

# **GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial**

---

In the format provided by the  
authors and unedited

## **SUPPLEMENTARY MATERIALS**

**Supplement 1: Example of a Management Case, Rubric, and a high scoring and low scoring participant case response**

**Supplement 2: GPT Prompt and Responses for Example Case**

**Supplement 3: Post-Hoc Analysis Adjusted for Character Count**

**Supplement 4: Sensitivity Analysis Adjusting for both Time and Character Count**

**Supplement 5: Sensitivity Analysis from Repeated ANOVA for the Primary and Secondary Outcomes by Physicians with GPT-4 and with Conventional Resources Only**

**Supplement 6: Ratings of Likelihood and Extent of Harm in Responses by Physicians with GPT-4 and with Conventional Resources Only**

### **Supplement 1: Example of a Management Case, Rubric, and a high scoring and low scoring participant case response**

#### **Case #4 23 total points possible**

##### **CLINICAL VIGNETTE (PART 1):**

A 72 year-old woman with a history of GERD, type 2 diabetes, and poorly controlled hypertension presents with fevers, chills, and abdominal pain. A CT of her abdomen shows perforated cholecystitis and peritonitis. She is started on broad spectrum antibiotics and taken for an urgent open cholecystectomy. The procedure went well, and it was without any immediate complications.

On post-operative day #1, you are paged by nursing after she was noted to have a brief run of asymptomatic atrial fibrillation picked up on telemetry. On detailed review of her telemetry data, in the last 24 hours she has experienced 3 brief episodes of atrial fibrillation, each lasting 15-20 seconds.

From the perspective of her cholecystitis and peritonitis, she is clinically improving, has had no fevers or chills since the surgery, and she is tolerating a liquid diet. Her blood pressure and heart rate are currently normal. Her labs are notable only for a downtrending leukocytosis. Both a preoperative ECG and a current ECG show normal sinus rhythm without any significant abnormalities. She has no prior echocardiogram. She is currently on no anticoagulation or antiplatelet medications.

##### **QUESTION 1:**

In addition to continuing telemetry monitoring, would you recommend any additional monitoring, testing, or treatment at this time? If so, what?

**This question has a total of 4 points.**

|                                                                                                                                                                                                                                          |                                |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|
| Prophylactic anticoagulation to prevent VTE/DVT (e.g. enoxaparin 40mg daily) – Credit for either correct labeling of the intervention, or correct med + dose combination                                                                 | 1 point                        |
| TSH                                                                                                                                                                                                                                      | 1 point                        |
| Inclusion of up to 2 of the following: <ul style="list-style-type: none"><li>• Echocardiogram</li><li>• Electrolyte assessment and/or repletion</li><li>• Evaluation for additional triggers (e.g. pain, constipation, etc...)</li></ul> | 2 points max<br>(1 point each) |

## CLINICAL VIGNETTE (PART 2):

There are no episodes of atrial fibrillation on either post-operative days #2 or #3. However, on post-operative day #4, the day prior to anticipated discharge, the patient again develops another episode on telemetry that is confirmed by ECG. This episode is also completely asymptomatic. It lasts two hours, during which the ventricular rate ranges from 70 to 105 beats/min. Her blood pressure is not affected. It spontaneously converts back to normal sinus rhythm.

### QUESTION 2:

Would you recommend to this patient that she be started on therapeutic anticoagulation?  
What factors would influence your decision?

**This question has a total of 9 points.**

Decision on anticoagulation

|                                        |          |
|----------------------------------------|----------|
| Yes, start therapeutic anticoagulation | 2 points |
|----------------------------------------|----------|

Factors which influence the decision

|                                                                                                                                                                                                                                                                                             |                                 |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------|
| Patient preference and/or risk tolerance                                                                                                                                                                                                                                                    | 2 points                        |
| <ul style="list-style-type: none"><li>CHA2DS2-VASc or CHADS2 score</li><li>Listing only 1-2 component of these scoring systems separately, without mentioning the score itself</li></ul>                                                                                                    | 2 points<br>1 point             |
| Inclusion of up to 3 of the following 5 factors: <ul style="list-style-type: none"><li>HAS-BLED score</li><li>Duration of a-fib episodes</li><li># of post-op episodes</li><li>Information potentially obtained from an echo</li><li>A-fib burden (i.e. % of time spent in a-fib)</li></ul> | 3 points max (1 point for each) |

### QUESTION 3:

If the decision was made to start the patient on therapeutic anticoagulation on hospital day #4 (i.e. immediately after the asymptomatic, 2 hour long episode of a-fib), what would be the best choice for anticoagulation?

