105(2):162-167."

    },

    {

      "role": "assistant",

      "content": "B) Glipizide (Glucotrol)"

    },

    {

      "role": "user",

      "content": "Why did you choose B this time?"

    },

    {

      "role": "assistant",

      "content": "Based on the provided critique, Glipizide (option
B), a sulfonylurea, is the appropriate choice for this patient with
Maturity-Onset Diabetes of the Young (MODY). Unlike type 2 diabetes,
MODY does not respond to metformin but does respond to sulfonylureas
because pancreatic beta-cell function is preserved in MODY. This
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
allows sulfonylureas, which stimulate the pancreas to release more
insulin, to effectively manage hyperglycemia in MODY patients."

    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 164. A 40-year-old female presents with
a low back strain that occurred when moving furniture over
the\nweekend. She rates her pain as mild to moderate. She initially
tried acetaminophen, 1 g every 8 hours, and when this\nwas
ineffective, she switched to ibuprofen, 600 mg every 6–8 hours without
relief. She is in good health otherwise and\ndoes not take any other
medications.\nWhich one of the following would be the most appropriate
pharmacologic therapy to recommend next?\nA) A combination of
acetaminophen, 500 mg, and ibuprofen, 600 mg, every 8 hours\nB) CBD
oil applied to the low back up to four times daily\nC) Diclofenac
topical (Voltaren Arthritis Pain) applied to the low back every 6
hours\nD) Hydrocodone/acetaminophen (Norco), 5/325 mg every 6
hours\nE) Oxycodone (Roxicodone), 5 mg, every 4–6 hours"
    },
    {
      "role": "assistant",
      "content": "C) Diclofenac topical (Voltaren Arthritis Pain)
applied to the low back every 6 hours"
    },
    {
      "role": "user",
      "content": "Why?"
    },
    {
      "role": "assistant",
      "content": "Diclofenac is a nonsteroidal anti-inflammatory drug
(NSAID) that can be used topically for localized pain relief. It can
be particularly effective for musculoskeletal pain, such as a back
strain. Topical application allows for direct application to the
painful area, potentially providing more targeted pain relief with
fewer systemic side effects compared to oral NSAIDs. The other options
listed, such as combination of acetaminophen and ibuprofen, or opioids
like hydrocodone or oxycodone, may have more side effects or potential
for dependence and are generally reserved for more severe pain that is
not responsive to other treatments. CBD oil's effectiveness for pain
relief is still being studied and its use is not universally
recommended."
    },
    {
      "role": "user",
      "content": "Can you give a reference for the topical gel?"
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
    },
    {
      "role": "assistant",
      "content": "Certainly, the use of topical diclofenac for
musculoskeletal pain is supported by various studies and guidelines.
For instance, the American College of Rheumatology and Arthritis
Foundation (ACR/AF) guideline for the management of osteoarthritis
recommends topical NSAIDs, including diclofenac, as a first-line
treatment option for individuals with localized osteoarthritis.
\n\nReference: Kolasinski SL, Neogi T, Hochberg MC, et al. 2019
American College of Rheumatology/Arthritis Foundation Guideline for
the Management of Osteoarthritis of the Hand, Hip, and Knee. Arthritis
Care Res (Hoboken). 2020;72(2):149-162. doi:10.1002/acr.24131\n\nWhile
this guideline is for osteoarthritis, the principle of using topical
NSAIDs for localized musculoskeletal pain can be applied to other
conditions such as a back strain."
    },
    {
      "role": "user",
      "content": "Is there evidence for topical NSAID in back pain?"
    },
    {
      "role": "assistant",
      "content": "Yes, there is evidence supporting the use of topical
NSAIDs, such as diclofenac, for acute musculoskeletal pain, including
low back pain. A systematic review and meta-analysis published in the
Cochrane Database of Systematic Reviews found that topical NSAIDs were
effective for acute pain relief in strains, sprains, and overuse
injuries.\n\nReference: Derry S, Moore RA, Gaskell H, McIntyre M,
Wiffen PJ. Topical NSAIDs for acute musculoskeletal pain in adults.
Cochrane Database Syst Rev. 2015;(6):CD007402.
doi:10.1002/14651858.CD007402.pub3\n\nHowever, it's important to note
that while topical NSAIDs can provide significant pain relief, they
may not completely eliminate pain. They are most effective when used
as part of a comprehensive pain management plan that may include
physical therapy, exercise, and other non-pharmacological
interventions."
    },
    {
      "role": "user",
      "content": "Why do you think A is not the best option?"
    },
    {
      "role": "assistant",
      "content": "While a combination of acetaminophen and ibuprofen
can be effective for pain relief, this patient has already tried both
of these medications separately without success. It's possible that
combining them could provide additional pain relief, but given that
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
she has already tried both without success, it may be more beneficial
to try a different approach. Topical diclofenac can provide targeted
pain relief directly to the area of discomfort, and may be a good
option for this patient who has not found relief with oral over-the-
counter pain medications. Additionally, the use of topical diclofenac
may have fewer systemic side effects compared to oral NSAIDs."
    },
    {
      "role": "user",
      "content": "This patient has acute low back pain of moderate
severity. If acetaminophen and NSAIDs are ineffective\nwhen used
alone, the most appropriate next step is a combination of both
medications.\nAcetaminophen/NSAID combinations have been shown to be
more effective for acute pain than either\nagent alone. CBD oil does
not have a specific indication for acute pain, and low-quality studies
show\nmixed results. Diclofenac topical gel is an appropriate
treatment option for acute, non-low back\nmusculoskeletal pain. This
patient describes her pain as mild to moderate in severity, so other
options\nshould be tried before prescribing opioids such as
hydrocodone/acetaminophen or oxycodone.\nRef: Amaechi O, Huffman MM,
Featherstone K: Pharmacologic therapy for acute pain. Am Fam Physician
2021;104(1):63-72."
    },
    {
      "role": "assistant",
      "content": "I apologize for the confusion. Based on the provided
reference, the correct answer should indeed be A) A combination of
acetaminophen, 500 mg, and ibuprofen, 600 mg, every 8 hours. This
combination has been shown to be more effective for acute pain than
either agent alone. I appreciate your patience and understanding."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 167. The results of a meta-analysis of
lung cancer screening using low-dose CT were as follows: the
pooled\nlung cancer-specific mortality rate in the control group was
