Artificial Intelligence (AI)-Assisted Early Data Insights and Literature Monitoring: A Case Study of Maintaining an Up-to-Date Reference Library in Metastatic Prostate Cancer (MPC)

Vidhi Patel,[1] Mihaela Musat,[2] Reza Jafar,[1] Stacy Grieve,[1] Maria Rizzo,[3]* Victoria Young[3]

[1]Cytel, Inc., Toronto, ON, Canada; [2]Cytel, Inc., Waltham, MA, USA, [3]Cytel, Inc., London, UK
*At the time of the study

Background

- The application of artificial intelligence (AI) tools for review and extraction of records has the potential to provide early insights during literature monitoring and reduce the burden of maintaining an up-to-date literature review.

- With the rapid development and involvement of AI in conducting, reporting, and reviewing research, medical and scientific organisations have been issuing guidelines to help apply this technology in a robust and transparent way.[1,2]

- In a recent position statement, the National Institute for Health and Care Excellence acknowledged the role of AI models to augment human involvement in the systematic review process, such as literature searches, study classification, screening, and visualisation of search results.[3]

Objective

To assess the efficiency, specificity, and accuracy of AI-assisted versus human-only record screening and extraction via a case study of up-to-date reference library maintenance in metastatic prostate cancer (MPC).

Methods

Data sources

- Studies in MPC from eight key oncology congresses reporting clinical, epidemiology, quality of life, and economic outcomes at any line of therapy were monitored from January 2023 to February 2024.

- The conferences searched were American Association for Cancer Research, American Society of Clinical Oncology (ASCO), ASCO Genitourinary Cancers Symposium, ASCO Quality Care Symposium, European Society for Medical Oncology, The Professional Society for Health Economics and Outcomes Research (ISPOR), and ISPOR Europe.

- Abstracts were assessed for eligibility according to the criteria listed in **Table 1**.

Table 1. Inclusion and exclusion criteria

	Inclusion criteria	Exclusion criteria
Population	• MPC with any prior treatment, any lines, including both CRPC and HSPC	• Non metastatic, oligo metastatic, localised • Preclinical studies • Other indications • Animal • Cell lines
Intervention	• Any types of systemic intervention for each indication treatment • No treatment is also allowed	• Treatments for AEs due to treatment • Infection prevention • Stem cell transplantations • Conditioning/bridging therapy • Surgery • Exercise/supplements/app • Screening/monitoring
Outcomes	• Any clinical outcomes • Any QOL/PRO outcomes • Any utilities • Any economic evaluation or BIM • Any cost or HCRU • Any RWE outcomes • MA, HTA review, HTA body decision review, pricing	• Exploring candidates for biomarkers, risk factors, prognostic factors in small sample sizes • Machine learning, AI • PK/PD
Study design	• Interventional studies • RWE studies • QOL studies • Economic studies • SLR/MA studies • ITC • MA/pricing studies • Guidelines/consensus	• Narrative review • Case report

model; CRPC, castration-resistant prostate cancer; HCRU, healthcare resource utilisation; HSPC, hormone sensitive prostate cancer; HTA, health technology assessment; ITC, indirect treatment comparison; MA, market access; MPC, metastatic prostate cancer; PK/PD, pharmacokinetic/pharmacodynamic; PRO, patient-reported outcome; QOL, quality of life; RWE, real-world evidence; SLR, systematic literature review.

Screening and extraction

- Abstracts were sequentially reviewed upon publication against predefined selection criteria separately by an AI model trained using over 65,000 human-annotated records (LiveSTART™) and human reviewers.

- Following human selection, another AI model, (LiveRef™) was used to extract the category of evidence, study type, and intervention (**Figure 1**). The same outcomes were also extracted by a human reviewer.

- The screening specificity and accuracy of LiveSTART™ and the extraction accuracy of LiveRef™ were assessed and compared with human review and extraction.

Methods (cont.)

