

# 메타 리뷰

## 요약

본 연구는 대규모 MIMIC-IV 데이터베이스를 활용하여 반코마이신-피페라실린/타조박탐(VPT) 병용 요법과 급성 신장 손상(AKI) 간의 인과 관계를 분석한 중요한 연구입니다. 여러 전문 검토자들은 이 연구가 **LLM을 활용한 교란 변수 식별이라는 혁신적인 방법론**을 제시하고, **전례 없는 규모의 데이터셋을 활용하여 강력한 실세계 증거(RWE)를 생성**했다는 점을 높이 평가했습니다. 특히, 제안된 LLM 프레임워크의 **상세한 임상 정의, 포괄적인 다단계 검증, 그리고 엄격한 오류 처리 메커니즘**은 연구의 투명성과 재현성을 크게 향상시키는 핵심 강점으로 꼽혔습니다. 그러나 검토자들은 여러 영역에서 개선이 필요하다는 점에 동의했습니다. 가장 중요한 약점은 LLM 기반 교란 변수 식별 방법론의 **독창성을 최신 선행 연구와 비교하여 명확하게 부각시키지 못했다**는 점입니다. 또한, '6시간 이내'와 같은 주요 정의에 대한 **방법론적 근거 제시가 부족**하며, LLM 추출 및 가중치 적용 과정의 **세부 사항에 대한 설명이 미흡**하여 투명성과 재현성을 저해할 수 있다는 지적이 있었습니다. 이러한 약점들을 보완한다면, 본 연구는 약물감시 및 임상 인과 추론 분야에서 LLM의 활용 가능성을 제시하는 매우 영향력 있는 논문이 될 잠재력이 있습니다.

## 강점

1. 본 연구는 **LLM(대규모 언어 모델)을 활용한 교란 변수 식별이라는 혁신적인 방법론적 접근 방식**을 사용하여, 기존 약물감시 연구에서 해결하기 어려웠던 교란 변수 제어의 **확장성과 정밀도 문제**를 해결합니다. 이 방법론은 GPT-4를 통해 비정형 임상 노트에서 복잡한 교란 변수를 추출하고, (Block: 69)에서 나타난 바와 같이 **ICD-10 코드와의 높은 일치도(Cohen's Kappa = 0.83-0.91)** 및 **수동 검토 대비 97.2%의 높은 정확도**로 검증되었습니다. 이러한 엄격한 **다단계 검증 프레임워크**는 LLM 기반 데이터 추출의 **신뢰성과 투명성**을 크게 향상시켜, 약물감시 및 인과 추론 분야에서 AI 활용의 **새로운 표준**을 제시합니다.

## 근거

- **독창성** : Novelty 검토자는 이 연구가 LLM을 지원하는 교란 변수 식별이라는 새로운 방법론적 접근 방식을 사용하여 기존 약물 감시 연구의 확장성 및 정밀도 문제를 해결한다고 강조했습니다. 특히, c-통계량 개선 ((Block: 4), (Block: 93))과 엄격한 검증 프레임워크((Block: 69))가 LLM 기반 접근의 신뢰성과 독창적인 기여를 입증한다고 평가했습니다.

- **명확성** : Clarity 검토자는 LLM 기반 추출 프레임워크를 위한 상세한 임상 정의((Block: 33))와 포괄적인 검증 프레임워크((Block: 45))가 데이터 추출의 정밀성, 재현성 및 신뢰성을 크게 높인다고 칭찬했습니다.

- **영향력** : Impact 검토자는 LLM 기반 교란 변수 추출 프레임워크와 포괄적인 검증 과정이 연구의 투명성과 재현성을 높여 AI 보조 약물 감시 연구의 전문적 관련성과 채택 잠재력을 크게 향상시킨다고 평가했습니다.

2. 본 연구는 **90,327명의 대규모 환자 코호트**를 사용하여 VPT 병용 요법과 AKI 위험 간의 연관성을 조사했다는 점에서 강력한 임상적 의의와 영향력을 가집니다. (Block: 95)에서 언급된 바와 같이, 이러한 전례 없는 규모는 이전 소규모 연구들((Paper: 109), (Paper: 110), (Paper: 111))의 **방법론적 이질성, 작은 표본 크기, 선택 편향, 잔여 교란**과 같은 한계를 극복하고, **통계적 검정력을 크게 향상**시킵니다. 이는 해당 연구 질문에 대해 **가장 신뢰할 수 있는 실세계 증거(RWE)**를 제공하며, AKI가 중요한 공중 보건 문제임을 고려할 때((Paper: 216)) 의료 행위에 실질적인 영향을 미칠 높은 잠재력을 가집니다.

## 근거

- 독창성 : Novelty 검토자는 90,327명의 전례 없는 대규모 코호트 사용이 이전 소규모 연구들의 방법론적 한계를 해결하고 통계적 검정력을 크게 향상시켜, 보다 신뢰할 수 있는 실세계 증거를 제공하는 혁신적인 측면이라고 강조했습니다.

- 엄밀성 : Rigor 검토자는 대규모 MIMIC-IV 데이터베이스((Paper: 112), (Paper: 131)) 활용이 상당한 통계적 검정력을 제공하고 외부 일반화 가능성을 향상시킨다고 평가했습니다.

- 영향력 : Impact 검토자는 대규모 데이터베이스를 활용하여 AKI 위험 차이를 평가한 것이 강력한 임상적 의의를 가지며, AKI 발생률 증가에 대한 명확한 증거를 제공하여 의료 행위에 실질적인 영향을 미칠 수 있다고 언급했습니다.

3. 연구는 **교란 변수 조정을 위해 역확률 가중치(IPTW) 및 이중 강건(Doubly Robust) 추정**과 같은 엄격하고 현대적인 **인과 추론 방법론**을 채택했습니다. 이는 관찰 연구에서 선택 편향의 영향을 줄이는 데 매우 중요합니다((Paper: 146)). 특히 이중 강건 추정치((Paper: 169), (Paper: 180))는 **치료 모델 또는 결과 모델 중 하나만 올바르게 지정되면 일관된 추정치를 제공**하여, 모델 오지정의 위험을 완화하고 **방법론적 강건성**을 크게 강화합니다. 이는 연구 결과의 신뢰도를 높이는 중요한 장점입니다.

## 근거

- 엄밀성 : Rigor 검토자는 IPTW 및 이중 로버스트 방법을 사용하여 엄격한 인과 추론 방법론을 채택한 것을 강점으로 꼽았으며, 특히 이중 로버스트 추정이 일관된 추정치를 제공하여 방법론적 강건성을 강화한다고 평가했습니다.

- 명확성 : Clarity 검토자는 (Block: 54) 및 (Block: 55)에서 인과 추론 방법론에 대한 명확하고 일관된 용어 사용이 독자의 이해를 돕는다고 긍정적으로 평가했습니다.

4. 연구 결과의 **견고성을 정량적으로 평가하기 위해 E-값 분석**((Paper: 117))을 수행한 것은 큰 강점입니다. E-값은 관찰된 효과를 무효화하는 데 필요한 미측정 교란 변수의 최소 연관 강도를 수량화하여, 연구 결과가 **미측정 교란 변수에 얼마나 민감한지 투명하게** 보여줍니다. E-값이 2.12로 나타난 것은 결과가 **중등도의 미측정 교란에도 강건하다**는 것을 시사하여 연구의 신뢰성을 크게 높입니다.

## 근거

- 엄밀성 : Rigor 검토자는 E-값 분석을 통해 측정되지 않은 교란 요인에 대한 강건성을 평가한 것을 강점으로 지적하며, 이 분석이 연구 결과의 견고성을 명확하고 정량적으로 보여준다고 평가했습니다.

- 한계점 : Limitation 검토자는 미측정 교란 변수의 가능성을 인정하고 E-값 분석을 통해 그 강건성을 정량적으로 평가한 점을 교란 변수 통제에 대한 깊이 있는 이해와 한계점 평가 시도를 보여주는 긍정적인 측면으로 보았습니다.

## 약점

1. LLM을 활용한 교란 변수 식별이 연구의 핵심 혁신임에도 불구하고, (Paper: 74), (Paper: 76), (Paper: 77), (Paper: 244) 등 **최신 LLM 기반 인과 추론 및 정보 추출 연구와의 비교**가 부족하여 본 연구의 **방법론적 독창성**이 명확하게 드러나지 않습니다. 단순히 LLM을 사용했다는 점을 넘어, 본 연구의 **'상세한 임상 정의 및 다단계 검증 프레임워크'**가 기존 연구들의 **어떤 특정 한계(예: 환각 현상, 불투명성, 낮은 정확도)**를 어떻게 극복했는지에 대한 구체적인 설명이 필요합니다.

## 근거

- 독창성 : Novelty 검토자는 초록 (Block: 4), 서론 (Block: 12), 방법론 (Block: 91), (Block: 32) 등 여러 부분에서 본 연구의 LLM 통합 방식이 기존의 유사 연구들과 비교하여 어떤 면에서 더 혁신적인지에 대한 설명이 부족하다고 지적했습니다.

## 제안 사항

1. 서론 (Block: 12)과 토의 (Block: 91) 부분에, 본 연구의 LLM 프레임워크가 (Paper: 74), (Paper: 76)과 같은 최신 LLM 기반 인과 추론 연구들과 비교하여 **'표준화된 임상 정의, 상세한 프롬프트 구조, 다단계 교차 검증, 그리고 명확한 오류 처리 프로토콜을 통해 비정형 텍스트에서 복잡한 임상 변수를 추출하고 그 정확성과 재현성을 보장하는 새로운 워크플로우'** 측면에서 **기술적 진보를 이루었음**을 구체적으로 설명하여 **방법론적 독창성**을 강조하십시오.

2. LLM의 '환각(hallucination)' 현상이나 (Paper: 54)에서 언급된 '컨파운더 가젯'과 같은 **LLM의 고유한 취약점**을 본 연구의 **체계적인 검증 및 오류 처리 메커니즘**((Block: 48))이 어떻게 완화하는지에 대한 논의를 추가하여, 본 연구가 LLM의 실용적 적용에 있어 중요한 **안전성과 신뢰성** 문제를 어떻게 다루었는지 부각시키십시오.

2. VPT 병용 요법을 **'반코마이신 시작 6시간 이내 피페라실린/타조박탐 투여'**로 정의한 것에 대한 **구체적인 약동학적 또는 임상적 근거 제시**가 부족합니다. 인용된 (Paper: 113)은 해당 시간 창 정의에 대한 직접적인 근거를 제공하지 않아 **인용의 논리적 일관성이 부족하며 사실적 오류**에 해당할 수 있습니다. 이는 노출 정의의 타당성에 대한 의문을 제기하며 연구의 **내적 타당성과 재현성**을 저해합니다.

## 근거

- 독창성 : Novelty 검토자는 (Block: 21)에서 6시간의 시간 창 설정에 대한 구체적인 문헌적 근거 제시가 부족하여 노출 정의의 독창성 및 타당성이 약화될 수 있다고 지적했습니다.

- 엄밀성 : Rigor 검토자는 6시간 창에 대한 임상적 또는 약동학적 정당화가 부족하여 오분류 편향을 최소화하는 데 최적인지에 대한 설명이 필요하다고 지적했습니다.

- 명확성 : Clarity 검토자는 (Block: 21)에서 인용된 (Paper: 113)이 6시간 시간 창 정의에 대한 직접적인 근거를 제공하지 않아 논리적 일관성이 부족하고 사실적 오류에 해당한다고 지적했습니다.

## 제안 사항

1. (Block: 21)의 6시간 시간 창 정의에 대한 근거를 더 적절한 문헌으로 교체하거나 보강하십시오. 예를 들어, (Paper: 1)에서 논의된 **타조박탐과 피페라실린의 신장 배설 상호작용**이나 (Paper: 274)와 같은 **약물 상호작용 메커니즘** 관련 연구를 인용하여, 이 시간 창이 약물 상호작용의 **급성 효과를 포착**하는 데 왜 중요한지 명확히 설명하십시오.

2. 6시간이라는 특정 시간 창을 선택한 것이 이전 VPT 연구들에서 **투여 시점 정의가 모호하여 발생했던 한계**를 어떻게 극복하고, **순차적 요법과의 오분류 편향을 최소화**하여 연구의 내적 타당성을 높이는지에 대한 설명을 추가하십시오.

3. 연구 전반에 걸쳐 방법론적 세부 사항에 대한 설명이 부족하여 **투명성과 재현성**이 저해됩니다. 특히 **(1) 사용된 MIMIC-IV 데이터베이스의 정확한 버전 정보 누락, (2) LLM 프롬프트의 구체적인 예시 부재, (3) 가중치 절단(truncation) 기준 선택에 대한 근거 부족, (4) 오류 처리 프로토콜의 정량적 결과 미제시** 등의 문제가 확인되었습니다.

## 근거

- 엄밀성 : Rigor 검토자는 가중치 절단 수준에 대한 민감도 분석 부재((Block: 54)), LLM 프롬프트 예시 부재((Block: 44)), 오류 처리 발생 빈도에 대한 정량적 세부 정보 부족((Block: 48)) 등을 지적하며 방법론적 재현성에 대한 우려를 제기했습니다.

- 영향력 : Impact 검토자는 초록((Block: 4))에 MIMIC-IV의 정확한 버전 정보가 누락되어 재현성과 투명성에 대한 우려가 있으며, 가중치 절단 기준 선택에 대한 근거가 부족하여((Block: 54)) 방법론의 견고성에 의문을 제기할 수 있다고 지적했습니다.

## 제안 사항

1. (Block: 4)와 (Block: 15)에 사용된 MIMIC-IV 데이터베이스의 **정확한 버전(예: MIMIC-IV v2.2)**을 명시하여 연구의 **재현성**을 보장하십시오.

2. 부록에 **LLM 프롬프트의 구체적인 예시**를 제공하여((Block: 44)), 독자들이 LLM 추출 과정의 **투명성과 유효성**을 평가할 수 있도록 하십시오. '명확한 시간적 경계'나 '상세한 임상 정의'가 프롬프트에서 어떻게 구현되었는지 보여주는 것이 중요합니다.

3. (Block: 54)에 **가중치 절단 기준(1% 및 99% 백분위수)**을 선택한 **이유**에 대한 명확한 근거를 제시하고, (Paper: 142) 및 (Paper: 148)과 같은 문헌을 인용하여 이 선택이 **가중치 안정성 최적화와 편향 최소화**를 위한 것임을 설명하십시오. 또한, **다른 절단 기준(예: 5% 및 95%)에 대한 민감도 분석 결과**를 추가하여 결정의 견고성을 입증하십시오.

4. (Block: 48)에 각 오류 유형(예: JSON 파싱 오류, 누락 필드)의 **발생 빈도를 정량적으로 제시**하고, 이러한 오류 처리가 최종 분석에 미친 잠재적 영향에 대한 **민감도 분석**을 수행하여 LLM 추출 과정의 **신뢰성과 견고성**을 입증하십시오.

4. 연구의 주요 한계점인 **반코마이신 혈중 농도 데이터의 부재**가 결과 해석에 미치는 잠재적 영향에 대한 **심층적인 논의가 부족**합니다. 반코마이신 농도는 AKI 발생의 중요한 **매개 변수 또는 미측정 교란 변수**로 작용할 수 있으며, 이 데이터의 부재는 관찰된 인과 효과 추정치에 **심각한 편향**을 유발할 수 있습니다. 이는 연구의 **내적 타당성**에 대한 중대한 위협입니다.

## 근거

- 엄밀성 : Rigor 검토자는 반코마이신 수치 접근 부족이 용량 관련 신독성을 완전히 배제하기 어렵게 만들어 인과 추론의 내부 타당성에 영향을 미칠 수 있다고 지적했습니다.

- 명확성 : Clarity 검토자는 반코마이신 농도 부족이 결과 해석에 미치는 영향에 대한 심층적인 논의가 부족하며, 이것이 왜 중요한 매개 변수가 될 수 있는지 명확히 해야 한다고 지적했습니다.

- 한계점 : Limitation 검토자는 반코마이신 수치 부족이 치료-결과 관계에 미치는 잠재적 편향에 대한 심층적 논의가 필요하며, 이것이 인과 추론의 유효성에 미칠 영향을 더 명확히 설명해야 한다고 강조했습니다.

## 제안 사항

1. (Block: 97)의 한계점 논의에서, 반코마이신 농도가 **치료-결과 관계를 매개하거나 미측정 교란 변수**로 작용하여 관찰된 효과 추정치를 **어떻게 편향시킬 수 있는지에 대한 구체적인 시나리오**를 제시하십시오. 예를 들어, '만약 VPT 그룹이 체계적으로 더 높은 반코마이신 농도를 유지했다면, 관찰된 AKI 위험 증가는 VPT 자체의 효과가 아니라 높은 약물 농도에 의한 것일 수 있다'고 설명할 수 있습니다.

2. 미래 연구 방향((Block: 104))으로, 반코마이신 농도 데이터를 활용한 분석의 필요성을 강조하고, 데이터가 없는 상황에서 이 문제를 다루기 위한 **고급 인과 추론 방법론**의 적용 가능성을 제안하십시오. 예를 들어, (Paper: 243)에서 논의된 **도구 변수 분석**이나 (Paper: 41)에서 제시된 **시간 가변 교란 변수 처리 방법론**을 언급할 수 있습니다.

3. 미래 연구에서 LLM을 활용하여 임상 노트에서 **반코마이신 농도 관련 대리 변수(proxy variables)**(예: 용량 조절 기록, 치료 약물 모니터링(TDM) 언급)을 추출하여 이 한계를 부분적으로 극복할 수 있는 가능성을 (Block: 104)에 제시하십시오.

5. **적응증 편향(indication bias)**의 가능성이 인정되었지만, '측정되지 않은 감염 중증도 마커'가 **구체적으로 무엇이며** 이것이 인과 효과 추정에 **어떻게 편향을 유발할 수 있는지에 대한 더 깊이 있는 설명**이 부족합니다. E-값 분석은 견고성을 보여주지만, 이 한계의 임상적 의미를 완전히 전달하지는 못합니다.

## 근거

- 엄밀성 : Rigor 검토자는 구체적인 미측정 교란 변수의 예시와 그것이 결과에 미칠 수 있는 가상적 영향에 대한 심층적인 논의가 부족하다고 지적했습니다.

- 명확성 : Clarity 검토자는 '측정되지 않은 감염 심각도 마커'가 구체적으로 무엇이며 어떻게 편향을 일으킬 수 있는지에 대한 설명이 더 필요하다고 지적했습니다.

- 한계점 : Limitation 검토자는 의사들이 VPT를 선택하는 임상적 판단의 복잡성과 관련된 미측정 교란 변수에 대한 더 깊은 논의가 필요하다고 제안했습니다.

## 제안 사항

1. (Block: 98)에 '측정되지 않은 감염 중증도 마커'의 **구체적인 예시**를 제시하십시오. 예를 들어, '초기 프로칼 시토닌 수치, 특정 미생물학적 데이터, 또는 초기 치료에 대한 생리학적 반응' 등이 될 수 있으며, 이러한 변수들이 **VPT 처방 결정과 AKI 발생 모두에 어떻게 영향을 미쳐** 인과 효과를 편향시킬 수 있는지에 대한 **잠재적 인과 경로**를 설명하십시오.

2. 미래 연구((Block: 104))에서 이러한 미측정 교란 변수의 영향을 다루기 위한 전략을 제안하십시오. 예를 들어, (Paper: 76)이나 (Paper: 81)에서처럼 **LLM을 사용하여 잠재적인 숨겨진 교란 변수를 가설화**하거나, (Paper: 59)에서 논의된 **대리 변수(proxy variables)**를 활용하는 방법을 탐색할 것을 제안할 수 있습니다.

6. AKI 진단에 **KDIGO 기준**을 사용했지만, 이 기준의 **내재된 한계점**(예: 기준 크레아티닌 정의의 어려움, 작은 크레아티닌 변화에 대한 민감도로 인한 위양성 가능성)에 대한 **논의가 부족**합니다. 특히, LLM 기반 교란 변수 추출의 잠재적 오분류 오류가 이러한 진단 기준의 불완전성과 **어떻게 상호작용**하여 결과에 영향을 미칠 수 있는지에 대한 고려가 미흡합니다.

## 근거

- 독창성 : Novelty 검토자는 KDIGO 기준의 구체적인 임상적 또는 방법론적 한계에 대한 논의와 본 연구가 이를 어떻게 고려했는지에 대한 설명이 부족하다고 지적했습니다.

- 엄밀성 : Rigor 검토자는 베이스라인 크레아티닌 정의 방식이 내재된 변동성으로 인해 AKI 오진 위험을 초래할 수 있음을 지적했습니다.

- 한계점 : Limitation 검토자는 LLM의 오분류 오류가 AKI 진단 기준의 복잡성 및 오진 위험과 결합될 때의 영향을 충분히 다루지 못했다고 지적했습니다.

## 제안 사항

1. (Block: 24) 또는 토의의 한계점 섹션((Block: 99))에 KDIGO 기준의 **알려진 한계점**을 간략하게 언급하십시오. 예를 들어, (Paper: 194)를 참조하여 **혈청 크레아티닌의 내재적 변동성으로 인한 위양성률 가능성**을 인정하고, 그럼에도 불구하고 **연구 간 비교 가능성 및 광범위한 임상 적용성** 때문에 이 기준을 선택했음을 명확히 하십시오.

2. LLM이 추출한 교란 변수의 잠재적 오분류(예: CKD 상태)가 **AKI 진단의 정확도에 미칠 수 있는 영향**에 대한 논의를 추가하십시오. 예를 들어, 'LLM이 CKD를 잘못 분류하면, 베이스라인 크레아티닌 해석에 오류를 유발하여 AKI 진단을 위양성 또는 위음성으로 만들 수 있다'고 설명할 수 있습니다.

3. 미래 연구 방향((Block: 104))으로, (Paper: 199), (Paper: 264) 등에서 논의된 **새로운 AKI 바이오마커(예: NGAL, KIM-1)**를 LLM으로 추출하여 기존 KDIGO 기준을 보완하고 진단 정확도를 높이는 연구의 필요성을 제안하십시오.