**This question has a total of 3 points.**

|                                                                     |          |
|---------------------------------------------------------------------|----------|
| Apixaban, rivaroxaban, edoxaban, dabigatran, therapeutic enoxaparin | 3 points |
| Unfractionated heparin gtt                                          | 1 point  |

#### QUESTION 4:

**This question has a total of 7 points.**

The patient is about to be discharged. Would you discharge her with an ambulatory ECG monitor?

|     |         |
|-----|---------|
| Yes | 1 point |
|-----|---------|

If so, what type of monitor would you choose?

|                                               |         |
|-----------------------------------------------|---------|
| Holter and/or Ziopatch (i.e. "14 day Holter") | 1 point |
|-----------------------------------------------|---------|

What factors should be considered in the choice of monitor for this patient?

|                                                                                                                                                                                                                                                            |                             |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|
| Inclusion of up to 2 of the following: <ul style="list-style-type: none"><li>• Cost of monitor</li><li>• Ease of insurance approval</li><li>• Duration of monitoring period</li><li>• Likelihood of the patient experiencing a lethal arrhythmia</li></ul> | 2 points max (1 point each) |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|

What would you do with the information from the monitor?

|                                                                                                                                         |                             |
|-----------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|
| Determine anticoagulation plan (i.e. whether or not the patient requires anticoagulation, or the duration of anticoagulation)           | 1 point                     |
| Inclusion of up to 2 of the following: <ul style="list-style-type: none"><li>• Determine if the patient requires rate control</li></ul> | 2 points max (1 point each) |

- |                                                                                                                                                                           |  |  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|
| <ul style="list-style-type: none"> <li>• Determine if the patient requires rhythm control</li> <li>• Determine if the patient requires more diagnostic testing</li> </ul> |  |  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|

## **Example of High-scoring vs. Low-scoring Responses**

### **High Scoring Response**

**Question 1. In addition to continuing telemetry monitoring, would you recommend any additional monitoring, testing, or treatment at this time? If so, what?**

I'd check her electrolytes and make sure hypoK or hypomag weren't contributing to her tachyarrhythmia. I'd also get tropes and an NT-proBNP to ensure there wasn't an inciting cardiac event triggering her afib, and TSH to ensure she wasn't hyperthyroid. I'd consider an echo. I'd ensure she was placed on tele as well. It's hard to say whether she is having paroxysmal afib in the setting of her acute illness or whether it has been happening at home before she was admitted. I would therefore discharge her with a ziopatch and cardiology follow-up to help make that distinction. She also has a CHA2DS2-VASc score of 4, which puts her at moderate-high risk of cardioembolic stroke and she thus warrants anticoagulation. I'd talk to her and see if she has any bleeding risk factors - if not, I'd likely start her on apixaban. I'd also provide situational awareness to the rest of her inpatient team that she could be started on metoprolol or amiodarone if she were to recur into afib but with RVR.

**Question 2. Would you recommend to this patient that she be started on therapeutic anticoagulation?**

Yes - she's at moderate-high risk of cardioembolic stroke. This could always be peeled off in the outpatient setting if she undergoes additional cardiac monitoring with no demonstrated afib in case her afib here is triggered by stress from recent surgery, but regardless of how permanent her afib is she is a high stroke risk and I would want to minimize this.

**Question 2b. What factors would influence your decision?**

Her willingness to take an anticoagulant, her ability to pay for a DOAC, her willingness to accept the hassles of being on warfarin or lovenox if she can't pay for a DOAC, her bleeding risk (does she fall, does she have any recent bleeds - could use HAS-BLED to better quantify this), her hemoglobin, whether she's had a stroke recently, whether I expect her surgical site to bleed, whether she can follow up with a physician after this hospitalization, whether anticoagulation would have a harmful interaction with any of her current medications.