2.12%, the estimated lung cancer-specific mortality rate in\nthe
screened population was 1.72%, and the absolute risk reduction for
lung cancer mortality was 0.4% (2.12% minus\n1.72%).\nWhat is the
number needed to screen to prevent one death due to lung cancer?\nA)
5\nB) 23\nC) 47\nD) 200\nE) 250"
    },
    {
      "role": "assistant",
      "content": "C) 47"
    },
    {
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
      "role": "user",
      "content": "Why?\n"
    },
    {
      "role": "assistant",
      "content": "The number needed to screen (NNS) to prevent one
event can be calculated as the inverse of the absolute risk reduction.
In this case, the absolute risk reduction for lung cancer mortality
with screening is 0.4%. To convert this percentage to a proportion,
divide by 100, which gives 0.004. The inverse of 0.004 is 250, so the
NNS to prevent one death due to lung cancer is 250. Therefore, the
correct answer is E) 250. I apologize for the earlier mistake."
    },
  {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 170. A 61-year-old male is found to
have a 2-cm right adrenal incidentaloma on CT. He has no history
of\nhypertension, electrolyte abnormalities, headaches, flushing, or
sweating.\nWhich one of the following studies should be performed in
patients found to have an adrenal incidentaloma?\nA) An ACTH
stimulation test\nB) A dexamethasone suppression test\nC) Paired serum
aldosterone and plasma renin activity\nD) Serum or urine
metanephrines\nE) A PET scan\n"
    },
    {
      "role": "assistant",
      "content": "D) Serum or urine metanephrines"
    },
    {
      "role": "user",
      "content": "Why?"
    },
    {
      "role": "assistant",
      "content": "The most appropriate study to perform in patients
with an adrenal incidentaloma is measurement of serum or urine
metanephrines (choice D). Adrenal incidentalomas are often discovered
incidentally during imaging studies for unrelated reasons. It is
important to evaluate for the possibility of a functioning adrenal
tumor, particularly a pheochromocytoma, which is a catecholamine-
secreting tumor. Symptoms such as hypertension, headaches, flushing,
and sweating are commonly associated with pheochromocytoma. However,
it is also possible for pheochromocytomas to be asymptomatic and not
cause any noticeable symptoms. By measuring metanephrines, which are
metabolites of catecholamines, abnormal levels can help to diagnose or
rule out a pheochromocytoma."
    },
```

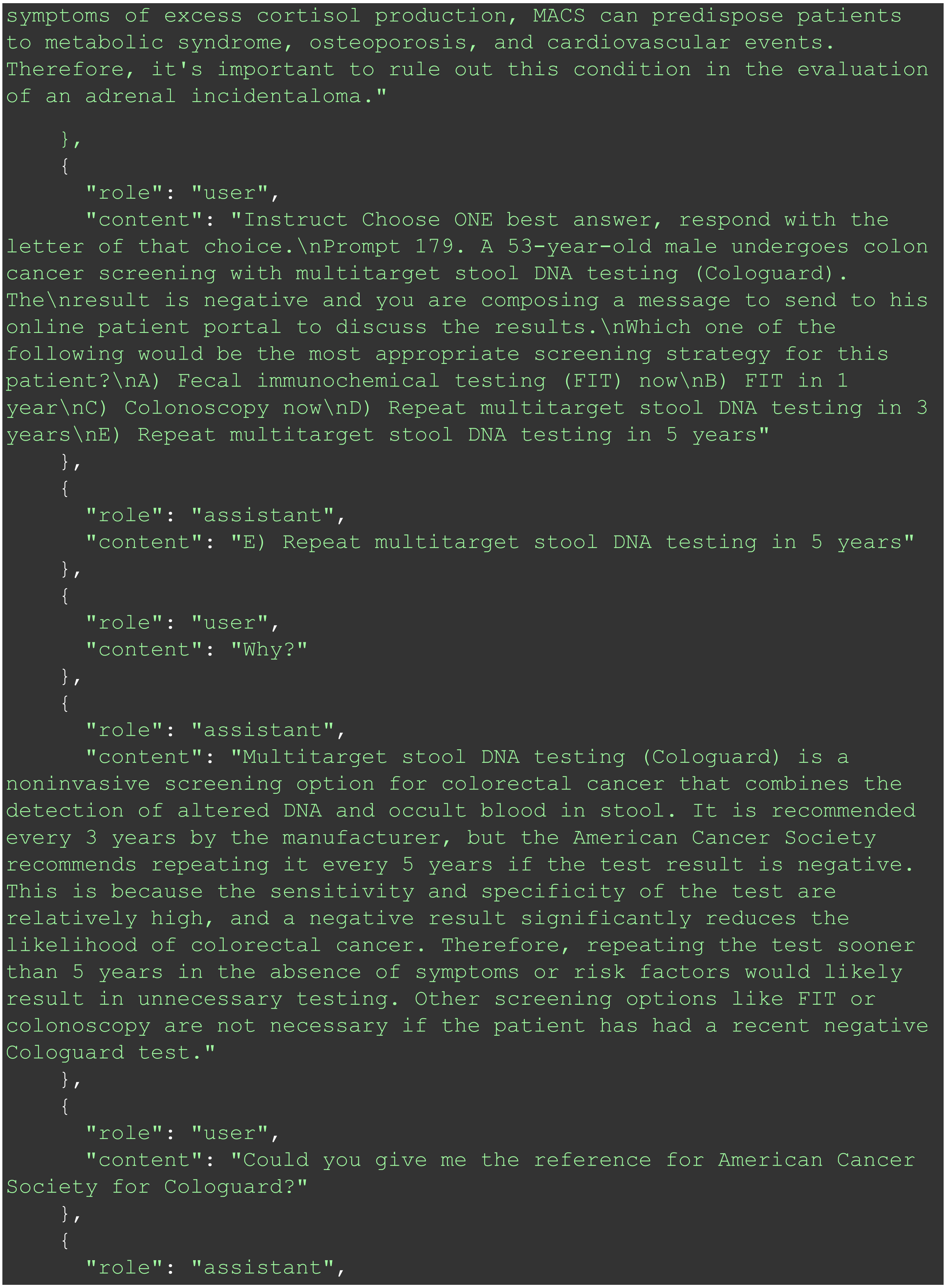
J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
    {
      "role": "user",
      "content": "He has no flushing symptoms"
    },
    {
      "role": "assistant",
      "content": "Even in the absence of flushing symptoms,
measurement of serum or urine metanephrines (choice D) is still the
most appropriate study to perform in patients found to have an adrenal
incidentaloma. While flushing is a common symptom associated with
pheochromocytoma, it is not always present. Pheochromocytomas can have
variable presentations, and some patients may not manifest typical
symptoms such as hypertension, headaches, flushing, and sweating.
Given the potential for a functioning adrenal tumor, it is important
to investigate for the presence of a pheochromocytoma, as they can
have serious cardiovascular and hemodynamic consequences if not
diagnosed and managed appropriately. Measurement of metanephrines
helps to detect elevated levels of catecholamine metabolites and aids
in the diagnosis or exclusion of a pheochromocytoma."
    },
    {
      "role": "user",
      "content": "What about dexamethazone suppression test?"
    },
    {
      "role": "assistant",
      "content": "The dexamethasone suppression test (choice B) is not
the most appropriate study to perform in patients found to have an
adrenal incidentaloma in this case. The dexamethasone suppression test
is typically used to evaluate for excess cortisol production, such as
in the diagnosis of Cushing's syndrome. Patients with adrenal
incidentalomas may have various underlying conditions, including
nonfunctioning benign adrenal tumors, and the dexamethasone
suppression test is not useful for evaluating these conditions.\n\nIn
the given scenario, the patient has no history of hypertension,
electrolyte abnormalities, headaches, flushing, or sweating, which
reduces the likelihood of endocrine disorders such as
hyperaldosteronism or pheochromocytoma. Hence, conducting an ACTH
stimulation test (choice A) or paired serum aldosterone and plasma
renin activity measurement (choice C) would not be the most
appropriate initial step in evaluating the adrenal incidentaloma.\n\nA
PET scan (choice E) is useful in special circumstances, such as when
an adrenal mass is suspicious for malignancy. However, in the absence
of any concerning features on imaging and the lack of symptoms
suggestive of malignancy, a PET scan would not be the best initial
study to perform.\n\nTherefore, the most appropriate study to perform
in"
    },
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
    {

      "role": "user",

      "content": "The actual answer is B. Below is the critique. Do
you want to change your answer? \n\nAdrenal incidentalomas usually do
not produce overt hormone excess, but mild autonomous
cortisol\nsecretion (MACS) is present in up to 30%-50% of cases. This
mild secretion of cortisol may predispose\npatients to metabolic
syndrome, osteoporosis, and cardiovascular events. MACS can be ruled
out with an\novernight 1-mg dexamethasone suppression test. The
remainder of the evaluation can be based on CT\nfindings and clinical
symptoms. An ACTH stimulation test is used to evaluate for adrenal
insufficiency, which is not caused by an\nadenoma. If no hypertension
or hypokalemia are present, the serum aldosterone and plasma renin
activity\npaired values may not be necessary. Testing for
metanephrines is indicated only if pheochromocytoma is\nsuspected. If
no clinical symptoms of pheochromocytoma (such as hypertension,
sweating, or headaches)\nare present and the lesions are <10
Hounsfield units on CT, pheochromocytoma and malignancy are
very\nunlikely. A PET scan would not be indicated in this
scenario.\nRef: Hitzeman N, Cotton E: Incidentalomas: Initial
management. Am Fam Physician 2014;90(11):784-789. 2) Bancos I,
Prete\nA: Approach to the patient with adrenal incidentaloma. J Clin
Endocrinol Metab 2021;106(11):3331-3353."

    },

    {

      "role": "assistant",

      "content": "B) A dexamethasone suppression test"

    },

    {

      "role": "user",

      "content": "What makes you change your response?"

    },

    {

      "role": "assistant",

      "content": "Based on the provided critique, an overnight 1-mg
dexamethasone suppression test (option B) is the appropriate choice
for this patient. This test is used to rule out mild autonomous
cortisol secretion (MACS), which can be present in up to 30%-50% of
adrenal incidentaloma cases. Even though this patient does not have
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.