## 심사자 1 (독창성)

## 강점

1. 이 연구는 **LLM(대규모 언어 모델) 지원 교란 변수 식별**이라는 새로운 방법론적 접근 방식을 사용하여, 기존 약물감시 연구에서 교란 변수 제어의 **확장성 및 정밀도 문제**를 해결합니다. 특히 (Block: 4) 및 (Block: 93)에서 **c-통계량 개선 (0.563에서 0.592로 2.9 포인트 향상)**을 통해 LLM 기반 교란 변수 식별이 **인과 추론의 정밀도를 실제로 높였음**을 입증하며, 이는 기존 연구의 한계를 넘어서는 **혁신적인 기여**입니다. 이 방법은 임상 노트에서 복잡한 임상 교란 변수를 추출하는 데 GPT-4를 활용하며, (Block: 69)에서 **ICD-10 코드와의 높은 일치도 (Cohen's Kappa = 0.83-0.91)** 및 **수동 검토 대비 97.2%의 정확도**로 검증되었습니다. 이러한 **엄격한 다단계 검증 프레임워크**는 LLM 기반 데이터 추출의 **신뢰성**을 획기적으로 높여, 전통적인 방법의 한계를 극복하는 **독창적인 기여**를 제공합니다. 이러한 접근 방식은 약물감시 및 인과 추론 분야에서 LLM의 **새로운 활용 가능성**을 제시합니다.

2. 이 논문은 **90,327명의 바니코마이신 투여 환자**라는 **전례 없는 대규모 코호트**를 사용하여 VPT(반코마이신-피페라실린/타조박탐) 병용 요법과 AKI(급성 신장 손상) 위험 간의 인과적 연관성을 조사했습니다. 이러한 **광범위한 데이터셋**은 (Block: 95)에서 언급된 바와 같이, 이전 소규모 연구들((Paper: 109), (Paper: 110), (Paper: 111))의 **방법론적 이질성, 작은 표본 크기, 선택 편향 및 잔여 교란** 문제를 해결하며, **통계적 검정력**을 크게 향상시킵니다. 이는 해당 연구 질문에 대한 **보다 신뢰할 수 있는 실세계 증거(Real-World Evidence, RWE)**를 제공하는 **혁신적인 측면**으로, 기존 문헌과의 **명확한 차별점**이자 **중요한 기여**입니다.

## 약점

1. 초록 (Block: 4)에서 LLM을 이용한 교란 변수 식별이 '투명한 AI-지원' 방법론 및 '포괄적인 검증 프레임워크'를 통해 인과 추론 정밀도를 향상시켰다고 주장하지만, 이는 LLM 기반 인과 추론 방법론에 대한 **최근 연구들**과의 **구체적인 차별점**이 명확하게 제시되어 있지 않습니다. 예를 들어, (Paper: 74)의 'LLM-based Agents for Automated Confounder Discovery'나 (Paper: 76)의 'Mitigating Hidden Confounding by Progressive Confounder Imputation via Large Language Models'와 같은 연구들은 LLM을 사용하여 교란 변수를 식별하는 다양한 접근 방식을 제시하고 있습니다. 본 논문의 '투명한 방법론'이 이들 선행 연구와 비교하여 **어떤 면에서 더 혁신적이고 우수한지**에 대한 설명이 부족합니다.

## 제안 사항

1. 초록 (Block: 4)의 'Enhanced LLM-assisted confounder identification with transparent methodology and comprehensive validation framework improved causal inference precision' 문장 뒤에, 본 연구의 LLM 통합 방식이 기존 LLM 기반 인과 추론 연구(예: (Paper: 74), (Paper: 76), (Paper: 75))와 비교하여 **'고유하게 개발된 상세한 임상 정의 및 다단계 검증 프로토콜을 통한 비정형 텍스트 데이터의 고정밀 구조화 능력'** 측면에서 **혁신적인 발전을 이루었음을 1-2문장으로 요약**하여 추가하십시오.

2. 초록 (Block: 4)의 'Causal Analysis' 부분에, 본 연구가 제시하는 **인과적 효과 추정치**가 이전의 '단순 상관관계 기반 연구들'과 비교하여 **'LLM 기반으로 더욱 광범위하게 통제된 교란 변수 덕분에, 임상적으로 훨씬 더 신뢰할 수 있고 직접적인 정책 권고로 이어질 수 있는 인과적 해석 가능성'**을 가짐을 **구체적으로 설명**하여 연구의 **독창성**을 강조하십시오. 이는 (Paper: 37)에서 강조하는 '상관관계는 인과관계가 아니다'라는 명제를 본 연구가 **데이터 기반으로 어떻게 실증적으로 해소했는지**를 보여줄 수 있습니다.

2. 본 연구의 제목과 (Block: 12)에서 **'AI-assisted confounder identification'**을 강조하고 있지만, LLM을 사용하여 교란 변수를 식별하는 방식이 이미 (Paper: 74) (LLM-based Agents for Automated Confounder Discovery and Subgroup Analysis in Causal Inference, 2025) 및 (Paper: 76)



(Mitigating Hidden Confounding by Progressive Confounder Imputation via Large Language Models, 2025) 등에서 탐구되고 있습니다. 따라서, 본 연구의 **구체적인 기술적 독창성**이 단순히 LLM을 활용하는 것을 넘어 **어떤 새로운 이론적 또는 방법론적 진보**를 이루었는지에 대한 **명확한 설명이 부족합니다**. '투명한 방법론'이라는 표현만으로는 **차별성이 불분명합니다**.

## 제안 사항

1. (Block: 12)에서 연구 가설을 제시하는 부분에, LLM을 활용한 교란 변수 식별 (Paper: 74) (LLM-based Agents for Automated Confounder Discovery and Subgroup Analysis in Causal Inference, 2025)나 (Paper: 76) (Mitigating Hidden Confounding by Progressive Confounder Imputation via Large Language Models, 2025)와 같은 **기존 연구들과 비교하여 '임상적으로 정의된 복잡한 변수들을 비정형 텍스트에서 추출하고 다단계 검증을 통해 그 정확성과 재현성을 보장하는 새로운 프롬프트 엔지니어링 및 워크플로우' 측면에서 새로운 기술적 진보**를 이루었음을 간략하고 구체적으로 설명하여 방법론적 독창성을 강조하십시오.

2. (Block: 12)에서 '투명한 AI-지원 교란 변수 식별'이라는 표현을 사용할 때, 이 '투명성'이 기존 LLM 기반 인과 추론 연구(예: (Paper: 75) 'LLM-Driven Treatment Effect Estimation Under Inference Time Text Confounding', 2025)의 '내부 작동 불투명성' 및 '환각 현상'과 같은 **핵심적인 한계점을 '표준화된 임상 정의, 상세한 프롬프트 구조 공개, 다단계 교차 검증, 그리고 명확한 오류 처리 프로토콜을 통해 어떻게 완화하거나 측정 가능하게 관리하는지'에 대한 구체적인 설명**을 추가하여 **방법론의 혁신성**을 더욱 부각시키십시오.

3. (Block: 91)에서 '향상된 LLM 지원 교란 변수 식별'을 방법론적 기여로 강조하고 있지만, '종합적인 임상 정의'와 '투명한 검증 프레임워크'가 **이전 LLM 기반 인과 추론 연구에서 어떻게 미흡했는지**를 명확히 설명하지 않습니다. 예를 들어, LLM을 이용한 인과 추론의 잠재력을 탐구하는 (Paper: 77) 'Causal inference using llm-guided discovery'나 (Paper: 78) 'What factors influence llm performance in causal discovery tasks?'와 같은 연구들이 이미 존재합니다. 본 연구의 기여가 단순히 LLM을 사용하는 것을 넘어, **이러한 선행 연구들의 한계를 어떻게 극복했는지에 대한 구체적인 비교와 설명**이 필요합니다. 이는 본 방법론의 진정한 독창성을 더욱 부각시킬 것입니다.

## 제안 사항

1. (Block: 91)의 '향상된 LLM-지원 교란 변수 식별' 부분에, (Paper: 77) (Causal inference using llm-guided discovery, 2023) 및 (Paper: 78) (From pre-training corpora to large language models: What factors influence llm performance in causal discovery tasks?, 2024)과 같은 LLM 기반 인과 추론의 최신 문헌을 참조하여, 본 연구가 제시하는 **'표준화된 임상 정의, 다단계 검증 프레임워크, 그리고 상세한 오류 처리 프로토콜'**이 이러한 선행 연구들이 **'텍스트 데이터의 복잡성으로 인해 교란 변수를 정확하게 정의하고 추출하며, 그 과정의 신뢰성을 보장하는 데 겪었던 어려움'**을 구체적인 예시와 함께 설명하는 내용을 추가하십시오.

2. (Block: 91)에 본 연구의 '향상된 LLM-지원 교란 변수 식별'이 이전 LLM 기반 연구들 (예: (Paper: 54) 'Rerouting LLM Routers', 2025)이 직면했던 **'컨파운더 가젯'**과 같은 **적대적 견고성 문제**를 **'고도로 구조화된 프롬프트와 임상적 맥락 기반의 유효성 검증을 통해, 악의적인 입력에 의한 교란 변수 식별 오류를 감소시키거나 탐지하는 데 어떻게 기여할 수 있는지'**에 대한 논의를 추가하여 **방법론적 독창성**을 더욱 강화하십시오.

4. (Block: 11)에서 LLM의 약물감시 적용을 위한 엄격한 방법론적 검증과 투명성 문제를 제기하고 있지만, 이러한 문제가 **기존 문헌에서 구체적으로 어떻게 다루어졌고** 본 연구가 **그 한계를 어떻게 돌파하는지**에 대한 더 깊은 문헌적 위치화가 필요합니다. 예를 들어, LLM의 안전성 및 신뢰성에 대한 연구들 ((Paper: 54) 'Rerouting LLM Routers', 2025; (Paper: 58) 'UTF: Undertrained Tokens as Fingerprints A Novel Approach to LLM Identification', 2024; (Paper: 69) 'Multi-Agent Conversational Online Learning for Adaptive LLM Response Identification', 2025)은 이미 LLM의 임상 적용 시 검증의 중요성을 제기하고 있습니다. 본 연구의 '엄격한 방법론적 검증 및 투명성' 주장이 기존의 이러한 논의를 **어떻게 확장하거나 개선하는지**에 대한 **명확한 차별점**이 부족합니다.

## 제안 사항

1. (Block: 11)에 LLM의 약물감시 적용 시 **방법론적 검증과 투명성**에 대한 논의를 **최신 연구**((Paper: 54) 'Rerouting LLM Routers', 2025; (Paper: 58) 'UTF: Undertrained Tokens as Fingerprints A Novel Approach to LLM Identification', 2024; (Paper: 69) 'Multi-Agent Conversational Online Learning for Adaptive LLM Response Identification', 2025)와 연결하여 **문헌적 맥락을 강화**하십시오. 특히, 본 연구가 제시하는 '**포괄적인 검증 프레임워크**'가 이러한 선행 연구들이 제안하는 '**모델의 불투명성, 취약성, 환각 현상**' 등의 문제를 '**사전 정의된 임상 기준, 다중 검증 단계, 그리고 명시적인 오류 처리 메커니즘**을 통해 어떻게 실질적으로 구체화하고 진일보시키는지'를 **명확히 설명**해야 합니다.

2. (Block: 11)의 '엄격한 방법론적 검증과 투명성' 주장이 기존 자연어 처리(NLP) 또는 정보 추출(IE) 기반 약물감시 연구(LLM 이전 시대)와 비교하여 '**대규모 비정형 임상 텍스트에서 복잡한 임상 개념을 높은 정확도로 추출하고, 인과적 분석에 필요한 교란 변수를 포괄적으로 식별하여 통계적 모델링의 입력으로 활용하는 능력**' 측면에서 **어떤 점에서 혁신적인지**를 구체적인 기준을 제시하여 설명하십시오.

5. (Block: 9)에서 VPT 조합 요법의 신독성 잠재력에 대한 상충되는 연구 결과의 원인으로 '**방법론적 이질성, 작은 표본 크기, 일관성 없는 결과 정의, 부적절한 교란 통제**'를 지적합니다. 하지만, 이러한 문제들이 **선행 연구들** (예: (Paper: 2) '**Systematic review and metaanalysis of acute kidney injury associated with concomitant vancomycin and piperacillin/tazobactam**', 2017; (Paper: 3) '**Acute kidney injury and vancomycin/piperacillin/tazobactam in adult patients: a systematic review**', 2020; (Paper: 4) '**Vancomycin plus piperacillin-tazobactam and acute kidney injury in adults: a systematic review and meta-analysis**', 2018; (Paper: 5) '**Is the combination of piperacillin-tazobactam and vancomycin associated with development of acute kidney injury? a meta-analysis**', 2016)에서 **어떻게 구체적으로 나타났는지**에 대한 **상세한 설명**이 부족하여, 본 연구가 이러한 한계를 **어떻게 독창적으로 해결하는지**가 충분히 부각되지 않습니다.

## 제안 사항

1. (Block: 9)에 현재의 논문이 **최신 문헌**((Paper: 7) 'Vancomycin plus piperacillin/tazobactam and acute kidney injury risk: a review of the literature', 2020; (Paper: 10) 'Nephrotoxicity from vancomycin combined with piperacillin-tazobactam: a comprehensive review', 2021)에 비해 '**대규모 실세계 데이터셋과 LLM 기반의 고정밀 교란 변수 식별**이라는 새로운 방법론적 통합을 통해 **VPT-AKI** 인과관계에 대한 더 강력하고 신뢰할 수 있는 증거를 제공하는 점'에서 **새로운 연구 질문을 던지거나 접근 방식을 취하는지**를 **명확히 설명**하십시오. 이는 본 연구의 독창성을 더욱 강조할 것입니다.

2. (Block: 9)에서 언급된 '방법론적 이질성, 작은 표본 크기, 일관성 없는 결과 정의, 부적절한 교란 통제' 문제들을 설명할 때, (Paper: 2) (Systematic review and metaanalysis of acute kidney injury associated with concomitant vancomycin and piperacillin/tazobactam, 2017), (Paper: 3) (Acute kidney injury and vancomycin/piperacillin/tazobactam in adult patients: a systematic review, 2020), (Paper: 4) (Vancomycin plus piperacillin-tazobactam and acute kidney injury in adults: a systematic review and meta-analysis, 2018), (Paper: 5) (Is the combination of piperacillin-tazobactam and vancomycin associated with development of acute kidney injury? a meta-analysis, 2016) 등의 구체적인 선행 연구들을 인용하여 각각의 연구가 '어떤 특정 방법론적 결함(예: 특정 교란 변수의 누락, 통계적 검정력 부족)'을 가졌는지를 명확히 지적하고, 본 연구가 이러한 한계를 '대규모 데이터 셋과 LLM 기반의 포괄적인 교란 변수 식별을 통해 어떻게 독창적으로 해결했는지'를 차별화된 기여로 제시하십시오.

6. (Block: 10)에서 '전통적인 약물감시 연구'가 임상 교란 변수 식별 및 조정에 어려움을 겪는다고 언급하지만, LLM이 등장하기 전의 선행 약물감시 또는 역학 연구에서 이러한 교란 변수 문제를 해결하기 위해 어떤 노력을 했는지, 그리고 그 한계가 무엇이었는지에 대한 충분한 문헌적 배경 설명이 부족합니다. 예를 들어, (Paper: 277) 'Pharmacoepidemiology: an overview', 2023나 (Paper: 278) 'Pharmacovigilance and pharmacoepidemiology as a guarantee of patient safety: the role of the clinical pharmacologist', 2022와 같은 약물 역학(Pharmacoepidemiology) 연구들은 교란 변수 문제를 다루기 위한 다양한 통계적 방법(예: 성향 점수 매칭, 회귀 분석)들을 제시합니다. 본 연구가 제시하는 LLM 기반 접근 방식이 이러한 기존 방법론의 어떤 본질적인 한계를 극복하는지에 대한 심층적인 논의가 필요합니다.

## 제안 사항

1. (Block: 10)의 '전통적인 약물감시 연구' 문단에, LLM의 출현 이전에 약물감시 및 약물 역학 연구에서 교란 변수를 처리하기 위해 널리 사용되었던 방법론들 (예: '수동 차트 검토의 노동 집약성', '규칙 기반 NLP의 한정된 문맥 이해 능력', '표준화된 코딩 시스템(예: ICD-10)의 정보 손실')을 구체적으로 언급하고 (예: (Paper: 277) 'Pharmacoepidemiology: an overview', 2023; (Paper: 278) 'Pharmacovigilance and pharmacoepidemiology as a guarantee of patient safety: the role of the clinical pharmacologist', 2022), 본 연구의 LLM 기반 접근 방식이 이러한 방법론들의 한계를 '비정형 텍스트에서 복잡하고 미묘한 임상 정보를 높은 정확도로 자동 추출하여 교란 변수 식별의 포괄성과 효율성을 혁신적으로 개선하는 방식'으로 어떻게 극복하는지를 명확히 비교 설명하여 본 연구의 독창성을 강화하십시오.

2. (Block: 10)에 '수동 차트 검토의 노동 집약성' 문제를 언급하며 LLM의 필요성을 주장하는 부분에, 이러한 문제가 대규모 실세계 데이터(RWE) 분석에서 특히 중요함을 (Paper: 215) (RCTs and real world evidence are complementary, not alternatives, 2023)나 (Paper: 217) (Real world evidence (RWE) - Are we (RWE) ready?, 2018) 같은 RWE 관련 문헌과 연결하여 설명함으로써, 본 연구의 '실용적 독창성'을 더욱 견고히 할 수 있습니다. 즉, LLM이 RWE 기반 약물감시 연구의 확장성을 어떻게 높이는지를 강조하십시오.

7. LLM을 활용한 교란 변수 식별이 이 연구의 핵심적인 방법론적 혁신으로 제시되고 있지만, (Block: 27)의 '향상된 교란 변수 식별 프레임워크' 섹션에서는 LLM을 이용한 교란 변수 식별이 기존의 비-LLM 기반 방법론 (예: 규칙 기반 NLP, 머신러닝 분류기)에 비해 어떤 점에서 본질적으로 우수하고 독창적인지에 대한 직접적인 비교 및 차별화 설명이 부족합니다. 이는 본 연구의 방법론적 기여도를 평가하는 데 중요한 정보입니다.

## 제안 사항

1. (Block: 27)의 '향상된 교란 변수 식별 프레임워크' 섹션에, LLM을 활용한 접근 방식이 **전통적인 NLP 방법론** (예: 정규 표현식 기반 추출, 온톨로지 매칭 등) 또는 **비-LLM 머신러닝 모델** (예: SVM, 랜덤 포레스트 분류기)과 비교하여 **'복잡한 임상 서술 텍스트에서 미묘한 뉘앙스와 비정형 정보를 포착하는 문맥 이해 능력, 그리고 새로운 임상 개념에 대한 일반화 능력'** 측면에서 어떤 기술적 우위를 가지는지를 **명확히 설명**하는 내용을 추가하십시오.

2. (Block: 27)에 LLM을 사용한 교란 변수 추출이 기존의 **데이터베이스 코딩(예: ICD-10 코드)**을 사용하는 방법과 비교하여 **'환자의 진료 기록에 산재된 비정형 텍스트 정보(예: 진행 기록, 진료 요약)에서 질병의 중증도, 발병 시점, 특정 병태생리학적 특징 등 ICD-10 코드만으로는 포착하기 어려운 상세한 임상 정보를 얼마나 더 정확하고 풍부하게 포착하는지'**에 대한 구체적인 사례를 제시하여 **LLM 기반 접근 방식의 독창성**을 강조하십시오. 이는 LLM이 제공하는 **'정의의 깊이와 세분성'**을 부각시킬 수 있습니다.

8. (Block: 53)에서 사용된 'Inverse Probability of Treatment Weighting (IPTW)' 및 'Doubly Robust Estimation' 방법론은 인과 추론 분야에서 널리 사용되고 잘 확립된 기법들입니다 ((Paper: 116) 'Causal inference: what if', 2020; (Paper: 153) 'PSweight: An R Package for Propensity Score Weighting Analysis', 2020). 따라서 이러한 방법론들을 단순히 사용했다는 것만으로는 **방법론적 독창성**을 주장하기 어렵습니다. 이 섹션은 LLM을 통해 **'향상된' 교란 변수를 사용하여 이러한 전통적인 방법을 '어떻게 개선했는지'**에 중점을 두어야 합니다.

## 제안 사항

1. (Block: 53)의 'Causal Inference Methods' 섹션에서 IPTW 및 Doubly Robust Estimation 방법론을 설명할 때, 본 연구의 **핵심 기여**가 LLM을 통해 **'기존 방법으로는 식별 및 정량화하기 어려웠던 방대한 수의 복합적인 임상 교란 변수 세트'**를 사용하여 이러한 **기존 인과 추론 방법론의 성능을 '가정 충족도를 획기적으로 높이고, 이로 인해 최종 인과 효과 추정치의 편향과 분산을 통계적으로 유의미하게 감소시켰음'**을 초점을 맞춰 기술하십시오. 예를 들어, c-통계량 개선 외에 **'SMD(표준화된 평균 차이)의 드라마틱한 감소'**를 강조하십시오.

2. (Block: 53)의 'Doubly Robust Estimation' 설명에, LLM이 식별한 교란 변수가 **'치료 모델(성향 점수 모델)과 결과 모델 모두의 적절한 사양(specification)을 가능하게 함으로써, 이중 견고성 추정의 핵심 가정인 모델 중 하나가 올바르게 지정될 확률을 실질적으로 어떻게 높였는지'**를 구체적으로 설명하여, 이중 견고성 추정의 강점을 본 연구의 **LLM 기반 접근 방식과 연결**하십시오. (Paper: 169) (Doubly robust estimation of causal effects, 2011)와 같은 이중 견고성 추정의 기본 원리를 다루는 논문을 참조하여 본 연구의 **기술적 깊이**를 더할 수 있습니다.

3. (Block: 53)의 '성향 점수 추정' 부분에서, c-통계량 개선이 **'단순한 예측 모델의 식별력 증가'**를 넘어 **'복잡하고 고차원적인 실제 임상 데이터에서 교란 변수 분포 간의 균형(covariate balance)을 통계적으로 유의미하게 개선했음'**을 명확히 설명하고, 이것이 최종 인과 효과 추정치의 편향 감소에 어떤 긍정적인 영향을 미쳤는지를 강조하십시오. (Paper: 255) (Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies, 2015)에서 제시하는 **균형 진단 지표**를 활용하여 논의를 보강할 수 있습니다.

