Validation of an Artificial Intelligence (AI) tool to identify PICO questions for EU Joint Clinical Assessment (JCA)

HTA64

Andrew Mumford[1,2], Georgia Roberts[1], Joanne Mumford[1],

1. Initiate Consultancy, London, UK. 2. Initiate Consultancy, Zug, Switzerland.

Study sponsored by Initiate Consultancy Ltd.

SUMMARY

OBJECTIVES

- The introduction of Joint Clinical Assessment (JCA) in Europe poses significant challenges for pharmaceutical developers, particularly in demonstrating the value of treatments against all existing management strategies across European member states.
- A key preparatory step for a successful JCA review is the provision of comparative data, either directly collected or synthesised. Given strict timelines, developers must ensure early mapping of PICO (Population, Intervention, Comparator, Outcomes) questions to facilitate evidence synthesis.
- An AI-powered algorithm, the **PICORadar® tool,** has been developed to identify PICO questions, but it requires validation against traditional targeted searches.

METHODS

- Three orphan diseases – cystinosis, urea cycle disorders, and tyrosinemia – were selected to test the algorithm. These diseases are all inborn errors of metabolism.
- A targeted review was conducted for each disease to determine PICO questions in each EU member state, which were then compared to those generated by the AI algorithm.
- The time spent developing PICO questions was recorded, with a maximum of 4 hours allocated for targeted searching across all 27 EU member states. For validation, the "intervention" in PICO was regarded as a new intervention.

FINDINGS

- The largest differences in PICO formulation were observed in the comparator (C) category.
- Both methods identified common elements for the patient (P) and outcome (O) categories. However, targeted searches failed to identify 25–33% of comparator strategies highlighted by the algorithm.
- Each AI-driven search was completed in under two minutes, with an additional 60 minutes per disease required for consistency checks.
- These findings suggest the algorithm achieves higher accuracy than traditional methods while significantly reducing time requirements.

BACKGROUND & AIMS

- The implementation of Joint Clinical Assessment (JCA) in Europe presents significant challenges for pharmaceutical developers. One major hurdle is demonstrating the value of new treatments compared to all existing management strategies across diverse European healthcare systems. This requires a deep understanding of current treatment protocols in each country.
- A crucial preparatory step is providing comparative data, either directly collected from clinical trials or synthesized from existing sources. This data is essential for a thorough JCA review, allowing detailed comparisons of new treatments against established methods.
- Given the strict timelines, developers must start early with mapping PICO (Population, Intervention, Comparator, Outcomes) questions to facilitate evidence synthesis. The PICO framework helps structure research questions and guide evidence collection efficiently.
- An AI-powered algorithm has been developed to identify PICO questions, streamlining the process by automatically pinpointing relevant questions from large data sets. However, this algorithm still needs validation against traditional targeted searches to ensure its accuracy and reliability.
- In summary, the introduction of JCA in Europe requires meticulous preparation and strategic planning. By providing robust comparative data and using advanced AI algorithms for PICO question identification, developers can better navigate the JCA review process and demonstrate the value of their treatments in a competitive healthcare landscape.

METHODS

- Three orphan diseases—cystinosis, urea cycle disorders, and tyrosinemia—were chosen to evaluate the algorithm. These diseases are all classified as inborn errors of metabolism, which are rare genetic disorders affecting the body's ability to metabolize certain substances. For each disease, a targeted review was conducted to identify PICO (Population, Intervention, Comparator, Outcomes) questions specific to each EU member state. These questions were then compared to those generated by the AI algorithm to assess its effectiveness.
- The process involved meticulously reviewing existing literature and clinical guidelines to formulate PICO questions for each disease in all 27 EU member states. The time spent on developing these PICO questions was carefully recorded, with a maximum of 4 hours allocated for targeted searching across all member states. This time constraint was set to simulate real-world conditions and ensure the feasibility of the approach in practical settings.
- For the purpose of validation, the "intervention" component of the PICO framework was considered as a new intervention. This allowed for a direct comparison between the traditional targeted search method and the AI algorithm in identifying relevant PICO questions. The goal was to determine whether the AI algorithm could match or exceed the accuracy and efficiency of manual searches.
- The results of this study are crucial for validating the AI algorithm's capability to streamline the PICO question identification process. By comparing the AI-generated questions with those derived from targeted reviews, researchers can assess the algorithm's reliability and potential to enhance the efficiency of evidence synthesis in clinical assessments. This validation step is essential to ensure that the AI tool can be confidently used in future clinical assessments, ultimately aiding in the development and evaluation of new treatments for rare metabolic disorders.