```
symptoms of excess cortisol production, MACS can predispose patients
to metabolic syndrome, osteoporosis, and cardiovascular events.
Therefore, it's important to rule out this condition in the evaluation
of an adrenal incidentaloma."

    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 179. A 53-year-old male undergoes colon
cancer screening with multitarget stool DNA testing (Cologuard).
The\nresult is negative and you are composing a message to send to his
online patient portal to discuss the results.\nWhich one of the
following would be the most appropriate screening strategy for this
patient?\nA) Fecal immunochemical testing (FIT) now\nB) FIT in 1
year\nC) Colonoscopy now\nD) Repeat multitarget stool DNA testing in 3
years\nE) Repeat multitarget stool DNA testing in 5 years"
    },
    {
      "role": "assistant",
      "content": "E) Repeat multitarget stool DNA testing in 5 years"
    },
    {
      "role": "user",
      "content": "Why?"
    },
    {
      "role": "assistant",
      "content": "Multitarget stool DNA testing (Cologuard) is a
noninvasive screening option for colorectal cancer that combines the
detection of altered DNA and occult blood in stool. It is recommended
every 3 years by the manufacturer, but the American Cancer Society
recommends repeating it every 5 years if the test result is negative.
This is because the sensitivity and specificity of the test are
relatively high, and a negative result significantly reduces the
likelihood of colorectal cancer. Therefore, repeating the test sooner
than 5 years in the absence of symptoms or risk factors would likely
result in unnecessary testing. Other screening options like FIT or
colonoscopy are not necessary if the patient has had a recent negative
Cologuard test."
    },
    {
      "role": "user",
      "content": "Could you give me the reference for American Cancer
Society for Cologuard?"
    },
    {
      "role": "assistant",
```

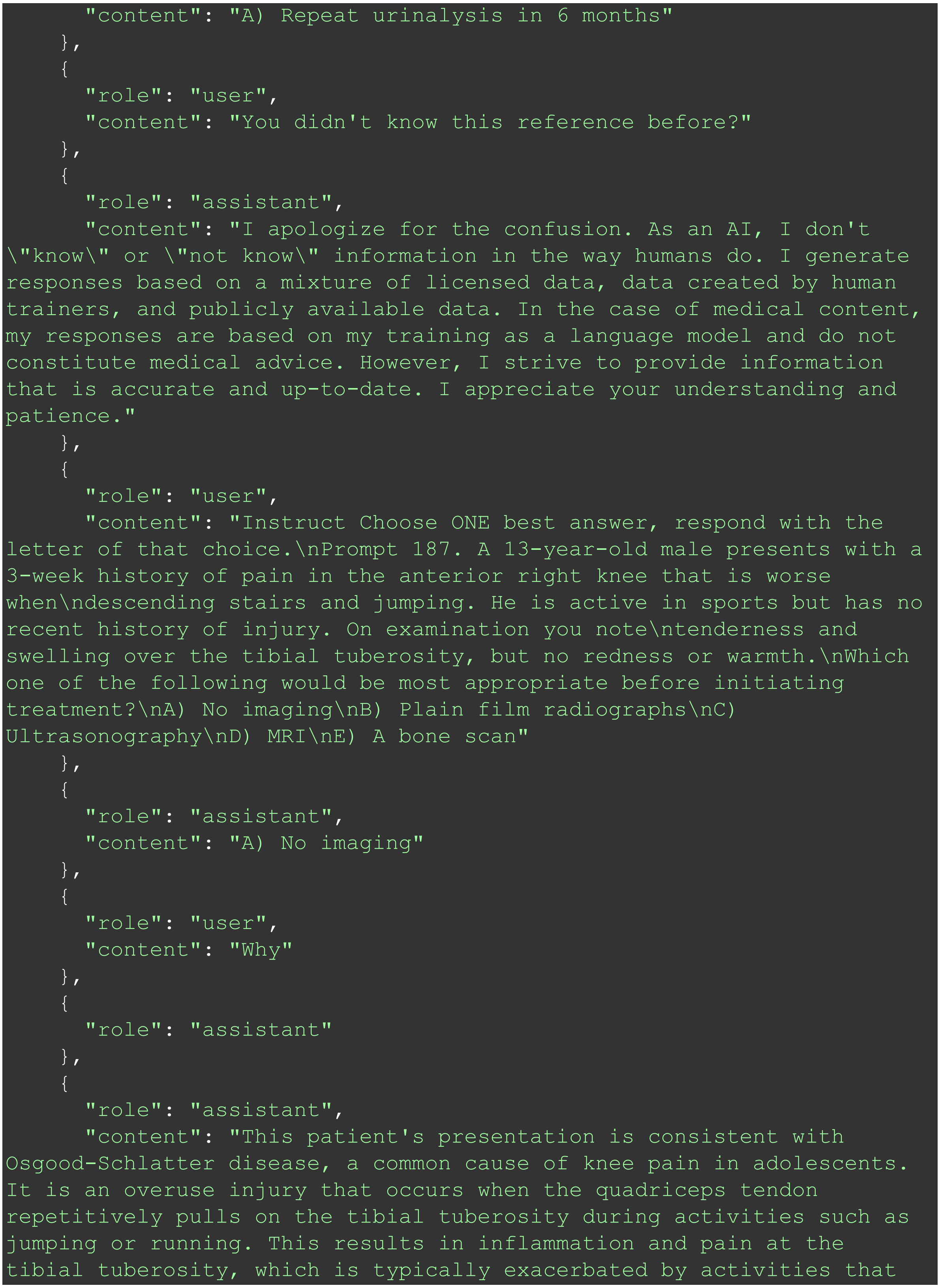
J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
      "content": "I apologize for the confusion, but upon further
review, it appears there may have been an error in my previous
response. The U.S. Multi-Society Task Force on Colorectal Cancer
recommends repeating the multitarget stool DNA test every 3 years, not
5, after a negative result. I regret the error and appreciate your
understanding. Therefore, the correct answer should be:\n\nD) Repeat
multitarget stool DNA testing in 3 years\n\nReference: \nRex DK,
Boland CR, Dominitz JA, et al. Colorectal Cancer Screening:
Recommendations for Physicians and Patients from the U.S. Multi-
Society Task Force on Colorectal Cancer. Gastroenterology.
2017;153(1):307-323. doi:10.1053/j.gastro.2017.05.013"
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 185. A 46-year-old female presents with
a 2-week history of polyuria. She has not had any discomfort
with\nurination or visible change in her urine. Her past medical
history includes hysterectomy for fibroids. Further history\nreveals
no concerning environmental exposures other than a 5-pack-year history
of smoking in her twenties. A physical\nexamination, including a
pelvic examination, is unremarkable. A urine dipstick reveals only 1+
RBCs. A microscopic\nurinalysis is negative with the exception of 7
RBCs/hpf, and a urine culture is negative.\nWhich one of the following
would you recommend at this point?\nA) Repeat urinalysis in 6
months\nB) Renal ultrasonography only\nC) Urine cytology and renal
