ARYSANA

HTA43

Artificial Intelligence-based Population, Intervention, Comparator, and Outcome (PICO) Prediction for European Union (EU) Joint Clinical Assessment (JCA)

CONTACT US

Gavin J. Outteridge, MA[1], Justin Yu, PharmD, MS[1], Catherine Chamoux, PharmD[2].
[1]Arysana Inc, NC, USA; [2]AESARA Europe, Paris, France.

INTRODUCTION

The European Union (EU) Health Technology Assessment (HTA) Regulation[1] has been in force in Europe since January 2025 for new oncology drugs and advanced therapy medicinal products (ATMPs). It introduced Joint Clinical Assessments (JCA) using the PICOs (Population, Intervention, Comparator, and Outcome) methodology. Each country is responsible for establishing its list of PICOs for each drug to be assessed and the scope of the dossier established by the EU HTA body is a consolidated list of these. The PICOs are based on unmet needs in relation to their corresponding disease states.

As PICOs are the ‘heart’ of EU HTA assessments, it is very important to anticipate what they could be and to identify relevant data to answer these questions. This can be done manually by looking at all available HTA reports in the individual countries or using an automated tool that provides a fast, first approach to generating this list. Given the frequent need for timely PICOs prediction, ARYSANA subsequently developed a multi-agent workflow (‘PICOs Predictor’) that utilizes OpenAI o3 and OpenAI GPT-4o large language models (LLMs) to enable the automated extraction of PICOs from national HTA reports of drugs. The ‘PICOs Predictor’ was previously refined by iteratively comparing its outputs with target information manually extracted by a human HTA expert from reports originating from >10 EU countries. Code updates, followed by retesting on all documents, were made until 100% agreement with the HTA expert was achieved.