**Question 3. If the decision was made to start the patient on therapeutic anticoagulation on hospital day #4 (i.e. immediately after the asymptomatic, 2 hour long episode of a-fib), what would be the best choice for anticoagulation?**

Apixaban

**Question 4a. The patient is about to be discharged. Would you discharge her with an ambulatory ECG monitor?**

Yes - I would want to monitor whether her afib is only present in the provoked post-op setting vs. if she has it recurrently at home, even when her body is no longer experiencing acute illness.

**Question 4b. If so, what type of monitor would you choose?**

Ziopatch

**Question 4c. What factors should be considered in the choice of monitor for this patient?**

Affordability and insurance coverage, accuracy of the monitor, whether the cardiologist she sees following hospitalization has the IT systems in place to receive the readout of the monitor, her ability to incorporate the monitor into her daily life and adhere to wearing it, the duration of monitoring that is most clinically appropriate (I'd go for 30 days).

**Question 4d. What would you do with the information from the monitor?**

Confirm whether she is having ongoing episodes of afib and if so better characterize its frequency, duration, and pattern; correlate any symptoms with her afib; guide decision-making around anticoagulation; selection/initiation of rhythm or rate control agents; and consideration for more permanent treatment options like cardiac ablation or Watchman device placement in case she wanted to get off of anticoagulation.

**Total score: 21 (of 23)**

### **Low Scoring Response**

**Question 1. In addition to continuing telemetry monitoring, would you recommend any additional monitoring, testing, or treatment at this time? If so, what?**

TTE and trop/CKMB. Can also consider a heart monitor such as a ziopatch upon d/c.

**Question 2. Would you recommend to this patient that she be started on therapeutic anticoagulation?**

Given recent surgery and current duration of episodes, at this time I would not, however I would place a monitoring device and insure soon outpt f/u w/ cardiology

**Question 2b. What factors would influence your decision?**

Medical hx, activity level, d/c labs, close f/u.

**Question 3. If the decision was made to start the patient on therapeutic anticoagulation on hospital day #4 (i.e. immediately after the asymptomatic, 2 hour long episode of a-fib), what would be the best choice for anticoagulation?**

Apixaban or Xarelto

**Question 4a. The patient is about to be discharged. Would you discharge her with an ambulatory ECG monitor?**

Yes

**Question 4b. If so, what type of monitor would you choose?**

I would consider a ziopatch or monitor such as that one.

**Question 4c. What factors should be considered in the choice of monitor for this patient?**

Risk factors, medical conditions, overall activity level.

**Question 4d. What would you do with the information from the monitor?**

Have it read by cardiology during pt f/u appt. It can also help evaluate for eventual need of AC or rhythm control medication

