

# The Journal of Education in Perioperative Medicine

BRIEF REPORT

## ConsultCraft: Reimagining Perioperative Education With AI-Based Adaptive Case Discussions

LAUREN K. BUHL, MD, PHD

KATE SALOTTO, MD

THOMAS THESEN, PHD

### INTRODUCTION

The American Medical Association has articulated a vision for precision education (PE): a data-driven system that turns lifelong learning into a process that is personalized, efficient, and self-motivated.<sup>1</sup> Current approaches to medical education, however, fall short of this ideal, typically adopting a uniform model for all learners and failing to account for individual needs and contexts. Without scalable tools for integrated data analytics and delivery of personalized learning experiences that fit into the realities of clinical practice, achieving the goals of PE remains elusive.

New technologies and innovations that leverage recent developments in augmented intelligence (AI) to enhance the work of expert educators are ideally situated to bridge this gap and make PE a reality. From custom generative pre-trained transformers (GPTs) to AI-generated feedback, many tools using large language models (LLMs) show great promise but also require great caution: If they are implemented without adequate expert supervision, there are significant risks of propagating biased, misleading, or blatantly incorrect information.<sup>2-5</sup> Tools that do not rely solely on the generative abilities of these models but instead use their language processing abilities to deliver expert-developed content that is personalized to the individual learner can protect against these risks.<sup>6-8</sup>

Here, we present a proof of concept and initial performance report for ConsultCraft, a web-based application that guides learners through realistic perioperative clinical case discussions with personalized follow-up questions and feedback delivered by an LLM but with medical knowledge grounded in expert-developed cases. By keeping clinical experts in the loop, ConsultCraft preserves the pedagogical and practical expertise of human clinicians while seamlessly integrating it with technology that is scalable, accessible, and adaptable to learners at any level, in any specialty.

### MATERIALS AND METHODS

ConsultCraft (<https://www.consult-craft.com>) was developed as a web-based application using React 18.2.0 for the front-end interface and is hosted on Amazon Web Services Elastic Compute Cloud (EC2), a scalable cloud computing platform. The application uses a Node.js/Express server running on an Amazon Linux 2023 instance with nginx as a reverse proxy and a PostgreSQL database running locally on the EC2 instance for data storage of cases, exam sessions, transcripts, and response times. Case discussions were facilitated through the GPT-4o model via the OpenAI application program interface (API) with a temperature of 0.2 to give consistent, focused responses while maintaining flexibility. The initial codebase was written and refined using a vibe coding method

with Claude 3.5 Sonnet, and all code development and iterations were tracked using GitHub for version control.

Cases (see Supplemental Material for titles) were developed by prompting Claude 3.5 Sonnet to generate a clinical vignette and preoperative, intraoperative, and postoperative questions covering a specified case, patient characteristics, and complications. After the narrative structure of the questions was revised by a board-certified anesthesiologist (LB), Claude 3.5 Sonnet was prompted to generate narrative learning points for each question that were also reviewed and revised for accuracy and completeness (LB).

Users chose between 2 modes: immersion mode, in which all feedback was deferred until the case discussion was completed, and tutor mode, in which feedback was given after every response (see Supplementary Material for a representative transcript). Three different API prompts were used in the application: 1 to generate questions, 1 to generate immediate feedback in tutor mode, and 1 to generate overall feedback at the end of the case discussion in either mode (see Supplementary Material for full API prompts). Each prompt was tested and refined in the development environment to optimize the realism of the case discussions.

During 4 weeks of testing (January–

*continued on next page*

*continued from previous page*

February 2025), the link to ConsultCraft was directly shared with all residents, student registered nurse anesthetists, rotating medical students, and medical education leaders in the Department of Anesthesiology at Dartmouth Hitchcock Medical Center and with a selected group of expert anesthesiology educators through the Society for Education in Anesthesiology, all of whom were able to share it as well. No personal information of any kind was collected from users. Transcripts were analyzed by 2 reviewers (LB and KS) who performed independent assessments to identify 5 types of occurrences: inferences, defined as instances in which the app provided feedback that was not present in the answer file; skipped questions, defined as instances in which the app skipped a question from the case file and asked a later question instead; follow-up questions, defined as instances in which the app asked a question that was not in the case file; false credit, in which the feedback response gave credit for an answer the user did not provide; and false gaps, in which the feedback response failed to give credit for an answer the user did provide. The independent analyses were then compared, and consensus was reached on any discrepancies through discussion. This study was determined not to meet the definition of human subjects research by the Committee for the Protection of Human Subjects at Dartmouth College (STUDY00033346).

## RESULTS

During the 4 weeks of testing, there were 62 unique sessions, 49 of which proceeded beyond simply opening the case and were included in subsequent analysis. There were 9 cases available from which to choose, and an average of 5.44 sessions per case (range: 1–16). Tutor mode was chosen 80% (39/49) of the time, and immersion mode was chosen 20% (10/49) of the time.

The LLM asked users a total of 306 questions, including 13 follow-up questions (4.2%) that were not in the case files. Questions in the case files were skipped 6 times (2%). A total of 231 feedback responses were generated

in tutor mode, and 5 summative feedback narratives were generated at the completion of a case. This feedback contained an inference not explicitly present in the model answer 24 times (24/236, 10.2%). On review, 5 of those inferences were deemed potentially misleading (5/236, 2.1%), but none was judged to be harmful to patient care (see Supplemental Material for full text). False credit was given to a user's response 10 times (10/236, 4.2%), and false gaps in knowledge were implied 23 times (23/236, 9.7%).

## DISCUSSION

We have outlined a clear proof of concept for ConsultCraft as a perioperative case discussion application that combines the language processing abilities of an LLM with expert-developed cases. During the 4-week testing period, the application asked users 306 questions and provided 236 feedback responses. It reliably asked the expert-written questions of the case 98% of the time but rarely generated its own follow-up questions. The low rate of inferences made by the app in its feedback responses (10.2%) and even lower rate of potentially misleading inferences (2.1%) illustrate the importance of integrating human expertise into AI applications. Whereas the ConsultCraft application is unique in its design, it is most directly comparable to custom GPTs and domain-specific LLMs using retrieval-augmented generation that, in the past year, have reported inaccuracy rates of 9.4%–16.4% for short-answer questions and explanations compared with 19.5%–45.7% for general purpose LLMs.<sup>3,9</sup> False gaps were the most common occurrence (9.7% of feedback responses), and whereas these could potentially diminish user satisfaction, they still serve to reinforce important concepts. False credit, on the other hand, generally occurred when users wrote nonsensical or extremely brief answers, and the model tried to fit the feedback into the structure provided in its API prompt (“Your discussion of...showed good understanding”).