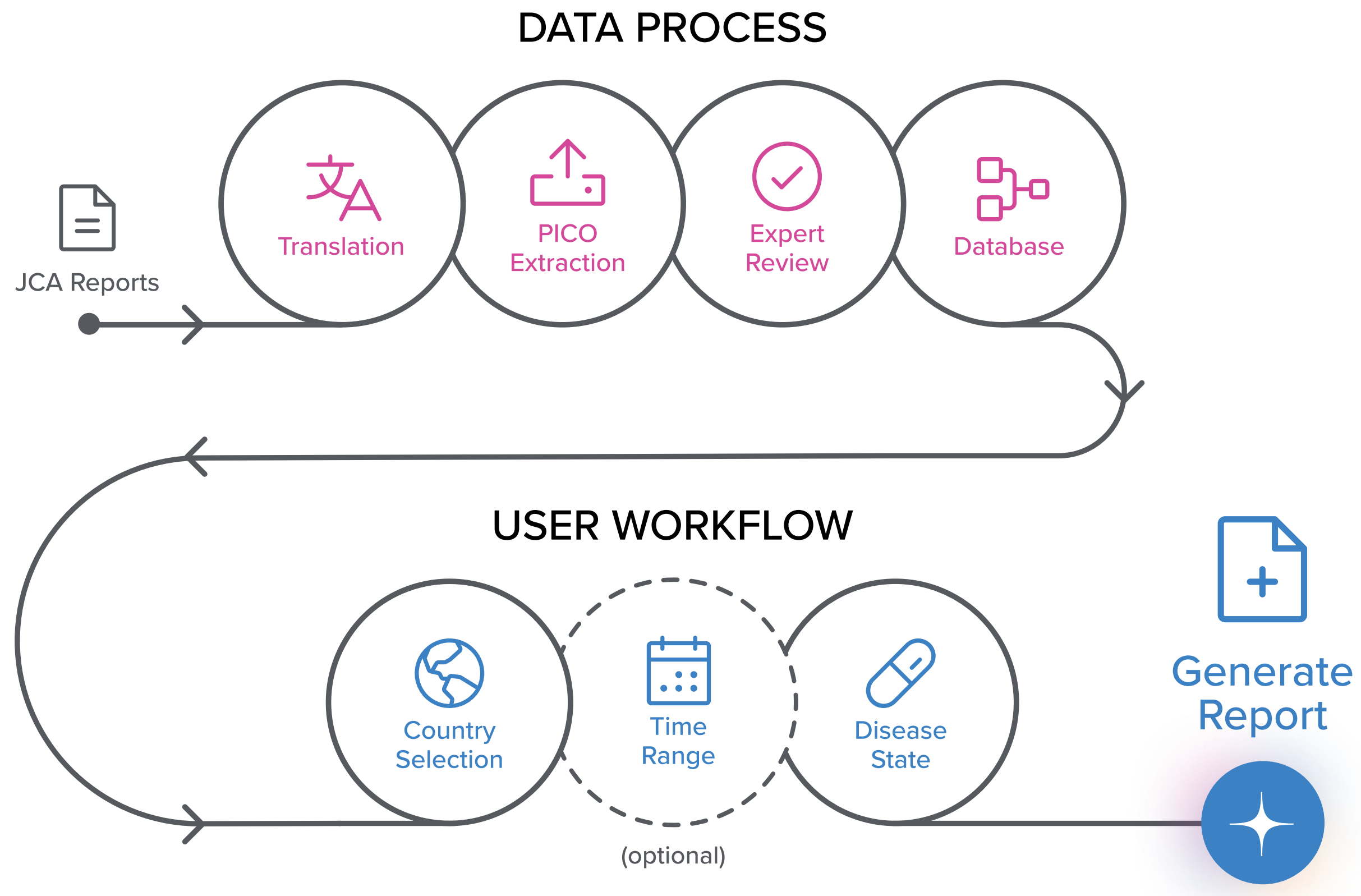
OBJECTIVES

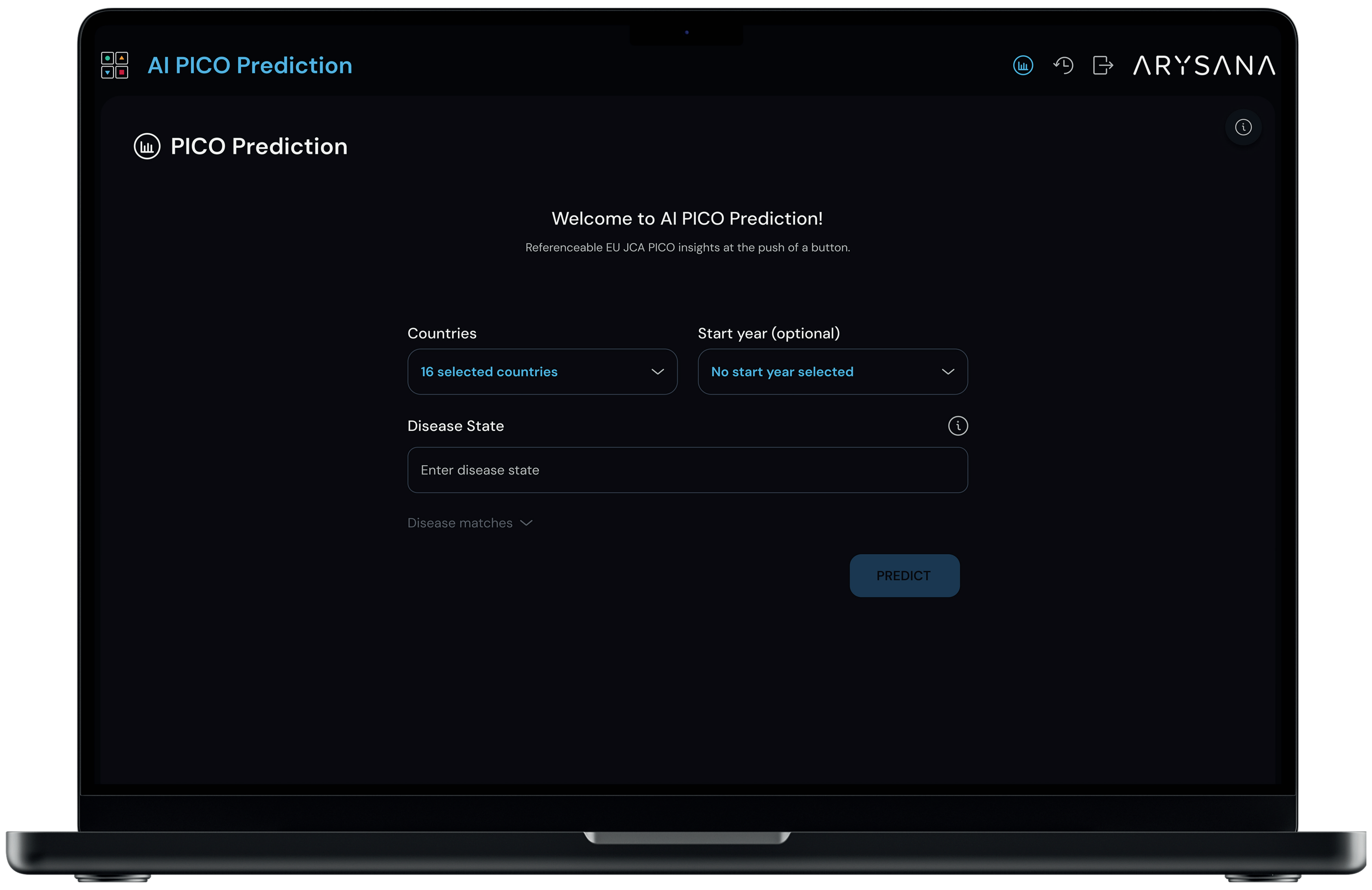
The objective of this work was to compare the results of the “PICOs Predictor” tool to those of leading LLMs (i.e., OpenAI o3, OpenAI GPT-4o, and Claude Sonnet 4), at the time of data collection, on the Population and Comparator items in terms of accuracy and completeness.