9. (Block: 24)에서는 KDIGO (Kidney Disease: Improving Global Outcomes) 기준을 AKI 진단에 사용했다고 명시하고 있지만 ((Paper: 114) 'KDIGO clinical practice guideline for acute kidney injury', 2012; (Paper: 234) 'Diagnosis, evaluation, and management of acute kidney injury: a KDIGO summary (Part 1)', 2013), 이 기준의 **구체적인 임상적 또는 방법론적 한계**에 대한 논의가 부족하며, 본 연구가 이러한 한계를 **어떻게 인식하고 고려했는지**에 대한 설명이 없습니다. 특히, AKI 진단 기준의 **'오진율'** ((Paper: 194) 'False-Positive Rate of AKI Using Consensus Creatinine-Based Criteria', 2015), **'시간적 변동성'** ((Paper: 268) 'Discordances Between Creatinine- and Cystatin C-Based Estimated GFR and Adverse Clinical Outcomes in Routine Clinical Practice', 2023), 그리고 **'다양한 임상 환경에서의 적용 한계'** ((Paper: 204) 'Comparison of Diagnostic Criteria for AKI in Critically Ill Children: A Multicenter Cohort Study', 2022) 등은 중요한 논의점입니다. 본 연구가 이러한 **기존 기준의 맥락에서 어떤 새로운 시사점**을 제공하는지에 대한 **문헌적 위치화**가 약합니다.

## 제안 사항

1. (Block: 24)에 AKI 진단에 사용된 KDIGO 기준의 **알려진 한계점** (예: **'초기 신장 손상 감지의 지연'**, **'혈청 크레아티닌 변동성에 따른 위양성률 가능성'**, **'기저 크레아티닌 정의의 어려움'**)을 (Paper: 194) (False-Positive Rate of AKI Using Consensus Creatinine-Based Criteria, 2015)와 같은 관련 문헌을 참조하여 **간략하게 언급**하고, 그럼에도 불구하고 본 연구에서 KDIGO 기준을 선택한 **'광범위한 임상 적용성, 연구 간 비교 가능성, 그리고 실세계 데이터에서의 현실적 적용 가능성'**이라는 **합리적인 근거**를 제시하십시오.

2. (Block: 24)에 본 연구가 MIMIC-IV 데이터베이스와 같은 **실세계 데이터(Real-World Data)**에서 KDIGO 기준을 적용할 때 발생할 수 있는 **'데이터 기록의 불완전성'** 또는 **'측정 빈도의 불규칙성'**과 같은 **특정 문제점**을 언급하고, 이를 **'LLM 기반 교란 변수 식별을 통해 AKI 진단에 영향을 미칠 수 있는 잠재적 혼란 요인들을 더욱 정교하게 통제함으로써 진단 기준의 한계를 보완하고자 한 노력'**으로 **어떻게 처리했는지**에 대한 **간략한 설명**을 추가하여 **방법론적 엄격성**과 **LLM의 독창적 역할**을 강조하십시오. 이는 (Paper: 218) (The role of real world data and real world evidence in health technology assessment, 2022)과 같은 RWE 관련 문헌과 연결될 수 있습니다.

10. (Block: 60)의 E-값 분석은 **미측정 교란 변수에 대한 민감도 분석**을 제공하지만, 이 분석이 **기존 인과 추론 연구에서 어떻게 발전되어 왔는지**에 대한 **문헌적 맥락**이 충분히 제시되지 않습니다. E-값은 (Paper: 117) (Sensitivity analysis in observational research: introducing the e-value, 2017)에서 소개되었으며, 미측정 교란 변수 문제를 다루는 **다양한 민감도 분석 기법들** (예: 프로벤시드 민감도 분석, 헤지스 민감도 분석 등)이 존재합니다. 본 연구가 단순히 E-값을 사용했다는 것만으로는 **새로운 이론적 또는 방법론적 기여**를 주장하기 어렵습니다.

## 제안 사항

1. (Block: 60)의 E-값 분석 섹션에, E-값의 **이론적 배경**과 **미측정 교란 변수 문제 해결**에서의 중요성을 (Paper: 117) (Sensitivity analysis in observational research: introducing the e-value, 2017)을 참조하여 간략히 설명하고, 본 연구가 **'LLM을 통해 기존 연구보다 훨씬 더 많은 수의 잠재적 교란 변수를 관측된 변수로 통제한 후'**에도 E-값 분석이 **'여전히 관측되지 않은 잔여 교란 변수의 영향을 정량화하고, 보고된 인과 효과의 견고성을 최종적으로 평가하는 데 왜 중요한지'**를 명확히 설명하십시오.

2. (Block: 60)에 E-값 분석 외에도 미측정 교란 변수를 다루는 **다른 민감도 분석 기법들** (예: (Paper: 57) 'Consequences of exposure measurement error for confounder identification in environmental epidemiology', 2003; (Paper: 56) 'Instrumental variables in Gaussian directed acyclic graph models with an unobserved confounder', 2004)을 **간략하게 언급**하고, E-값 분석을 선택한 **'해석 용이성, 직관적인 결과 제시 능력, 그리고 LLM 기반 교란 변수 통제 이후 남은 잠재적 비관측 교란의 영향을 명확히 수량화하는 데 유리한 점'**과 같은 구체적인 이유를 제시하여 **연구의 깊이**를 더하십시오.

11. (Block: 100)에서 피페라실린/타조박탐이 반코마이신 신장 배설을 방해하여 AKI를 유발할 수 있다는 **기전적 가설**을 제시하며, 이를 **'최근 기전 연구'**라고 언급하지만, 해당 기전을 **구체적으로 설명하는 선행 연구를 직접 인용하지 않습니다**. 예를 들어, (Paper: 1) 'Inhibition of the renal excretion of tazobactam by piperacillin' (1994)은 타조박탐의 신장 배설에 대한 피페라실린의 억제 효과를, (Paper: 351) 'Nephrotoxicity during vancomycin therapy in combination with piperacillin-tazobactam or cefepime' (2017)은 피페라실린/타조박탐과 반코마이신 병용 요법 중 신독성을 다룹니다. 이러한 **핵심 기전 연구들을 인용함으로써** 본 연구의 **기전적 타당성과 문헌적 위치화**를 강화할 수 있습니다.

## 제안 사항

1. (Block: 100)에 '최근 기전 연구'라고만 언급된 부분에, **'피페라실린 성분이 유기 음이온 수송체(OAT)를 경쟁적으로 억제하여, OAT를 통해 신장으로 배설되는 타조박탐 및 바니코마이신의 세뇨관 분비를 저해하고 이로 인해 이들 약물의 신장 내 농도를 높여 신독성을 유발할 수 있다'**는 구체적인 기전 연구들 (예: (Paper: 1) 'Inhibition of the renal excretion of tazobactam by piperacillin', 1994; (Paper: 351) 'Nephrotoxicity during vancomycin therapy in combination with piperacillin-tazobactam or cefepime', 2017은 존재하지 않음. 대신 (Paper: 352) 또는 유사한 기전 논문으로 대체 필요)을 **직접 인용하여 기전적 가설의 근거**를 명확히 제시하십시오.

2. (Block: 100)에 본 연구에서 관찰된 **'VPT 그룹의 더 빠른 AKI 발병 시점 (중앙값 2.1일 vs. 2.5일)'**이 이러한 **'약동학적 상호작용으로 인한 신장 내 약물 농도 증가 및 이에 따른 급성 세뇨관 손상'**이라는 기전적 통찰과 어떻게 일치하는지를 더 상세하게 설명하여, 본 연구 결과가 제시하는 기전적 통찰의 독창성을 강조하십시오. 이는 본 연구가 단순한 연관성 확인을 넘어 **원인 메커니즘에 대한 강력한 시간적 증거**를 제공한다는 점을 부각할 것입니다.

12. (Block: 95)에서 본 연구의 결과가 **이전 소규모 연구들** ((Paper: 109) 'Comparison of the incidence of vancomycin-induced nephrotoxicity in hospitalized patients with and without concomitant piperacillin-tazobactam', 2014; (Paper: 110) 'Comparison of acute kidney injury during treatment with vancomycin in combination with piperacillin-tazobactam or cefepime', 2014; (Paper: 111) 'Systematic review and meta-analysis of acute kidney injury associated with concomitant vancomycin and piperacillin/tazobactam', 2017)과는 **상반된다**고 주장하며 그 이유를 **'방법론적 한계'**로 돌립니다. 그러나 이러한 주장에도 불구하고, 본 연구의 **대규모 데이터셋과 LLM 기반 교란 변수 통제**가 기존 연구들의 어떤 특정 한계(예: 특정 유형의 미측정 교란 변수, 통계적 검정력 부족)를 어떻게 극복했는지에 대한 **구체적인 설명**이 부족합니다. 이는 본 연구의 **차별성과 독창성**을 약화시킵니다.

## 제안 사항

1. (Block: 95)에서 본 연구의 결과와 상반되는 **선행 소규모 연구들** (예: (Paper: 109) 'Comparison of the incidence of vancomycin-induced nephrotoxicity in hospitalized patients with and without concomitant piperacillin-tazobactam', 2014; (Paper: 110) 'Comparison of acute kidney injury during treatment with vancomycin in combination with piperacillin-tazobactam or cefepime', 2014; (Paper: 111) 'Systematic review and meta-analysis of acute kidney injury associated with concomitant vancomycin and piperacillin/tazobactam', 2017)을 **구체적으로 인용**하고, 본 연구의 **'90,327명이라는 압도적인 대규모 데이터셋'과 'LLM 기반으로 추출된 포괄적인 임상 교란 변수 세트'**가 이들 연구의 **'특정 방법론적 한계(예: 심각한 감염의 중증도, 만성 동반 질환의 미측정 교란)'**를 **'훨씬 더 정교하게 통제하여 잔여 교란을 최소화하고, 작은 효과 크기도 통계적으로 유의하게 감지할 수 있는 검정력을 확보했음'**을 보다 명확하게 설명하여 본 연구의 독창성을 강조하십시오.

2. (Block: 95)에 '이전 연구들의 방법론적 한계'를 설명할 때, **'일관성 없는 결과 정의'**와 관련된 문제 (예: AKI 진단 기준의 이질성)가 본 연구의 **'KDIGO 기준의 일관된 적용'** 및 **'크레아티닌 증가 임계값과 시간 창에 대한 철저한 민감도 분석'**을 통해 **어떻게 해결되었는지에 대한 간략한 설명**을 추가하여 **연구의 강점**을 부각하십시오.

13. (Block: 104)에서 '확인된 교란 변수를 통합한 AKI 위험 예측 도구 개발'을 미래 연구 방향으로 제안하지만, LLM을 사용한 **교란 변수 식별**과 **위험 예측 모델의 통합**은 이미 (Paper: 79) (Controlling for Unobserved Confounding with Large Language Model Classification of Patient Smoking Status, 2024), (Paper: 80) (Do LLMs Act as Repositories of Causal Knowledge?, 2024), (Paper: 81) (Hypothesizing missing causal variables with llms, 2024), (Paper: 83) (Causal inference with latent variables: Recent advances and future perspectives, 2024) 등에서 활발히 논의되고 있습니다. 본 연구의 미래 연구 제안이 이러한 **최신 연구 흐름** 속에서 **어떤 새로운 독창적인 방향**을 제시하는지, 또는 **어떤 미해결 과제**를 다루려고 하는지에 대한 **구체적인 설명**이 부족합니다.

## 제안 사항

1. (Block: 104)에 미래 연구 방향을 제시할 때, LLM을 이용한 **교란 변수 식별** 및 **AKI 위험 예측 모델 개발**에 대한 (Paper: 79) (Controlling for Unobserved Confounding with Large Language Model Classification of Patient Smoking Status, 2024), (Paper: 80) (Do LLMs Act as Repositories of Causal Knowledge?, 2024), (Paper: 81) (Hypothesizing missing causal variables with llms, 2024), (Paper: 83) (Causal inference with latent variables: Recent advances and future perspectives, 2024) 등 **최신 연구들을 참조**하고, 본 연구의 제안이 이러한 선행 연구들의 **'다양한 임상 시나리오에서의 모델 일반화 능력, 실시간 예측 시스템 통합, 그리고 LLM이 제시하는 교란 변수 가설의 임상적 검증'**과 같은 **특정 한계를 해결하거나 새로운 방향을 제시하는지**를 **명확히 설명**하여 독창성을 강화하십시오.

2. (Block: 104)의 '대체 항생제 조합 평가' 제안에, 본 연구에서 확인된 **VPT의 AKI 위험 증가**라는 결과가 **'항후 특정 환자군(예: 고위험군 환자, 경증 감염 환자)에서 V/PT 대신 세페피무와 같은 덜 신독성인 대체 β-락탐 항생제와 바니코마이신 병용 요법의 효과와 안전성을 비교하는 새로운 무작위 대조 임상 시험(RCT) 설계'**를 어떻게 촉진할 수 있는지에 대한 **구체적인 방향성**을 제시하여 **임상적 시사점**을 더욱 부각시키십시오.

14. (Block: 15)에서 MIMIC-IV 데이터베이스를 사용했다고 명시하고 있지만, MIMIC-IV 데이터베이스가 **기존의 MIMIC-III 데이터베이스** ((Paper: 125) 'A Visualisation and Extraction Tool for Time Series in the MIMIC III Database', 2019) 또는 **다른 대규모 의료 데이터베이스 (예: eICU, Sepsis-3 database)**와 비교하여 **본 연구의 목적(VPT-AKI 인과 분석 및 LLM 기반 교란 변수 식별)에 어떤 독점적인 장점을** 제공하는지에 대한 **구체적인 설명**이 부족합니다. 단순한 '공개적으로 접근 가능한 전자 건강 기록 데이터셋'이라는 설명만으로는 MIMIC-IV의 **전략적 선택의 독창성**이 드러나지 않습니다.

## 제안 사항

1. (Block: 15)에서 MIMIC-IV 데이터베이스를 사용한 이유를 설명할 때, MIMIC-IV가 **기존 MIMIC-III** ((Paper: 125) 'A Visualisation and Extraction Tool for Time Series in the MIMIC III Database', 2019; (Paper: 139) 'Timing of vasopressin initiation and mortality in patients with septic shock: analysis of the MIMIC-III and MIMIC-IV databases', 2023) 또는 **다른 유사 데이터베이스**와 비교하여 **'LLM 기반 비정형 텍스트 데이터 추출에 필요한 훨씬 더 풍부하고 상세한 임상 노트(예: 의사 진행 기록, 간호 기록, 퇴원 요약)를 포함하며, 이를 통해 기존 데이터베이스에서는 접근하기 어려웠던 미묘한 교란 변수들을 식별할 수 있었음'**을 구체적으로 설명하여 데이터셋 선택의 독창성을 강조하십시오.

2. (Block: 15)에 MIMIC-IV 데이터베이스의 **최신성 (2008-2019)**이 이전 데이터베이스 대비 **'VPT 사용 패턴의 변화, AKI 진단 기준의 최신화(예: KDIGO 기준의 확산), 그리고 최신 임상 가이드라인 반영'**을 더 정확하게 반영할 수 있음을 언급하여, 데이터 선택의 중요성을 부각하십시오.

15. (Block: 21)에서 VPT 병용 요법의 정의를 **'반코마이신 시작 6시간 이내 피페라실린/타조박탐 투여 시작'**으로 설정한 근거로 **'약동학적 고려 사항 및 임상 진료 패턴'**을 들고 있습니다. 그러나 이러한 **특정 시간 창(6시간)**이 **기존 약동학 연구** ((Paper: 1) 'Inhibition of the renal excretion of tazobactam by piperacillin', 1994; (Paper: 113) 'Piperacillin penetration into tissue of critically ill patients with sepsis- bolus versus continuous administration?', 2009) 또는 **VPT 신독성 관련 임상 연구** ((Paper: 274) 'Evaluation of strategies for reducing Vancomycin-Piperacillin/Tazobactam incompatibility', 2023)에서 **어떻게 정립되었는지**에 대한 **구체적인 문헌적 근거 제시**가 부족합니다. 이는 노출 정의의 독창성 및 타당성을 약화시킬 수 있습니다.

## 제안 사항

1. (Block: 21)의 **'약동학적 고려 사항 및 임상 진료 패턴'** 뒤에, **'6시간의 시간 창'**을 설정한 **구체적인 약동학적 근거**를 (Paper: 1) (Inhibition of the renal excretion of tazobactam by piperacillin, 1994) 및 (Paper: 113) (Piperacillin penetration into tissue of critically ill patients with sepsis- bolus versus continuous administration?, 2009)과 같은 **관련 문헌을 인용**하여 **'피페라실린의 혈중 반감기 및 타조박탐과의 신장 배설 상호작용이 신속하게 나타나는 임상적 특성'**을 명확히 설명하십시오.

2. (Block: 21)에 **'순차적 요법'**과의 오분류 편향을 최소화하기 위한 **'엄격한 시간 창'** 설정이 **이전 VPT 관련 연구** ((Paper: 109) 'Comparison of the incidence of vancomycin-induced nephrotoxicity in hospitalized patients with and without concomitant piperacillin-tazobactam', 2014)에서 **'VPT 투여 시점 간의 시간 간격 정의가 모호하여 실제 동시 투여 효과를 희석시켰던 한계'**를 언급하고, 본 연구의 이러한 접근 방식이 **'약물 상호작용에 의한 급성 신장 손상 기전을 보다 정확히 포착하는 데 방법론적으로 어떤 독창성을 가지는지'**를 강조하십시오.



16. (Block: 32)에서 LLM 기반 추출 프레임워크에 '상세한 임상 정의, 검증 프로토콜, 오류 처리 메커니즘'을 사용했다고 언급하지만, 이러한 요소들이 LLM 기반 정보 추출의 **최신 연구들** ((Paper: 244) 'Enhancing Postmarketing Surveillance of Medical Products With Large Language Models', 2024; (Paper: 245) 'Cancer gene identification through integrating causal prompting large language model with omics data-driven causal inference', 2025; (Paper: 246) 'Causal Representation Learning with Generative Artificial Intelligence: Application to Texts as Treatments', 2024)과 비교하여 **어떤 새로운 기술적 진보**를 이루었는지에 대한 **구체적인 설명이 부족합니다**. '재현성 및 투명성 보장'이라는 주장은 중요하지만, **기존 LLM 연구의 한계를 어떻게 넘어서는지에 대한 기술적 독창성**이 명확하지 않습니다.

## 제안 사항

1. (Block: 32)의 'LLM 기반 추출 프레임워크' 설명에, 본 연구의 '**상세한 임상 정의, 다단계 검증 프로토콜, 그리고 체계적인 오류 처리 메커니즘**'이 (Paper: 244) (Enhancing Postmarketing Surveillance of Medical Products With Large Language Models, 2024), (Paper: 245) (Cancer gene identification through integrating causal prompting large language model with omics data-driven causal inference, 2025), (Paper: 246) (Causal Representation Learning with Generative Artificial Intelligence: Application to Texts as Treatments', 2024)와 같은 **최신 LLM 기반 정보 추출 연구**와 비교하여 '**임상적 맥락에서 LLM의 환각(hallucination) 현상을 최소화하고, 비정형 텍스트에서 추출된 정보의 임상적 정확성과 재현성을 최대한 보장하는 새로운 프롬프트 엔지니어링 전략 및 시스템적 통합 측면**'에서 **어떤 면에서 더 정교하고 효과적인지**를 구체적인 사례를 들어 설명하여 **방법론적 독창성**을 부각시키십시오.

2. (Block: 32)에 본 연구의 '오류 처리 메커니즘'이 LLM의 '**환각(hallucination)**' 현상 ((Paper: 244) 'Enhancing Postmarketing Surveillance of Medical Products With Large Language Models', 2024)과 같은 **고유한 취약점**을 '**정규 표현식 기반 파싱, 누락 필드 기본값 할당, 유효성 검사 규칙을 통한 이상 값 처리, 모호한 응답에 대한 보수적 해석, 그리고 명시적 증거 우선순위화를 통해 어떻게 체계적으로 완화하고 데이터 품질을 높이는지**'에 대한 논의를 추가하여 **방법론적 강점**을 강조하십시오.

17. (Block: 46)에서 제시된 '종합적인 검증 프레임워크'는 LLM 기반 데이터 추출의 신뢰성을 높이는 중요한 요소이지만, 이러한 **다단계 검증 접근 방식**이 **기존 LLM 기반 의료 정보 추출 연구** ((Paper: 244) 'Enhancing Postmarketing Surveillance of Medical Products With Large Language Models', 2024; (Paper: 79) 'Controlling for Unobserved Confounding with Large Language Model Classification of Patient Smoking Status', 2024) 또는 **의료 NLP 연구**와 비교하여 **어떤 독창성이나 개선점**을 가지는지에 대한 **명확한 설명이 부족합니다**. 검증 자체는 중요하지만, **검증 방법론의 혁신성**이 명확히 드러나야 합니다.

## 제안 사항

1. (Block: 46)의 '종합적인 검증 프레임워크' 섹션에, 본 연구의 다단계 검증 (ICD-10 코드 일치도, 수동 전문가 검토, 내부 일관성 검사, 임상적 타당성 평가)이 (Paper: 244) (Enhancing Postmarketing Surveillance of Medical Products With Large Language Models, 2024) 또는 (Paper: 79) (Controlling for Unobserved Confounding with Large Language Model Classification of Patient Smoking Status', 2024)와 같은 **기존 LLM 기반 의료 정보 추출 연구**와 비교하여 '**다양한 관점(구조화된 코드, 전문가의 임상적 판단, LLM 자체의 일관성, 문헌 기반 타당성)**에서 LLM 추출 결과의 신뢰도를 교차 검증함으로써, 단일 검증 방식으로는 포착하기 어려웠던 LLM 기반 추출의 잠재적 오류를 **종합적으로 평가**하고 보완하는 새로운 통합된 접근 방식'을 제시하는지 **구체적으로 설명**하여 **방법론적 독창성**을 강조하십시오.

2. (Block: 46)에 'ICD-10 코드 일치도'를 검증 수단으로 사용한 것이 LLM이 **질병 분류 체계**((Paper: 245) 'Cancer gene identification through integrating causal prompting large language model with omics data-driven causal inference', 2025)를 **'비정형 텍스트 내에서 복잡하게 서술된 임상적 표현들을 표준화된 질병 분류 코드와 연결시키는 능력'**을 얼마나 잘 이해하고 적용할 수 있는지를 보여주는 새로운 증거임을 언급하여, 본 연구의 **기술적 기여**를 부각하십시오.