**Total score: 8 (of 23)**



## Supplement 2: GPT Prompt and Responses for Example Case

| GPT-4 Prompt                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | GPT-4 Responses                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>You are an experienced internal medicine physician undergoing a test of your medical management skills. I am going to give you a case with multiple questions. These are not real patients but are examples purely to test your management reasoning. Please answer these questions as best as possible. If the question is a yes or no question, please answer it directly. Please number your responses the same way they are numbered in the question stem.</p> <p>A 72 year-old woman with a history of GERD, type 2 diabetes, and poorly controlled hypertension presents with fevers, chills, and abdominal pain. A CT of her abdomen shows perforated cholecystitis and peritonitis. She is started on broad spectrum antibiotics and taken for an urgent open cholecystectomy. The procedure went well, and it was without any immediate complications.</p> <p>On post-operative day #1, you are paged by nursing after she was noted to have a brief run of asymptomatic atrial fibrillation picked up on telemetry. On detailed review of her telemetry data, in the last 24 hours she has experienced 3 brief episodes of atrial fibrillation, each lasting 15-20 seconds.</p> <p>From the perspective of her cholecystitis and peritonitis, she is clinically improving, has had no fevers or chills since the surgery, and she is tolerating a liquid diet. Her blood pressure and heart rate are currently normal. Her labs are notable only for a downtrending leukocytosis. Both a preoperative ECG and a current ECG show normal sinus rhythm without any significant abnormalities. She has no prior echocardiogram. She is currently on no anticoagulation or antiplatelet medications.</p> <p><b>QUESTION 1:</b> In addition to continuing telemetry monitoring, would you recommend any additional monitoring, testing, or treatment at this time? If so, what?</p> | <p><b>Question 1:</b> Yes, additional monitoring, testing, and potentially treatment might be recommended given the episodes of atrial fibrillation (AF), even if brief and asymptomatic. Additional testing could include an echocardiogram to assess her cardiac structure and function, as AF can be associated with structural heart disease. Electrolyte monitoring and correction of any abnormalities, as well as assessing thyroid function, would also be prudent.</p> <p><b>Question 2a:</b> The decision to start therapeutic anticoagulation would be influenced by her risk of stroke, assessed by tools like the CHA2DS2-VASc score, and balanced against her bleeding risk (assessed by tools like HAS-BLED). Given the brief and asymptomatic nature of her AF episodes, and without further risk factor assessment, it's challenging to make a definitive recommendation without additional patient-specific information.</p> <p><b>Question 2b:</b> Factors influencing this decision include the duration and frequency of AF episodes, her stroke risk (based on age, presence of hypertension, diabetes, heart failure, or history of stroke/TIA), and her bleeding risk.</p> <p><b>Question 3:</b> If the decision was made to start anticoagulation, the best choice would depend on her renal function, risk of bleeding, and potential for drug-drug interactions. Given her stable renal function (CKD stage II), a direct oral anticoagulant (DOAC) like apixaban or dabigatran could be preferred for their ease of use and favorable risk profile compared to warfarin, assuming no contraindications.</p> <p><b>Question 4a:</b> Yes, discharging her with an ambulatory ECG monitor could be advisable to better characterize the frequency, duration, and potential triggers of her AF episodes.</p> <p><b>Question 4b:</b> A type of monitor chosen could be a Holter monitor for continuous recording over a few days or an event recorder for longer monitoring if episodes are infrequent.</p> |

There are no episodes of atrial fibrillation on either post-operative days #2 or #3. However, on post-operative day #4, the day prior to anticipated discharge, the patient again develops another episode on telemetry that is confirmed by ECG. This episode is also completely asymptomatic. It lasts two hours, during which the ventricular rate ranges from 70 to 105 beats/min. Her blood pressure is not affected. It spontaneously converts back to normal sinus rhythm.

**QUESTION 2a:** Would you recommend to this patient that she be started on therapeutic anticoagulation?

**QUESTION 2b:** What factors would influence your decision?

**QUESTION 3:** If the decision was made to start the patient on therapeutic anticoagulation on hospital day #4 (i.e. immediately after the asymptomatic, 2 hour long episode of a-fib), what would be the best choice for anticoagulation?

**QUESTION 4a:** The patient is about to be discharged. Would you discharge her with an ambulatory ECG monitor?

**QUESTION 4b:** If so, what type of monitor would you choose?

**QUESTION 4c:** What factors should be considered in the choice of monitor for this patient?

**QUESTION 4d:** What would you do with the information from the monitor?

**Question 4c:** Considerations include the frequency of her AF episodes, the patient's ability to use the device properly, and the clinical question being answered (e.g., burden of AF, symptom-rhythm correlation).

**Question 4d:** Information from the monitor would guide further management, such as the need for and type of anticoagulation, the potential benefit of antiarrhythmic drugs, and whether any lifestyle modifications or further interventions are warranted.

### **Supplement 3: Post-Hoc Analysis Adjusted for Character Count**

| <b>Characteristic</b>                               | <b>Bet<br/>a</b> | <b>95%<br/>CI<sup>1</sup></b> | <b>p-valu<br/>e</b> |
|-----------------------------------------------------|------------------|-------------------------------|---------------------|
| <b>Group.final3</b>                                 |                  |                               |                     |
| <b>Physicians + Conventional Resources<br/>only</b> | —                | —                             |                     |
| <b>Physicians+GPT-4</b>                             | 3.7              | 0.7, 6.7                      | 0.016               |
| <b>Length (per 100 characters)</b>                  | 0.3              | 0.2, 0.4                      | <0.001              |