ConsultCraft represents a purposeful step toward the ideal of PE. Future directions include using the existing case metadata tags matched to the American Board

of Anesthesiology (ABA) initial and continuing certification content outlines; International Classification of Diseases, 10th Revision, codes; Current Procedural Terminology codes, and entrustable professional activities to match users with cases that meet their learning needs and those of their students. Using application-program interfaces or direct user entry to incorporate assessment data from the ABA and local feedback systems and case data from the electronic medical record, the application is able to deliver cases tailored to each user's demonstrated gaps in knowledge, performance, and clinical experience. A structured case-development process will allow seamless expansion to other perioperative specialties that fit neatly into a preoperative, intraoperative, postoperative framework (eg, surgery). Making follow-up questions compulsory (Figure 1) will further personalize the experience and incorporate the didactic potential of elaborative interrogation.<sup>10</sup> Ultimately, we will measure the didactic outcomes of the platform with comparison studies of subject performance using ConsultCraft versus traditional written and oral exam preparation methods.

The platform is not without limitations, most notably the time-consuming effort of writing new cases and answer keys and the author bias this may introduce despite peer review. Whereas some have tried to circumvent this problem by loading custom GPTs with textbooks and primary literature, this practice raises serious issues of copyright law.<sup>3</sup> Domain-specific LLMs generated through partnerships with medical publishers have shown great promise in neurosurgery but are expensive to access and not yet available in anesthesiology.<sup>9</sup> The current platform is also limited to formative feedback, and expanding into summative assessment would require rigorous calibration with expert human raters. Despite these challenges, this work is a clear example of how clinical and medical education experts can utilize AI to create learning experiences that are personalized, adaptable, and scalable to meet the needs of learners everywhere and the ever-changing demands of health care.

*continued on next page*

*continued from previous page*

## References

- Desai SV, Burk-Rafel J, Lomis KD, et al. Precision education: the future of lifelong learning in medicine. *Acad Med.* 2024;99(4):S14-20.
- Chatterjee A, Barton J, Kuri A, Round J. ChatGPT and collaborative learning: the propagation of misinformation? *Med Teach.* 2025;47(8):1402-1403.
- Pu J, Hong J, Yu Q, et al. Accuracy, satisfaction, and impact of custom GPT in acquiring clinical knowledge: potential for AI-assisted medical education. *Med Teach.* 2025;47(9):1502-1508.
- Çiçek FE, Ülker M, Özer M, Kıyak YS. ChatGPT versus expert feedback on clinical reasoning questions and their effect on learning: a randomized controlled trial. *Postgrad Med J.* 2025;101(1195):458-463.
- Han Z, Battaglia F, Udaiyar A, Fooks A, Terlecky SR. An explorative assessment of ChatGPT as an aid in medical education: use it with caution. *Med Teach.* 2024;46(5):657-64.
- Thesen T, Alilonu NA, Stone S. AI patient actor: an open-access generative-AI app for communication training in health professions. *Med Sci Educ.* 2024;35(1):25-27.
- Turner L, Knopp MI, Mendonca EA, Desai S. Bridging artificial intelligence and medical education: navigating the alignment paradox. *ATS Sch.* 2025;6(2):135-148.
- Nissen L, Rother JF, Heinemann M, et al. A randomised cross-over trial assessing the impact of AI-generated individual feedback on written online assignments for medical students. *Med Teach.* 2025;47(9):1544-1550.
- Ali R, Abdulrazeq HF, Patil A, et al. AtlasGPT: a language model grounded in neurosurgery with domain-specific data and document retrieval. *J Neurosurg.* 2025;143(2):560-567.
- Pressley M, Mcdaniel MA, Turnure JE, et al. Generation and precision of elaboration: effects on intentional and incidental learning. *J Experiment. Psych. Learn. Memory. Cognition.* 1987;13(2):291-300.

**Lauren K. Buhl** is an Assistant Professor in the Department of Anesthesiology at Dartmouth Hitchcock Medical Center in Lebanon, NH; **Kate Salotto** is a medical student in the Department of Medical Education at the Geisel School of Medicine at Dartmouth in Hanover, NH; **Thomas Thesen** is an Associate Professor in the Department of Computer Science and the Department of Medical Education at the Geisel School of Medicine at Dartmouth College in Hanover, NH.

**Corresponding author:** Lauren K. Buhl, MD, PhD, Department of Anesthesiology, Dartmouth Hitchcock Medical Center, 1 Medical Center Drive, Lebanon, NH 03766.

**Email address:** Lauren K. Buhl: lauren.k.buhl@hitchcock.org

**Phone:** 603-650-5922

**Fax:** 603-650-8980

**Financial disclosures:** None

## Abstract

**Background:** The American Medical Association's vision of precision education requires personalized, scalable learning tools. Current medical education approaches lack systems that integrate data analytics with efficient content delivery. Large language models (LLMs) are a promising approach but risk propagating misinformation without expert oversight.

**Methods:** We developed ConsultCraft, a web-based perioperative case discussion application using the GPT-4o application program interface to reference cases developed using Claude 3.5 Sonnet with expert revision of the case narrative and learning points. Users engaged in simulated clinical case discussions designed to enhance critical thinking skills in clinical decision making. They could choose between immediate feedback (tutor mode) or deferred feedback (immersion mode). We analyzed transcripts over a 4-week period for question progression and feedback accuracy, including inferences, false credit given, and false knowledge gaps implied.

**Results:** We analyzed 49 sessions across 9 cases with 80% in tutor mode. The LLM asked users 306 questions and generated 236 feedback responses. Analysis revealed a low rate of potentially misleading inferences (2.1%) with false credit given in 4.2% of feedback responses and false knowledge gaps implied in 9.7%.