MATERIALS & METHODS

We chose 3 countries where there are multiple HTA reports published (France, Germany, Spain) and metastatic Colorectal Cancer (mCRC) as the target indication. The information was extracted from reports on individual drugs issued between Jan. 2010-Aug. 2025: 19 from HAS (France), 9 from G-BA (Germany), and 8 from AEMPS (Spain). All reports were translated to English using Google Translate.

“PICOs Predictor”’ outputs, generated from multiple prompts within country-specific workflows, on Populations and Comparators were compared to those from GPT-4o (non-reasoning), OpenAI o3 (reasoning), and Sonnet 4 (hybrid reasoning) using one-shot prompts (i.e., detailed prompts comprised of the same content from all prompts used in the corresponding multi-agent workflow) and with initial validation focused on HAS reports. All outputs were analyzed manually by an HTA expert in terms of accuracy (complete population and comparators descriptions) and deviations grouped by nature: duplicate populations (same population with different wording), wrong statements (population not in scope, treatment combination missing for example), or incomplete statements (ECOG missing for ex).





PICO Prediction

Welcome to AI PICO Prediction!

Referenceable EU JCA PICO insights at the push of a button.

Countries: 16 selected countries

Start year (optional): No start year selected

Disease State: Enter disease state

Disease matches

PREDICT

RESULTS

“PICOs Predictor” accurately identified populations and comparators in all cases except 1, whereas GPT-4o, o3, and Claude Sonnet 4 accurately identified the population(s)/comparators of interest for each report in 16/36 (44%), 31/36 (86%) and 23/36 (64%) of cases, respectively (Table 1).

