

RAG-based Medical QA

Oct 21, 2024

Outline

- Ways to Evaluate Machine Generated Text
- A Small GraphRAG Experiment
- Results
- Next Steps

Evaluating Generated Texts

Question: **Where is Gross Hall?**

Ground Truth:

It is located on Science Drive.

Machine Generated Answer:

It is a research building at Duke University.

- **BLEU Score** (precision based)
 - Precision: Num of matched words/ Num of generated words: $2/8 = \underline{.25}$
- **ROUGE Score** (F1-score-based)
 - Recall: Num of matched words/ Num of reference words: $2/6$.
 - F1 Score: $2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall}) = 1/7 = \underline{.14}$

A Small Experiment

Using the following two questions (from Wiki Journal), we used different number of relevant documents (1, 3, 5) to build the GraphRAG model and test the results.

Question	Answer	Source
In patients with moderate to severe ARDS , does the early use of continuous neuromuscular blockade with cisatracurium improve mortality when used with current light sedation protocols ?	Not only does early neuromuscular blockade not reduce mortality when using modern light sedation protocols , it was associated with an increase in ICU-acquired weakness and serious adverse cardiovascular events .	ROSE
Among patients with septic shock and relative adrenal insufficiency , do corticosteroids reduce 28-day mortality ?	Among patients with septic shock and relative adrenal insufficiency, administration of corticosteroids reduces 28-day mortality , although this finding was not confirmed in the follow-up CORTICUS, HYPRESS, or ADRENAL trials, but confirmed in APROCCHSS.	Annane Trial