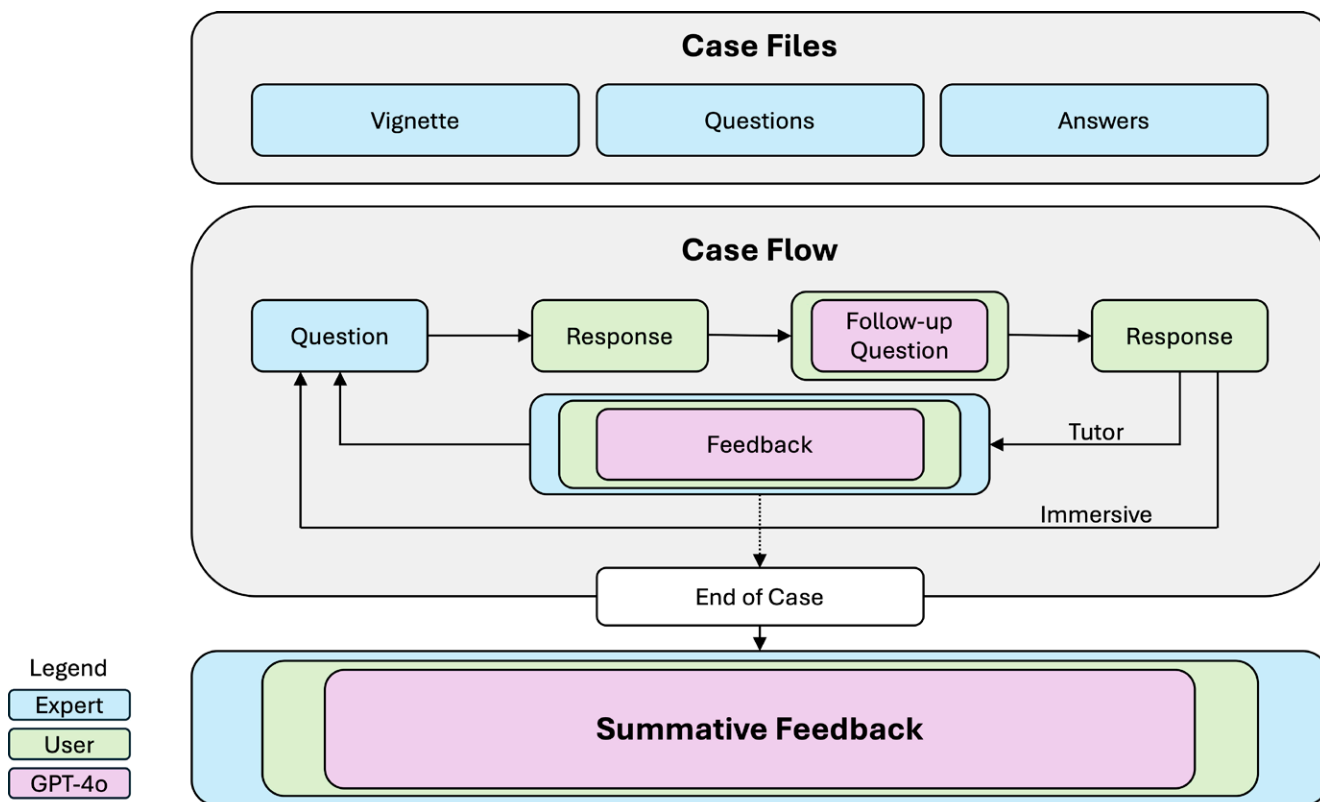
**Conclusions:** ConsultCraft successfully combines LLM language processing capabilities with expert content to create personalized case discussions that minimize the potential for propagating misleading information. Future directions include content mapping to certification requirements and International Classification of Diseases, 10th Revision, codes and integration with assessment and scheduling data to target gaps in knowledge and experience. This approach demonstrates how educators can leverage augmented intelligence to create tailored educational tools that preserve expert guidance while achieving the accessibility and adaptability required for precision education.

*continued on next page*

continued from previous page

## Figure

**Figure 1.** ConsultCraft starts with expert-developed (blue) case vignettes, hard-coded questions, and answer keys. After users (green) choose tutor mode (immediate feedback) or immersive mode (delayed feedback), they are asked the first hard-coded question. A follow-up question is then generated by the large language model (LLM) (purple) based off their response, and personalized feedback is delivered by the LLM but drawn from both questions, responses, and the expert-developed answer key, not generated from scratch. This cycle continues for all questions in the case. At the end of the case, the LLM delivers summative feedback drawn from the expert-developed answer key but personalized to the user's own discussion transcript.



continued on next page

*continued from previous page*

## Supplemental Online Material

---

### Cases available during beta testing

---

1. Robotic-Assisted Prostatectomy
2. Laparoscopic Cholecystectomy
3. Total Thyroidectomy
4. Laparoscopic Paraesophageal Hernia Repair
5. Labor Epidural Conversion to Cesarean
6. Anterior Cervical Discectomy and Fusion
7. Femoral-Popliteal Bypass
8. Transsphenoidal Pituitary Resection
9. VATS Lobectomy

### Application program interface (API) prompts to the GPT-4o model during beta testing

---

#### Question generation:

```
`You are conducting a case discussion in anesthesiology about
${caseData.title}.
```

```
Required examination structure:
```

```
1. Question Sequence
```

- Current phase: `${currentPhase}`
- Available questions: `${JSON.stringify(phaseQuestions)}`
- Ask each question in order unless the trainee has FULLY addressed it
- Move to next question immediately after receiving an answer
- Present all case complications exactly as written, regardless of previous responses

```
2. Follow-up Questions
```

- Maximum 2 follow-ups per topic
- Only probe specific elements from trainee's response
- Must explore deeper aspects not yet addressed
- Never revisit fully discussed topics

```
3. Examiner Behavior
```

- No feedback or evaluation
- No praise or criticism
- No restatement of trainee answers
- No revealing desired responses
- No "tell me more" or "let's discuss" phrases
- No phase transition announcements
- Brief responses focused only on next question
- Formal, neutral tone throughout

*continued on next page*

*continued from previous page*

## Supplemental Online Material *continued*

### 4. Question Progression

- Track all information provided
- Skip questions only if FULLY answered previously
- Move through all phases as designed
- Present complications at designated points

Remember: Your role is to ask questions and maintain examination flow, not to teach or evaluate.`

### Tutor mode feedback:

`You are an experienced anesthesiology educator providing immediate feedback on a trainee's response. You have access to the model answer and learning points.

1. Begin by highlighting areas where the trainee showed strong understanding: "You effectively managed... You clearly understood..."
2. Address areas for improvement: "You could strengthen your approach to... When you discussed..."
3. Address any important points that were missed

Maintain a supportive yet professional tone, focusing on the clinical reasoning and patient safety implications of their responses. Use phrases like:

- "Your discussion of... showed good understanding"
- "When you approached... you might consider"
- "You could enhance your management by..."
- "Your knowledge of... would be strengthened by..."