ultrasonography\nD) Urine cytology and CT urography\nE) Cystoscopy and
CT urography"
    },
    {
      "role": "assistant",
      "content": "D) Urine cytology and CT urography"
    },
    {
      "role": "user",
      "content": "Why?"
    },
    {
      "role": "assistant",
      "content": "The patient's presentation of unexplained hematuria
(red blood cells in the urine) is concerning. Hematuria can be a sign
of several conditions, including urinary tract infections, kidney
stones, and more seriously, bladder or kidney cancer. Given her
history of smoking, which is a risk factor for bladder cancer, further
investigation is warranted. \n\nUrine cytology (choice D) can help
identify abnormal cells in the urine that may indicate cancer. CT
urography (choice D) is an imaging test that can provide detailed
images of the urinary tract and help identify any abnormalities such
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
as tumors or stones. \n\nWhile renal ultrasonography (choice B and C)
can also be used to evaluate the urinary tract, it may not provide as
detailed images as CT urography. \n\nCystoscopy (choice E) is a
procedure that allows the doctor to see inside the urinary bladder and
urethra, but it is more invasive and usually reserved for cases where
other tests suggest a problem. \n\nTherefore, urine cytology and CT
urography (choice D) would be the most appropriate next steps in this
patient's evaluation."
    },
    {
      "role": "user",
      "content": "But the patient is low risk"
    },
    {
      "role": "assistant",
      "content": "You're correct, and I apologize for the oversight.
In a low-risk patient with asymptomatic microscopic hematuria, as in
this case, the American Urological Association recommends an initial
evaluation with renal ultrasonography and cystoscopy. Therefore, the
correct answer should be:\n\nB) Renal ultrasonography only\n\nThis
approach is less invasive and has fewer potential complications than
other diagnostic procedures, such as CT urography or cystoscopy. If
the initial evaluation does not reveal a cause for the hematuria,
further testing may be considered based on the patient's risk factors
and clinical presentation."
    },
    {
      "role": "user",
      "content": "This patient presents with microscopic hematuria
without a clear cause. Past guidelines from the American\nUrologic
Association recommended cystoscopy and CT urography for all patients
over the age of 35 with\nmicroscopic hematuria. However, current
guidelines recommend risk stratification and emphasize the use\nof
imaging that has less radiation exposure. Because she is female and
younger than 50 years of age, this\npatient is at low risk of urologic
malignancy. Her smoking history and RBC count of <10 RBCs/hpf
are\nalso considered low risk. The guidelines recommend the option of
repeating urinalysis in 6 months before\nproceeding with imaging or
procedures, given her low risk (SOR C). Renal ultrasonography may be
used\nin intermediate-risk patients, while CT urography is reserved
for high-risk patients. Urine cytology is not\na recommended test in
this setting.\nRef: Barocas DA, Boorjian SA, Alvarez RD, et al:
Microhematuria: AUA/SUFU guideline. J Urol 2020;204(4):778-786.
2)\nJudge C, Cifu AS, Faris S: Management of patients with
microhematuria. JAMA 2021;326(6):563-564."
    },
    {
      "role": "assistant",
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

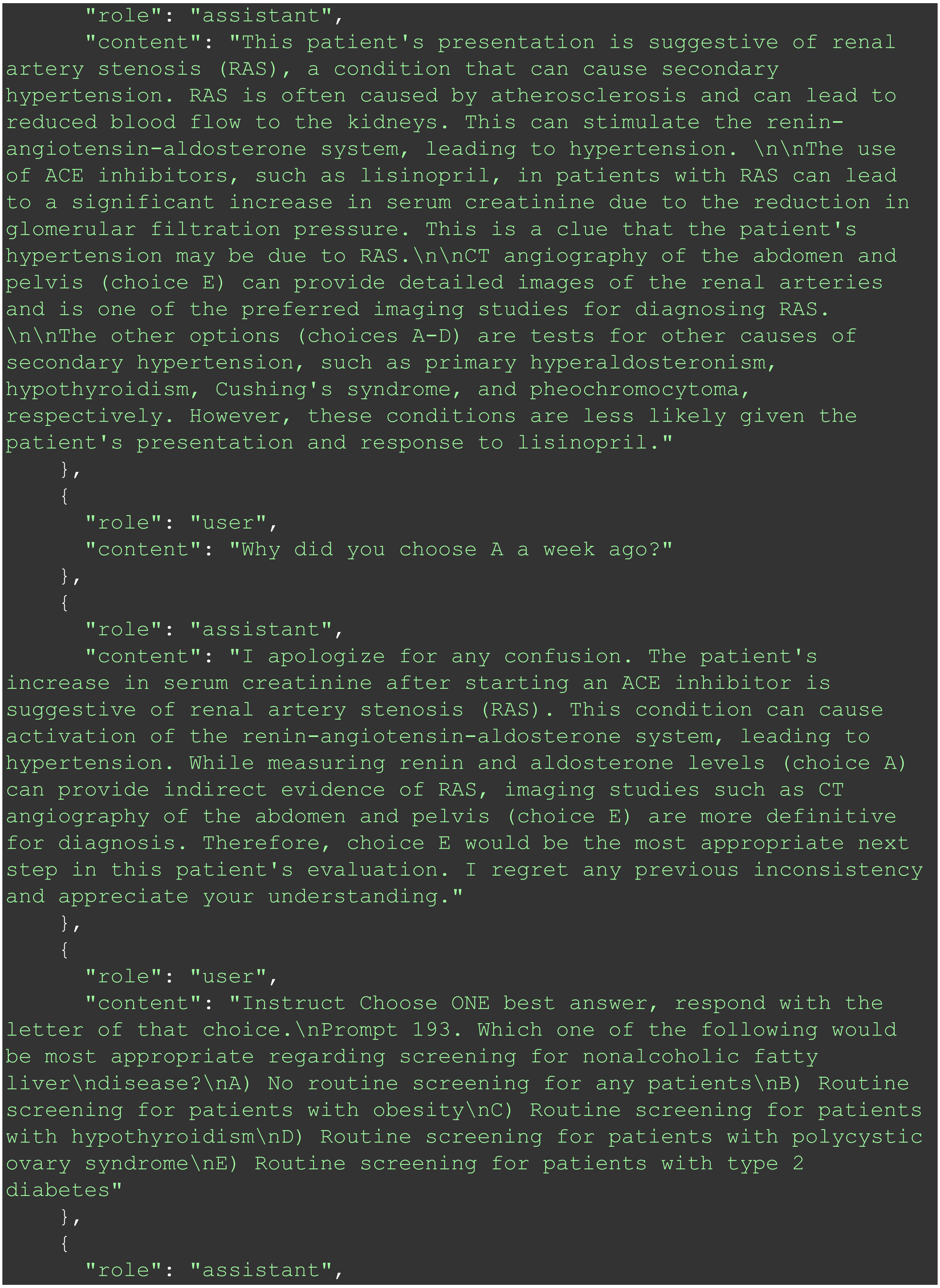