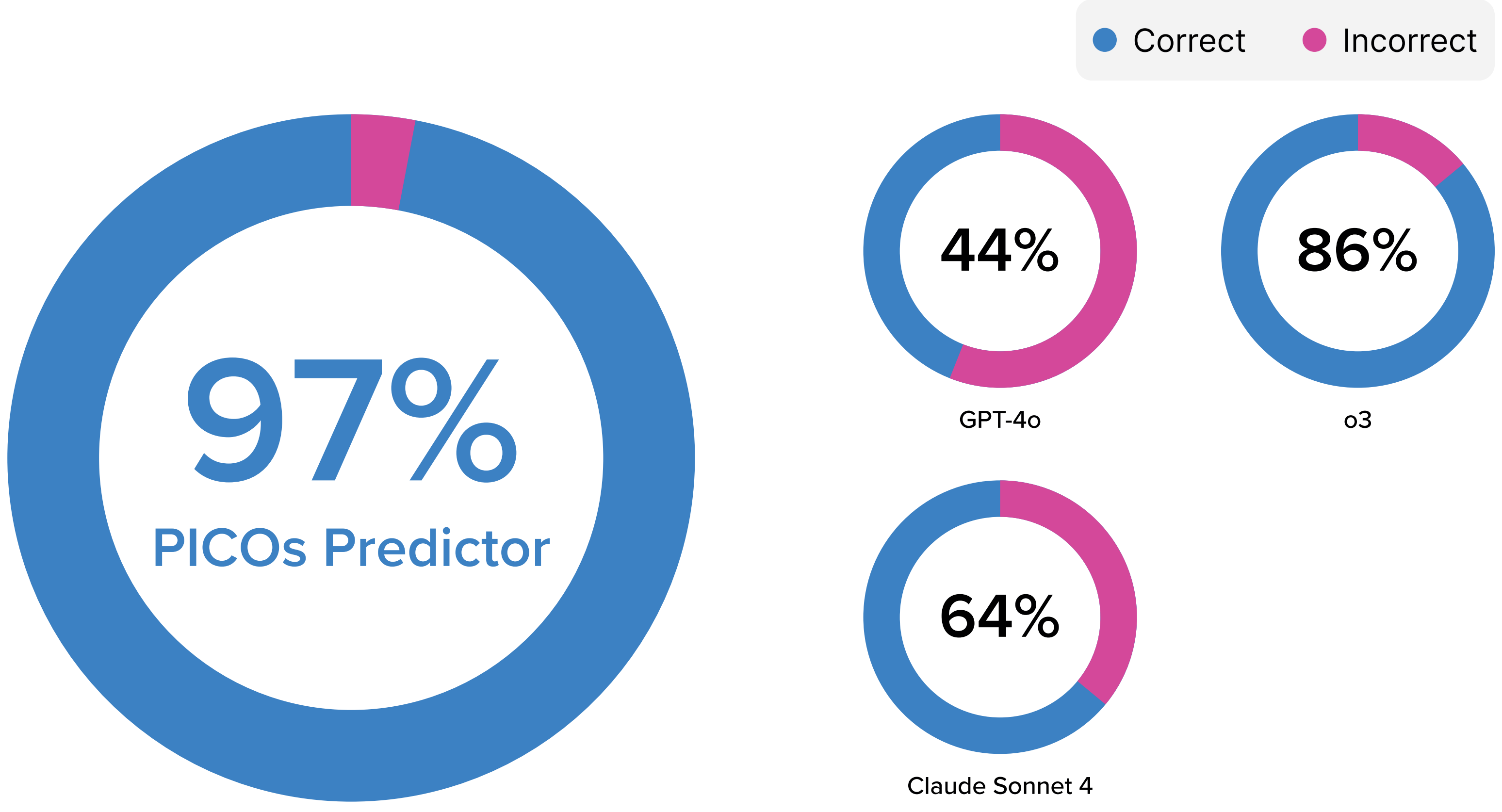


Table 1: Summary of analysis results

If we look at the results by country, GPT-4o, o3, and Claude Sonnet 4 had their best score analyzing the G-BA reports (77%, 100% and 78% of reports respectively).

GPT-4o and Claude Sonnet 4 had the lowest score for HAS reports with 26% & 53% respectively.

The most frequent deviation for populations was duplicates. Populations out of scope were extracted by GPT 4o and Claude Sonnet 4 in 3 and 2 cases. Outputs were empty in 1 case each for GPT-4o and Claude Sonnet 4. Two outputs were not translated into English by Claude Sonnet 4.

	OpenAI GPT-4o				OpenAI o3				Anthropic Sonnet-4				PICOs Predictor			
	No issues		Issues identified		No issues		Issues identified		No issues		Issues identified		No issues		Issues identified	
HAS (n=19)	5	26%	14	74%	17	89%	2	11%	10	53%	9	47%	19	100%	0	0%
GBA (n=9)	6	67%	3	33%	9	100%	0	0%	7	78%	2	22%	9	100%	0	0%
AEMPS (n=8)	5	63%	3	38%	5	63%	3	38%	6	75%	2	25%	7	88%	1	13%
Total (n=36)	16	44%	20	56%	31	86%	5	14%	23	64%	13	36%	35	97%	1	3%

CONCLUSIONS

To date, HTA assessments in the EU have been purely national, and not all EU countries have been using the PICOs methodology. The format of national reports is very different from one country to another, as is the wording of the different reports, making automated extraction difficult.

“PICOs Predictor” had the best results. o3, GPT-4o, and Claude Sonnet 4 showed very different performances, with o3 (pure reasoning LLM) being the most accurate of the 3 tested.

AI-enabled PICOs prediction is a promising approach to support evidence generation and access strategy for treatments subject to the JCA process. The “PICOs Predictor” can accelerate manual review of HTA documents and generate high-quality PICO predictions when manual review of relevant documents is not viable, for example in tight timeline business development processes or when affiliate teams lack capacity for primary research. The “PICOs Predictor” is both more comprehensive and more reliable than off-the-shelf LLMs. Further evaluation of the “PICOs predictor” across other disease states would demonstrate its generalizability.

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 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32021R2282

ACKNOWLEDGEMENT

Kateryna Horblyuk developed the graphics for this poster.