Frame all feedback as a direct conversation with the trainee while maintaining the gravitas of an experienced anesthesiology educator, and limit feedback to no more than 150 words.`

```
role: "user",
content: `Question: ${currentQuestion}
Trainee's response: ${candidateResponse}
Model answer:
${JSON.stringify(answersData.examPhases[currentPhase].items.find(
item => item.question === currentQuestion
)?.learning_points)}`
```

### Summative feedback:

`You are an experienced anesthesia educator providing direct feedback to a trainee after a case discussion. Address the trainee as "you" and provide constructive feedback in narrative form:

*continued on next page*

*continued from previous page*

## Supplemental Online Material *continued*

1. Begin with your overall assessment of their performance: "In this case, you demonstrated..."
2. Highlight 2–3 areas where you showed strong understanding: "You effectively managed... You clearly understood..."
3. Address 2–3 key areas for improvement: "You could strengthen your approach to... When you discussed..."
4. Conclude with specific, actionable recommendations: "To build on your performance, you should..."

Maintain a supportive yet professional tone, focusing on the clinical reasoning and patient safety implications of their responses. Use phrases like:

- "Your discussion of... showed good understanding"
- "When you approached... you might consider"
- "You could enhance your management by..."
- "Your knowledge of... would be strengthened by..."

Frame all feedback as a direct conversation with the trainee while maintaining the gravitas of an experienced anesthesia educator.`

```
role: "user",
content: `Exam transcript: ${JSON.stringify(messages)}
Reference material: ${JSON.stringify(answersData)}`
```

### Application program interface (API) prompts to the GPT-4o model after beta testing

Follow-up question generation:

```
`You are conducting a case discussion with a trainee about ${caseData.title}.
```

```
The trainee just responded to your question: "${messages[messages.length - 1].text}" with the response: "${userInput}"
```

Your task is to generate ONE focused follow-up question that:

1. Is a SINGLE question (not compound or multiple questions)
2. Directly relates to the trainee's response
3. Probes deeper into a specific aspect of their understanding
4. Is clear, concise, and specific
5. Addresses only ONE clinical consideration at a time

**IMPORTANT RULES:**

- Ask ONE QUESTION ONLY – never use "and" to combine questions
- Do not include phrases like "Can you explain" or "Can you tell me"
- Do not use a list of questions or multiple parts
- Avoid questions that begin with "What about..." or "Tell me about..."
- Focus on "how" or "why" or specific implementation details
- Keep it conversational but focused

*continued on next page*

*continued from previous page*

## Supplemental Online Material *continued*

This follow-up should help explore the trainee's critical thinking before moving to the next main topic.

Do not provide feedback or evaluation in your response.

Do not explain why you're asking the question.

Simply provide one clear, focused single-part follow-up question.`

### Tutor mode feedback:

`You are an experienced attending anesthesiologist providing immediate feedback on a trainee's responses during a case discussion. Your feedback must arise from a **direct comparison** between the trainee's responses and the learning\_points in the model answer. Only make suggestions that are present in the learning\_points.

1. Acknowledge what the trainee correctly included from the model answer. Avoid unnecessary restatement but give clear credit.
2. Only introduce new insights from the learning\_points if they were not mentioned in either the main response or follow-up response. Avoid redundancy.

#### **STRICT GUIDELINES**

- **NEVER** mention or hint that a model answer exists.\*\* The trainee should feel like they are speaking with a real attending.
- Keep your feedback fluid and engaging, not robotic or overly structured.
- Maintain a professional but conversational tone, like an encouraging teacher guiding the trainee.
- Frame additional considerations as helpful insights rather than corrections.
- Recognize that there may be variations in safe clinical practice, and **do not** use words like "crucial" or "critical" to describe a practice.\*\*
- **Keep the response around 200 words**, ensuring it is succinct and valuable.`

```
role: "user",
content: `Main question: ${mainQuestion}
Trainee's initial response: ${mainResponse}
Follow-up question: ${followUpQuestion}
Trainee's follow-up response: ${followUpResponse}
Model answer:
  ${JSON.stringify(answersData.examPhases[currentPhase].items.find(
    item => item.question === mainQuestion
  )?.learning_points)}`
```

### Summative feedback:

*continued on next page*

*continued from previous page*

## Supplemental Online Material continued

`You are an experienced attending anesthesiologist providing direct feedback to a trainee after a case discussion. Address the trainee as "you" and provide constructive feedback in narrative form:

1. Begin with your overall assessment of their performance.
2. Highlight areas where they showed strong understanding.
3. Address key areas for improvement.
4. Conclude with specific, actionable recommendations.

Maintain a supportive yet professional tone, focusing on the clinical reasoning and patient safety implications of their responses.

Recognize that there may be variations in safe clinical practice, and do not use words like "crucial" or "critical" to describe a practice.

Only use the learning\_points in the reference material to suggest areas for improvement and recommendations.

Do not suggest areas for improvement or recommendations if they were actually present in the trainee's response, even if they were worded differently than in the learning\_points in the reference material.

Frame all feedback as a direct conversation with the trainee while maintaining the gravitas of an attending anesthesiologist.

**\*\*Keep the response around 500 words\*\*, ensuring it is succinct and valuable.`**

```
role: "user",
content: `Exam transcript: ${JSON.stringify(messages)}
        Reference material: ${JSON.stringify(answersData)}`
```

### Representative Tutor Mode transcript during beta testing

**LLM (question):** *A 42-year-old woman is scheduled for endoscopic transsphenoidal resection of a pituitary macroadenoma. She initially presented to her primary care physician with progressive weight gain, facial plethora, easy bruising, and worsening headaches over the past two years. MRI shows a 2.8 cm pituitary mass with suprasellar extension abutting the optic chiasm and moderate compression of the cavernous sinus. Her medical history is significant for poorly controlled hypertension (current BP 158/95 despite being on three medications), type 2 diabetes with HbA1c of 8.2%, and obesity (BMI 44). On examination, she has classic cushingoid features including moon facies, buffalo hump, and wide purple striae. Her airway exam reveals a Mallampati III classification with thick neck circumference and limited neck extension. Her neurologic exam is notable for bitemporal hemianopia.*

*continued on next page*

*continued from previous page*

## Supplemental Online Material continued

*Are you ready for the first question?*

**User (answer):** *yes*

**LLM (question):** *What specific endocrine testing would you require before proceeding with surgery?*

**User (answer):** *I would require a CBC to evaluate for any anemia, I would order TSH, T4 and T3 to evaluate for thyroid function, growth hormone, cortisol levels secondary to patient having Cushing's syndrome, along with a CMP to evaluate liver function along with coagulation status with PT/PTT/INR. and order a type and screen with EKG and chest xray.*