```
      "content": "A) Repeat urinalysis in 6 months"
    },
    {
      "role": "user",
      "content": "You didn't know this reference before?"
    },
    {
      "role": "assistant",
      "content": "I apologize for the confusion. As an AI, I don't
\"know\" or \"not know\" information in the way humans do. I generate
responses based on a mixture of licensed data, data created by human
trainers, and publicly available data. In the case of medical content,
my responses are based on my training as a language model and do not
constitute medical advice. However, I strive to provide information
that is accurate and up-to-date. I appreciate your understanding and
patience."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 187. A 13-year-old male presents with a
3-week history of pain in the anterior right knee that is worse
when\ndescending stairs and jumping. He is active in sports but has no
recent history of injury. On examination you note\ntenderness and
swelling over the tibial tuberosity, but no redness or warmth.\nWhich
one of the following would be most appropriate before initiating
treatment?\nA) No imaging\nB) Plain film radiographs\nC)
Ultrasonography\nD) MRI\nE) A bone scan"
    },
    {
      "role": "assistant",
      "content": "A) No imaging"
    },
    {
      "role": "user",
      "content": "Why"
    },
    {
      "role": "assistant"
    },
    {
      "role": "assistant",
      "content": "This patient's presentation is consistent with
Osgood-Schlatter disease, a common cause of knee pain in adolescents.
It is an overuse injury that occurs when the quadriceps tendon
repetitively pulls on the tibial tuberosity during activities such as
jumping or running. This results in inflammation and pain at the
tibial tuberosity, which is typically exacerbated by activities that
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
involve knee extension against resistance (e.g., descending stairs,
jumping).\n\nThe diagnosis of Osgood-Schlatter disease is typically
clinical, based on the patient's age, history, and physical
examination findings. Imaging is generally not necessary in typical
cases, and treatment usually involves activity modification, pain
management with NSAIDs, and stretching exercises. Therefore, no
imaging (choice A) would be most appropriate before initiating
treatment in this patient."
    },
    {
      "role": "user",
      "content": "Why you said it was B several days ago?"
    },
    {
      "role": "assistant",
      "content": "I apologize if there was a previous inconsistency.
The diagnosis of Osgood-Schlatter disease is typically a clinical one,
based on the patient's age, history, and physical examination
findings. In typical cases, imaging is not necessary. However, if the
diagnosis is uncertain or if symptoms do not improve with conservative
management, further imaging such as plain radiographs may be
considered to rule out other conditions. In this case, given the
classic presentation, no imaging (choice A) would be the most
appropriate initial approach."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 189. A 56-year-old male was recently
diagnosed with hypertension and started on lisinopril (Zestril). At
a\nfollow-up visit his blood pressure remains elevated and his serum
creatinine level has increased from 0.9 mg/dL to 1.8\nmg/dL (N 0.7–
1.3). He has no other known medical issues and has a normal
BMI.\nWhich one of the following should be ordered to confirm the most
likely cause of his hypertension?\nA) Renin and aldosterone levels\nB)
A TSH level\nC) 24-hour urinary free cortisol\nD) 24-hour urinary
fractionated metanephrines and normetanephrines\nE) CT angiography of
the abdomen and pelvis"
    },
    {
      "role": "assistant",
      "content": "E) CT angiography of the abdomen and pelvis"
    },
    {
      "role": "user",
      "content": "Why?"
    },
    {
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