<sup>1</sup> CI = Confidence Interval

**Supplement 4: Sensitivity Analysis Adjusting for both Time and Character Count**

| Characteristic                           | Beta | 95% CI <sup>1</sup> | p-value |
|------------------------------------------|------|---------------------|---------|
| <b>Group.final3</b>                      |      |                     |         |
| Physicians + Conventional Resources only | —    | —                   |         |
| Physicians+GPT-4                         | 3.3  | 0.4, 6.1            | 0.027   |
| time_spent_min                           | 0.4  | 0.2, 0.6            | <0.001  |
| Char100                                  | 0.3  | 0.2, 0.3            | <0.001  |

<sup>1</sup> CI = Confidence Interval

**Supplement 5: Sensitivity Analysis from Repeated ANOVA for the Primary and Secondary Outcomes by Physicians with GPT-4 and with Conventional Resources Only**

| Variable                     | Comparisons between Physicians+GPT-4 and Physicians+Conventional Resources Only (Group Effect)              |                                                                                 |                     |                     |                    |                    |
|------------------------------|-------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------|---------------------|--------------------|--------------------|
|                              | Primary Analysis<br><br>(based on Generalized Mixed Effect Model, 95% Confidence Interval, <i>p-value</i> ) | Sensitivity Analysis from Repeated Measure ANOVA<br><br>(F-Statistics, P value) |                     |                     |                    |                    |
|                              |                                                                                                             | Group<br><br>(Main Effect)                                                      | Case                |                     | Group X Case       |                    |
|                              |                                                                                                             |                                                                                 | Between             | Within              | Between            | Within             |
| <b>Total Score</b>           | 6.5 (2.7 to 10.2), $p<0.001$                                                                                | 14.3,<br>$p<0.001$                                                              | 2.32,<br>$p=0.06$   | 102.3,<br>$p<0.001$ | 0.29,<br>$p=0.88$  | 2.11,<br>$p=0.08$  |
| <b>Management</b>            | 6.1 (2.5 to 9.7),<br>$p=0.001$                                                                              | 13.8,<br>$p<0.001$                                                              | 4.01,<br>$p=0.048$  | 36.7,<br>$p<0.001$  | 0.052,<br>$p=0.82$ | 0.18,<br>$p=0.68$  |
| <b>Factual</b>               | 9.6 (-3.1 to 22.3), $p=0.14$                                                                                | 1.99,<br>$p=0.162$                                                              | 1.57,<br>$p=0.21$   | 2.38,<br>$P=0.13$   | 0.17,<br>$p=0.68$  | 1.71,<br>$p=0.20$  |
| <b>Diagnostic</b>            | 12.1 (3.1 to 21.0), $p=0.009$                                                                               | 6.10,<br>$p=0.016$                                                              | 22.7<br>$p<0.001$   | 62.3,<br>$p<0.001$  | 0.31,<br>$p=0.58$  | 0.85,<br>$p=0.36$  |
| <b>Specific</b>              | 6.2 (2.4 to 9.9),<br>$p=0.002$                                                                              | 14.7,<br>$p<0.001$                                                              | 8.41,<br>$p=0.005$  | 91.3,<br>$p<0.001$  | 0.02,<br>$p=0.88$  | 0.07,<br>$p=0.79$  |
| <b>General</b>               | 3.3 (-1.3 to 7.9),<br>$p=0.2$                                                                               | 1.59,<br>$p=0.21$                                                               | 11.24,<br>$p=0.001$ | 39.7,<br>$p<0.001$  | 0.64,<br>$p=0.42$  | 4.18,<br>$p=0.046$ |
| <b>Time Spent in Seconds</b> | 119.3 (17.4 to 221.2), $p=0.022$                                                                            | 4.79,                                                                           | 2.00,<br>$p=0.16$   | 58.7,               | 0.37,              | 0.12,<br>$p=0.74$  |

|  |  |               |  |               |              |  |
|--|--|---------------|--|---------------|--------------|--|
|  |  | $p=0.03$<br>1 |  | $p<0.00$<br>1 | $p=0.5$<br>4 |  |
|--|--|---------------|--|---------------|--------------|--|

## **Supplement 6: Ratings of Likelihood and Extent of Harm in Responses by Physicians with GPT-4 and with Conventional Resources Only**