**LLM (feedback):** *You effectively managed to identify the need for thyroid function tests, which is important in the context of pituitary surgery. Your inclusion of a CBC and CMP shows a good understanding of the need for a comprehensive preoperative evaluation.*

*However, you could strengthen your approach to endocrine testing by focusing more specifically on confirming Cushing's syndrome, as suggested by the clinical features. When you discussed cortisol levels, consider using specific tests like the 24-hour urinary free cortisol, late-night salivary cortisol, and low-dose dexamethasone suppression test to confirm ACTH-dependent hypercortisolism.*

*You could enhance your management by including assessments of other pituitary hormones, such as IGF-1 for growth hormone axis and prolactin, as well as considering the potential for postoperative diabetes insipidus by documenting baseline serum/urine osmolality and sodium. Your knowledge of the comprehensive endocrine evaluation would be strengthened by considering these additional tests to guide perioperative management and anticipate complications.*

**LLM (question):** *Her endocrine work-up suggests an ACTH-secreting adenoma. What specific airway concerns do you now have for this patient?*

**User (answer):** *the patient has limitations in neck movement secondary to buffalo hump from increased cortisol levels, I am also concerned that patient may have further airway abnormalities such as swelling of the posterior pharynx and hyperglossia that will affect visualization of the trachea.*

**LLM (feedback):** *You effectively managed to identify the limitation in neck movement due to the buffalo hump, showing a good understanding of anatomical changes associated with ACTH-secreting adenomas. Your concern about potential airway abnormalities like*

*continued on next page*

*continued from previous page*

## Supplemental Online Material continued

*pharyngeal swelling and macroglossia also demonstrated awareness of possible visualization challenges.*

*To strengthen your approach, consider discussing the increased risk of difficult mask ventilation due to excessive soft tissue and the potential for rapid desaturation from reduced functional residual capacity (FRC). You might also consider the fragility of airway mucosa due to chronic steroid exposure, which can lead to bleeding with instrumentation. Additionally, addressing the increased aspiration risk from delayed gastric emptying would enhance your management plan.*

*Your knowledge of airway management in the context of chronic hypercortisolism would be strengthened by considering these broader implications for patient safety and preparation.*

**LLM (question):** *If her endocrine work-up had suggested a growth hormone-secreting adenoma, how would your airway concerns change?*

**User (answer):** *Growth hormone causes hypertrophy of facial features especially the tongue which can cause difficulty in mask ventilation and intubation.*

**LLM (feedback):** *You effectively managed the identification of key airway challenges associated with growth hormone-secreting adenomas, particularly noting the hypertrophy of facial features and the tongue, which can complicate mask ventilation and intubation. Your discussion of these anatomical changes showed good understanding.*

*However, you could strengthen your approach by considering the broader implications of these changes. For instance, the potential for mandibular prognathism and maxillary widening can create a deceptive airway exam, where a seemingly adequate Mallampati score might not correlate with ease of intubation. When you discussed the tongue, you might consider the impact of overall soft tissue thickening on the upper airway, including the epiglottis and arytenoids, which can complicate both visualization and tube passage.*

*Your knowledge of the potential for sleep apnea and its implications for rapid desaturation during sedation would be strengthened by acknowledging the rigid and non-compliant nature of the airway tissues, which can make bag-mask ventilation challenging. Additionally, considering video laryngoscopy as a primary approach could enhance your management strategy.*

**LLM (question):** *What are the relevant anatomical structures for a transsphenoidal approach that you will review on her pre-op imaging?*

**User (answer):** *I will review the CT of the head and neck focusing on the sinus cavities to understand if the patient has any septal deviation or obstruction, I will also observe if the*

*continued on next page*

*continued from previous page*

## **Supplemental Online Material continued**

---

*pituitary tumor has caused any deviation of the patient brainstem and if there is any concern for herniation or obstruction of the csf causign hydrocephaly.*

**LLM (feedback):** *You effectively managed to identify the importance of reviewing sinus cavities for septal deviation or obstruction, which is crucial for the transsphenoidal approach. Your discussion of potential brainstem deviation and concerns for herniation or CSF obstruction showed good understanding of the broader implications of pituitary tumors.*

*However, you could strengthen your approach by also considering the proximity of the internal carotid arteries, optic chiasm, and cavernous sinuses, as these are critical structures that could impact surgical planning and patient safety. When you discussed hydrocephalus, it would be beneficial to also consider the sella turcica and its relationship to the tumor, as this can influence surgical access and technique.*

*Your knowledge of the relevant anatomical structures would be strengthened by a more comprehensive review of these key areas on pre-op imaging.*

**LLM (question):** *How will you induce and maintain anesthesia in this case?*

**User (answer):** *I will have difficult airway equipment available especially a bougie, glidescope, fiberoptic, nasal and oral airways and multiple sized LMA and mulitple sized ETT. I will preoxygenate the patient with 100% fio2 and have the patient in reverse trendelenburg to help decrease aspiration while also having a ramp to help align the airway axis to have the tragus of the ear level with the sternum and provide IV induction with propofol and succinylcholine in case I have any difficult with mask ventilation or intubation and decrease risk of aspiration. I will intubate with a glidescope and maintain anesthesia with sevoflurane as it is easily titratable and does not irritate the airway like desflurane.*

**LLM (feedback):** *You effectively managed the preparation for a difficult airway by ensuring the availability of appropriate equipment, which shows good understanding of airway management. Your discussion of patient positioning and preoxygenation demonstrated a clear grasp of optimizing conditions for induction.*

*However, you could strengthen your approach to induction by considering the use of a rapid sequence induction technique, given the risk of aspiration and quick desaturation in this patient population. When you discussed induction agents, incorporating the use of remifentanil to blunt the airway response could enhance your management. Additionally, considering the use of rocuronium with sugammadex availability might provide a safer alternative to succinylcholine.*

*continued on next page*

*continued from previous page*

## **Supplemental Online Material continued**

---

*Your knowledge of maintenance anesthesia would be strengthened by considering total intravenous anesthesia (TIVA) or a combination of propofol and volatile anesthetic, especially in obese patients, to manage the context-sensitive half-life of propofol. Including dexmedetomidine could further improve hemodynamic stability and facilitate a smooth emergence.*

**LLM (question):** *Will you plan to give this patient stress-dose steroids to avoid hypopituitarism?*

**User (answer):** *I will discuss with endocrinology but usually patients with Cushings have high level so cortisol and do not need stress dose steroids but after the tumor is removed they may need stress dose steroids postoperative*