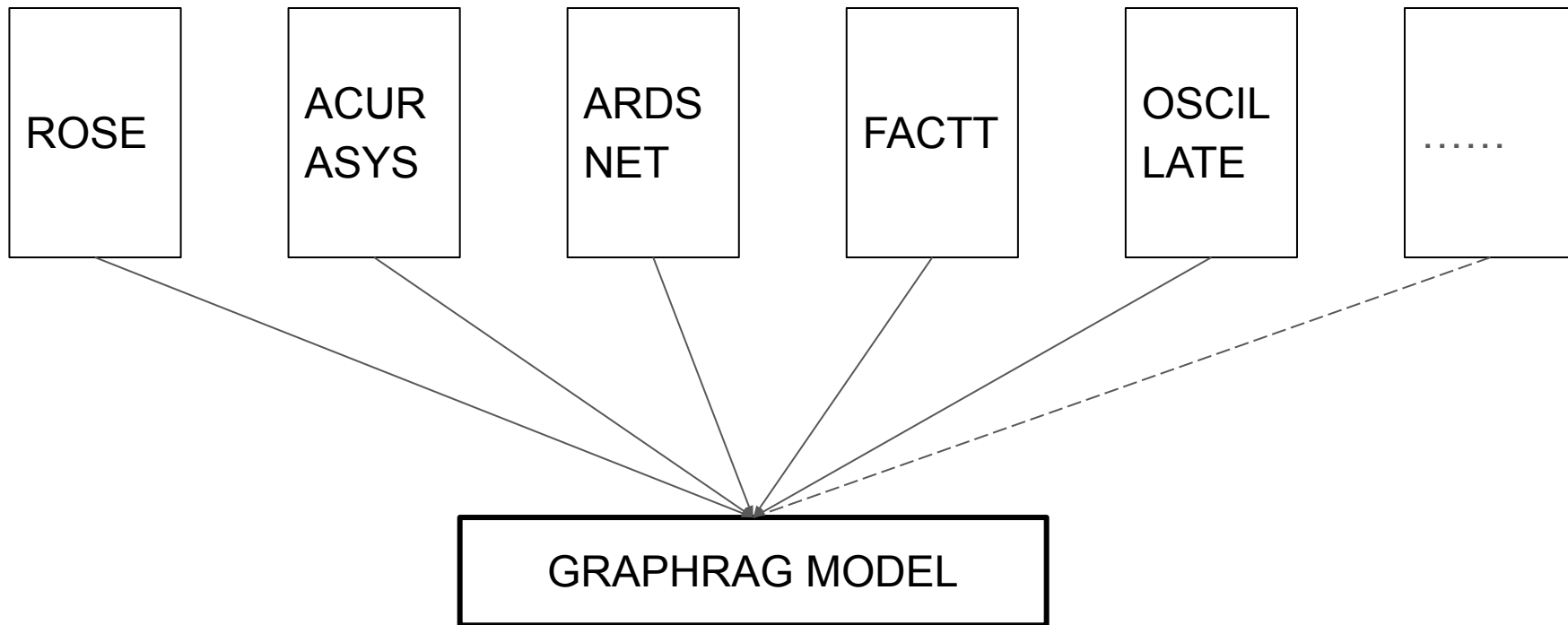
Model performance evaluation

- What results does the model generate?
- How does the result change when the relevant docs increase?
- How does the result change when there exist confusing docs?

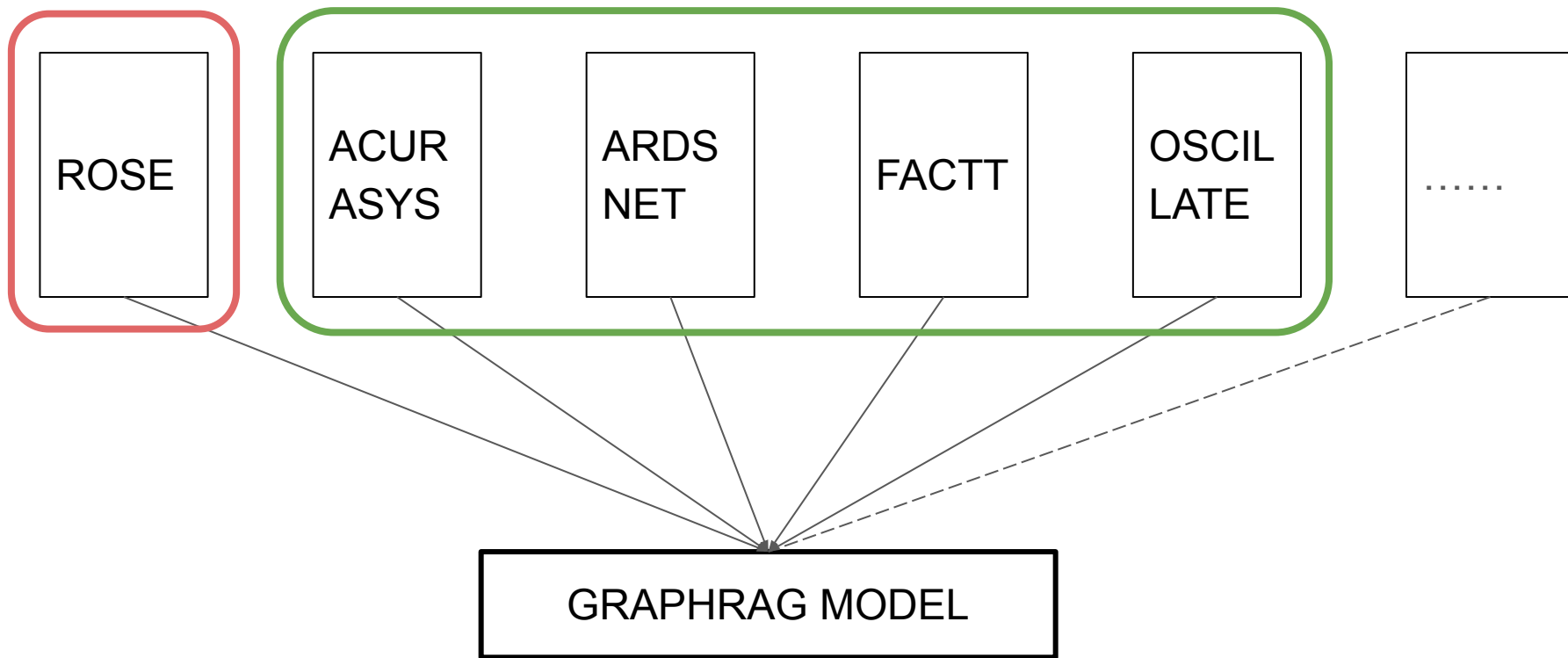
Metrics for assessing model performance:

BLEU, ROUGE , GPT, Professional Rating

Example



Example



Example result from GraphRAG (1 doc local)

```
python -m graphrag.query \  
--root ./ragtest \  
--method local \  

```

"In patients with moderate to severe ARDS, does the early use of continuous neuromuscular blockade with cisatracurium improve mortality when used with current light sedation protocols?"

The question of whether the early use of continuous neuromuscular blockade with cisatracurium improves mortality in patients with moderate to severe Acute Respiratory Distress Syndrome (ARDS) when used alongside current light sedation protocols is a nuanced one. The data provided does not directly address the specific impact of cisatracurium or any neuromuscular blockade agents on mortality rates in ARDS patients under light sedation protocols. However, insights can be drawn from the broader context of ARDS management and treatment strategies as discussed in the provided reports and studies.

ARDS is a critical condition that requires intensive care, often involving mechanical ventilation to support breathing. The management of ARDS has evolved over time, with strategies such as prone positioning and the use of sedatives and neuromuscular blocking agents being part of the therapeutic arsenal aimed at improving oxygenation and patient outcomes [Data: Reports (2, 0); Entities (18)].

Prone positioning, a method where patients are positioned face down, has been shown to improve oxygenation and potentially reduce mortality in severe ARDS cases. This strategy is highlighted in the PROSEVA trial, which demonstrated significant benefits in terms of reduced 28-day and 90-day mortality rates for patients with severe ARDS who were managed with prone positioning [Data: Entities (3); Sources (0)]. The trial's protocol included the use of neuromuscular blocking agents to facilitate prone positioning, suggesting that these agents play a role in the management of severe ARDS, particularly in conjunction with prone positioning [Data: Sources (1)].

The use of neuromuscular blocking agents, such as cisatracurium, is often considered in ARDS management to improve ventilator synchrony, reduce oxygen consumption, and potentially mitigate ventilator-induced lung injury. However, the specific impact of these agents on mortality, especially when used early and in combination with light sedation protocols, is not explicitly detailed in the provided data. It's worth noting that the approach to sedation has shifted towards lighter sedation protocols in critical care, aiming to reduce sedation-related complications and potentially improve outcomes by facilitating earlier mobilization and reducing the duration of mechanical ventilation [Data: Reports (2)].

Given the complexity of ARDS management and the multifactorial nature of interventions aimed at improving patient outcomes, the direct impact of continuous neuromuscular blockade with cisatracurium on mortality rates within the context of light sedation protocols requires further investigation. Clinical guidelines and treatment strategies continue to evolve based on emerging evidence from randomized controlled trials and meta-analyses. Therefore, while the early use of neuromuscular blocking agents is a component of ARDS management, its efficacy and impact on mortality as part of a comprehensive treatment strategy, including light sedation protocols, should be interpreted within the broader context of individual patient needs and evolving clinical evidence [Data: Entities (18, 1); Reports (0, 2)].

Results- Question 1

Q1: In patients with **moderate to severe ARDS**, does the **early use of continuous neuromuscular blockade with cisatracurium improve mortality** when used with **current light sedation protocols**?

Document	Type	Human Rating	GPT	BLEU	ROUGE-L
1 (ROSE)	Global	0	0	0	0
3 (ROSE, ACURASYS, ARDSNET)		0.1	<u>0.1</u>	0.01	0.06
5 (ROSE, ACURASYS, ARDSNET, FACTT, OSCILLATE)		<u>0.1</u>	0.1	0.01	0.05
1 (ROSE)	Local	<u>0.2</u>	<u>0.5</u>	0.01	0.05
3 (ROSE, ACURASYS, ARDSNET)		0.1	0.2	0.00	0.07
5 (ROSE, ACURASYS, ARDSNET, FACTT, OSCILLATE)		0.1	0.1	0.01	0.06

GPT Rating (1 doc global)

Reference summary = “Not only does early neuromuscular blockade not reduce mortality when using modern light sedation protocols, it was associated with an increase in ICU-acquired weakness and serious adverse cardiovascular events.”