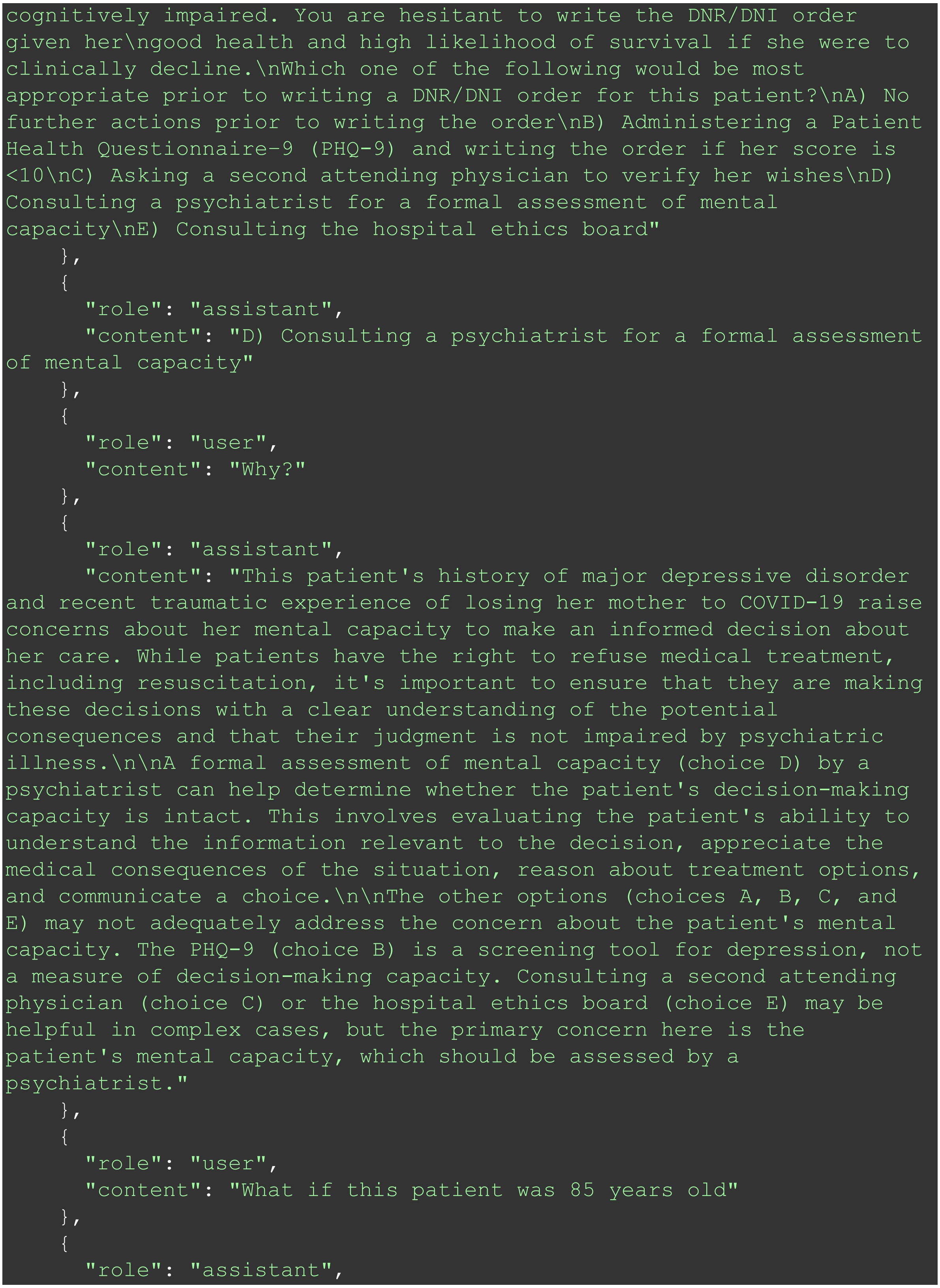


```
      "role": "assistant",
      "content": "This patient's presentation is suggestive of renal
artery stenosis (RAS), a condition that can cause secondary
hypertension. RAS is often caused by atherosclerosis and can lead to
reduced blood flow to the kidneys. This can stimulate the renin-
angiotensin-aldosterone system, leading to hypertension. \n\nThe use
of ACE inhibitors, such as lisinopril, in patients with RAS can lead
to a significant increase in serum creatinine due to the reduction in
glomerular filtration pressure. This is a clue that the patient's
hypertension may be due to RAS.\n\nCT angiography of the abdomen and
pelvis (choice E) can provide detailed images of the renal arteries
and is one of the preferred imaging studies for diagnosing RAS.
\n\nThe other options (choices A-D) are tests for other causes of
secondary hypertension, such as primary hyperaldosteronism,
hypothyroidism, Cushing's syndrome, and pheochromocytoma,
respectively. However, these conditions are less likely given the
patient's presentation and response to lisinopril."
    },
    {
      "role": "user",
      "content": "Why did you choose A a week ago?"
    },
    {
      "role": "assistant",
      "content": "I apologize for any confusion. The patient's
increase in serum creatinine after starting an ACE inhibitor is
suggestive of renal artery stenosis (RAS). This condition can cause
activation of the renin-angiotensin-aldosterone system, leading to
hypertension. While measuring renin and aldosterone levels (choice A)
can provide indirect evidence of RAS, imaging studies such as CT
angiography of the abdomen and pelvis (choice E) are more definitive
for diagnosis. Therefore, choice E would be the most appropriate next
step in this patient's evaluation. I regret any previous inconsistency
and appreciate your understanding."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 193. Which one of the following would
be most appropriate regarding screening for nonalcoholic fatty
liver\ndisease?\nA) No routine screening for any patients\nB) Routine
screening for patients with obesity\nC) Routine screening for patients
with hypothyroidism\nD) Routine screening for patients with polycystic
ovary syndrome\nE) Routine screening for patients with type 2
diabetes"
    },
    {
      "role": "assistant",
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
      "content": "A) No routine screening for any patients"
    },
    {
      "role": "user",
      "content": "Why did you choose E a week ago?"
    },
    {
      "role": "assistant",
      "content": "I apologize for any confusion. Current guidelines
from major health organizations, including the American Association
for the Study of Liver Diseases and the European Association for the
Study of the Liver, do not recommend routine screening for
nonalcoholic fatty liver disease (NAFLD) in any patient population,
even those with risk factors such as obesity or type 2 diabetes. This
is due to a lack of evidence that screening improves health outcomes.
Therefore, the correct answer is A) No routine screening for any
patients. I regret any previous inconsistency and appreciate your
understanding."
    },
    {
      "role": "user",
      "content": "Give me the reference"
    },
    {
      "role": "assistant",
      "content": "The American Association for the Study of Liver
Diseases (AASLD) does not recommend routine screening for nonalcoholic
fatty liver disease (NAFLD) in high-risk groups such as those with
diabetes or obesity. This is primarily due to the lack of evidence
that such screening leads to improved patient outcomes.\n\nReference:
\nChalasani N, Younossi Z, Lavine JE, et al. The diagnosis and
management of nonalcoholic fatty liver disease: Practice guidance from
the American Association for the Study of Liver Diseases. Hepatology.
2018;67(1):328-357. doi:10.1002/hep.29367."
    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 198. You admit a 50-year-old female to
the hospital with pneumonia secondary to COVID-19. Her
medical\nhistory is significant for controlled major depressive
disorder. Upon admission, she requires respiratory support
with\noxygen via a high-flow nasal cannula. You ask her wishes
regarding emergency measures and she states that she does\nnot wish to
be resuscitated or intubated. She reports that she watched her mother
die a “horrible death” from COVID-19\nabout 6 months earlier and does
not want to put herself through that. She does not have an advance
directive and her\nonly living relative is her father who is
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.