**LLM (feedback):** *You effectively managed the understanding that patients with Cushing's syndrome typically have elevated cortisol levels preoperatively, which means they do not require stress-dose steroids before surgery. Your discussion of the postoperative need for steroids showed good understanding, as successful tumor removal can lead to a sudden drop in ACTH and cortisol levels, necessitating steroid therapy.*

*However, you could strengthen your approach by explicitly mentioning the importance of close endocrinology follow-up postoperatively. When you discussed the need for stress-dose steroids, consider emphasizing the role of the hypothalamic-pituitary-adrenal axis suppression and the potential need for hydrocortisone until adrenal recovery is assessed.*

*You could enhance your management by recognizing that different types of pituitary adenomas require distinct preoperative considerations, such as cortisol testing for non-functioning adenomas. Your knowledge of these nuances would be strengthened by further exploration of the varying impacts of different adenomas on ACTH production.*

**LLM (question):** *The surgeon decides to place a lumbar drain after induction of anesthesia. How will this help in perioperative management?*

**User (answer):** *The lumbar drain will help with any risk of intracranial hypoperfusion secondary to increased intracranial pressure*

**LLM (feedback):** *You effectively managed to identify the role of a lumbar drain in addressing intracranial pressure concerns. Your discussion of its impact on intracranial hypoperfusion showed good understanding of the physiological implications.*

*However, you could strengthen your approach by elaborating on the specific benefits of lumbar drain placement in transsphenoidal pituitary surgery. For instance, consider how modulating CSF pressure can aid in tumor resection by promoting the descent of*

*continued on next page*

*continued from previous page*

## **Supplemental Online Material continued**

---

*suprasellar tumor components into the surgical field. This is particularly useful when dealing with tumors that compress the optic chiasm.*

*Additionally, your knowledge of CSF leak management would be strengthened by recognizing the diagnostic and therapeutic roles of the lumbar drain in this context. Enhancing your understanding of these aspects will improve your comprehensive perioperative management skills.*

**LLM (question):** *What specific strategies can you use to minimize the risk of CSF leak?*

**User (answer):** *I can maintain hemodynamic control along with normocarbia to decrease the risk of intracranial pressure leading to csf leak, also during extubation is limiting positive pressure ventilation or increased nasal pressure.*

**LLM (feedback):** *You effectively managed the concept of maintaining hemodynamic control, which is crucial in minimizing the risk of CSF leak. Your understanding of limiting positive pressure ventilation during extubation is also well noted.*

*However, you could strengthen your approach by considering the importance of patient positioning, such as a slight head-up tilt, to reduce CSF pressure. When you discussed ventilation, maintaining a slightly elevated PaCO<sub>2</sub> (40-45 mmHg) is beneficial, as opposed to normocarbia, to avoid brain relaxation issues. Additionally, you might consider strategies to prevent coughing or straining during emergence, such as using low-dose remifentanil and dexmedetomidine infusions, or deep extubation in suitable patients. Your knowledge of blood pressure management would be strengthened by recognizing its role in preventing disruption of surgical closure.*

*Overall, integrating these strategies will enhance your management of CSF leak risk.*

**LLM (question):** *The surgeon requests SBP < 100 to minimize bleeding and improve surgical visualization with an endoscopic approach. How would you achieve this safely in this patient?*

**User (answer):** *I will discuss the concerns of limiting intracranial perfusion with low blood pressures but would be okay with temporary SBP decrease to help with visualization, I will utilize the remifentanil drip and if needed vasodilators and vasopressors to maintain an SBP less than 100 while continuing to monitor the patients EKG for any ST changes for cardiac strain.*

**LLM (feedback):** *You effectively managed the initial discussion with the surgeon about the risks of low blood pressure, showing good understanding of the need for careful communication. Your plan to use remifentanil for controlled hypotension is appropriate, given its rapid titratability.*

*continued on next page*

*continued from previous page*

## Supplemental Online Material continued

*However, you could strengthen your approach by considering the patient's chronic hypertension, which shifts the cerebral autoregulation curve, increasing the risk of cerebral hypoperfusion. When you discussed using vasodilators, consider clevidipine for its ultra-short half-life and minimal intracranial pressure effects.*

*You could enhance your management by incorporating non-pharmacological strategies, such as positioning and minimizing PEEP, to optimize surgical conditions without systemic hypotension. Additionally, monitoring with processed EEG could help detect cerebral hypoperfusion, although its sensitivity is limited.*

*Your knowledge of contraindications to controlled hypotension would be strengthened by considering factors like elevated ICP or significant carotid stenosis."*

**LLM (question):** *The surgeon reports difficulty accessing the suprasellar component of the tumor. What interventions might improve the surgical field?*

**User (answer):** *The lumbar drain can be utilized to help exposure along with managing any hypertension which can lead to intracranial hypertension, along with raising the head of the bed and decreasing peep*

**LLM (feedback):** *You effectively managed the concept of using a lumbar drain to improve surgical exposure, demonstrating a good understanding of its potential role. Your awareness of managing hypertension to prevent intracranial hypertension is also commendable.*

*However, you could strengthen your approach by considering the importance of patient positioning. Increasing the reverse Trendelenburg position to 15-20 degrees can significantly aid in surgical exposure, provided the patient's hemodynamics allow for it. Additionally, while you mentioned decreasing PEEP, increasing PEEP up to 10 mmHg can actually help promote tumor descent.*

*Your knowledge of the risks associated with lumbar drain use would be strengthened by understanding the potential complications, such as infection or inadvertent injection errors. Maintaining open communication with the surgical team about the effectiveness of these interventions is crucial for patient safety and optimal surgical outcomes.*

**LLM (question):** *When the resection is complete, how will you assess for and manage a possible CSF leak?*

**User (answer):** *I will discuss with the surgeon if there is any concern for increased CSF leak but it is usually a suspected complications although minimal after the surgery. I will make*

*continued on next page*

*continued from previous page*

## Supplemental Online Material continued

*sure to have the patients head of the bed elevated at least 30 degrees along with preventing any coughing or gagging during extubation as long as the patient meets extubation criteria and limit any positive pressure to the nasal and oral cavity to decrease the risk of increasing intracranial pressure along with maintaining hemodynamic stability.*

**LLM (feedback):** *You effectively managed the patient's positioning and extubation considerations, demonstrating a clear understanding of minimizing intracranial pressure and hemodynamic stability. Your discussion of preventing coughing and gagging during extubation showed good understanding of patient safety.*