Generated summary = “I am sorry but I am unable to answer this question given the provided data.”

Based on the information above, **on a scale from 0 to 1**, how well do you think the generated summary compares to the reference summary? Keep your answer concise.

The generated summary would **score 0** on a scale from 0 to 1. It fails to capture any content from the reference, providing no relevant information, making it entirely unaligned with the intended summary.

GPT Rating (1 doc local)

Reference summary = “Not only does early neuromuscular blockade not reduce mortality when using modern light sedation protocols, it was associated with an increase in ICU-acquired weakness and serious adverse cardiovascular events.”

Generated summary = “Given the complexity of ARDS management and the multifactorial nature of interventions aimed at improving patient outcomes, the direct impact of continuous neuromuscular blockade with cisatracurium on mortality rates within the context of light sedation protocols requires further investigation. Clinical guidelines and treatment strategies continue to evolve based on emerging evidence from randomized controlled trials and meta-analyses. Therefore, while the early use of neuromuscular blocking agents is a component of ARDS management, its efficacy and impact on mortality as part of a comprehensive treatment strategy, including light sedation protocols, should be interpreted within the broader context of individual patient needs and evolving clinical evidence”

Based on the information above, **on a scale from 0 to 1**, how well do you think the generated summary compares to the reference summary? Keep your answer concise.

On a scale from 0 to 1, I would rate the generated summary as **0.5**. It does not directly contradict the reference summary but rather takes a more cautious, nuanced approach. While it does not explicitly state the negative outcomes (increased ICU-acquired weakness and adverse cardiovascular events) mentioned in the reference, it acknowledges the need for further investigation and evolving clinical evidence. However, it fails to accurately reflect the key findings of the reference, especially the lack of mortality benefit and associated risks, making the comparison incomplete.

GPT Rating (2 docs global)

Reference summary = “Not only does early neuromuscular blockade not reduce mortality when using modern light sedation protocols, it was associated with an increase in ICU-acquired weakness and serious adverse cardiovascular events.”

Generated summary = “The early use of continuous neuromuscular blockade with The evidence strongly supports the early use of continuous neuromuscular blockade with cisatracurium in patients with moderate to severe ARDS, particularly when used alongside current light sedation protocols. The benefits include a significant reduction in mortality, an increase in ventilator-free days, and no significant increase in the risk of ICU-acquired paresis. These findings suggest that cisatracurium is a valuable addition to the treatment regimen for ARDS, potentially improving outcomes for a challenging patient population.”

Based on the information above, **on a scale from 0 to 1**, how well do you think the generated summary compares to the reference summary? Keep your answer concise.

On a scale from 0 to 1, I would rate the generated summary as **0**. It directly contradicts the reference summary, which states that early neuromuscular blockade does not reduce mortality and is linked to increased ICU-acquired weakness and adverse cardiovascular events. The generated summary inaccurately claims benefits like reduced mortality and no increase in ICU-acquired paresis, making it entirely misaligned with the reference information.

GPT Rating (2 docs local)

Reference summary = “Not only does early neuromuscular blockade not reduce mortality when using modern light sedation protocols, it was associated with an increase in ICU-acquired weakness and serious adverse cardiovascular events.”

Generated summary = “The early use of continuous neuromuscular blockade with cisatracurium in patients with moderate to severe ARDS, when used in conjunction with current light sedation protocols, appears to improve mortality rates without increasing the risk of ICU-acquired paresis. These findings underscore the potential benefits of incorporating cisatracurium into the treatment regimen for ARDS in appropriately selected patients [Data: Reports (3)]. However, it's crucial to tailor treatment strategies to individual patient needs and the evolving clinical context, considering the balance between benefits and potential risks.”