RESULTS

Table 1: PICOs developed from AI and Human input. **BOLD** shows additional information provided by AI tool

	Patient (P)	Intervention(I)	Comparator(C)	Outcome(O)
Urea Cycle Disorders	• Adults and Children • Non acute UCD patients • Not in HAC (Hyperammonaemic Crisis) • **Individuals with Urea Cycle Disorder (UCD) across European Union member states (including paediatric and adult populations)**	• Theoretical intervention of interest	• Ravicti® • Low protein diet • Haparesc® • Ammunol® • Ammonaps® • **Liver Transplantation**	• Reduction in blood ammonia levels • Prevention of Hyperammonaemic episodes and associated complications • **Improvement in neurological outcomes (e.g., cognitive function, seizure frequency)** • **Survival rates and hospitalizations** • **Long-term quality of life (e.g., physical, cognitive, and social functioning)**
Cystinosis	• Adults with Cystinosis • **Adults and children with Cystinosis**	• Theoretical intervention of interest	• cysteamine hydrochloride • Cystagon® • **Cysteamine ophthalmic gel** • **Renal Support**	• Reduction in Cystine levels • **Kidney function** • **Improvement in ocular health** • **Growth and development** • **Improved QoL**
Tyrosinemia	• Adults and children with Tyrosinemia	• Theoretical intervention of interest	• Orfadin® • Low Protein diet • **Nitisinone** • **Liver transplant** • **Amino acid supplementation**	• Reduction of tyrosine and phenylalanine • **Prevention of liver damage** • **Improvement of neurological outcomes**

- The most significant differences in PICO formulation were observed in the comparator (C) category. While both the AI algorithm and traditional targeted searches identified common elements for the patient (P) and outcome (O) categories, the targeted searches fell short in identifying 25–33% of the comparator strategies that the algorithm highlighted. This discrepancy underscores the algorithm's ability to uncover a broader range of comparator strategies, which is crucial for comprehensive clinical assessments.
- **Table 1** above shows the data provided, with information in **bold** that the AI tool provided in addition to human searching.
- Each AI-driven search was remarkably efficient, completing in under two minutes. This rapid processing time is a stark contrast to the traditional method, which is considerably more time-consuming. Additionally, an extra 60 minutes per disease was required for consistency checks to ensure the accuracy and reliability of the AI-generated PICO questions. Despite this additional step, the overall time required for the AI-driven approach was significantly less than that of traditional methods.
- These findings suggest that the AI algorithm not only achieves higher accuracy but also significantly reduces the time required for PICO formulation. The ability to quickly and accurately identify relevant comparator strategies is particularly valuable in the context of clinical assessments, where timely and precise data is essential. By leveraging the AI algorithm, researchers and developers can streamline the evidence synthesis process, making it more efficient and comprehensive.
- In summary, the AI algorithm's superior performance in identifying comparator strategies and its rapid processing time highlight its potential to revolutionize the PICO formulation process. This advancement could lead to more accurate and timely clinical assessments, ultimately benefiting the development and evaluation of new treatments.

CONCLUSIONS

- The AI-powered algorithm demonstrates potential as a more accurate and resource-efficient alternative to traditional methods. However, further validation is necessary.
- It should be used alongside human-led searches to ensure robustness, with subsequent testing of PICO outputs by clinicians in relevant markets to confirm accuracy.
- **The PICORadar® appears to be an accurate and useful tool in mapping relevant PICOs for EU JCA planning**

Presented at ISPOR, 13 – 16 May 2025

initiateconsultancy.com hello@initiateconsultancy.com