Poster: MSR218

Faster, Mostly Faithful: Can AI Reproduce Manual PICO Extraction in Preparation for EU JCA?

F. SHOUGHARI[1], P. ADAMS[1], A. CICHEWICZ[2], L. COX[1], M. SOOD[1] and K. KALLMES[2]

[1]Acumetis; [2]Nested Knowledge

INTRODUCTION

Under the EU HTA Regulation, new oncology medicines and ATMPs have undergone **Joint Clinical Assessment (JCA)** since January 2025.[1]

During scoping, the PICOs requested by each Member State are consolidated into a single assessment scope, which determines the comparisons and outcomes the manufacturer must provide evidence for.[2] Anticipating these PICOs early is critical for evidence generation and indirect comparison planning.

This currently relies on manual extraction of PICO elements from HTA reports and clinical guidelines, a labour-intensive process repeated for every asset.

AI tools are increasingly used to gain efficiency in evidence synthesis, but their accuracy compared to expert PICO extraction for JCA preparation is not yet well characterised.[3]

OBJECTIVE

Use a second-line NSCLC asset as a case study to assess whether complete AI extraction can reproduce manual PICO extraction from HTA reports and clinical guidelines, identify where they diverge, and compare turnaround time.

AI EXTRACTION SET-UP

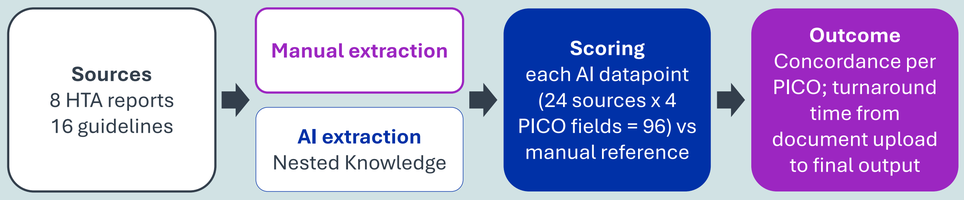
Tool: Nested Knowledge's Adaptive Smart Tags, an LLM-based module that extracts user-defined data from full-text documents

Configuration: Tag table with one column per PICO element, including individual column prompts written by the study team

Model: OpenAI GPT-4.1

METHOD

PICO elements relevant to a second-line NSCLC asset were extracted from 8 HTA reports and 16 clinical guidelines sources across several European countries in various languages. **Manual extraction** was conducted by 2 analysts and underwent an external review. **AI extraction** was run on the same full texts using defined prompts for the Nested Knowledge's Adaptive Smart Tags.



RESULTS

88%

of HTA report datapoints concordant (28/32)

44%

of guideline sources concordant (7/16), driven by scope mismatch

~40×

faster: two hours with AI vs two weeks manually

National HTA reports

28/32 datapoints (87.5%) were concordant (Figure 1). Intervention matched perfectly; population and comparator were strong, with gaps where AI truncated comparator lists (France, Germany) or dropped population qualifiers.

The single true error was one report where AI replaced the clinical endpoint (OS) with economic outcomes (Table 1).

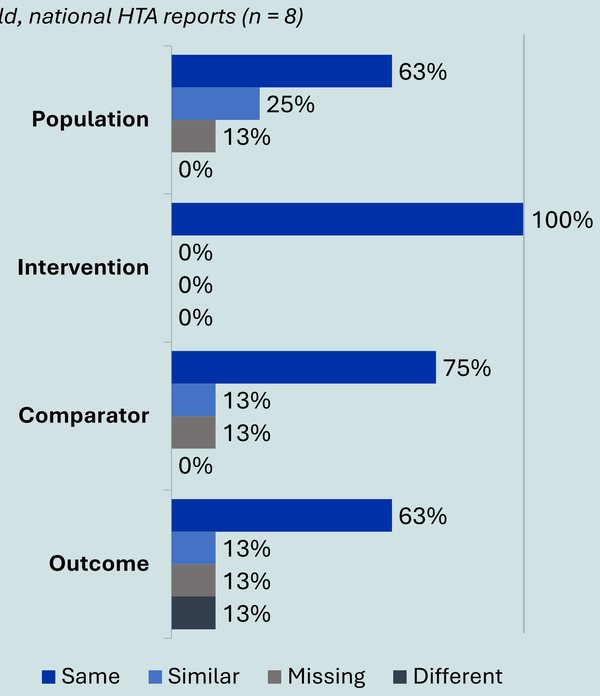
Clinical guidelines

7/16 sources (44%) were concordant (Figure 2). This mainly reflected **scope mismatch**: the manual reference was restricted to second-line-plus data, whereas AI prompts did not specify line of therapy and returned data from the full guidelines (Table 1).