Based on the information above, **on a scale from 0 to 1**, how well do you think the generated summary compares to the reference summary? Keep your answer concise.

On a scale from 0 to 1, I would rate the generated summary as **0.2**. While it acknowledges the importance of balancing benefits and risks, it still contradicts the reference summary by stating that early neuromuscular blockade improves mortality rates and does not increase the risk of ICU-acquired paresis. This is the opposite of what the reference summary conveys, which mentions no mortality benefit and increased ICU-acquired weakness.

Interesting outcome based on Q1

Ground truth: Not only does early neuromuscular blockade **NOT** reduce mortality when using modern light sedation protocols, it was associated with an **INCREASE** in ICU-acquired weakness and **serious ADVERSE** cardiovascular events.

Model result (1 doc global): I am sorry but I am **unable to answer** this question given the provided data.

Model result (1 doc local): ... the direct impact of continuous neuromuscular blockade with cisatracurium on mortality rates within the context of light sedation protocols **requires further investigation**. ... its efficacy and impact on mortality ... should be **interpreted within the broader context** of individual patient needs and evolving clinical evidence

Interesting outcome based on Q1

Ground truth: Not only does early neuromuscular blockade **NOT reduce mortality** when using modern light sedation protocols, it was associated with an **INCREASE in ICU-acquired weakness** and **serious ADVERSE** cardiovascular events.

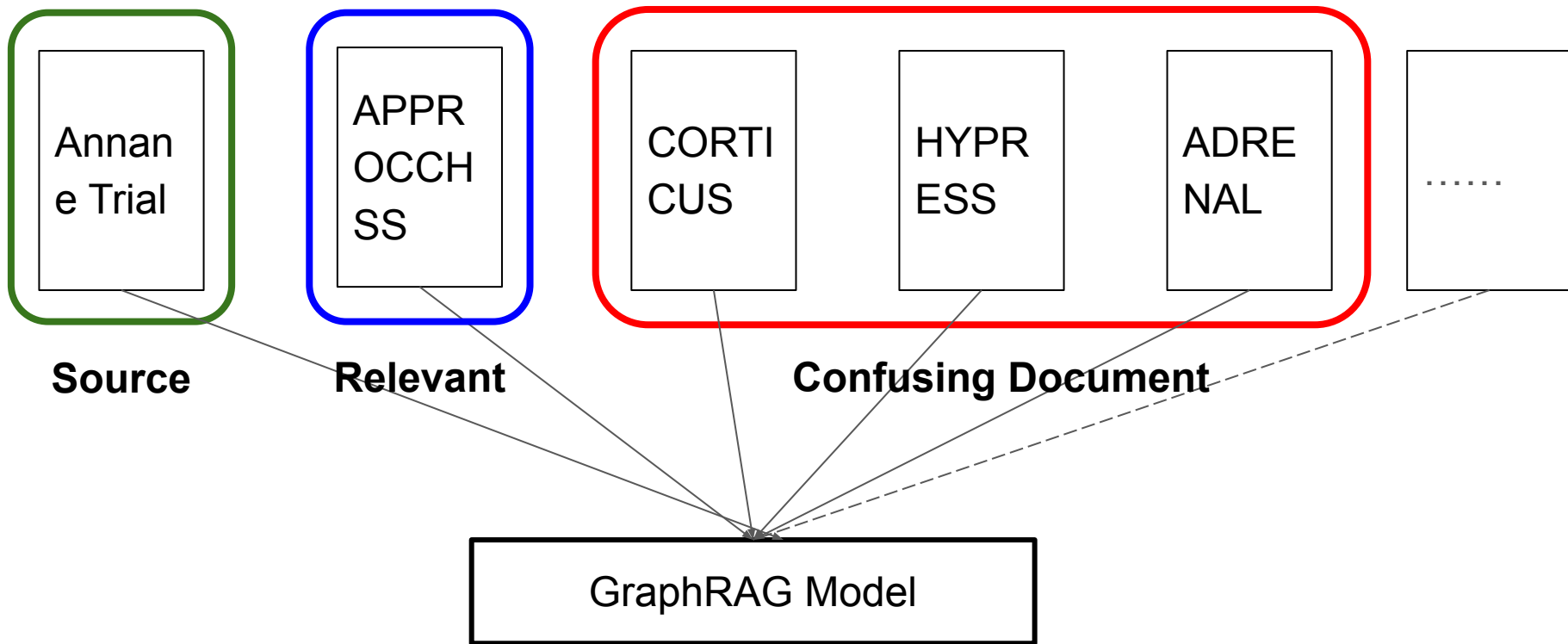
Model result (3 docs global): In conclusion, the early use of continuous neuromuscular blockade with cisatracurium, when used in conjunction with current light sedation protocols, appears to **improve mortality** in patients with moderate to severe ARDS. The **findings support the safety of cisatracurium**, showing **no increased risk of muscle weakness**, and highlight the importance of adherence to established sedation protocols and monitoring guidelines to maximize patient outcomes.