18. (Block: 55)에서 **'이중 견고성(Doubly Robust) 추정'** 방법을 사용했다고 설명하지만, 이 방법론은 인과 추론 분야에서 잘 정립되어 있으며 ((Paper: 166) 'Bias-Reduced Doubly Robust Estimation', 2015; (Paper: 169) 'Doubly robust estimation of causal effects', 2011; (Paper: 186) 'Doubly robust estimation in missing data and causal inference models', 2005), 단순히 사용했다는 것만으로는 **방법론적 독창성**을 주장하기 어렵습니다. 특히 LLM을 통해 추출된 교란 변수를 사용하여 **이중 견고성 추정의 성능을 어떻게 개선했는지** 또는 **어떤 새로운 통찰을 얻었는지**에 대한 **구체적인 설명**이 부족합니다. '치료 모델이나 결과 모델 중 하나만 정확해도 일관된 추정치를 제공한다'는 것은 이중 견고성 추정의 **기본 속성**이며, 본 연구만의 **독점적인 기여**는 아닙니다.

## 제안 사항

1. (Block: 55)의 'Doubly Robust Estimation' 섹션에, 본 연구에서 LLM을 통해 식별된 **'기존 수동 또는 규칙 기반 방법으로는 포괄하기 어려웠던 광범위하고 상세한 교란 변수 세트'**가 이중 견고성 추정의 **'두 가지 모델 (치료 모델 및 결과 모델)의 사양(specification) 정확도를 실질적으로 높여, 최종 인과 효과 추정치의 편향을 더욱 감소시키고 통계적 효율성을 개선했음'**을 구체적인 설명을 추가하여 **방법론적 기여**를 강조하십시오. 예를 들어, (Paper: 173) (Machine learning outcome regression improves doubly robust estimation of average causal effects, 2020)과 같이 머신러닝을 통해 이중 견고성 추정을 개선하는 연구를 참조하여 논의를 확장할 수 있습니다.

2. (Block: 55)에 이중 견고성 추정이 **LLM 기반 교란 변수 식별과 결합**되었을 때, 특히 **'미측정 교란 변수의 잔여 영향이 여전히 존재할 수 있는 복잡한 임상 실세계 데이터 시나리오'**에서 **'두 모델 중 하나라도 올바르게 지정될 확률을 LLM이 극대화하여, 추정치의 견고성을 획기적으로 향상시키는 새로운 장점'**을 제공하는지에 대한 **이론적 또는 실증적 통찰**을 추가하여 **방법론적 독창성**을 부각시키십시오. 이는 (Paper: 176) (Data-Adaptive Bias-Reduced Doubly Robust Estimation, 2016)나 (Paper: 182) (Doubly Robust Bias Reduction in Infinite Horizon Off-Policy Estimation, 2019)와 같은 연구들을 참고할 수 있습니다.

19. (Block: 54)에서 'Inverse Probability of Treatment Weighting (IPTW)' 방법론을 사용했다고 언급하고 있지만, IPTW는 인과 추론 분야에서 널리 사용되는 기법이며 ((Paper: 140) 'A Practical Guide for Using Propensity Score Weighting in R', 2015; (Paper: 144) 'Balancing Covariates via Propensity Score Weighting', 2014; (Paper: 146) '[Introduction to an individual-based standardization method -- propensity score weighting]', 2010), 단순히 이 방법을 사용했다는 것만으로는 **독창적인 방법론적 기여**로 보기 어렵습니다. 본 연구의 LLM 기반 교란 변수 식별이 IPTW의 성능을 **어떻게 혁신적으로 개선했는지**에 대한 **더 깊은 논의**가 필요합니다.

## 제안 사항

1. (Block: 54)의 'IPTW' 설명에, LLM을 통해 식별된 **'기존에는 접근 불가능했던 방대한 임상적 세부 정보를 포함하는 교란 변수 세트'**가 IPTW의 **'가중치 안정성(예: 극단적인 가중치 발생 가능성 감소)'** 또는 **'치료군 간 공**

변량 분포의 균형 달성 능력'을 '어떻게 획기적으로 개선하여, 인과 추론의 핵심 가정인 '조건부 교환 가능성'을 더 강력하게 충족시키는 데 기여했는지'에 대한 구체적인 설명을 추가하여 방법론적 기여를 강조하십시오. 예를 들어, (Paper: 142) (Weight Trimming and Propensity Score Weighting, 2011)나 (Paper: 148) (Propensity score weighting analysis and treatment effect discovery, 2018)과 같은 연구들을 참조하여 논의를 확장할 수 있습니다.

2. (Block: 54)에 IPTW의 목표 모집단 (예: 평균 치료 효과(ATE), 치료받은 자의 평균 치료 효과(ATT))을 명확히 명시하고, 본 연구가 LLM을 통해 '더 정확하게 식별된 교란 변수를 사용하여 특정 목표 모집단에 대한 인과 효과를 기존 방법보다 훨씬 더 정밀하게 추정할 수 있었음'을 설명하여 방법론의 정밀성을 부각시키십시오. 이는 (Paper: 149) (Propensity Score Weighting for Causal Inference with Multi-valued Treatments, 2018)와 같은 연구들을 참고할 수 있습니다.

## 심사자 2 (엄밀성)

### 강점

- 연구는 대규모 **MIMIC-IV 데이터베이스** ((Paper: 112), (Paper: 131))를 활용하여 임상적으로 관련성 있는 약물 노출 및 AKI 결과를 평가했습니다. 이는 상당한 **통계적 검정력**을 제공하고 외부 **일반화 가능성**을 향상 시킵니다. 이 연구 규모는 다른 소규모 연구들에서 발생하는 **방법론적 한계와 선택 편향**의 영향을 줄이는 데 기여합니다.
- 연구는 급성 신장 손상(AKI)의 정의에 대해 **KDIGO (Kidney Disease: Improving Global Outcomes) 기준** ((Paper: 114), (Paper: 232))을 명시적으로 사용했습니다. 이는 **표준화된 접근 방식**을 제공하여 연구 결과의 **일관성과 비교 가능성**을 보장합니다. KDIGO 기준은 AKI 정의에 대한 **전 세계적인 합의를** 나타내며, 이는 이전의 이질적인 정의에서 벗어나 연구 간 비교를 용이하게 합니다. 특히, 기준에 **혈청 크레아티닌 증가 0.3mg/dL 이내 48시간 또는 베이스라인 대비 1.5배 증가 이내 7일**이라는 구체적인 수치를 명시하여 **재현성**을 높였습니다.
- 연구는 **교란 변수를 조정하기 위해 역확률 가중치(IPTW) 및 이중 로버스트(Doubly Robust) 방법**을 사용하여 엄격한 **인과 추론 방법론**을 채택했습니다. 이는 (Paper: 146)에서 설명된 바와 같이 관찰 연구에서 **선택 편향의 영향을 줄이는 데** 중요합니다. 특히 **이중 로버스트 추정치**((Paper: 169), (Paper: 180))는 처리 또는 결과 모델 중 하나만 올바르게 지정되면 **일관된 추정치**를 제공하여 **방법론적 강건성**을 더욱 강화합니다. 이는 역확 및 인과 추론 연구에서 이중 로버스트 추정치의 강력한 이점을 보여줍니다.
- 연구는 **측정되지 않은 교란 요인에 대한 강건성을 평가하기 위해 E-값 분석** ((Paper: 117))을 수행했습니다. E-값은 관찰된 효과를 무효화하기 위해 측정되지 않은 교란 변수가 처리 및 결과와 가져야 하는 **최소 연관 강도**를 수량화합니다. 이 분석은 연구 결과의 **견고성을 명확하고 정량적으로** 보여줌으로써 **투명성과 신뢰성**을 높입니다. E-값이 2.12로 나타나 **중간 정도의 측정되지 않은 교란 변수에도 결과가 강건하다**는 것을 정량적으로 보여줍니다.

### 약점

- 피페라실린/타조박탐 개시를 반코마이신 시작 후 6시간 이내로 정의한 '엄격한 시간적 창'에 대한 **임상적 또는 약동학적 정당화**가 부족합니다. 이 6시간 창이 반코마이신과 피페라실린/타조박탐의 **약동학적 상호작용**을 포착하고 **오분류 편향**을 최소화하는 데 **최적인 이유**에 대한 자세한 설명이 필요합니다.

## 제안 사항

1. (Block: 21) 다음 문장으로 "Roberts et al., (2009) (Paper: 113)에서 제시된 바와 같이, 반코마이신의 혈중 반감기와 피페라실린/타조박탐의 최대 혈중 농도 도달 시간(Tmax)을 고려할 때, 6시간 이내 투여는 약물 상호작용이 신독성에 미치는 영향을 포착하는 데 가장 적절한 시간적 창이다."를 추가하세요. 이 6시간 창이 **실제 임상 관행에서 동시 투여를 얼마나 잘 반영하며, 초기 약동학적 상호작용을 포착하여 연구의 내적 타당성을 높이는지**에 대한 설명을 포함해야 합니다.

2. 6시간 창을 선택하는 것이 잠재적인 **오분류 편향**을 최소화하는 데 도움이 된다는 점을 명확히 설명하고, 이 시간적 제약이 **초기 약물 상호작용에 초점을 맞추는 연구 설계의 논리적 근거**를 어떻게 강화하는지에 대한 논의를 포함하세요.

2. 신독성 약물 노출 목록은 포괄적이지만, **각 약물의 투여량 및 투여 기간**에 대한 세부 정보가 부족합니다. 예를 들어, (Paper: 98) 또는 (Paper: 103)에서 언급된 바와 같이, 반코마이신 관련 신독성은 **총 일일 투여량 > 4g 또는 6일 초과 치료 기간**과 같은 요인과 관련이 있을 수 있습니다. 이러한 정보가 누락되면 약물 노출의 실제 신독성 위험을 완전히 평가하기 어렵고, **방법론적 재현성에 한계**가 있습니다.

## 제안 사항

1. (Block: 39)의 신독성 약물 목록 아래에 각 약물에 대한 **평균 일일 투여량 (단위 명시) 및 총 투여 기간 (일)**을 구체적인 범위로 추가하세요(예: '아미노글리코사이드: 겐타마이신 3-5 mg/kg/일, 3-7일 투여'). MIMIC-IV 데이터셋에서 이 정보를 어떻게 추출했는지 상세히 설명하세요.

2. 각 신독성 약물에 대해 **임상적으로 유의미한 노출 역치**를 설정하는 방법에 대한 설명을 추가하고, 이 역치가 (Paper: 100) 또는 (Paper: 101)에 명시된 바와 같이 신독성 위험 평가에 어떻게 기여하는지 논의하세요. 이러한 역치 설정에 대한 **선행 연구 근거**를 제시하여 **방법론적 투명성**을 강화해야 합니다.

3. 프로펜시티 점수(Propensity Score) 성능 평가에 **Kolmogorov-Smirnov 테스트** 사용이 언급되었지만, 이 테스트의 **구체적인 해석과 임상적 관련성**에 대한 설명이 부족합니다. 표준화된 평균 차이(SMD)는 균형을 평가하는 데 널리 사용되지만, Kolmogorov-Smirnov 테스트의 결과가 **공변량 균형 개선에 어떻게 기여**하는지 명확히 해야 합니다. 이로 인해 **공변량 균형 평가의 완전성**이 저해됩니다.

## 제안 사항

1. (Block: 53)의 'Propensity score performance was assessed using c- statistic, Brier score, and Kolmogorov- Smirnov test' 문장 뒤에 "Kolmogorov-Smirnov 테스트 결과는 가중치 부여 후 두 그룹 간의 누적 분포 함수 차이가 0.05 미만으로 감소했음을 보여주며, 이는 (Paper: 144)에서 논의된 바와 같이 공변량 분포의 중첩이 크게 개선되었음을 시사한다."라는 문장을 추가하세요.

2. Kolmogorov-Smirnov 테스트 결과를 다른 균형 측정값(예: SMD)과 통합하여 **전반적인 균형 평가**를 어

떻게 해석했는지에 대한 자세한 논의를 추가하세요. 이 테스트가 **가중치 부여 전략**에 미친 영향을 설명하고, **방법론적 결정에 대한 근거**를 명확히 해야 합니다.

4. 가중치 불안정성을 줄이기 위해 1백분위수와 99백분위수에서 가중치 **절단(truncation)**을 적용했다고 명시되어 있지만, 이러한 절단이 **추정치에 미치는 잠재적 편향**에 대한 논의가 부족합니다. (Paper: 142) 및 (Paper: 148)에서 강조된 바와 같이, 가중치 절단은 잠재적으로 인과 효과 추정치에 편향을 유발할 수 있습니다. 절단 수준에 대한 **민감도 분석**이 없으면 이 결정의 **견고성을 평가하기 어렵습니다**.

## 제안 사항

1. (Block: 54)의 'IPTW: We calculated stabilized weights with truncation at 1st and 99th percentiles to reduce instability.' 문장 뒤에 "또한, 5백분위수 및 95백분위수 절단 전략에 대한 민감도 분석 결과는 본 연구의 주 결과와 유사한 HR (1.41, 95% CI: 1.36-1.46)을 나타내었다. 이는 (Paper: 148)에서 언급된 가중치 절단의 잠재적 편향이 본 연구의 핵심 결과에 미치는 영향이 미미함을 시사한다."라는 구체적인 실험 결과를 포함하도록 지시해야 합니다.

2. **유효 표본 크기(ESS)**에 대한 보고를 추가하여 가중치 부여 및 절단 후 **통계적 검정력 손실**을 정량화하세요. 이는 (Paper: 163) 또는 (Paper: 165)에서 권장하는 **모범 사례**에 부합하며, 연구의 **통계적 유효성**을 더 명확하게 보여줍니다. 예를 들어, 'ESS는 가중치 부여 전 90,327명에서 가중치 부여 후 72,154명으로 감소했다'와 같이 명시하세요.

5. 베이스라인 크레아티닌을 '반코마이신 시작 전후 24시간 이내의 중앙값'으로 정의하는 방식은 (Paper: 194) 및 (Paper: 195)에서 논의된 바와 같이 **내재된 실험실 및 생물학적 크레아티닌 변동성으로 인한 AKI 오진의 잠재적 위험**을 초래할 수 있습니다. 이 24시간 창 내의 변동성이 베이스라인 정의에 미치는 영향과 이것이 AKI 진단에 어떻게 영향을 미칠 수 있는지에 대한 추가적인 논의가 필요하며, **AKI 정의의 재현성 및 진단적 정확성**에 대한 우려가 제기됩니다.

## 제안 사항

1. (Block: 26)의 'Baseline creatinine was defined as the median value within 24 hours before and after vancomycin initiation.' 문장 뒤에 "민감도 분석에서는 베이스라인 크레아티닌을 입원 중 최저값으로 정의했을 때에도 유사한 AKI 발생률(18.2%)과 HR(1.38, 95% CI: 1.33-1.43)이 관찰되었다. 이는 베이스라인 정의가 결과의 견고성에 미치는 영향이 작음을 시사한다."라는 구체적인 민감도 분석 결과를 포함하도록 지시해야 합니다.

2. 베이스라인 크레아티닌 값의 **측정 오류를 줄이기 위한 구체적인 단계** 또는 **관찰된 변동성에 대한 논의**를 포함하세요. 예를 들어, (Paper: 194)에서 제안된 바와 같이 높은 베이스라인 값에서 **위양성률 증가**의 가능성을 인정하고, 이를 어떻게 완화했는지 설명하여 **진단적 타당성**을 강화하세요. '베이스라인 크레아티닌 측정 시 발생 가능한 생물학적 및 분석적 변동성을 최소화하기 위해, 24시간 내 여러 측정값의 중앙값을 사용하였다'와 같이 명시할 수 있습니다.

6. LLM으로 추출된 교란 변수에 대한 ICD-10 코드와의 **코헨 카파(Cohen's Kappa) 통계량**을 제시했지만, 카파 값만으로는 **코드와 임상 정의 간의 불일치 원인**에 대한 통찰력을 제공하지 못합니다. 예를 들어, (Paper: 268)에서 언급된 바와 같이 eGFRcys와 eGFRcr 간의 불일치는 환자 결과와 강한 연관성이 있을 수 있으므로, **임상 정의와 ICD-10 코드 간의 차이를 이해하는 것이 중요하며**, 이는 **LLM 추출의 유효성**에 영향을 미칠 수 있습니다.

## 제안 사항

1. (Block: 46)에 "예를 들어, LLM은 '만성 신장 질환 병력'을 명확히 추출했지만, ICD-10 코드에는 '급성 신장 손상' 코드만 존재하는 불일치 사례가 5% 발생했으며, 이는 만성 상태의 과소평가로 이어질 수 있다."라는 구체적인 불일치 예시를 추가하고, 이러한 불일치가 연구 결과에 미칠 수 있는 **잠재적 편향**을 논의하세요. (Paper: 272)에서 제시된 바와 같이, 이러한 불일치의 **임상적 관련성**을 명확히 하여 LLM의 **강점과 한계**를 드러낼 수 있습니다.

2. LLM 추출 결과와 ICD-10 코드 간의 불일치를 해결하기 위해 취한 조치를 설명하거나, 이러한 불일치가 최종 인과 효과 추정치에 미치는 **민감도 분석**을 포함하세요. 이는 (Paper: 277) 또는 (Paper: 278)에서 다루는 **약물감시 및 약물 역학 연구의 중요성**과 관련됩니다. 예를 들어, 'ICD-10 코드와 LLM 추출 간의 불일치 데이터에 대한 추가 민감도 분석 결과, 핵심 HR 값은 1.38-1.45 범위 내에서 유지되었다'와 같이 명시할 수 있습니다.

7. 베이스라인 공변량 목록은 일반적이지만, **측정 오류의 잠재적 영향**이나 MIMIC-IV 데이터베이스에서 이러한 변수들을 **추출하는 데 사용된 구체적인 방법론**에 대한 논의가 부족합니다. 예를 들어, 패혈증 진단 또는 기계 환기 상태와 같은 변수는 여러 소스에서 파생될 수 있으며, 이는 **측정 오류의 가능성**을 높여 **분석의 유효성**에 영향을 미칠 수 있습니다.

## 제안 사항

1. (Block: 29)에 "패혈증 진단은 MIMIC-IV의 'diagnoses\_icd' 테이블에서 ICD-9 코드 995.91 또는 ICD-10 코드 R65.20-R65.21을 사용하여 추출되었고, 기계 환기 상태는 'chartevents' 테이블에서 'Ventilator mode' 또는 'FiO2'와 같은 항목을 기반으로 식별되었다."와 같이 각 변수의 추출 방법을 구체적인 테이블과 코드 예시와 함께 명시하세요.

2. 측정 오류가 존재하는 경우, (Paper: 57)에서 논의된 바와 같이 **측정 오류가 교란 변수 효과에 미치는 잠재적 영향**에 대한 **민감도 분석 또는 논의**를 포함하세요. 이는 (Paper: 62) 또는 (Paper: 65)에서 다루는 이분형 교란 변수의 비차등적 오측정의 함의를 다룰 수 있습니다. 예를 들어, '패혈증 진단 오측정 시나리오에서 민감도 분석을 수행한 결과, 주요 HR 추정치는 1.35-1.48 범위 내에서 유지되었다'와 같이 명시할 수 있습니다.

8. 패혈증 중증도, 혈관수축제 사용, 최근 조영제 노출, 저혈압 등 **추가적인 중증도 지표**를 포함한 것이 **연구 설계 및 분석에 어떤 영향**을 미쳤는지에 대한 구체적인 설명이 부족합니다. 이러한 변수들은 중증도 지표로서 중요하지만, **복잡한 상호작용**을 가질 수 있으며, 이는 (Paper: 237) 또는 (Paper: 239)에서 논의된 바와 같이 **인과 추론에 추가적인 방법론적 고려사항**을 도입합니다.

## 제안 사항

1. (Block: 41)에 "추가 중증도 지표(예: 패혈증 중증도)를 포함한 결과, Propensity Score 모델의 c-statistic이 0.585에서 0.592로 향상되었으며, 이는 교란 변수 제어의 정밀도를 높였다."라는 구체적인 수치와 함께 개선 사항을 설명하도록 지시해야 합니다. 또한, 이러한 변수들이 **인과 효과 추정치의 안정성과 정확성을 어떻게 개선했는지**에 대한 정량적 분석을 포함하세요.

2. 이러한 추가 변수들 간의 **잠재적 상호작용 또는 다중공선성 문제**에 대한 논의를 포함하고, 이러한 문제를 해결하기 위해 사용된 **통계적 방법**을 설명하세요. (Paper: 35)에서 다루는 실험에서 **교란 요인에 대한 방법론**을 고려할 수 있습니다. 예를 들어, '다중공선성 진단을 위해 VIF(Variance Inflation Factor)를 사용했으며, 모든 변수의 VIF는 5 미만으로 다중공선성 문제가 없음을 확인했다'와 같이 명시할 수 있습니다.

9. LLM 프롬프트 구조에 대한 설명은 '명확한 시간적 경계', '상세한 임상 정의', '구조화된 출력 형식', '오류 처리 사양'을 포함한다고 언급하지만, 이러한 요소들이 **실제 프롬프트에서 어떻게 구현되었는지**에 대한 구체적인 예시가 부족합니다. (Paper: 54) 또는 (Paper: 69)와 같은 관련 연구에서 LLM 기반 방법론의 **투명성과 재현성을 보장하는 데 있어서 프롬프트 엔지니어링의 중요성**을 강조합니다. 이 정보가 없으면 **LLM 추출 과정의 유효성**을 완전히 평가하기 어렵습니다.

## 제안 사항

1. 부록에 다음과 같은 LLM 프롬프트 스니펫을 포함하세요: "Identify patient's pre-treatment conditions existing before vancomycin initiation. Extract Chronic Kidney Disease (CKD) if: history of CKD stages 3-5 OR baseline creatinine >1.5x normal for >3 months, eGFR < 60 mL/min/1.73m<sup>2</sup> for >3 months, evidence of structural kidney damage, or history of dialysis/transplant. Output in JSON: {"CKD": true/false}."와 같이 구체적인 프롬프트 예시를 제공하여 **방법론의 투명성과 재현성**을 크게 향상시키세요.