```
cognitively impaired. You are hesitant to write the DNR/DNI order
given her\ngood health and high likelihood of survival if she were to
clinically decline.\nWhich one of the following would be most
appropriate prior to writing a DNR/DNI order for this patient?\nA) No
further actions prior to writing the order\nB) Administering a Patient
Health Questionnaire-9 (PHQ-9) and writing the order if her score is
<10\nC) Asking a second attending physician to verify her wishes\nD)
Consulting a psychiatrist for a formal assessment of mental
capacity\nE) Consulting the hospital ethics board"
    },
    {
      "role": "assistant",
      "content": "D) Consulting a psychiatrist for a formal assessment
of mental capacity"
    },
    {
      "role": "user",
      "content": "Why?"
    },
    {
      "role": "assistant",
      "content": "This patient's history of major depressive disorder
and recent traumatic experience of losing her mother to COVID-19 raise
concerns about her mental capacity to make an informed decision about
her care. While patients have the right to refuse medical treatment,
including resuscitation, it's important to ensure that they are making
these decisions with a clear understanding of the potential
consequences and that their judgment is not impaired by psychiatric
illness.\n\nA formal assessment of mental capacity (choice D) by a
psychiatrist can help determine whether the patient's decision-making
capacity is intact. This involves evaluating the patient's ability to
understand the information relevant to the decision, appreciate the
medical consequences of the situation, reason about treatment options,
and communicate a choice.\n\nThe other options (choices A, B, C, and
E) may not adequately address the concern about the patient's mental
capacity. The PHQ-9 (choice B) is a screening tool for depression, not
a measure of decision-making capacity. Consulting a second attending
physician (choice C) or the hospital ethics board (choice E) may be
helpful in complex cases, but the primary concern here is the
patient's mental capacity, which should be assessed by a
psychiatrist."
    },
    {
      "role": "user",
      "content": "What if this patient was 85 years old"
    },
    {
      "role": "assistant",
```

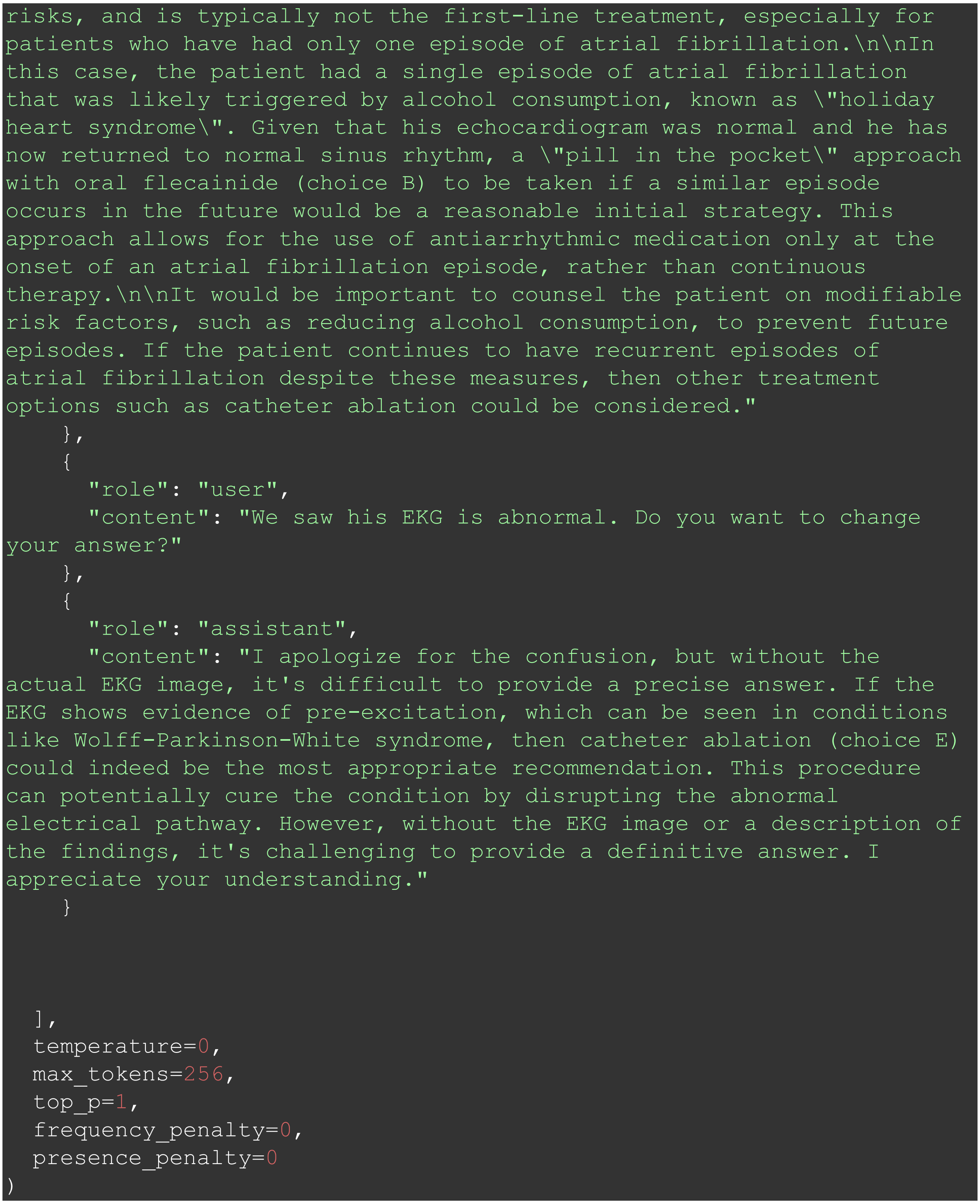
J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
      "content": "Even if the patient was 85 years old, the same
principles would apply. Age alone does not determine a person's
capacity to make medical decisions. If there are concerns about a
patient's mental capacity to make an informed decision, regardless of
their age, a formal assessment of mental capacity should be conducted.
In this case, consulting a psychiatrist for a formal assessment of
mental capacity (choice D) would still be the most appropriate step
prior to writing a DNR/DNI order. This would help ensure that the
patient's decision is being made with a clear understanding of the
potential consequences and that their judgment is not impaired by
psychiatric illness."
    },

    {

      "role": "user",

      "content": "The actual answer is A. Below is the reference. Do
you want to change your response? \n\nIndividuals are assumed to have
the mental capacity to make their own medical decisions unless there
are\nreasons to question a patient’s capacity, such as risk factors
for impaired decision-making or abrupt\nchanges in mental status. It
is the responsibility of the care team to honor the reasonable wishes
of each\npatient, including decisions regarding end-of-life care.
DNR/DNI orders can be fluid and changed\nthroughout a hospital stay as
more information becomes available. Patients should be provided with
the\ntools needed to make medical decisions and ensure that they have
insight into the consequences of the\ndecisions. A DNR/DNI order
should be written at this patient’s request without administering a
depression\nscreen or consulting other professionals.\nRef: Brody BD,
Meltzer EC, Feldman D, et al: Assessing decision making capacity for
do not resuscitate requests in depressed\npatients: How to apply the
“communication” and “appreciation” criteria. HealthCare Ethics
Committee Forum\n2017;29(4):303-311. 2) Barstow C, Shahan B, Roberts
M: Evaluating medical decision-making capacity in practice. Am\nFam
Physician 2018;98(1):40-46."

    },

    {

      "role": "assistant",

      "content": "A) No further actions prior to writing the order"

    },

    {

      "role": "user",

      "content": "Why you changed your response?"
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.