*However, you could strengthen your approach by incorporating the intraoperative assessment for a CSF leak. Performing a graded Valsalva maneuver with specific airway pressures, as outlined in the model answer, is crucial for identifying leaks. Additionally, understanding the role of the anesthesia team in protecting surgical repairs and the use of lumbar drains is essential.*

*Your knowledge of postoperative management would be strengthened by considering the use of medications like remifentanyl and dexmedetomidine for smooth emergence, and the importance of avoiding nasal trumpets or non-invasive positive pressure ventilation post-surgery. This comprehensive approach ensures optimal patient outcomes and safety.*

**LLM (question):** *About 20 minutes into recovery, you are called to the bedside because the patient remains very sleepy with an oxygen saturation of 85%. The nurse asks if you can order BiPAP or high-flow nasal cannula oxygen to improve her oxygenation. How do you proceed?*

**User (answer):** *I would advise the nurse not to start any positive pressure of bipap or high flow nasal cannula due to the opening of the nasal cavity and risk of pneumocephalus. I will go to the patient bedside and place the patient on 100% fio2 with a nonrebreather and auscultate the lungs to listen for wheezing for bronchospasm or decreased breath sounds secondary to atelectasis or concern for pneumothorax. I would also assess if the patient was awake and following commands and perform a neuro exam with concern for stroke or intracranial hemorrhage or herniation. I will also assess the patient's medications if there is any hypoventilation due to increase in opioid use and if there is need for narcan.*

**LLM (feedback):** *You effectively managed the situation by recognizing the risks associated with positive pressure ventilation post-transsphenoidal surgery, showing a good understanding of the potential for pneumocephalus. Your decision to use a non-rebreather mask and assess for reversible causes like opioid oversedation was appropriate.*

*However, you could strengthen your approach by considering the immediate positioning of the patient upright to aid ventilation and reduce the risk of aspiration. When you discussed*

*continued on next page*

*continued from previous page*

## **Supplemental Online Material continued**

---

*assessing the patient, including a more detailed plan for monitoring and possibly administering naloxone or flumazenil could enhance your management. Additionally, early communication with the surgical team is crucial, as they may need to intervene if intracranial complications are suspected. Your knowledge of the importance of imaging in this context would be strengthened by considering the need for urgent CT if conservative measures fail.*

**LLM (question):** *Four hours postoperatively, the patient's urine output increases to 300mL/hr for 2 consecutive hours. How will you evaluate the possible causes?*

**User (answer):** *I am concerned about diabetes insipidus and will assess the patient's vitals to make sure there is not a need for vasopressors along with sending for urinary creatinine and sodium along with a serum BMP to assess their serum sodium and if there is diabetes insipidus to replace fluid losses and increase the patient's blood pressure.*

**LLM (feedback):** *You effectively managed to identify diabetes insipidus as a potential cause of postoperative polyuria, showing good understanding of its relevance in the context of pituitary surgery. Your plan to assess vital signs and consider fluid replacement demonstrates awareness of patient safety and hemodynamic stability.*

*However, you could strengthen your approach by broadening your differential diagnosis to include cerebral salt wasting and osmotic diuresis. When you discussed laboratory evaluations, consider including urine specific gravity, urine osmolality, and serum osmolality to differentiate between these conditions.*

*Additionally, reviewing the patient's fluid balance and operative details can provide valuable context for your assessment. Your knowledge of the nuances in diagnosing postoperative polyuria would be strengthened by incorporating these additional diagnostic steps.*

**LLM (question):** *Her serum sodium has increased from 138 to 149 mEq/L, and she appears hypovolemic. How will you treat her?*

**User (answer):** *I will give vasopressors and replace any fluid loss with D5 1/2 normal saline and titrate vasopressin to aid with the increase in fluid secondary to diabetes insipidus*

**LLM (feedback):** *You effectively recognized the need to address both the fluid loss and the underlying condition, indicating an understanding of the patient's hypovolemic state. However, you could strengthen your approach to fluid management by initially prioritizing isotonic crystalloid solutions for volume resuscitation, as hypotension from hypovolemia is a more immediate threat than hypernatremia. When you discussed using vasopressors, consider that they are not typically first-line in this context unless there is refractory*

*continued on next page*

*continued from previous page*

## Supplemental Online Material continued

*hypotension. Your knowledge of diabetes insipidus management would be strengthened by incorporating the use of DDAVP as both a diagnostic and therapeutic measure. Additionally, monitoring urine output, specific gravity, and serum sodium closely is crucial to guide ongoing treatment and avoid complications such as cerebral edema. Your approach would benefit from calculating the free water deficit to guide the correction of hypernatremia safely.*

**LLM (question):** *On postoperative day 2, the patient develops severe headache and nausea. What complications must be considered?*

**User (answer):** *I am concerned about a csf leak causing tension of the dura leading to headaches along with increased risk of herniation or stroke. I would perform a neuro exam and look intranasally for any csf and can test with a glucose strip to confirm if there is csf or not as csf will test positive for glucose. I am also concerned fo pneumocephalus or stroke and would contact neursorgery but also get a CT head. I will check the vitals to monitor the EKG for any arrhythmias and if the patient is having any chest pain or shortness of breath and provide oxygen if there is any hypoxia.*

**LLM (feedback):** *You effectively managed the initial assessment by considering a CSF leak and the associated risks, demonstrating a good understanding of potential complications. Your plan to perform a neuro exam and check for CSF intranasally was well thought out. Additionally, your consideration of pneumocephalus and the decision to obtain a CT head showed good clinical reasoning.*

*However, you could strengthen your approach by considering the specific characteristics of headaches associated with different complications, such as postural changes in CSF leaks or supine worsening in pneumocephalus. When you discussed monitoring vitals, consider focusing on orthostatic blood pressures to assess for adrenal insufficiency. You might also enhance your management by considering the possibility of meningitis and the need for CSF studies if symptoms suggest infection. Your knowledge of these complications would be strengthened by incorporating these additional diagnostic considerations.*

Potentially misleading feedback (underlined)

---

1)

**LLM (question):** *How does the history of bilateral inguinal hernia repairs with mesh affect your preoperative planning?*

*continued on next page*

*continued from previous page*

## Supplemental Online Material continued

**User (answer):** *Prior surgery with mesh, may imply likely adhesions/scar tissue hence technical difficulties for peritoneal cavity entry. Also would mean longer case duration hence more stress of surgery to patient, increased bleeding risk, and increased risk of ACS. Hence may aim for two peripheral IV access. Ensuring adequate paralysis during case.*