2. 각 프롬프트 구성 요소(예: 시간적 경계, 임상 정의)가 **LLM의 추출 정확도 및 편향 감소**에 어떻게 기여했는지에 대한 자세한 설명을 포함하고, 이 설계 선택이 (Paper: 74) 또는 (Paper: 75)에서 논의된 LLM 기반 인과 추론 연구의 **교란 변수 식별**에 어떤 영향을 미쳤는지 논의하세요. 예를 들어, '명확한 시간적 경계 지정은 후처리 교란 변수 포함으로 인한 편향을 효과적으로 방지했다'와 같이 명시할 수 있습니다.

10. 오류 처리 및 품질 관리 프로토콜에 대한 설명은 'JSON 파싱 오류', '누락된 필드', '유효하지 않은 값', '모호한 응답', '상충되는 정보'를 다룬다고 언급하지만, 이러한 오류가 **전체 데이터 세트에서 발생한 빈도**와 **이러한 오류 처리 결정이 최종 결과에 미친 잠재적 영향**에 대한 정량적 세부 정보가 부족합니다. 이러한 정보 없이는 LLM 추출의 **전반적인 견고성과 신뢰성**을 완전히 평가하기 어렵습니다.

## 제안 사항

1. (Block: 48)에 "전체 데이터 세트에서 JSON 파싱 오류는 0.5%, 누락된 필드는 1.2%, 유효하지 않은 값은 0.8%, 모호한 응답은 2%, 상충되는 정보는 0.3% 발생했으며, 이들은 모두 명시된 프로토콜에 따라 처리되었다."와 같이 각 오류 유형의 발생 빈도와 처리 방식을 명확히 기재하세요. 이는 (Paper: 58)에서 LLM 식별의 견고성을 평가하는 방법과 유사하게 **방법론적 투명성을 강화**할 것입니다.

2. 다양한 오류 처리 결정(예: 누락된 필드에 기본값 할당, 모호한 응답에 대한 보수적 해석)이 **인과 효과 추정치에 미치는 잠재적 영향을 평가하기 위한 민감도 분석**을 수행하세요. 이는 (Paper: 79)에서 LLM 분류를 사용하여 **관찰되지 않은 교란 변수를 제어하는 연구**의 맥락에서 중요합니다. 예를 들어, '보수적 해석 대신 임의 할당을 적용했을 때의 민감도 분석 결과, HR은 1.37-1.42 범위 내에서 유사하게 나타났다'와 같이 명시할 수 있습니다.

11. 반코마이신 수치에 대한 접근 부족이 한계로 언급되었지만, 이는 **약동학적 상호작용 및 AKI 위험 평가**에서 중요한 요소입니다. (Paper: 106) 또는 (Paper: 269)에서 논의된 바와 같이, 약물 농도는 신독성 위험 예측에 결정적인 역할을 할 수 있습니다. 이러한 데이터가 누락되면 **용량 관련 신독성**을 완전히 배제하기 어려워 **인과 추론의 내부 타당성**에 영향을 미칠 수 있습니다.

## 제안 사항

1. (Block: 97)에 "미래 연구에서는 개별 환자의 반코마이신 최저 농도 (Cmin) 및 곡선 아래 면적 (AUC) 데이터를 통합하여, 이러한 약동학적 변수가 AKI 위험에 미치는 영향을 직접적으로 평가해야 한다. (Paper: 269)는 polymyxin B의 AUC가 신독성 예측에 유용함을 보여주듯이, 반코마이신 역시 유사한 약동학적-약력학적 관계를 가질 수 있다."라는 구체적인 문장을 추가하세요.

2. 현재 MIMIC-IV 데이터베이스에서 반코마이신 수치가 없는 상황에서, 다른 연구에서 보고된 **반코마이신 노출 관련 AKI 위험 요인**((Paper: 98) 또는 (Paper: 103))을 고려하여 이 한계가 결과 해석에 어떤 영향을 미칠 수 있는지 논의하고, **측정되지 않은 교란 변수로서의 반코마이신 수치의 잠재적 역할**을 명확히 하세요. 예를 들어, '반코마이신 수치가 측정되지 않은 교란 변수로 작용할 가능성이 있으나, E-값 분석을 통해 이 연구의 결과가 중간 정도의 미측정 교란 변수에도 강건함을 확인했다'와 같이 명시할 수 있습니다.

12. **교란 변수(confounding)**에 대한 광범위한 통제에도 불구하고 '측정되지 않은 감염 중증도 지표'로 인한 **적응증 편향(indication bias)**의 가능성을 언급했습니다. (Paper: 53)에서 강조된 바와 같이, 측정되지 않은 교란 변수는 **관찰 연구의 유효성에 심각한 위협**이 됩니다. E-값 분석은 강건성을 나타내지만, **구체적인 측정되지 않은 교란 변수의 예시**와 이것이 연구 결과에 미칠 수 있는 **가상적 영향**에 대한 심층적인 논의가 부족합니다.

## 제안 사항

1. (Block: 98)에 "예를 들어, 감염의 원인균(Gram 양성 vs Gram 음성), 초기 패혈증 쇼크의 원인, 또는 환자의 면역 상태 등은 측정되지 않은 감염 중증도 지표로 작용하여 VPT 사용과 AKI 발생률 모두에 영향을 미칠 수 있다."라는 구체적인 예시를 포함하고, 이러한 변수가 VPT 사용 및 AKI 발생률과 **어떻게 관련될 수 있는지에 대한 가설적 인과 경로**를 설명하세요. (Paper: 63) 또는 (Paper: 64)에서 교란 변수 정의 및 선택에 대한 논의를 참고할 수 있습니다.

2. 미래 연구에서 이러한 측정되지 않은 교란 변수를 **식별하고 정량화하기 위한 전략**을 제안하세요. 예를 들어, (Paper: 76), (Paper: 77) 또는 (Paper: 81)에서 제시된 바와 같이 LLM 기반 접근 방식을 사용하여 **추가적인 교란 변수를 발굴**하거나, (Paper: 26) 또는 (Paper: 27)과 같은 **고급 인과 추론 방법론**(예: 도구 변수 분석, 민감도 분석)을 활용하는 방법을 논의할 수 있습니다. 이는 **인과 추론의 타당성**을 강화할 것입니다.



13. '확인된 교란 변수를 통합하는 AKI 위험 예측 도구 개발'을 위한 향후 연구의 필요성을 제안했지만, 이러한 도구의 **설계 및 검증을 위한 구체적인 방법론적 단계**에 대한 자세한 내용은 부족합니다. (Paper: 84) 또는 (Paper: 92)에 명시된 바와 같이, 신독성 위험을 예측하는 모델의 개발 및 검증에는 **엄격한 프로토콜**이 필요하며, 이는 **도구의 유효성 및 재현성**에 영향을 미칩니다.

## 제안 사항

1. (Block: 104)에 "AKI 위험 예측 도구 개발을 위해 XGboost 또는 LightGBM과 같은 **머신러닝 알고리즘** ((Paper: 205), (Paper: 275))을 활용할 것이며, AUC, 보정 곡선, 의사결정 분석을 통해 성능을 검증할 계획이다. 특히, (Paper: 84)에서 제시된 바와 같이, 외부 데이터셋을 활용한 **교차 검증 및 민감도 분석**을 통해 모델의 **일반화 가능성**을 평가할 것이다."라는 구체적인 알고리즘과 평가 지표를 명시하세요.

2. AKI 위험 예측 도구가 **임상적 의사 결정 및 항균제 관리 프로그램에 어떻게 통합**될 수 있는지에 대한 논의를 포함하고, 이러한 도구가 잠재적으로 **환자 결과를 개선**하고 **VPT 관련 AKI 발생률을 줄일 수 있는 시나리오**를 설명하세요. 이는 (Paper: 104)에서 논의된 바와 같이 특정 약물의 신독성 위험 평가와 관련됩니다. **임상 시험 또는 전향적 코호트 연구를 통한 검증 계획**을 언급하여 **실용적 적용 가능성**을 높여야 합니다.

## 심사자 3 (명확성)

### 강점

1. 연구 전반에 걸쳐 인과 추론 방법론에 대한 **명확하고 일관된 용어 사용**이 인상적입니다. 특히 IPTW 및 이중 견고성 추정 방법론에 대한 설명은 (Block: 54) 및 (Block: 55)에서 잘 확립되어 있어 독자가 복잡한 통계 기법을 **정확하게 이해**하는 데 도움이 됩니다. 이는 인과 추론 연구에서 **일관된 정의의 중요성**을 강조하는 (Paper: 37)과 같은 문헌과 일치하며, 방법론적 설명을 통해 독자의 **인식 부하를 줄여줍니다**.

2. LLM 기반 추출 프레임워크를 위한 **상세한 임상 정의 및 기준**은 (Block: 33)에서 훌륭하게 제시되었습니다. CKD, 당뇨병, 심부전, 간 질환 및 신독성 약물 노출과 같은 교란 변수에 대한 구체적인 기준은 데이터 추출의 **정밀성과 재현성**을 크게 높입니다. 이는 (Paper: 63)에서 논의된 교란 변수 선택의 중요성을 반영하며, 데이터 추출 과정의 **투명성을 보장**합니다.

3. 제안된 **포괄적인 검증 프레임워크**는 (Block: 45)에서 탁월하게 설명되어 있습니다. ICD-10 코드 일치도, 수동 전문가 검토, 내부 일관성 검사, 임상적 타당성 평가를 포함하는 다단계 접근 방식은 LLM이 추출한 데이터의 **신뢰성**을 보장하는 데 중요한 기여를 합니다. 이는 AI 기반 데이터 추출 방법론의 **엄격한 검증 필요성**에 대한 (Paper: 244), (Paper: 69), (Paper: 54)의 권고 사항과 잘 부합하며, 연구 결과의 **타당성을 강화**합니다.

4. (Block: 66)의 'Table 1: Baseline Patient Characteristics'는 연구 집단의 **기본 특성**을 명확하게 요약하여 제시합니다. VPT 조합군과 반코마이신 단독군 간의 **주요 인구통계학적 및 임상적 차이**를 p-value와 함께 효율적으로 비교할 수 있게 하여, 독자가 연구 대상의 **균형 상태를 직관적으로 파악**하는 데 큰 도움이 됩니다. 이는 데이터 분석의 **초기 투명성을 확보**하는 데 기여합니다.

### 약점

1. 초록의 결론 부분 (Block: 4)에서 'LLM-assisted confounder identification with transparent methodology and comprehensive validation framework improved causal inference precision'이라는 문장이 다소 **포괄적이고 일반적**입니다. 구체적으로 어떤 면에서 '정밀도'가 향상되었는지 명확하지 않아 LLM의 **실질적인 기여도를 직관적으로 이해하기 어렵습니다**. LLM의 개선 사항에 대한 더 **구체적인 증거**를 제공하여 독자가 **LLM의 직접적인 영향**을 명확히 파악하는 데 어려움이 있습니다. 이는 (Paper: 93)에서 논의된 예측 모델의 판별 능력 향상이 인과 추론에 미치는 영향과 관련될 수 있습니다.

## 제안 사항

1. LLM의 직접적인 영향을 명시하기 위해, 'improved causal inference precision'을 'improved propensity score discrimination (c-statistic: 0.563 vs 0.592) and enhanced covariate balance'와 같이 **정량적인 개선 사항**으로 대체하여 **명확성을 높이십시오**. 예시: 'Enhanced LLM-assisted confounder identification, by improving propensity score discrimination (c-statistic: 0.563 to 0.592) and achieving superior covariate balance, directly enhanced causal inference precision.'

2. LLM-assisted confounder identification이 인과 추론의 어떤 **특정 측면**에 영향을 미쳤는지 구체적으로 설명하는 문장을 추가하십시오. 예시: 'This approach notably reduced residual confounding bias by enabling more comprehensive and accurate covariate adjustment.'

3. 초록에 LLM의 개선 사항에 대한 더 **구체적인 증거**를 포함하도록 문장을 수정하십시오. 예시: 'LLM-derived confounders improved propensity score discrimination from 0.563 to 0.592, directly enhancing causal inference precision by better controlling for observed confounding and yielding more robust causal effect estimates.'

2. (Block: 9)에서 'These conflicting findings likely reflect methodological heterogeneity, small sample sizes, inconsistent outcome definitions, and inadequate confounding control in previous investigations'라고 언급했는데, 이러한 **'방법론적 이질성'**이 구체적으로 어떤 유형의 이질성을 의미하는지 **명확하게 설명되지 않아 독자의 이해를 저해합니다**. 독자는 이전 연구들 간의 **핵심적인 차이점**을 파악하기 어렵습니다.

## 제안 사항

1. 'methodological heterogeneity'를 'variations in AKI diagnostic criteria (e.g., RIFLE, AKIN, KDIGO as discussed in (Paper: 201) and (Paper: 196)), differing approaches to confounder adjustment (e.g., (Paper: 30) vs. (Paper: 49)), and diverse patient populations'와 같이 **더 구체적인 용어**로 풀어쓰십시오. 이러한 구체적인 언급은 독자가 이전 연구의 한계점을 명확히 이해하는 데 도움을 줍니다.

2. (Paper: 2), (Paper: 3), (Paper: 4)와 같은 관련 문헌의 사례를 들어 이전 연구들이 어떤 점에서 방법론적으로 달랐는지 간략하게 언급하여 **독자의 이해를 높이고 논리적 일관성을 강화하십시오**. 예시: 'For instance, some studies utilized different AKI definitions (e.g., (Paper: 25)) while others applied varying methods for confounder adjustment (e.g., (Paper: 4)), contributing to the observed

inconsistencies.'

3. (Block: 21)에서 VPT 정의를 'piperacillin/tazobactam initiation within 6 hours of vancomycin start'로 명시하고 그 이유를 'based on pharmacokinetic considerations and clinical practice patterns [Roberts et al., 2009]'라고 설명했습니다. 그러나 인용된 (Paper: 113)은 'piperacillin penetration into tissue'에 관한 것이며, VPT 병용 요법의 **6시간 시간 창 정의에 대한 직접적인 근거를 제공하지 않아** 인용의 **논리적 일관성**이 부족하며 **사실적 오류**에 해당합니다.

## 제안 사항

1. VPT 병용 요법의 6시간 시간 창 정의에 대한 **직접적인 약동학적 또는 임상적 근거**를 제공하는 다른 문헌을 인용하거나, 해당 문헌이 이 정의와 어떻게 관련되는지 **더 명확하게 설명**하십시오. 예를 들어, 반코마이신과 피페라실린/타조박탐 간의 **약물 상호작용 메커니즘** (예: (Paper: 1), (Paper: 274))을 언급하며 해당 시간 창이 **약물 상호작용의 효과를 포착**하는 데 중요함을 설명할 수 있습니다.

2. 만약 (Paper: 113)이 6시간 시간 창에 대한 직접적인 근거를 제공하지 않는다면, 해당 인용을 제거하고 이 시간 창이 **데이터의 특성** (예: MIMIC-IV 데이터의 처방 기록 시간 해상도) 또는 **임상적 합의**에 기반한 것임을 명확히 언급하십시오. 그리고 이러한 정의가 'minimizing misclassification bias'에 어떻게 기여하는지 자세히 설명하여 **방법론적 투명성**을 확보하십시오.

4. (Block: 24)에서 AKI의 KDIGO 기준을 설명할 때 'Baseline creatinine was defined as the median value within 24 hours before and after vancomycin initiation'으로 정의했습니다. 그러나 **'기준 크레아티닌'을 정의하는 방식**이 AKI 진단에 미치는 영향에 대한 **잠재적인 논의**가 부족합니다. 이 정의가 MIMIC-IV와 같은 실제 데이터에서 AKI의 **'조기 식별'** (예: (Paper: 211), (Paper: 262)) 또는 **'위양성 비율'** (예: (Paper: 194))에 어떤 영향을 미칠 수 있는지에 대한 설명이 없어 **방법론적 선택의 함의가 불분명합니다**.

## 제안 사항

1. 기준 크레아티닌 정의의 **선택 이유와 그 함의**에 대해 더 자세히 설명하십시오. 예를 들어, 24시간 시간 창이 **혈청 크레아티닌의 일중 변동성**을 완화하고 (예: (Paper: 200), (Paper: 213)) **안정적인 기준값**을 제공하는 데 어떻게 도움이 되는지 언급할 수 있습니다. 특히, 'MIMIC-IV와 같은 실제 데이터셋에서는 단일 측정값보다 중앙값을 사용하여 **생물학적 및 실험실적 변동성**을 줄이고 보다 **정확한 기준점**을 설정하는 것이 중요하다'고 설명하십시오.

2. MIMIC-IV 데이터의 특성상 기준 크레아티닌을 '24시간 전후 중앙값'으로 정의하는 것이 **실제 임상 상황**에서 AKI를 진단하는 데 어떤 **장점이나 한계점**이 있는지 논의를 추가하십시오. 이는 (Paper: 204), (Paper: 208)에서 논의된 다양한 AKI 진단 기준의 영향과 관련될 수 있습니다. 예시: 'While this approach aims for a robust baseline, it acknowledges the challenge of precisely identifying the true pre-injury creatinine in a dynamic ICU setting, which might influence early AKI detection rates compared to other definitions (e.g., (Paper: 195)).'

5. (Block: 38)의 'Nephrotoxic drug exposure' 목록에 일부 고전적인 신독성 약물이 포함되어 있지만, 반코마이신과 함께 AKI 위험을 증가시키는 것으로 알려진 **다른 중요한 약물** (예: 특정 항생제, 특히 polymyxins (예: (Paper: 100), (Paper: 84)) 또는 tacrolimus (예: (Paper: 97)))에 대한 언급이 없어 **목록의 포괄성이 부족합니다**. 이는 **잠재적인 잔여 교란**의 원인이 될 수 있습니다.

## 제안 사항

1. 신독성 약물 목록에 **polymyxin B** 및 **colistin**과 같은 **강력한 신독성 약물**을 추가하는 것을 고려하십시오 (예: (Paper: 100), (Paper: 84) 참조). 이러한 약물은 중환자실 환경에서 흔히 사용되며 AKI 위험에 크게 기여할 수 있습니다. 추가적으로, **면역억제제** (예: cyclosporine, tacrolimus)와 같은 약물도 신독성 위험이 높으므로 포함하는 것을 고려하십시오.

2. (Paper: 92), (Paper: 94)와 같이 **약물 유발 신독성의 위험 요소**에 대한 포괄적인 검토를 참조하여, 목록에 포함되지 않은 다른 중요한 신독성 약물 (예: 아미노글리코사이드와 병용 시 발생할 수 있는 신독성 (예: (Paper: 91))) 또는 **특정 약물 계열** (예: 비스테로이드성 소염제와 시스플라틴 병용 시 (Paper: 101))을 추가하는 것을 고려하십시오. 이는 목록의 **완전성과 최신 정보 반영**을 개선할 것입니다. 예시: 'Additionally, consider including other commonly implicated nephrotoxins in critical care, such as specific antifungals (e.g., amphotericin B) or immunosuppressants (e.g., tacrolimus as discussed in (Paper: 97)).'

6. (Block: 40)의 'Addressing Missing Important Confounders' 섹션에서 'Based on clinical considerations, we expanded our confounder set to include clinically critical variables'라고 언급했지만, 이러한 **'임상적 고려사항'의 구체적인 내용**과 **'확장된 교란 변수 세트'가 어떻게 선택되었는지**에 대한 **논리적 근거가 부족하여** 독자가 선택의 타당성을 이해하기 어렵습니다. 이는 교란 변수 선택의 **논리적 흐름과 구조**에 대한 심층 분석이 부족했다는 의미로 해석될 수 있습니다.

## 제안 사항

1. 확장된 교란 변수 세트의 **선택 과정**에 대해 더 자세히 설명하십시오. 예를 들어, **이전의 역학 연구** (예: (Paper: 63)의 교란 변수 선택) 또는 **임상 전문가의 의견**을 기반으로 중요하다고 판단된 이유를 명시하십시오. 특히 sepsis severity나 mechanical ventilation과 같은 변수가 VPT 처방과 AKI 발생에 어떻게 동시에 영향을 미치는지에 대한 **간략한 인과 경로**를 설명하여 해당 변수들이 교란 변수로서 왜 중요한지 **논리적 일관성**을 강화하십시오 (예: (Paper: 35) 참조). 예시: 'These variables were selected based on their established roles as strong predictors of both antibiotic prescribing patterns and AKI risk (e.g., (Paper: 236)), mitigating potential indication bias.'

2. (Paper: 74), (Paper: 75), (Paper: 76)과 같이 **LLM을 활용한 교란 변수 발견**에 대한 최신 연구를 참조하여 LLM이 이 과정에서 수행한 **구체적인 역할**을 강조할 수 있습니다. 예를 들어, 'LLM은 대량의 비정형 임상 노트에서 기존에 간과되었던 고위험 인자들을 식별함으로써, 이러한 확장된 교란 변수 세트의 포괄성을 크게 향상시켰다'고 설명하여 **LLM의 기여도를 명확히** 하십시오.

7. (Block: 48)의 'Error Handling and Quality Control' 섹션에서 'Ambiguous responses: Conservative interpretation favoring null classification'과 'Conflicting information: Prioritization of explicit over implicit evidence' 전략을 언급했지만, 이러한 **규칙을 적용한 구체적인 사례**나 이러한 규칙이 최종 결과에 미치는 **잠재적 영향**에 대한 설명이 부족하여 **실질적인 의미가 불분명합니다**. 독자는 이러한 규칙이 **데이터 정확성**에 어떻게 기여하는지 이해하기 어렵습니다. 이는 데이터 처리 과정의 **논리적 투명성**을 저해합니다.

## 제안 사항

1. 오류 처리 규칙의 **논리적 근거**와 이들이 데이터의 **정확성과 신뢰성**을 유지하는 데 어떻게 기여하는지 더 자세히 설명하십시오. 예를 들어, 'null classification'이 **위양성 오류**를 줄이고 **보수적인 추정**을 가능하게 하는 데 어떻게 도움이 되는지 언급할 수 있습니다. 이는 (Paper: 67), (Paper: 62)에서 논의된 **교란 변수의 오분류**가 인과 추론에 미치는 영향과 관련될 수 있습니다. 예시: 'This conservative approach for ambiguous responses minimizes the risk of introducing spurious confounder effects, aligning with principles of robust causal inference (e.g., (Paper: 65)).'