Turnaround

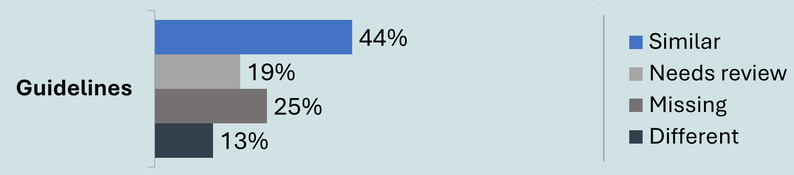
Manual extraction took **two weeks** (~80 person-hours); AI took **two hours**, covering document upload, prompt development and extraction, approximately 40-fold faster.

Figure 1: Concordance of AI with manual extraction by PICO field, national HTA reports (n = 8)



Note: None of the AI HTA report extractions were scored as "Needs review"

Figure 2: Overall concordance of AI with manual extraction, clinical guidelines (n = 16)



Note: None of the AI clinical guidelines extractions were scored as "Same"

Table 1: Where AI and manual extraction diverged, and how to close the gap

Source	Type	What AI did differently	Proposed refinement
HTA: population	Omission	Dropped population qualifiers	Prompt for the full population definition, including qualifiers
HTA: comparator	Omission	Truncated comparator lists (France, Germany)	Prompt to extract each comparator individually
HTA: outcomes	Error	Reported economic outcomes instead of OS	Specify clinical outcomes only in prompts and provide examples, explicitly excluding economic outcomes
Guidelines: population	Scope	Returned the full guideline; reference covered 2L+ only	Specify line of therapy in prompts
Guidelines: comparators	Scope	Returned comparators from all populations in the guideline	Filter by population and line of therapy before extracting comparators

CONCLUSIONS

AI closely reproduced manual PICO extraction from HTA reports while reducing turnaround by ~40-fold.

- **Most discordance was omission of detail or context.** AI rarely reported incorrect content but sometimes missed comparators or population qualifiers. Source traceability makes it quick to verify what AI extracts, but not what it misses, so human review should focus on completeness.
- **Lower concordance on guidelines largely reflected prompt scope.** Prompts did not specify line of therapy, so AI returned data beyond the 2L+ reference. Line-specific prompting is essential for JCA scoping.
- **These results support AI as a first-pass engine for JCA PICO extraction,** shifting human effort to verification and strategic review.

Abbreviations: 2L+, second line and beyond; AI, artificial intelligence; ATMP, advanced therapy medicinal product; EU, European Union; HTA, health technology assessment; JCA, Joint Clinical Assessment; LLM, large language model; NSCLC, non-small cell lung cancer; OS, overall survival; PICO, population, intervention, comparator, outcome.

REFERENCES

1. Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on Health Technology Assessment and Amending Directive 2011/24/EU (Text with EEA Relevance). Vol 458. 2021. Accessed September 2026. http://data.europa.eu/eli/reg/2021/2282/oj
2. Guidance on the scoping process - Public Health - European Commission. Accessed September 2026. https://health.ec.europa.eu/publications/guidance-scoping-process_en
3. General principles on the use of Artificial Intelligence in the preparation of dossiers for Joint Clinical Assessment. Accessed September 2026. https://health.ec.europa.eu/document/download/ca29446a-654d-48e9-aac0-dcaa500a1d2c_en?filename=hta_20260716_jca_ai-gen-principles_en.pdf

CONTACT INFORMATION

Want to know more? Fadel Shoughari: fadel.shoughari@acumetisglobal.com