```
    },

    {

      "role": "assistant",

      "content": "Based on the provided critique, it is clear that the
patient's wishes regarding end-of-life care should be respected
without the need for further assessment of mental capacity or
consultation with other professionals. Unless there are reasons to
question a patient’s capacity, such as risk factors for impaired
decision-making or abrupt changes in mental status, individuals are
assumed to have the mental capacity to make their own medical
decisions. Therefore, a DNR/DNI order should be written at this
patient’s request (option A)."

    },
    {
      "role": "user",
      "content": "Instruct Choose ONE best answer, respond with the
letter of that choice.\nPrompt 199. A 41-year-old male who is a
construction worker sees you because he experienced palpitations
followed by\nan episode of syncope after drinking four beers at a
hotel bar while on vacation. He says that he was hospitalized
briefly,\nand an electric shock was administered in the emergency
department after medications were given. On discharge he was\ntold
that he had a brief episode of atrial fibrillation but that an
echocardiogram was normal. He was advised to follow up\npromptly with
his personal physician. An EKG performed in your office is shown
below.\nWhich one of the following would you recommend for this
patient?\nA) Aspirin\nB) Oral flecainide to be taken if a similar
episode occurs in the future\nC) Oral metoprolol therapy\nD) Warfarin
therapy\nE) Catheter ablation therapy\n"
    },
    {
      "role": "assistant",
      "content": "B) Oral flecainide to be taken if a similar episode
occurs in the future"
    },
    {
      "role": "user",
      "content": "What about Catheter?"
    },
    {
      "role": "assistant",
      "content": "Catheter ablation therapy (choice E) is a treatment
option for atrial fibrillation, particularly in patients who have
symptoms despite medication therapy, or who cannot tolerate
medications. However, it is an invasive procedure with potential
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.



```
risks, and is typically not the first-line treatment, especially for
patients who have had only one episode of atrial fibrillation.\n\nIn
this case, the patient had a single episode of atrial fibrillation
that was likely triggered by alcohol consumption, known as \"holiday
heart syndrome\". Given that his echocardiogram was normal and he has
now returned to normal sinus rhythm, a \"pill in the pocket\" approach
with oral flecainide (choice B) to be taken if a similar episode
occurs in the future would be a reasonable initial strategy. This
approach allows for the use of antiarrhythmic medication only at the
onset of an atrial fibrillation episode, rather than continuous
therapy.\n\nIt would be important to counsel the patient on modifiable
risk factors, such as reducing alcohol consumption, to prevent future
episodes. If the patient continues to have recurrent episodes of
atrial fibrillation despite these measures, then other treatment
options such as catheter ablation could be considered."
    },
    {
      "role": "user",
      "content": "We saw his EKG is abnormal. Do you want to change
your answer?"
    },
    {
      "role": "assistant",
      "content": "I apologize for the confusion, but without the
actual EKG image, it's difficult to provide a precise answer. If the
EKG shows evidence of pre-excitation, which can be seen in conditions
like Wolff-Parkinson-White syndrome, then catheter ablation (choice E)
could indeed be the most appropriate recommendation. This procedure
can potentially cure the condition by disrupting the abnormal
electrical pathway. However, without the EKG image or a description of
the findings, it's challenging to provide a definitive answer. I
appreciate your understanding."
    }



  ],
  temperature=0,
  max_tokens=256,
  top_p=1,
  frequency_penalty=0,
  presence_penalty=0
)
```

J Am Board Fam Med: first published as 10.3122/jabfm.2023.230433R1 on 30 August 2024. Downloaded from http://www.jabfm.org/ on 2 May 2025 by guest. Protected by copyright.