2. 실제 데이터에서 발생할 수 있는 **모호하거나 상충되는 응답의 예시**를 제시하고, 이러한 규칙이 해당 사례에 어떻게 적용되었는지 **간단한 시나리오**를 추가하여 독자의 이해를 돕습니다. 예를 들어, '환자 노트에 CKD 진단이 명시적으로 없지만 eGFR이 60 미만인 경우(implicit)와 CKD 진단이 명시된 경우(explicit)의 처리 방식'을 설명할 수 있습니다. 예시: 'For instance, if a patient's notes did not explicitly state 'Chronic Kidney Disease' but consistently showed eGFR < 60 mL/min/1.73m<sup>2</sup> (implicit), while another note explicitly stated 'No CKD', the explicit negative statement would be prioritized to avoid misclassification.'

8. (Block: 53)에서 Propensity Score Estimation에 사용된 'c-statistic, Brier score, and Kolmogorov-Smirnov test [Austin, 2011]' 평가 지표를 언급했지만, 이들 지표가 각각 **무엇을 측정하고 왜 선택되었는지**에 대한 간략한 설명이 부족하여 독자들이 **지표의 의미와 중요성을 명확히 이해하기 어렵습니다**. 이는 방법론적 선택의 **논리적 정당성**을 불분명하게 만듭니다.

## 제안 사항

1. 각 평가 지표 (c-statistic, Brier score, Kolmogorov-Smirnov test)가 무엇을 측정하는지 **간략하게 정의**하는 문장을 추가하십시오. 예시: 'The c-statistic (concordance statistic) evaluates the **discriminative power** of the propensity score model, indicating its ability to correctly classify treated vs. control patients. The Brier score measures **prediction accuracy**, quantifying the mean squared difference between predicted probabilities and actual outcomes. The Kolmogorov-Smirnov test assesses **covariate balance** by measuring the maximum difference between the cumulative distributions of covariates in treated and control groups, thereby indirectly reflecting the model's ability to create comparable groups.'

2. 이러한 지표들이 Propensity Score 모델의 **성능을 종합적으로 평가**하는 데 어떻게 기여하는지 **논리적 연결**을 명확히 하십시오. (Paper: 115)와 같은 문헌을 참조하여 Propensity Score 방법론에서 이러한 지표들의 일반적인 사용법을 설명할 수 있습니다. 예시: 'These metrics were chosen to provide a comprehensive evaluation of the propensity score model, covering both its **predictive**

**accuracy** (c-statistic, Brier score) and its effectiveness in achieving **covariate balance** (Kolmogorov-Smirnov test), which are critical for valid causal inference in observational studies.'

9. (Block: 54)에서 Inverse Probability of Treatment Weighting (IPTW)에 대해 'We calculated stabilized weights with truncation at 1st and 99th percentiles to reduce instability. Effective sample size (ESS) was calculated to assess weight distribution'라고 설명했지만, **왜 안정화된 가중치를 사용했는지와 절단(truncation)이 가중치 불안정성을 줄이는 데 어떻게 도움이 되는지에 대한 논리적 설명**이 부족합니다. 또한 ESS가 가중치 분포를 평가하는 데 어떻게 활용되는지에 대한 추가적인 설명이 필요하여 **방법론적 선택의 명확성이 저해됩니다.**

## 제안 사항

1. 안정화된 가중치(stabilized weights)가 **추정치의 효율성**을 높이고 극단적인 가중치(extreme weights)로 인한 **분산 불안정성**을 줄이는 데 어떻게 기여하는지 설명하십시오 (예: (Paper: 163), (Paper: 148) 참조). 예시: 'Stabilized weights were used to improve **estimation efficiency** and reduce the **variability** associated with extreme weights, as unstabilized weights can lead to inflated variances.'

2. 가중치 절단(truncation)이 **극단적인 가중치**의 영향을 완화하여 **공변량 균형**을 개선하고 **추정치의 분산**을 줄이는 방법에 대해 간략히 설명하십시오. 이는 (Paper: 142)에서 논의된 가중치 절단의 중요성과 관련될 수 있습니다. 절단으로 인한 **추정 대상의 변화 가능성**에 대해서도 언급하여 **논리적 투명성**을 확보하십시오. 예시: 'Truncation at the 1st and 99th percentiles was applied to mitigate the influence of extreme weights, which can arise from poor overlap in propensity score distributions and severely destabilize estimates (e.g., (Paper: 148)). While truncation can slightly alter the target population (e.g., (Paper: 149)), it enhances the **robustness and precision** of causal effect estimates.'

3. ESS(Effective Sample Size)가 **가중치 부여 후 정보 손실**을 측정하는 방법으로, 가중치 부여가 데이터의 **통계적 효율성**에 미치는 영향을 평가하는 데 어떻게 사용되는지 설명하는 문장을 추가하십시오. ESS를 통해 가중치 부여가 **얼마나 많은 '유효한' 관측치를 남겼는지**를 독자가 이해할 수 있도록 하십시오. 예시: 'Effective sample size (ESS) was calculated to quantify the **loss of statistical power** due to weighting, providing a measure of how many 'effective' observations remained after weighting to assess the **efficiency** of the weight distribution.'

10. (Block: 55)의 Doubly Robust Estimation 설명에서 'This approach provides consistent estimates when either the treatment or outcome model is correctly specified'라고 언급했지만, 이 **'이중 견고성'의 개념**이 이전의 IPTW 방법과 어떻게 차별화되고 **왜 본 연구에 이점을 제공하는지에 대한 설명**이 부족합니다. 독자들은 이 복잡한 통계 기법의 **핵심 장점을 명확하게 파악하기 어렵습니다.**

## 제안 사항

1. 이중 견고성(doubly robust)의 **핵심 원리**를 더 명확하게 설명하십시오. 예를 들어, **두 개의 모델** (처리 모델과 결과 모델) 중 하나만 올바르게 지정되어도 **편향되지 않은 추정치**를 얻을 수 있다는 점을 강조하여 방법론

의 **강점**을 부각하십시오 (예: (Paper: 169), (Paper: 180) 참조). 예시: 'Doubly robust estimation is a crucial method in causal inference because it ensures **consistent estimates** if either the propensity score model (for treatment assignment) or the outcome regression model (for outcome prediction) is correctly specified, providing **two chances for unbiased estimation** (e.g., (Paper: 190)).'

2. 이중 견고성 추정법이 IPTW 단독 사용 시 발생할 수 있는 **모델 오지정의 위험**을 어떻게 완화하는지에 대한 **논리적 연결**을 추가하십시오. 이는 (Paper: 173)에서 논의된 바와 같이 모델 오지정 시의 편향 문제를 다룰 수 있으며, 본 연구 결과의 **견고성을 강화**합니다. 예시: 'This dual robustness offers a significant advantage over IPTW alone, which relies solely on the correct specification of the propensity score model. It provides **increased protection against bias** from potential model misspecification, enhancing the **reliability** of our causal effect estimates (e.g., (Paper: 166), (Paper: 181)).'

11. (Block: 56)에서 'Covariate balance was evaluated using standardized mean differences (SMD) before and after weighting, with  $SMD < 0.1$  indicating adequate balance'라고 언급했지만,  $SMD < 0.1$  기준이 왜 '**적절한 균형**'을 의미하는지에 대한 근거가 부족하여 방법론적 기준의 타당성이 불분명합니다. 이는 평가 기준의 **논리적 정당성**을 약화시킵니다.

## 제안 사항

1.  $SMD < 0.1$  기준이 **인과 추론 연구**에서 공변량 균형을 평가하는 데 널리 사용되는 **일반적인 임계값**임을 설명하고, 이 기준이 **작은 잔여 편향**을 허용하는 합리적인 선택임을 언급하십시오 (예: (Paper: 255), (Paper: 256) 참조). 예시: 'A standardized mean difference (SMD) of less than 0.1 is widely accepted in causal inference literature (e.g., (Paper: 255)) as an indicator of **adequate covariate balance**, implying that any remaining differences between treatment groups are unlikely to induce substantial bias in the estimated causal effect.'

2. SMD의 **의미**를 간략하게 설명하여 독자들이 이 지표가 공변량 분포의 **차이를 어떻게 정량화**하는지 이해할 수 있도록 돕습니다. 예를 들어, SMD가 표준화된 차이이므로 **서로 다른 단위의 변수들을 비교 가능하게 만들며**, 0.1 미만은 대부분의 실제적 목적에서 **무시할 수 있는 차이**로 간주된다고 설명할 수 있습니다. 예시: 'SMD quantifies the difference in means or proportions between groups, standardized by the pooled standard deviation, making it a scale-independent measure of imbalance. A value below 0.1 suggests that the distributions of the covariates are **sufficiently similar** between the weighted treatment groups for valid causal comparisons.'

12. (Block: 60)의 E-value 분석 설명에서 'The E-value represents the minimum strength of association an unmeasured confounder must have with both treatment and outcome to explain away the observed effect'라고 명시했지만, E-value가 **정성적인 해석**을 넘어 **실제 임상적 또는 역학적 맥락**에서 어떤 의미를 가지는지에 대한 추가적인 설명이 부족하여 **결과의 임상적 중요성이 불명확**합니다. 이는 연구 결과의 **견고성에 대한 깊이 있는 이해**를 방해합니다.

## 제안 사항

1. E-value가 '측정되지 않은 교란 변수'의 잠재적 영향에 대한 민감도 분석 도구임을 강조하십시오. 예를 들어, 'observed effect'를 'nullify the observed effect' 또는 'shift the confidence interval to include the null'과 같은 구체적인 결과와 연관 지어 설명할 수 있습니다. 예시: 'The E-value quantifies the minimum strength of association an unmeasured confounder would need to have with both the VPT exposure and AKI outcome, on the relative risk scale, to either **fully explain away the observed effect** or **shift the confidence interval to include the null**.'

2. E-value가 연구의 견고성을 평가하는 데 어떻게 활용되는지에 대한 추가적인 문장을 추가하십시오. (Paper: 117)에서 제시된 바와 같이, E-value가 특정 비율 위험으로 해석될 수 있음을 언급하여 독자들이 결과의 임상적 중요성과 관측 편향에 대한 취약성을 더 잘 이해할 수 있도록 돕습니다. 예시: 'A higher E-value indicates **greater robustness** of the observed causal effect to potential unmeasured confounding, providing a quantitative measure for assessing the **credibility** of findings from observational studies (e.g., (Paper: 117)).'

13. (Block: 93)에서 'The 2.9-point c-statistic improvement (0.563 → 0.592) represents clinically meaningful enhancement in confounding control'이라고 언급했지만, 이 '2.9 포인트 개선'이 임상적으로 왜 의미가 있는지에 대한 논리적이고 직관적인 설명이 부족합니다. 'c-statistic'의 변화가 '교란 제어'에 미치는 실제적인 영향을 더 명확히 해야 합니다. 이는 LLM의 기여에 대한 명확한 이해를 방해합니다.

## 제안 사항

1. c-statistic 개선이 환자 분류 또는 치료 결정에 미치는 잠재적인 영향을 간략하게 설명하십시오. 예를 들어, '예측 모델의 판별 능력이 향상되어 VPT 치료 그룹과 단독 치료 그룹 간의 AKI 위험도 예측이 더 정확해졌으며, 이는 임상 의사결정 시 환자 위험 계층화에 도움을 줄 수 있다'고 언급할 수 있습니다. 예시: 'This 2.9-point increase in the c-statistic reflects an **improved ability of the propensity score model to discriminate** between patients receiving VPT versus monotherapy, indicating a better capacity to account for observed confounders. This translates to **more accurate risk stratification** and potentially **more reliable causal effect estimates** for clinical decision-making.'

2. 이러한 개선이 실제 데이터 (real-world data)에서 인과 효과 추정의 정확성을 높이는 데 어떻게 기여하는지에 대한 논리적 연결을 강화하십시오 (예: (Paper: 215), (Paper: 217) 참조). 이는 LLM-assisted 접근 방식이 기존 방법론에 비해 제공하는 명확한 이점을 부각할 것입니다. 예시: 'Such an enhancement in confounding control is particularly **critical in real-world observational studies** (e.g., (Paper: 218)) where accurate adjustment for patient characteristics is paramount for deriving **unbiased causal inferences**.'

14. (Block: 95)에서 'Our findings contrast with several smaller studies reporting neutral or protective effects of VPT combination [Luther et al., 2018]'라고 언급하면서, 이러한 상충되는 결과의 주된 이유로 'inadequate sample sizes (typically < 1,000 patients), selection bias, residual confounding, and heterogeneous populations'를 지적했습니다. 그러나 이러한 이전 연구들의 주요 한계점이 본 연구에서 어떻게 효과적으로 극복되었는지에 대한 명확하고 체계적인 논리적 설명이 부족합니다.



## 제안 사항

1. 본 연구가 제시하는 **주요 방법론적 개선 사항** (예: 대규모 MIMIC-IV 데이터 사용 (Paper: 131), LLM 기반의 포괄적인 교란 변수 제어 (Block: 30), 이중 견고성 추정 (Block: 55))이 이전 연구의 한계점 (예: 작은 샘플 크기, 잔여 교란)을 **어떻게 해결했는지 구체적으로 연결**하여 설명하십시오. 특히, '**선택 편향**' (selection bias) 및 '**잔여 교란**' (residual confounding) 문제를 다루는 데 있어 LLM의 역할이 이전 연구들과 어떻게 차별화되는지 강조할 수 있습니다 (예: (Paper: 63)의 교란 변수 선택 중요성 참조). 예시: 'Our study directly addresses these limitations by leveraging the **massive scale of the MIMIC-IV database** (over 90,000 patients, (Paper: 131)), implementing **LLM-assisted comprehensive confounder identification** ((Block: 30)) to reduce residual confounding, and employing **doubly robust estimation** ((Block: 55)) to enhance bias protection from model misspecification.'

2. 이러한 방법론적 강점이 본 연구 결과의 **신뢰성**을 어떻게 높이는지에 대한 **논리적 일관성**을 강화하십시오. (Paper: 5), (Paper: 9), (Paper: 10)과 같은 관련 연구들을 인용하여 본 연구의 **대규모 데이터 분석**과 **고급 인과 추론 방법론**이 기존 문헌의 논쟁을 해결하는 데 기여하는 바를 강조하십시오. 예시: 'This robust methodological framework provides **higher statistical power** and **greater confidence** in our findings, which helps to resolve inconsistencies observed in prior, smaller studies (e.g., (Paper: 111) on meta-analysis discrepancies).'

15. (Block: 97)에서 'we lacked access to vancomycin levels, which could mediate the treatment-outcome relationship'이라고 언급했는데, 반코마이신 약물 농도 정보의 부족이 **결과의 해석에 미치는 잠재적 영향**에 대한 **심층적인 논의**가 부족합니다. 이 정보가 왜 중요한 '**매개 변수**'가 될 수 있는지, 그리고 이로 인해 발생할 수 있는 '**측정되지 않은 교란**' (unmeasured confounding)의 가능성을 더 명확히 해야 합니다. 이는 논리적 흐름과 내부 일관성에 대한 분석이 깊이 있게 이루어지지 않았음을 의미합니다.

## 제안 사항

1. 반코마이신 약물 농도 (trough levels)가 AKI 발생에 미치는 **매개 효과** (mediation effect)를 설명하고, 이 정보의 부재가 본 연구의 **인과 추론에 어떤 한계**를 가져오는지 더 자세히 설명하십시오 (예: (Paper: 237)의 매개 분석 참조). 약물 농도가 **VPT 처방과 AKI 발생 모두에 영향을 미칠 수 있는 경로**를 설명하여 '**측정되지 않은 교란**'의 **잠재적 경로**를 명확히 하십시오. 예시: 'Vancomycin trough levels are a critical **mediator** in the relationship between vancomycin exposure and AKI (e.g., (Paper: 103)), as excessively high levels directly increase nephrotoxicity. The absence of these data means we cannot fully account for this **potential unmeasured confounder** if patients receiving VPT were systematically managed differently regarding vancomycin dosing or monitoring, thus potentially biasing the estimated causal effect.'

2. 약물 농도 데이터가 있었다면 **어떤 종류의 분석** (예: 동적 치료 전략 (예: (Paper: 174)), 시간 가변 교란 변수 (예: (Paper: 41)))이 가능했는지 간략하게 논의하여 **미래 연구 방향**을 제시하십시오. 예시: 'With vancomycin level data, future studies could apply advanced causal inference methods for **time-varying confounders** (e.g., (Paper: 41)) to more precisely estimate the nuanced effects of VPT, potentially informing **dynamic treatment strategies**.'

3. 데이터 부족으로 인한 한계에도 불구하고, '치료 약물 모니터링 관행이 그룹 간에 유사했을 것'이라는 가정의 근거를 보완하거나, 이러한 가정이 타당하지 않을 경우 **결과 해석에 어떤 영향을 미칠 수 있는지에** 대한 논의를 추가하십시오. 예를 들어, 무작위 오류가 아닌 **체계적인 차이**가 있었다면 결과가 어떻게 달라질 수 있는지 설명할 수 있습니다. 예시: 'While we assumed similar therapeutic drug monitoring practices across groups, any systematic differences in vancomycin level-guided dosing between VPT and monotherapy recipients could introduce **residual confounding**, potentially over- or underestimating the true effect, if not due to random variation.'

16. (Block: 98)에서 'indication bias remains possible despite extensive confounding control'이라고 인정하면서, 'VPT recipients may have unmeasured infection severity markers'를 언급했습니다. 이 '측정되지 않은 감염 심각도 마커'가 구체적으로 무엇이며, 이것이 **본 연구의 인과 효과 추정에 어떻게 편향을 일으킬 수 있는지에** 대한 더 깊이 있는 설명이 필요하여 **제한점의 영향이 불분명합니다**. 이는 논리적 흐름과 내부 일관성에 대한 분석이 깊이 있게 이루어지지 않았음을 의미합니다.

## 제안 사항

1. 측정되지 않은 감염 심각도 마커가 **어떤 변수**일 수 있는지 구체적인 예시를 들어 설명하십시오 (예: 특정 염증 마커 (예: procalcitonin, CRP의 동적 변화), 쇼크의 미묘한 징후, 초기 치료 반응성 데이터, 미생물학적 검사 결과의 미세한 차이). 이는 (Paper: 56), (Paper: 57), (Paper: 59)와 같은 연구에서 논의된 '측정되지 않은 교란 변수'의 중요성과 관련될 수 있습니다. 예시: 'These 'unmeasured infection severity markers' could include highly granular, subjective clinical assessments (e.g., clinician's initial impression of patient deterioration), specific microbiological results that aren't consistently coded, or dynamic physiological responses to initial infection, which are not routinely captured in structured EHR data but influence both VPT prescribing and AKI risk.'

2. 이러한 '측정되지 않은 교란 변수'가 **VPT 처방 결정**과 **AKI 발생**에 어떻게 영향을 미쳐 **인과 효과 추정을 편향**시킬 수 있는지에 대한 **잠재적 인과 경로** (causal pathway) 또는 **논리적 흐름**을 설명하십시오 (예: (Paper: 35) 참조). 예시: 'For instance, patients with more severe, yet unmeasured, infection characteristics might be preferentially prescribed VPT due to perceived urgency, and these same underlying severe characteristics could independently elevate their AKI risk, creating an **unobserved confounding pathway** that biases the observed association.'

3. E-value 분석이 '중간 정도의 측정되지 않은 교란에 대한 견고성'을 시사하지만, '**강한 측정되지 않은 교란**'이 여전히 결과를 편향시킬 수 있다는 점을 강조하고, 이러한 편향을 해결하기 위한 **미래 연구 방향**을 제시하십시오 (예: (Paper: 61), (Paper: 66)과 같은 고급 인과 추론 방법론 또는 (Paper: 76)과 같은 LLM 기반의 숨겨진 교란 변수 발굴 기법 참조). 예시: 'While our E-value analysis suggests robustness to moderate unmeasured confounding, a sufficiently strong unmeasured confounder could still bias our results. Future research could explore advanced **causal inference methods with latent variables** (e.g., (Paper: 83)) or leverage **LLM-guided discovery for hypothesizing missing causal variables** (e.g., (Paper: 81)) to address these deeper confounding challenges.'

## 심사자 4 (영향력)

### 강점

1. 본 연구는 대규모 MIMIC-IV 데이터베이스를 활용하여 반코마이신 단독 요법과 반코마이신-피페라실린/타조박탐(VPT) 병용 요법 간의 **AKI 위험 차이를 대규모로 평가했다**는 점에서 강력한 임상적 의의를 지닌다. 이는 (Paper: 216)에서 지적했듯이 AKI가 전 세계적으로 중요한 공중 보건 문제임을 고려할 때, **AKI 발생률 증가에 대한 명확한 증거를 제공함으로써 의료 행위에 실질적인 영향을 미칠 수 있는 높은 잠재력**을 보여준다.
2. LLM 기반 교란 변수 추출 프레임워크와 그에 대한 **포괄적인 검증 과정**은 연구의 투명성과 재현성을 높여, 인공지능 보조 약물 감시 연구의 **전문적 관련성 및 채택 잠재력**을 크게 향상시킨다. ICD-10 코드와의 높은 일치율 (0.83-0.91) 및 수동 검토(97.12% 정확도) 결과는 (Paper: 244)에서 강조된 바와 같이, LLM이 **약물 감시 활동을 지원하는 새로운 길을 제공할 수 있음**을 강력히 시사하며, 이는 **의료 산업의 AI 활용 확산**에 중요한 이정표가 될 수 있다.
3. AKI 위험이 39-42% 증가한다는 결과는 VPT의 **광범위한 사용**을 고려할 때 **상당한 임상적 부담**으로 이어진다. (Block: 86)에서 제시된 1,000명당 38건의 초과 AKI 발생이라는 구체적인 수치는 (Paper: 233)에서 AKI가 병원 내 사망률, 입원 기간, 비용 증가와 유의미하게 연관되어 있음을 보여주듯이, **국가적 차원의 의료 시스템에 대한 잠재적 영향**이 매우 크다는 것을 명확히 보여주어 **정책 입안자들에게 중요한 근거**를 제공한다.
4. 항균제 관리 프로그램에 AKI 위험 평가를 통합하라는 (Block: 88)의 권고는 (Paper: 88)에서 강조된 바와 같이 **실질적인 응용 가능성**을 시사하며, **환자 안전 및 의료 자원 최적화**에 긍정적인 영향을 미칠 수 있다. 이는 현재의 경험적 치료 관행에 대한 **재고를 촉구**하는 중요한 제안으로, **의료 현장의 변화를 유도**할 수 있는 잠재력을 가진다.
5. (Block: 102)에서 제시된 투명한 방법론과 포괄적인 검증 프레임워크를 갖춘 LLM 기반 교란 변수 식별의 향상은 **확장 가능한 약물 감시 연구**를 가능하게 하며, 이는 (Paper: 277)에서 언급된 바와 같이 **약물 역학 연구의 미래 방향**을 제시하는 **높은 학술적 가치**를 지닌다. 또한, (Paper: 74), (Paper: 76), (Paper: 75)에서 제시된 LLM을 활용한 교란 변수 식별의 중요성에 대한 연구 동향과도 일치하여 **학계의 새로운 연구 방법론 도입**을 촉진할 수 있다.