**LLM (feedback):** *You effectively managed to identify the potential for adhesions and scar tissue due to prior surgery with mesh, which can indeed lead to technical difficulties during peritoneal cavity entry. You also correctly anticipated a longer case duration and the associated increased stress on the patient, as well as the heightened risk of bleeding and acute compartment syndrome (ACS). Your plan to aim for two peripheral IV accesses and ensuring adequate paralysis during the case shows good foresight in managing these risks.*

*You could strengthen your approach to preoperative planning by considering the need for large-bore IV access and possibly an arterial line for continuous blood pressure monitoring and intraoperative lab draws, given the increased risk of bleeding. When you discussed the implications of previous surgery, you might also consider including the need for a type and screen and having blood products readily available.*

*Your knowledge of anesthetic management would be strengthened by considering the potential need for adjustments in the timing of antibiotic prophylaxis, pain medication administration, and titration of infusions with context-sensitive half-lives. Additionally, emphasizing the importance of padding all pressure points to prevent nerve injury during the prolonged surgery is crucial. Lastly, it's important to prepare for the possibility of conversion to an open procedure if the robotic approach is impeded by scarring.*

2)

**LLM (question):** *You decide to perform a spontaneously breathing asleep video laryngoscopy. How will you induce this patient?*

**User (answer):** *preoxygenate with 100% fio2, keep patients head of bed elevated, perform IV induction with ketamine to maintain spontaneous respirations with additional inhalation of sevoflurane as needed for sedation, have difficult airway cart in place and have ENT on standby for potential emergent need for invasive airway*

*continued on next page*

*continued from previous page*

## Supplemental Online Material continued

**LLM (feedback):** *You effectively managed to outline a strategy for inducing anesthesia while maintaining spontaneous respiration, which is crucial in a patient with a potentially difficult airway. You clearly understood the importance of preoxygenation and keeping the patient's head elevated to optimize the airway. Your choice of ketamine to maintain spontaneous respirations shows good understanding of its pharmacological profile. Having a difficult airway cart ready and ENT on standby demonstrates excellent preparedness for potential complications.*

*You could strengthen your approach to induction by considering the use of a high-flow nasal cannula to extend safe apnea time, as mentioned in the model answer. When you discussed using sevoflurane for sedation, you might consider the potential for upper airway irritation and coughing, which could be counterproductive in a patient where you're trying to maintain spontaneous breathing. Your knowledge of induction agents would be strengthened by considering the use of low-dose remifentanyl or dexmedetomidine as alternatives, which can provide sedation with minimal respiratory depression. Additionally, premedication with glycopyrrolate to manage ketamine-induced secretions was an important point that was missed. Remember, the goal is to achieve adequate sedation without compromising respiratory effort, so careful titration of medications is key.*

3)

**LLM (question):** *The surgeon reports significant venous bleeding obscuring the surgical field. How do you respond?*

**User (answer):** *I would decrease IV propofol and remifentanyl to minimum doses to maintain sedation remifentanyl 0.1 mcg/kg/hr and propofol 200mg/kg/hr as well as titrating epinephrine to aid with vasoconstriction after reviewing the monitors if patient is hypotensive.*

**LLM (response):** *You effectively managed the concept of reducing mean arterial pressure (MAP) to help with surgical exposure, which is a key point in managing venous bleeding. You also showed good understanding of using vasoactive agents like epinephrine to aid with vasoconstriction.*

*However, you could strengthen your approach to hemodynamic management by considering the use of short-acting agents like esmolol or nicardipine, which offer rapid titratability and are preferred for maintaining MAP at the lower end of normal.*

*continued on next page*

*continued from previous page*

## Supplemental Online Material continued

*When you discussed reducing the doses of propofol and remifentanyl, you might consider the importance of maintaining adequate anesthesia depth while managing hemodynamics, as under-dosing could lead to patient awareness.*

*You could enhance your management by considering the patient's position, such as placing them in slight reverse Trendelenburg to reduce venous pressure, and avoiding excessive PEEP which can increase venous pressure. Additionally, your knowledge of intraoperative monitoring would be strengthened by discussing the potential need to temporarily cease nerve monitoring to allow for more aggressive hemostasis if needed. Remember to have a low threshold for monitoring hemoglobin levels, especially if the blood loss is significant.*

4)

**LLM (question):** *The patient is extremely anxious in the preoperative holding area and requests medication to help her relax. What can you offer her?*

**User (answer):** *She could be offered versed, however I would want to avoid if possible given effect on mental status at end of surgery. Precedex is another option or small dose of fentanyl for sedation. otherwise I would provide verbal reassurance*

**LLM (feedback):** *You effectively managed the patient's anxiety by considering both pharmacological and non-pharmacological options. Your suggestion to offer verbal reassurance showed good understanding of non-pharmacological interventions. However, when you discussed medication options, you might consider the specific effects and duration of action of each drug. Midazolam (Versed) is commonly used for its anxiolytic properties, but your concern about postoperative mental status is valid. Precedex can be useful for its sedative and anxiolytic effects without significant respiratory depression, but it may not be ideal for short procedures due to its longer duration. Fentanyl is not typically used for preoperative anxiety due to its primary role as an analgesic. You could enhance your management by considering the patient's specific needs and the surgical context when selecting a sedative. Your knowledge of the pharmacokinetics and dynamics of these medications would be strengthened by further review.*

5)

**LLM (question):** *How will you induce and maintain anesthesia in this case to optimize IONM signals?*

*continued on next page*

*continued from previous page*

## Supplemental Online Material continued

---

**User (answer):** *I would induce with propofol, Ketamine, and remifentanyl. Alternatively, I could use succinylcholine in place of remifentanyl. My goal would be to obtain optimal intubating conditions while avoiding prolonged paralysis*

**LLM (feedback):** *You effectively managed the induction phase by selecting propofol and ketamine, which can enhance MEP and SSEP signals. Your understanding of avoiding prolonged paralysis is crucial for optimizing IONM signals.*

*However, you could strengthen your approach by reconsidering the use of succinylcholine in place of remifentanyl. While succinylcholine provides rapid intubation, it may not be ideal for maintaining conditions conducive to IONM. Instead, consider using rocuronium with sugammadex for reversal if needed.*

*Your knowledge of anesthetic maintenance would be strengthened by focusing on TIVA with propofol and remifentanyl, as volatile agents can suppress MEPs. Additionally, incorporating ketamine can mitigate opioid-induced hyperalgesia.*

*You could enhance your management by ensuring regular communication with the neurophysiologist to adjust anesthetic depth and optimize signal quality. This collaboration is vital for patient safety and effective monitoring.*