Q2 Experiment

Question: Among patients with septic shock and relative adrenal insufficiency, do corticosteroids reduce 28-day mortality?

Ground Truth: Among patients with septic shock and relative adrenal insufficiency, administration of corticosteroids reduces 28-day mortality, although this finding was not confirmed in the follow-up CORTICUS, HYPRESS, or ADRENAL trials, but confirmed in APROCCHSS.

Example



Example result from GraphRAG (1 doc local)

```
python -m graphrag.query \  
--root ./ragtest \  
--method local \  

```

"Among patients with septic shock and relative adrenal insufficiency, do corticosteroids reduce 28-day mortality?"

Overview

The clinical study conducted in France on the efficacy of corticosteroids in treating septic shock, particularly in patients with relative adrenal insufficiency, provides significant insights into the role of corticosteroids in reducing 28-day mortality rates. This placebo-controlled, randomized, double-blind trial focused on adult patients who met the criteria for septic shock and were subjected to a short corticotropin test to determine their eligibility based on their adrenal response [Data: Sources (0)].

Efficacy of Corticosteroids

The study's findings indicate that corticosteroids, specifically hydrocortisone and fludrocortisone, significantly reduced the risk of death at 28 days in patients with septic shock and relative adrenal insufficiency. Among the nonresponders to the corticotropin test (those with relative adrenal insufficiency), there were fewer deaths in the corticosteroid group compared to the placebo group. Specifically, 60 deaths (53%) occurred in the corticosteroid group, as opposed to 73 deaths (63%) in the placebo group. This difference translates to a hazard ratio of 0.67, indicating a significant reduction in mortality risk for patients treated with corticosteroids [Data: Sources (0, 3)].

Mortality Rates and Vasopressor Therapy

The study also assessed the impact of corticosteroid treatment on the withdrawal of vasopressor therapy within 28 days. In the corticosteroid group, a higher percentage of patients had vasopressor therapy withdrawn compared to the placebo group, highlighting an additional benefit of corticosteroid treatment in managing septic shock [Data: Sources (4)].

Adverse Events

Importantly, the rates of adverse events were similar between the corticosteroid and placebo groups. This finding suggests that the administration of low doses of hydrocortisone and fludrocortisone for 7 days does not increase the risk of adverse events, thereby supporting the safety profile of corticosteroid treatment in this patient population [Data: Sources (4)].

Conclusion

In conclusion, the clinical study provides compelling evidence that a 7-day treatment with low doses of hydrocortisone and fludrocortisone significantly reduces the 28-day mortality rate among patients with septic shock and relative adrenal insufficiency, without increasing the incidence of adverse events. These results underscore the potential life-saving benefits of corticosteroid treatment in a critically ill patient population and highlight the importance of identifying patients with relative adrenal insufficiency for targeted therapy [Data: Sources (0, 3, 4)].

