GENERATIVE AI TO IDENTIFY COMPARATIVE CRITIQUES IN HTA REPORTS FROM MULTIPLE COUNTRIES

Turgay Ayer[1,2], Tugce Ermis[1], I. Fatih Yildirim[1], Deniz Balta[1], Sumeyye Samur[1], Jagpreet Chhatwal[1,3]

[1] Value Analytics Labs, Boston, MA, USA, [2] Georgia Institute of Technology, Atlanta, Georgia, USA, [3] Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA,

KEY FINDINGS

The deployment of Generative AI for the analysis of HTA reports demonstrates a significant potential to streamline the extraction and synthesis of comparative critiques.

ValueGen.AI's ability to process multilingual documents and produce comprehensive summaries underscores its value as a crucial tool for navigating the complex landscape of global health technology assessments.

By leveraging Generative AI, stakeholders in healthcare economics and outcomes research can enhance their decision-making processes, ensuring a more informed, data-driven approach to assessing medical interventions.

BACKGROUND

- Health Technology Assessment (HTA) reports are pivotal for evaluating the clinical and economic value of health interventions. These reports, generated by various national agencies, offer comprehensive insights into treatment effectiveness, safety, and cost-effectiveness.
- However, manual review and cross-analysis of HTA reports across different countries is a labor-intensive and time-consuming process, posing challenges for health economists, policymakers, and decision-makers.
- The emergence of Generative AI presents an opportunity to revolutionize the analysis of these reports by automating data extraction, synthesis, and comparative critique identification.

OBJECTIVE

- To explore the potential of Generative AI to efficiently process HTA reports from multiple countries, extract relevant critiques, and provide synthesized insights, thereby improving the review process and supporting informed healthcare decisions.
- Specifically, we focused on assessing how AI could be leveraged to identify shared and unique critiques related to the evaluation of tofacitinib for the treatment of ulcerative colitis.

METHODS

We developed and deployed a customized Generative AI tool, named ValueGen.AI1, designed to conduct comprehensive analyses of HTA reports. Our methods included several key phases:

Data Source: We utilized 13 HTA reports from three major agencies: the National Institute for Health and Care Excellence (NICE)2 in the United Kingdom, the Haute Autorité de Santé (HAS) 3 in France, and the Gemeinsamer Bundesausschuss (G-BA) 4 in Germany.

Text Processing: The AI framework utilized language-agnostic capabilities, enabling the processing of documents written in English and native languages (French and German). This functionality was vital for ensuring that the AI accurately extracted critiques in different languages without requiring translations that might lose nuances.

Data Extraction and Synthesis: We employed GPT-4 integrated with Python's LangChain5 library for comprehensive data extraction and synthesis. The AI model was tasked with identifying critique themes, identifying unique and shared aspects, and summarizing key points raised in the reports.

Data Conversion: Extracted data were formatted using LaTeX templates through Jinja6 to ensure compatibility for structured reporting.

Data Summarization: MiKTeX7 was utilized to compile summaries into PDF documents, enhancing the clarity and usability of the output.

RESULTS

1. Comparative Critiques by Agency

ValueGen.AI synthesized critiques from NICE, HAS, and G-BA, highlighting shared and unique observations (Figure 1). This cross-comparison enabled a clearer understanding of shared concerns as well as agency-specific perspectives.

RESULTS (cont.)

Figure 1. *Summary of Shared and Unique Critiques by HTA Agency*

CURRENT MANAGEMENT			
Critiques against current management include not fully complying with licensing, inadequacies in providing detailed information on comparator therapy, and adjustments not considered in dossiers.	●	●	●
Classification, reporting, and analysis of adverse events in studies were highlighted as concerns.	●		
TRIAL DESIGN			
Absence of adequate comparisons and direct comparative trials with relevant competitors.	●	●	
Weaknesses in study design, data collection, and analysis methodologies that could impact the reliability and applicability of the data.	●	●	●
Concerns regarding the accuracy of reporting trial outcomes and analysis methods, and the need for clearer communication and justification of trial design choices.	●		
Methodological robustness concerning bias potential, but some limitations or concerns were not fully detailed.	●	●	
COMPARATORS			
Absence of direct comparative data with key competitors, limitations, and potential areas of improvement in the choice and justification of comparators used.	●	●	
UNMET NEEDS			
Limited additional impact on public health, positioning as a later-line treatment option, and challenges in substantiating significant unmet needs, especially in comparison to existing therapies.		●	
Substantial symptom burden and disability risk with current therapies, expressing a high unmet need in terms of treatment efficacy, safety, and maintenance of improvements in health-related quality of life.	●		
EFFICACY			
Absence of robust comparative data, making it challenging to accurately assess the drug's efficacy relative to other established treatments.	●	●	
Highlighted limitations of current therapeutic options and the novel contributions of the drug under review in addressing treatment goals effectively.	●		
SAFETY			
Increased risk profile concerning tolerance and significant risks compared to alternatives, alongside contraindications and the need for careful monitoring.		●	
Tofacitinib's safety profile discussed extensively with mention of substantial clinical data supporting its use but highlighted common adverse events, especially infections.	●		
QoL			
Unmet needs specifically pointed out the considerable symptom burden and the high risk of disability, indicating an ongoing struggle with maintaining quality of life despite current therapies.	●		
ECONOMIC ANALYSIS			
Concerns over the choice and justification of comparators used in economic analysis and limitations in including all relevant comparators for a comprehensive evaluation.	●		

2. Insights on Common Critiques

The review of HTA reports revealed several recurring critiques across agencies regarding tofacitinib's assessment for ulcerative colitis:

Trial Design: A predominant critique noted by all three agencies pertained to the design and methodology of clinical trials. Concerns were raised about data collection methods and reliability, suggesting a need for more robust trial frameworks.

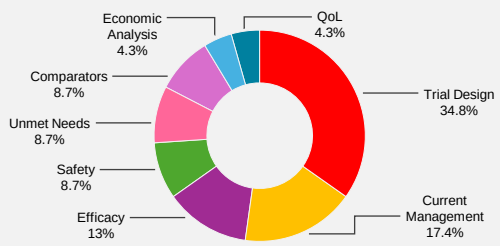
Comparator Arm: All three agencies pointed out limitations related to insufficient details on the comparator arm used in clinical studies, which undermined the robustness of comparative analyses.

Safety: NICE uniquely raised issues around safety and adverse events, while HAS focused on public health impact.

Economic Analysis: Only NICE raised concerns about the economic analysis conducted.

The majority of the critiques were related to trial design, which was followed by the current management, including lack of details on the comparator arm or analysis of adverse events (Figure 2).

Figure 2. *Critique Distribution Across Agencies*



REFERENCES

[1] ValueGen.AI, https://valuegen.ai/
[2] NICE, https://www.nice.org.uk/
[3] HAS, https://www.has-sante.fr/jcms/pprd_2986129/en/home
[4] G-BA, https://www.g-ba.de/
[5] LangChain, https://python.langchain.com/docs/how_to/qa_citations/?form=MG0AV3
[6] Jinja, https://jinja.palletsprojects.com/en/stable/
[7] MiKTeX, https://miktex.org/

Contact: Jagpreet Chhatwal, PhD: jagchhatwal@mgh.harvard.edu