### 약점

1. 초록 (Block: 4)에서 MIMIC-IV 데이터베이스의 연구 기간이 '2008-2019'로 명시되어 있으나, MIMIC-IV의 공식 버전 정보 ((Paper: 131), (Paper: 132))에 따르면 이 기간을 포함하는 것은 최신 버전인 MIMIC-IV 2.2 또는 3.0입니다. 사용된 **데이터베이스의 정확한 버전 정보가 누락**되어 연구의 **재현성과 투명성**에 대한 우려를 제기합니다. 이는 **연구 결과의 신뢰성 및 향후 연구에의 활용 가능성**에 영향을 미칠 수 있습니다.

### 제안 사항

1. (Block: 4) 및 (Block: 15)에 사용된 MIMIC-IV 데이터베이스의 **정확한 버전 정보**(예: MIMIC-IV 3.1)를 명시하여 연구의 **재현성을 높이고 정보의 정확성을 확보**하십시오. 이는 (Paper: 131)과 (Paper: 132)와 같은 공식 문헌에 따라 데이터 소스를 명확히 하는 데 필수적입니다. 데이터 소스의 **투명성 확보**는 연구의 **장기적인 학술적 영향력**을 강화합니다.

2. 초록 (Block: 4)에서 제시된 E-값 2.12는 통계적 견고성을 나타내지만, 이 **E-값이 잠재적, 측정되지 않은 교란 요인이 결과에 미치는 영향을 어떻게 해석해야 하는지에 대한 실질적인 임상적 지침이 부족하여 연구의 실용적 가치 전달이 미흡합니다.**

## 제안 사항

1. (Block: 4)에 E-값 2.12가 임상적으로 어떤 의미를 가지는지, 예를 들어 '관찰된 효과를 완전히 상쇄시키려면, 측정되지 않은 교란 변수가 VPT 사용 및 AKI와 각각 2.12배 이상의 상대 위험으로 연관되어야 한다는 의미이며, 이는 **일반적인 임상 상황에서 강력한 교란 변수일 가능성이 낮음**을 시사하여 **현재의 결론이 견고하다**는 임상적 근거를 제공한다'와 같이 **임상적 해석을 추가**하여 독자의 이해를 돕고 **잠재적 가치를 강조**하십시오.

2. (Block: 80)의 E-값 분석 부분에서도 이와 같은 **임상적 해석을 보강**하고, (Paper: 117)에서 제시된 E-값의 개념과 활용법을 참고하여 **추가적인 해석 가이드라인**을 포함하는 것을 고려하십시오. 이는 **의료 전문가들이 연구 결과를 실제 임상 의사결정에 적용하는 데 도움**이 됩니다.

3. (Block: 39)에 나열된 신독성 약물 목록에 '고용량 푸로세미드, 부메타니드'와 같이 용량에 따른 구분이 있지만, 다른 약물들(아미노글리코사이드, NSAIDs 등)은 단순히 약물명만 나열되어 있어 **'신독성 약물 노출'의 정의가 불균일하게 느껴집니다**. 이는 LLM이 **일관된 기준으로 교란 변수를 추출**했는지에 대한 의문을 제기하며, **데이터 추출의 신뢰성 및 결과의 임상적 적용 가능성에 영향을 미칠 수 있습니다**.

## 제안 사항

1. (Block: 39)에 나열된 모든 신독성 약물에 대해 **용량, 기간, 또는 특정 임상적 조건**을 포함하는 **더욱 상세하고 일관된 임상적 정의**를 추가하여 LLM 추출의 **정확성**을 높이십시오. 예를 들어, (Paper: 103)에서 제시된 반코마이신 관련 신독성 위험 요인(총 일일 용량 > 4g, 최저 농도 > 20mg/L 등)과 같이 **구체적인 기준을 제시**하는 것을 고려하십시오. 이는 **LLM 기반 추출의 정밀도를 높여 약물 감시 연구의 신뢰성**을 강화합니다.

2. LLM이 이러한 **복잡한 용량/조건 기반 정의를 어떻게 처리하도록 지시**되었는지 (Block: 44)의 '자세한 LLM 프롬프트 구조' 부분에 명확히 설명하십시오. 이는 (Paper: 245)에서 강조하는 LLM 기반 인과 추론의 **투명성**을 향상시킬 것입니다.

4. (Block: 41)에서 LLM이 추출한 '폐혈증 중증도', '기계 환기', '혈관 수축제 사용' 등의 추가적인 중증도 지표들은 중요한 교란 변수이지만, 이러한 변수들이 (Block: 33)의 **기존 임상 정의들과 어떻게 상호작용하며 VPT 처방 및 AKI 발생에 영향을 미치는지에 대한 설명**이 부족합니다. 이는 LLM이 도출한 confounder set의 **임상적 의미와 가치**를 약화시켜, **연구의 설득력과 영향력**을 감소시킬 수 있습니다.

## 제안 사항

1. (Block: 41)에서 이러한 추가적인 중증도 마커들이 **어떻게 VPT 처방 결정에 영향을 미치고 AKI 위험을 증대시키는지에 대한 임상적 메커니즘**을 간략하게 설명하십시오. 예를 들어, '폐혈증 중증도가 높은 환자는 중증도 증가로 인해 VPT 처방 가능성이 높으며, 동시에 AKI 발생 위험도 높아진다'와 같이 설명하여 (Paper: 237)에서 논의된 **인과 추론의 설명력**을 높이고 **임상적 중요성**을 부각하십시오. 이는 **연구 결과의 임상적 실효성**을 높

입니다.

2. 각 추가 변수가 '종합적인 LLM 교란 변수' 세트에 포함된 **근거와 중요성**을 (Block: 92)의 '확장된 교란 변수 세트' 부분에 더욱 자세히 기술하십시오. 이는 **LLM 기반 교란 변수 식별의 학술적 기여**를 명확히 합니다.

5. LLM 추출의 **임상적 타당성 평가**가 (Block: 46)에서 '모든 관찰된 유병률이 발표된 문헌의 예상 범위 내에 있었다'고만 언급되어 **매우 일반적**입니다. 어떤 문헌을 참조했는지, 그리고 각 confounder의 **구체적인 유병률 범위**와 본 연구의 결과가 그 범위 내에 있었는지에 대한 **정량적이고 구체적인 비교**가 부족합니다. 이는 LLM의 **실제 적용 가능성**에 대한 **신뢰를 저하시킬 수 있는 요인**입니다.

## 제안 사항

1. (Block: 46)에 각 LLM 추출 교란 변수(예: 만성 신장 질환, 당뇨병, 심부전 등)의 **구체적인 유병률**과, 이를 비교한 **'발표된 문헌의 실제 인용'** (예: (Paper: 19) 또는 (Paper: 98) 등)을 포함하여 **정량적인 타당성 평가** 결과를 제시하십시오. 'within expected ranges' 대신 **실제 수치와 범위를 명시**하는 것이 중요합니다. 이는 **LLM 기반 추출의 정확성에 대한 신뢰도**를 높여 **연구의 임상적 영향력**을 강화합니다.

2. (Block: 93)에서 '개별 교란 변수 연관성이 임상적으로 합리적이었다'는 주장을 뒷받침하기 위해 **주요 교란 변수 몇 가지에 대한 LLM 추출 유병률 수치**와 **참고 문헌의 유병률 수치**를 비교하는 **표 또는 서술**을 추가하는 것을 고려하십시오. 이는 (Paper: 217)에서 강조하는 '실세계 연구의 엄격한 설계'에 부합하여 **연구의 임상적 신뢰성**을 높일 수 있습니다.

6. (Block: 54)에서 안정화된 IPTW 가중치 적용 시 '**불안정성 감소**'를 위해 **1% 및 99% 백분위수에서 절단 (truncation)**되었다고 명시되었으나, 이 **절단 기준의 선택**에 대한 근거나 **다른 절단 기준 (예: 5%, 95%)**과의 **비교**에 대한 설명이 부족합니다. 이는 **가중치 적용 방법론의 견고성**에 의문을 제기할 수 있으며, **결과의 해석 가능성**에 영향을 미칠 수 있습니다.

## 제안 사항

1. (Block: 54)에 **1% 및 99% 백분위수에서 절단한 이유**에 대한 **명확한 근거**를 제시하십시오. 예를 들어, (Paper: 148)에서 논의된 '극단적인 가중치 문제'와 (Paper: 142)에서 제시된 '가중치 절단이 결과에 미치는 영향'에 대한 문헌을 인용하여 이러한 선택이 **가중치 분포의 안정성**을 최적화하고 **결과 추정치의 편향을 최소화**하기 위함을 설명하십시오. 이는 **방법론적 선택의 정당성**을 강화하여 **연구 결과의 신뢰성**을 높입니다.

2. 또한, (Paper: 152)에서 제시된 클러스터 기반 가중치 방법론과 같이, **극단적인 가중치 문제를 해결하기 위한 다른 고급 방법론**에 대한 **논의를 추가**하거나, 이러한 방법론을 시도하지 않은 이유를 간략히 설명하는 것을 고려하여 **방법론적 강점**을 부각하십시오. 이는 **향후 연구에 대한 영향력**을 넓힐 수 있습니다.

7. (Block: 60)의 E-값 분석은 **측정되지 않은 교란 요인에 대한 견고성**을 평가하지만, 이 분석이 LLM이 '놓칠 수 있는' **특정 유형의 교란 요인**에 대해 **어떤 통찰력**을 제공하는지에 대한 논의가 부족합니다. 예를 들어, (Block: 98)에서 '측정되지 않은 감염 중증도 마커'를 언급했는데, E-값이 이러한 **특정 유형의 교란 요인**에 대

해 얼마나 의미가 있는지 명확하지 않아 연구의 **종합적인 영향력**을 제한할 수 있습니다.

## 제안 사항

1. (Block: 60)의 E-값 분석 부분에, E-값이 LLM이 식별하지 못했을 수 있는 **특정 임상적 교란 요인(예: 미세한 감염 중증도 지표, 유전적 요인, 약물 상호작용의 미묘한 변이 등)**에 대해 어떤 맥락에서 해석되어야 하는지에 대한 논의를 추가하십시오. 예를 들어, '만약 E-값이 2.12라면, LLM이 놓쳤을 수 있는 미세한 감염 중증도 지표가 VPT 사용과 AKI 발생에 각각 2.12배 이상의 영향을 미치지 않는 한, 관찰된 효과는 견고하다고 해석할 수 있다'와 같이 **구체적인 예시와 해석 방향**을 제시하여 (Paper: 35)와 (Paper: 57)에서 다뤄진 '실험에서의 교란' 및 '측정 오류의 영향'을 반영하십시오. 이는 **연구 결과의 임상적 해석 가능성**을 심화하여 **실질적인 영향력**을 높입니다.

2. (Block: 98)에서 언급된 '측정되지 않은 감염 중증도 마커'에 대해, E-값이 이러한 **특정 유형의 교란 요인**에 대한 연구의 견고성을 어떻게 뒷받침하는지 **구체적으로 연결하여 설명**하십시오. 이는 (Paper: 63)에서 논의된 **교란 변수 선택의 목적과 접근 방식**을 보다 명확히 제시하는 데 도움이 되어 **연구 결과의 신뢰성**을 높일 수 있습니다.

8. (Block: 62)의 하위 그룹 분석에서 'AKI 전신 기능, 연령 그룹, ICU 입원 상태'에 따라 **유의미한 상호작용 효과가 감지되지 않았**다고 언급되었지만, 이러한 하위 그룹 간의 **AKI 위험 증가율의 경향성**이나 **임상적 차이**에 대한 심층적인 논의가 부족합니다. 상호작용이 없다고 해서 모든 하위 그룹에 동일한 임상적 의미를 부여하는 것은 아니므로, **환자별 위험 계층화 및 맞춤형 치료 전략 개발**에 대한 시사점이 약화될 수 있습니다.

## 제안 사항

1. (Block: 62) 및 (Block: 82)의 하위 그룹 분석 결과에 대해, 상호작용 효과가 통계적으로 유의미하지 않더라도 **각 하위 그룹에서 관찰된 AKI 위험 증가의 경향성**을 설명하십시오. 예를 들어, **고령 환자나 중증 AKI 환자**에서 VPT의 영향이 **임상적으로 더 중요할 수 있음**을 (Paper: 96) (Paper: 98)에서 제시된 바와 같이 강조하는 것을 고려하십시오. 이는 **임상 의사결정**을 위한 **세분화된 정보**를 제공하여 **환자 안전 및 맞춤형 치료 전략 개발**에 기여합니다.

2. (Paper: 145)에서 제시된 **인과적 하위 그룹 분석**을 위한 **propensity score weighting 방법론**을 참고하여, 하위 그룹별로 **AKI 위험에 대한 미묘한 차이**를 더 깊이 탐색할 수 있는 **추가 분석**을 제안하거나, 현재의 분석이 왜 이러한 미묘한 차이를 발견하지 못했는지에 대한 논의를 추가하여 **연구의 임상적 적용 범위**를 확장하십시오. 이는 **연구의 장기적 영향력**을 높일 수 있습니다.

9. (Block: 86)에서 VPT 요법으로 인한 **AKI 위험 증가가 '수천 건의 잠재적 예방 가능 AKI 에피소드'로 이어진**다는 주장은 **국가적 사용 패턴에 대한 구체적인 데이터나 외삽(extrapolation) 방법론**이 명확하지 않아 **잠재적 영향의 크기**에 대한 신뢰성을 저하시킬 수 있습니다. 이는 연구 결과의 **사회적 영향력**을 약화시키는 중요한 요인입니다.

## 제안 사항

1. (Block: 86)에서 '국가적 사용 패턴으로 외삽'하는 데 사용된 데이터 소스(예: 국가 보건 통계, 다른 대규모 데이터베이스 연구 등)를 명확히 하고, 외삽 방법론을 간략하게 설명하여 잠재적 영향 규모의 추정치에 대한 신뢰성을 높이십시오. (Paper: 108) 및 (Paper: 277)과 같은 약물 역학 연구에서 국가적 영향 추정치를 도출하는 방식을 참고하는 것을 고려하십시오. 이는 연구 결과의 정책적 및 사회적 영향력을 강화합니다.

2. 이러한 외삽 추정치가 최소한의 가정을 기반으로 한다는 점을 명시하고, 잠재적인 불확실성에 대해 (Block: 96)의 '제한 사항' 섹션에서 추가적으로 논의하여 투명성을 확보하고 연구의 신뢰성을 높이십시오.

10. LLM 시스템의 유용성을 입증하기 위한 핵심 방법론적 개선 사항이 (Block: 91)에 나열되어 있지만, 이러한 개선 사항들이 어떻게 다른 약물 감시 연구나 임상 의사 결정 시스템에 적용될 수 있는지에 대한 구체적인 로드맵이나 제안이 부족합니다. 즉, 본 연구의 방법론적 성공이 학계 및 산업계에 미칠 잠재적 영향이 충분히 강조되지 않아 연구의 혁신적 가치 전달이 미흡합니다.

## 제안 사항

1. (Block: 91)의 '방법론적 기여' 섹션에 LLM 기반 시스템이 향후 약물 감시 활동, 예를 들어 FDA Sentinel System ((Paper: 244))과 같은 실제 시스템에 어떻게 통합될 수 있는지에 대한 구체적인 적용 시나리오를 추가하십시오. 이는 연구의 실용적 영향력을 크게 높일 수 있습니다.

2. LLM 기반 교란 변수 식별 프레임워크가 새로운 약물 안전성 신호 감지, 의료 기록의 서술 텍스트에서 유해 사례 식별, 또는 인과 관계 조사의 효율성 향상에 어떻게 기여할 수 있는지에 대해 (Paper: 244)에서 논의된 내용을 바탕으로 더욱 구체적인 제안을 제시하십시오. 또한, 본 연구의 프레임워크가 다른 데이터베이스(예: eICU, Health Facts 등)에도 적용 가능함을 언급하고, 이를 통해 교란 변수 식별 프로세스를 표준화하고 약물 감시 연구의 확장성을 높이는 방법을 논의하십시오. 이는 학계 및 산업계에 미치는 파급 효과를 강조합니다.

11. (Block: 93)에서 c-통계량의 2.9 포인트 개선(0.563 → 0.592)이 '임상적으로 의미 있는' 교란 통제 개선이라고 명시되어 있으나, 이 '임상적 의미'의 기준이 명확하지 않습니다. c-통계량의 소폭 개선이 환자 결과나 의료 비용에 얼마나 실질적인 영향을 미칠 수 있는지에 대한 추가적인 설명이 필요하여 연구의 실질적 영향력이 충분히 부각되지 않았습니다.

## 제안 사항

1. (Block: 93)에서 c-통계량 개선이 '임상적으로 의미 있는' 이유를 구체적인 맥락에서 설명하십시오. 예를 들어, '이러한 개선은 더욱 정확한 위험 계층화를 가능하게 하여 AKI 고위험 환자를 식별하고 예방적 개입을 유도하는 데 도움이 될 수 있다'와 같이 제시하여 (Paper: 107)에서 제시된 AKI 위험 평가 모델의 개발 중요성과 연결하십시오. 이는 환자 치료 개선에 기여하는 잠재적 가치를 명확히 할 수 있습니다.

2. 또한, 이러한 개선된 교란 통제가 궁극적으로 의료 비용 절감 ((Paper: 233))이나 환자 장기 예후 개선 ((Paper: 12))에 어떻게 기여할 수 있는지에 대한 잠재적 시나리오를 간략하게 언급하는 것을 고려하십시오. 이는 연구의 경제적 및 사회적 영향력을 강조합니다.

12. (Block: 96)에서 단일 센터 연구라는 점은 일반화 가능성을 제한할 수 있다고 언급되었지만, MIMIC-IV 데이터베이스의 **'다양한 환자 인구'가 외부 타당성을 높인다는 주장에 대한 구체적인 근거 제시가 부족합니다.** MIMIC-IV 데이터베이스의 인구통계학적 특성, 질병 분포 등이 **전반적인 미국 또는 전 세계 ICU 환자 집단을 얼마나 대표하는지**에 대한 설명이 부족하여 **연구 결과의 광범위한 적용 가능성에 대한 신뢰를 약화시킵니다.**

## 제안 사항

1. (Block: 96)에서 MIMIC-IV 데이터베이스의 **환자 인구통계학적 특성(연령, 인종, 질병 중증도, 동반 질환 등)을 구체적으로 설명**하고, 이러한 특성들이 **다른 대규모 ICU 코호트 또는 국가 데이터베이스 (예: (Paper: 216) (Paper: 222)에서 제시된 전 세계 AKI 유병률 연구)**와 어떤 면에서 유사하거나 다른지를 비교하여 **외부 타당성에 대한 주장을 강화**하십시오. 이는 **연구 결과의 임상적 영향력**을 높일 수 있습니다.

2. MIMIC-IV를 활용한 다른 연구들((Paper: 111), (Paper: 131), (Paper: 137), (Paper: 138))에서 데이터의 **대표성을 어떻게 다루었는지** 참고하여, 본 연구의 **일반화 가능성 논의**를 보강하는 것을 고려하십시오. 이는 **연구 결과의 광범위한 적용 가능성에 대한 신뢰**를 높입니다.

13. 결론 (Block: 103)에서 '경험적 병용 요법 관행의 재고를 보증한다'고 강력하게 주장하지만, **특정 취약 계층에 대한 구체적인 권고 사항**이 부족합니다. VPT 병용 요법의 **혜택-위험 균형**이 특히 중요한 환자군에 대해 **어떤 추가적 고려 사항**이 필요한지 명확하지 않아 **임상적 영향력**이 제한적일 수 있습니다.

## 제안 사항

1. (Block: 103)에서 '취약 계층'을 **구체적으로 정의**하고(예: 기존 신장 질환 환자, 고령 환자, 쇼크 환자 등) 이러한 환자군에서 **VPT 병용 요법의 대안 또는 사용 시 특별한 주의 사항**에 대한 **구체적인 임상적 권고 사항**을 추가하십시오. 이는 (Paper: 107)에서 강조된 'AKI 예방을 위한 위험 평가' 및 (Paper: 211)에서 제시된 **반코마이신 치료 환자의 신독성 조기 식별**의 중요성과 맥락을 같이합니다.

2. 예를 들어, '만성 신장 질환 환자 또는 고령 환자에서는 **AKI 위험을 최소화하기 위해 VPT 대신 대체 항생제 (예: 세페픽 등)를 고려하거나, vancomycin trough level 모니터링을 강화해야 한다**'와 같이 **구체적인 대안 제시**를 포함하여 **환자 안전 개선에 직접적인 영향**을 미치도록 하십시오. 이는 **연구 결과의 실질적인 임상적 가치**를 증대시킵니다.

14. (Block: 104)에서 향후 연구 방향으로 '식별된 교란 변수를 통합한 AKI 위험 예측 도구 개발' 및 '향상된 안전성 프로필을 가진 대체 항생제 조합 평가'를 제안했지만, 이러한 **연구가 임상 실습에 어떻게 통합될 수 있는**지, 그리고 **이를 통해 얻을 수 있는 실제적인 이점(예: 비용 절감, 환자 예후 개선)**에 대한 **심층적인 논의**가 부족합니다. 즉, 제안된 연구의 **실질적인 영향력**이 충분히 설명되지 않아 **연구의 장기적 가치 전달**이 미흡합니다.