Overview

Septic shock, a severe and life-threatening condition characterized by circulatory, cellular, and metabolic abnormalities due to an infection, often leads to high mortality rates among critically ill patients. A significant aspect of its management involves addressing relative adrenal insufficiency, a condition where the adrenal glands do not produce sufficient amounts of hormones, notably cortisol, which is crucial for the body's response to stress and illness. This has led to the exploration of corticosteroids as a potential treatment to improve outcomes in septic shock patients, particularly those with relative adrenal insufficiency.

Corticosteroids in Septic Shock Treatment

Corticosteroids, including hydrocortisone and fludrocortisone, have been studied for their role in treating septic shock, aiming to mitigate the effects of relative adrenal insufficiency. The rationale behind using corticosteroids is their potential to modulate the immune response, reduce inflammation, and improve hemodynamic stability, which are critical factors in managing septic shock.

Clinical Evidence

The impact of corticosteroids on 28-day mortality among septic shock patients with relative adrenal insufficiency has been the subject of clinical research. One pivotal study, as detailed in the provided data, assessed the effectiveness of low doses of hydrocortisone and fludrocortisone on 28-day survival in patients with septic shock and relative adrenal insufficiency [Data: Sources (0)]. This placebo-controlled, randomized, double-blind, parallel-group trial found that a 7-day treatment with low doses of hydrocortisone and fludrocortisone significantly reduced the risk of death in these patients without increasing adverse events [Data: Sources (0)]. Specifically, in nonresponders to the corticotropin test (indicative of relative adrenal insufficiency), there were fewer deaths in the corticosteroid group compared to the placebo group, suggesting a beneficial effect of corticosteroids on survival [Data: Sources (0)].

Conclusion

The evidence from clinical trials indicates that corticosteroids can reduce 28-day mortality in patients with septic shock and relative adrenal insufficiency. This finding supports the use of corticosteroids as part of the treatment regimen for this specific subgroup of septic shock patients, highlighting the importance of identifying relative adrenal insufficiency to tailor therapy appropriately. However, it is crucial to consider the potential risks and benefits of corticosteroid therapy, including the possibility of adverse events and the need for careful patient monitoring [Data: Sources (0)]. Further research may provide additional insights into optimizing corticosteroid therapy for septic shock, including the determination of the most effective doses, treatment duration, and patient selection criteria.

Results- Question 2

Q2: Among patients with **septic shock** and **relative adrenal insufficiency**, do **corticosteroids** reduce **28-day mortality**?

Document	Type	Human Rating	GPT	BLEU	ROUGE-L
1 (Annane Trial)	Global	0.6	0.60	0.02	0.09
3 (Annane Trial, ADRENAL, APROCCHSS)		0.7	0.85	0.02	0.10
5 (Annane Trial, ADRENAL, APROCCHSS, CORTICUS, HYPRESS)		0.85	0.85	0.04	0.12
1 (Annane Trial)	Local	0.7	0.80	0.03	0.09
3 (Annane Trial, ADRENAL, APROCCHSS)		0.8	0.90	0.03	0.08
5 (Annane Trial, ADRENAL, APROCCHSS, CORTICUS, HYPRESS)		0.85	0.90	0.03	0.09

Challenges

1. **What is the ground truth?** Metrics designed for model evaluation mostly compares the generated answer to the reference answer provided by experts.
2. Graphrag model returns results based on majority and relevancy of documents it has; **more dimensions should be considered when building the graph**, for example chronological recentancy, citation factor, clinical trial results etc.
3. **Evaluation** metrics and method. BLEU and ROUGE are still questionable. Human rating is not scalable for increasing amount of questions and answers.

Next Steps

- Other ways to evaluate answers:
 - RAGAS
 - BERT Score
- Adjust the Graph model to enhance answers based on the current 10 documents.
- Graph storage (entities, relationships, weights etc) and model explainability (local and global, SHAP, LIME and other feature attribution methods)

Backup