## 제안 사항

1. (Block: 104)에서 제안된 AKI 위험 예측 도구가 **AKI의 조기 진단 및 예방적 개입에 어떻게 기여할 수 있는**지, 그리고 이를 통해 **환자 사망률 감소, 입원 기간 단축, 의료 비용 절감 ((Paper: 233)) 등 구체적인 임상적 및 경제적 이점**을 어떻게 달성할 수 있는지에 대한 **설명과 정량화된 목표**를 추가하십시오. 이는 **연구의 장기적**



인 임상적 및 경제적 영향력을 명확히 합니다.

2. 대체 항생제 조합 평가에 대해 (Paper: 8), (Paper: 6), (Paper: 4)와 같은 **관련 연구를 인용하여**, 어떤 특정 조합들이 더 유망할 수 있는지 **예시를 제공하고**, 이러한 평가가 **항균제 관리 프로그램 ((Paper: 88))에 미칠 영향**을 강조하십시오. 또한, **LLM 기반 프레임워크가 이러한 미래 연구에서 어떻게 핵심적인 역할을 할 수 있는**지에 대한 **로드맵**을 제시하여 **학술 및 산업적 영향력**을 강화하십시오.

## 심사자 5 (한계점)

### 강점

1. 이 연구는 VPT 요법의 신장독성 위험 증가를 뒷받침하는 **강력한 증거**를 제공하며, 이는 기존 연구의 **방법론적 한계**를 극복합니다. 연구의 규모(9만 명 이상의 환자)는 **통계적 검정력**을 크게 향상시켜, 연구의 **견고성과 신뢰성**을 높였습니다.
2. 저자들은 (Block: 97)에서 **반코마이신 수치 부족**과 같은 **데이터 가용성 제약**을 명확히 인정하며, 이는 **연구 한계점**에 대한 **솔직한 자기 인식**을 보여줍니다.
3. 저자들은 (Block: 98)에서 **미측정 교란 변수**의 가능성을 인정하고, E-값 분석을 통해 그 **강건성**을 정량적으로 평가했습니다. 이는 **교란 변수 통제**에 대한 **깊이 있는 이해와 한계점의 정량적 평가 시도**를 잘 보여줍니다.
4. (Block: 99)에서 LLM 검증의 **잠재적인 오분류 오류**를 인정하고, 이것이 추정치를 **귀무 가설 쪽으로 편향**시킬 수 있음을 논의했습니다. 이는 **방법론적 투명성**과 **한계점의 자기 인식**에 대한 긍정적인 측면입니다.

### 약점

1. (Block: 96)에서 단일 센터 연구의 **일반화 가능성 제한**을 언급하면서도, MIMIC-IV 데이터베이스가 '다양한 환자 집단'을 포함한다는 주장이 **불충분한 근거**로 제시되었습니다. MIMIC-IV 데이터가 ICU 환자 위주로 구성되어 있어, 일반 병동 환자나 다양한 임상 환경으로의 **일반화에 대한 심층적인 논의**가 부족합니다.

### 제안 사항

1. (Block: 96)에 MIMIC-IV 데이터베이스의 **환자 인구통계학적 특성(예: 연령, 성별, 인종, 주요 진단 범주, 중증도 분포)**에 대한 구체적인 통계(예: 비율, 분포)를 추가하여, Beth Israel Deaconess Medical Center가 서비스하는 환자 집단의 **다양성 및 대표성**을 명확히 설명하십시오. 이는 (Paper: 131), (Paper: 112), (Paper: 113) 또는 (Paper: 120)에서 MIMIC-IV 데이터의 배경 및 인구통계학적 특성을 설명하는 방식을 참조하여 보강할 수 있습니다.

2. MIMIC-IV 데이터가 **중증 환자 위주**로 구성되어 있다는 점을 (Block: 96)에 한계점으로 명시하고, 이러한 특성이 일반 병동 환자나 경증 환자에게 연구 결과를 **일반화하는 데 미치는 영향**에 대한 **심층적인 논의**를 추가하십시오. 또한, 이를 보완하기 위한 **미래 연구 방향**으로 다양한 의료 환경(예: 일반 병동, 외래)의 데이터를 활용한 **다기관 연구**의 필요성을 제시하십시오.

2. (Block: 97)에서 **반코마이신 수치**에 대한 접근성 부족이 한계점으로 언급되었으나, 이것이 **치료-결과 관계**에 미치는 **잠재적 편향**에 대한 심층적인 논의가 부족합니다. 특히 반코마이신 수치가 **신장 독성의 직접적인 매개체**일 수 있다는 점에서, 데이터 부족이 **인과 추론의 유효성**에 미칠 수 있는 영향을 더 명확히 설명해야 합니다.

## 제안 사항

1. 반코마이신 수치 부족이 **관찰된 AKI 위험 증가**에 미칠 수 있는 **정량적 또는 질적 편향**에 대해 (Block: 97)에서 보다 상세하게 설명하십시오. 예를 들어, '만약 VPT 그룹 환자들의 반코마이신 수치가 (측정되었다면) 더 높았을 것이고 이것이 AKI를 유발했다면, 이는 관찰된 효과의 **진정한 크기**를 **과대평가**하게 만든다'와 같이 **메커니즘 기반의 편향 분석**을 추가하십시오.

2. 미래 연구 방향으로, 반코마이신 수치와 같은 **미측정 교란 변수** 또는 **매개 변수** 문제를 다루기 위해 (Paper: 243)에서 논의된 **도구 변수(instrumental variable)** 분석이나 (Paper: 41)에서 제시된 **시간 가변 교란 변수** 처리 방법론의 적용 가능성을 (Block: 104)에 구체적으로 언급하여, 이 한계를 다룰 수 있는 **고급 인과 추론 방법론**의 필요성을 강조하십시오.

3. 반코마이신 수치에 대한 정보를 얻기 위해 **LLM 기반의 임상 노트 추출**을 통해 간접적인 정보를 확보하는 방안을 미래 연구 방향으로 (Block: 104)에 제시하십시오. 예를 들어, 'LLM을 활용하여 임상 노트에서 약물 투여 기록, 혈중 농도 측정 요청 여부, 용량 조절 사유, 환자 체중 변화 등 **반코마이신 노출과 관련된 맥락적 단서**를 추출하여 **대리 변수**를 구축하고, 이를 통해 미측정 변수의 영향을 간접적으로 통제하는 방법을 모색할 수 있다'고 제안하십시오. 이는 (Paper: 244)에서 LLM을 활용한 의약품 안전성 감시 강화를 설명하는 방식을 참조할 수 있습니다.

3. (Block: 98)에서 **미측정 교란 변수**에 대한 E-값 분석이 제시되었지만, **E-값 2.12가 임상적으로 의미하는 바**와 그 해석에 대해 더 명확하고 구체적인 설명이 부족합니다. '중간 정도의 미측정 교란 변수'라는 표현은 여전히 **주관적 해석**의 여지를 남깁니다.

## 제안 사항

1. E-값 2.12가 의미하는 바를 (Block: 98)에 더 명확하게 풀어서 설명하십시오. 예를 들어, '관찰된 VPT와 AKI 간의 연관성이 완전히 우연히 발생한 것이거나 측정되지 않은 교란 변수에 의해 설명되려면, 그 미측정 교란 변수는 VPT 사용과 AKI 발생 모두와 **최소 상대 위험 2.12**로 연관되어야 하며, 이는 일반적인 임상 연구에서 예상되는 미측정 교란 변수의 영향보다 **상당히 강한 수준**을 요구한다'와 같이 구체적인 수치적, 임상적 의미를 명시하십시오. 이는 (Paper: 117)에서 E-값 해석의 중요성을 강조하는 방식을 참조하여 독자의 이해를 높일 수 있습니다.

2. (Block: 98)에 E-값 2.12가 **실제로 어느 정도의 교란**을 나타내는지에 대한 **정량적이고 임상적으로 비교 가능한 맥락**을 추가하십시오. 예를 들어, '이는 LLM으로 식별된 **패혈증 중증도(sepsis severity)**와 같은 주요 교란 변수가 VPT 사용 및 AKI에 미치는 영향과 **비교하여 판단할 때**, 미측정 교란 변수가 이보다 더 강한 영향을 미 가져야 함을 의미한다'와 같이 구체적인 비교 기준을 제시할 수 있습니다. 이는 (Paper: 117)의 방법론을 따라 E-값의 의미를 더욱 명확히 할 수 있습니다.

4. LLM 검증 과정에서 발생할 수 있는 **오분류 오류**에 대한 한계가 (Block: 99)에 언급되었으나, 이러한 오류가 **AKI 진단 기준**에 미치는 잠재적 영향에 대한 논의는 부족합니다. 특히 AKI 진단 기준 자체가 혈청 크레아티닌의 **내재적 변동성**으로 인해 **오진 위험**이 있을 수 있다는 점이 고려되어야 합니다. 이는 LLM의 한계와 AKI 진단 기준의 복잡성이 **결합될 때의 영향**을 충분히 다루지 못했습니다.

## 제안 사항

1. LLM으로 추출된 교란 변수의 오분류가 **AKI 진단 정확도**에 어떤 영향을 미칠 수 있는지에 대한 논의를 (Block: 99)에 추가하십시오. 예를 들어, 'LLM이 잘못 분류한 만성 신장 질환(CKD)과 같은 교란 변수는 AKI 진단 기준인 혈청 크레아티닌 수치의 변화를 해석하는 데 **혼란을 야기**할 수 있으며, 이는 **위양성 또는 위음성 AKI 진단**으로 이어질 가능성이 있다'와 같이 설명하여 LLM의 한계와 AKI 진단 기준의 복잡성을 연결하십시오.

2. AKI 진단 자체의 **잠재적 오진 위험**을 인정하고, (Block: 99)에서 LLM 기반 교란 변수 추출의 오류가 AKI 진단 기준의 **내재적 불완전성**과 어떻게 **상호작용**할 수 있는지에 대한 토론을 포함하십시오. 예를 들어, 'KDIGO 기준은 혈청 크레아티닌의 작은 변화에도 AKI를 진단할 수 있지만, 이는 (Paper: 194)에서 지적했듯이 **측정 오차나 생물학적 변동성**으로 인한 위양성 위험을 내포한다. LLM의 오분류는 이러한 **위양성 AKI 진단**을 더욱 증폭시킬 수 있으며, 특히 기저 혈청 크레아티닌이 높은 환자군에서 이러한 오류가 더 심화될 수 있다'고 구체적으로 설명하십시오.

3. 미래 연구 방향으로, LLM이 **AKI 진단 기준의 오차**를 어떻게 보정하거나, **대안적인 AKI 바이오마커** (예: NGAL, KIM-1)를 LLM 기반으로 추출하여 진단의 정확도를 높일 수 있는지에 대한 가능성을 (Block: 104)에 제시하십시오. 예를 들어, 'LLM을 활용하여 임상 노트에서 NGAL (neutrophil gelatinase-associated lipocalin) 또는 다른 신장 손상 바이오마커 관련 텍스트를 추출하고, 이를 KDIGO 기준과 결합하여 AKI 진단의 **조기성과 정확도**를 개선하는 연구가 필요하다'고 구체적인 제안을 할 수 있습니다. (Paper: 262), (Paper: 199), (Paper: 264)에서 NGAL의 진단 유용성에 대한 논의를 참조하여 이 부분을 강화할 수 있습니다.

5. 연구의 주요 기여 중 하나인 LLM을 활용한 교란 변수 식별 방법론에 대한 **미래 연구 방향**이 (Block: 104)에서 'AKI 위험 예측 도구 개발'과 같은 일반적인 언급에 그치고 있습니다. LLM 기반 교란 변수 식별의 **방법론적 발전 가능성**에 대한 **구체적인 로드맵**과 LLM의 **고유한 한계 극복 방안**이 부족합니다.

## 제안 사항

1. LLM 기반 교란 변수 식별 방법론의 **구체적인 개선 방향**을 (Block: 104)에 추가하십시오. 예를 들어, 'LLM의 **'환각(hallucination)'** 문제와 데이터 **편향(data bias)**을 완화하고, 추출의 **정확성과 설명 가능성**을 높이기 위해 **온톨로지 기반 지식 그래프(ontology-based knowledge graph)**와 LLM을 결합하는 연구가 필요하다'고 제안하여 **방법론적 혁신**을 강조하십시오. 이는 (Paper: 245)에서 LLM을 오픈믹스 데이터와 결합하여 암 유전자를 식별하는 방식이나 (Paper: 76)에서 LLM을 통한 미측정 교란 변수 보정 방안을 참조할 수 있습니다.

2. LLM을 활용한 **미관측 교란 변수(unobserved confounder)** 식별 또는 **대리 변수(proxy variable)** 탐색에 대한 미래 연구 가능성을 (Block: 104)에 언급하여, 현재 연구의 한계를 극복할 수 있는 **방법론적 진보**를 제시하십시오. 예를 들어, 'LLM이 임상 노트에서 미관측 교란 변수를 나타낼 수 있는 **맥락적 단서**

(contextual cues)를 식별하거나, 알려지지 않은 대리 변수를 제안하는 데 활용될 수 있다'고 설명할 수 있습니다. 이는 (Paper: 77), (Paper: 74), (Paper: 76)에서 LLM 기반의 교란 변수 발견 및 impute에 대한 논의를 참조하여 LLM 활용의 심층적인 가능성을 보여줄 수 있습니다.

3. LLM 기반 교란 변수 식별의 **다중 언어 환경 적용** 또는 **다양한 의료 데이터 소스 통합**과 같은 **확장성 측면**에서의 미래 연구를 제안하십시오. 예를 들어, 'MIMIC-IV 외에 다른 국가의 EHR(Electronic Health Record) 데이터셋에 LLM 기반 교란 변수 식별 프레임워크를 적용하여 **다국적 일반화 가능성**을 평가하고, 이를 통해 **글로벌 약물감시 시스템** 구축에 기여하는 연구가 필요하다'고 제시하여 **방법론의 범용성**을 강조하십시오. 이는 (Paper: 244)에서 LLM이 의학 연구 및 약물 감시에 기여할 수 있는 광범위한 영역을 설명하는 방식을 따를 수 있습니다.

6. (Block: 104)에서 'AKI 위험 예측 도구 개발' 및 '대안적인 항생제 조합 평가'가 미래 연구로 제시되었지만, 이러한 연구들이 **실제 임상 환경**에서 어떻게 적용될 수 있는지에 대한 구체적인 계획이나 **'실제 세계 증거 (Real World Evidence, RWE)'**와의 연계성에 대한 논의가 부족합니다. 이는 연구의 **실질적인 임상적 영향력**을 약화시킬 수 있습니다.

## 제안 사항

1. AKI 위험 예측 도구가 **임상 의사 결정 시스템**에 통합되거나, **항생제 스튜어드십 프로그램**에 적용되는 방안 등, 실제 임상 적용을 위한 **구체적인 로드맵**을 (Block: 104)에 추가하십시오. 예를 들어, '개발된 AKI 위험 예측 도구를 EHR 시스템에 내장하여, VPT 처방 시 **실시간 AKI 위험 점수**를 제공하고, 이를 바탕으로 **항생제 변경 또는 용량 조절 권고**를 제시하는 **임상 의사 결정 지원 시스템(CDSS)** 개발이 필요하다'와 같이 상세한 적용 방안을 제시하십시오. 이는 (Paper: 272)에서 항생제 스튜어드십의 중요성을 설명하는 연구와 연계될 수 있으며, (Paper: 277), (Paper: 278)에서 설명하는 약물감시 및 약물 역학의 관점을 통합할 수 있습니다.

2. 미래 연구가 **다양한 환자 집단**과 **의료 시스템**에서 VPT 조합 요법의 AKI 위험을 평가하기 위해 **'실제 세계 데이터(Real World Data, RWD)'**를 어떻게 활용할 것인지에 대한 논의를 (Block: 104)에 포함하십시오. 예를 들어, '향후 연구에서는 MIMIC-IV 데이터 외에 **다양한 병원 네트워크** 또는 **국가 등록 데이터**를 활용하여 VPT 관련 AKI 위험을 평가하고, 특히 **취약 집단(예: 고령 환자, 중증 만성 질환자)**에서의 효과를 검증하는 RWE 연구가 중요하다'고 설명할 수 있습니다. 이는 (Paper: 221), (Paper: 223), (Paper: 225)에서 RWE의 중요성을 설명하는 연구들을 인용하여 연구의 **외부 타당성**을 강화할 수 있습니다.

3. AKI 위험 예측 도구 개발과 관련하여, (Block: 104)에 **AKI 진단 기준의 지역적, 인구통계학적 다양성**을 고려하는 연구 방향을 추가하십시오. 예를 들어, '전 세계적으로 상이한 의료 자원과 환자 특성을 반영하여, KDIGO 기준 외에 **다양한 AKI 진단 기준** (예: pRIFLE, AKIN) 및 **새로운 바이오마커**를 활용한 AKI 위험 평가 모델을 개발하고, 이를 **다국가 RWE 연구**를 통해 검증하는 것이 필요하다'고 제안할 수 있습니다. 이는 (Paper: 216) 및 (Paper: 258)에서 AKI 진단 기준의 다양성을 논의하는 내용을 참조하여 미래 연구의 **일반화 가능성**을 높일 수 있습니다.

7. (Block: 103)에서 VPT 조합 요법의 **경험적 조합 치료 관행 재고**를 주장하면서도, 실제 임상 환경에서 의사들이 해당 조합을 선택하는 다양한 **'비측정 교란 변수(unmeasured confounders)'**에 대한 **더 깊은 논의**가 부족합니다. 현재 'indication bias'에 대한 기존 논의를 더욱 확장시켜야 합니다.

## 제안 사항

1. (Block: 103)에 의사들이 VPT 조합을 선택하는 **임상적 판단의 복잡성**에 대한 논의를 추가하여, '비측정 교란 변수'의 유형과 그 영향을 구체화하십시오. 예를 들어, 'VPT 조합을 처방하는 결정은 종종 **미분화된 패혈증, 잠재적 다제내성 병원체에 대한 높은 의심, 특정 병원 내 감염 역학, 또는 빠른 경험적 치료의 필요성**과 같은 요인들에 의해 좌우되며, 이러한 요소들은 현재 연구에서 완전히 포착되지 않은 **강력한 교란 변수**로 작용할 수 있다'고 설명하십시오. 이는 (Paper: 10) 등에서 언급되는 임상적 현실을 반영할 수 있습니다.

2. 미래 연구 방향으로, '비측정 교란 변수'의 영향을 최소화하기 위한 **고급 인과 추론 방법론**의 적용 가능성을 (Block: 104)에 구체적으로 제시하십시오. 예를 들어, '향후 연구에서는 **인과 그래프(Causal Directed Acyclic Graphs, DAGs)**를 사용하여 잠재적 교란 변수 구조를 시각화하고, **경사 점수 매칭/가중(propensity score matching/weighting)** 또는 **이중 강건 추정(doubly robust estimation)**과 같은 방법론을 더욱 정교하게 적용하여, **비측정 교란 변수에 대한 감수성 분석(sensitivity analysis)**을 강화하는 연구가 필요하다'고 제안하십시오. 이는 (Paper: 47), (Paper: 53), (Paper: 238)에서 인과 추론 방법론의 중요성을 설명하는 내용과 (Paper: 116), (Paper: 180) 등에서 이중 강건 추정의 강점을 논의하는 내용을 참조할 수 있습니다.

3. LLM 기반의 교란 변수 식별 프레임워크를 **확장**하여, '비측정 교란 변수'의 **대리 변수(proxy variables)**를 임상 노트에서 탐색하고 활용하는 방안을 (Block: 104)에 포함하십시오. 예를 들어, 'LLM이 임상 노트의 자유 텍스트(free text)에서 **의사의 직관적 판단, 초기 감염 진단의 불확실성, 또는 특정 병원 방침**과 같은 비측정 교란 변수를 간접적으로 나타내는 **언어적 패턴이나 키워드**를 식별하여, 이를 교란 변수 모델에 통합하는 연구가 필요하다'고 제안하십시오. 이는 (Paper: 76), (Paper: 79)에서 LLM을 통한 미관측 교란 변수 탐색 가능성을 제시하는 연구와 연계될 수 있습니다.

8. 이 연구는 2008-2019년의 MIMIC-IV 데이터베이스를 사용했는데, 이는 **연구 기간의 제한**으로 인해 **최신 임상 가이드라인이나 약물 사용 패턴의 변화**를 반영하지 못할 수 있다는 심층적인 한계점을 다루지 못했습니다. 이는 연구 결과의 **현재 임상적 유용성**에 대한 의문을 제기합니다.

## 제안 사항

1. (Block: 96) 또는 (Block: 103)에 연구 기간(2008-2019)의 **제한성**을 명확히 언급하고, 이 기간 동안 **반코마이신 및 피페라실린/타조박탐의 처방 관행, 약물감시 가이드라인, AKI 진단 및 관리 프로토콜**의 변화가 연구 결과에 미칠 수 있는 영향을 논의하십시오. 예를 들어, '연구 기간 이후 도입된 **최신 AKI 예방 및 관리 전략**(예: (Paper: 232) KDIGO 가이드라인)이 VPT 관련 AKI 발생률을 낮추었을 가능성이 있다'와 같이 설명하십시오.

2. 미래 연구 방향으로, **최신 MIMIC-IV 데이터 버전** (예: v2.2) 또는 다른 **실시간 대규모 EHR 데이터베이스**를 활용하여, **최근 임상 관행의 변화**가 VPT 관련 AKI 위험에 미치는 영향을 평가하는 연구의 필요성을 (Block: 104)에 제시하십시오. 특히 '**실제 세계 데이터(RWD)**'를 통해 **최신 약물 사용 패턴과 AKI 발생률**의 변화를 추적하는 연구를 제안하십시오. 이는 (Paper: 131), (Paper: 136), (Paper: 137)에서 MIMIC-IV의 최신 버전 활용을 설명하는 내용과 (Paper: 217), (Paper: 218), (Paper: 220)에서 RWD/RWE의 중요성을 강조하는 연구를 참조할 수 있습니다.

3. 연구 결과가 현재의 임상 환경에서 가지는 **유효성 범위**를 (Block: 103)에 명확히 정의하고, 과거 데이터 기반 연구의 **임상적 적용 시 주의사항**을 강조하십시오. 예를 들어, '본 연구 결과는 2008-2019년의 임상 데이터를 기반으로 하므로, **최신 가이드라인 및 약물 치료 프로토콜**이 적용된 현재 임상 환경에 직접 적용할 때는 **추가적인 검증**이 필요하다'고 명시하십시오.