Figure 1. Screening process

Congress scraping → AI-enabled screening (LiveSTART™) → AI-enabled extraction (LiveRef™)
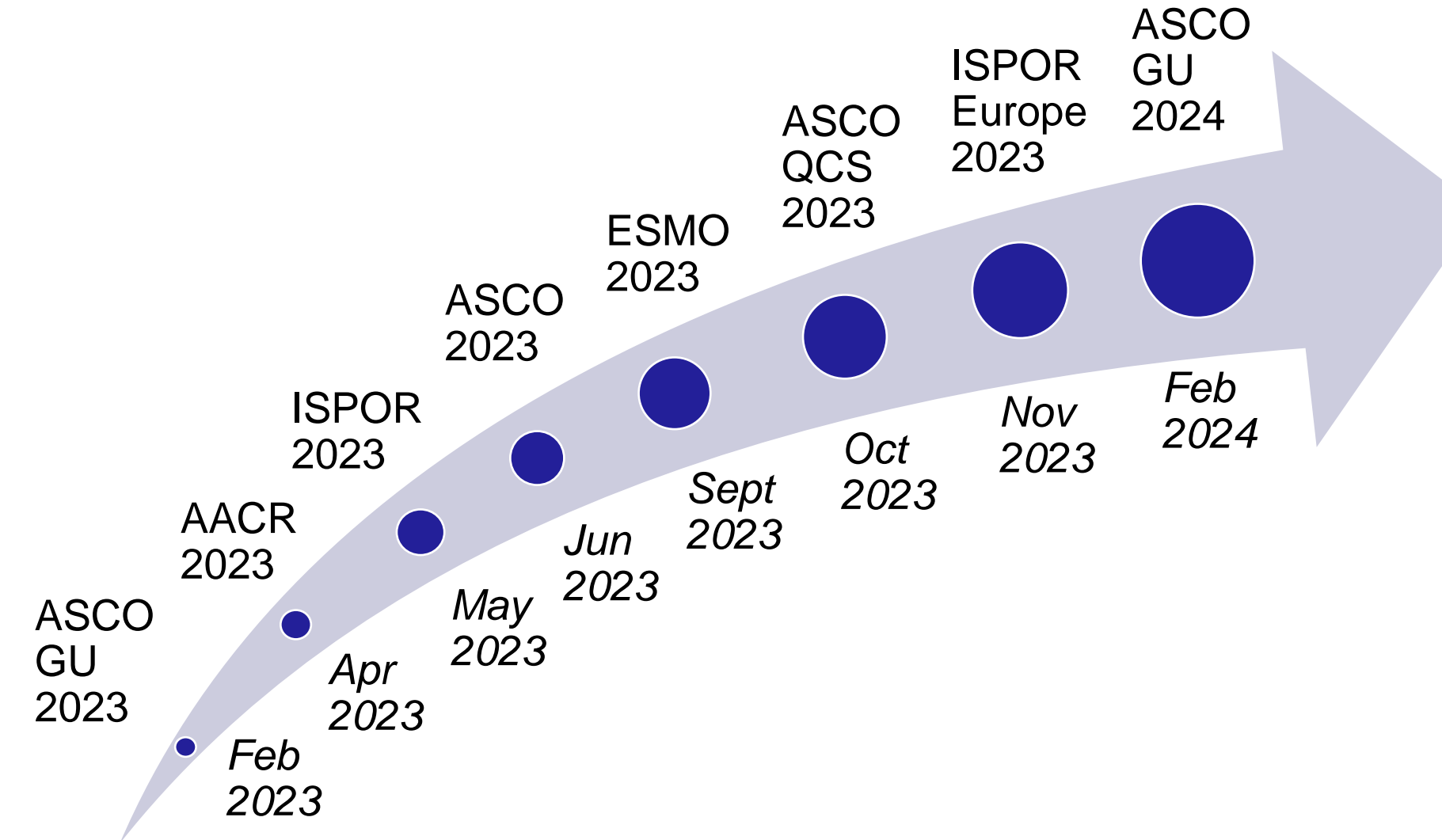
2 WEEKS

Abbreviations: AI, artificial intelligence.

Results

- Over one year, a total of 15,668 abstracts from eight congresses were reviewed (**Figure 2**).

- Of these, 1,773 were related to the MPC indication and 427 met the inclusion criteria.

Figure 2. Monitoring schedule of conference abstracts

ASCO GU 2023 — Feb 2023
AACR 2023 — Apr 2023
ISPOR 2023 — May 2023
ASCO 2023 — Jun 2023
ESMO 2023 — Sept 2023
ASCO QCS 2023 — Oct 2023
ISPOR Europe 2023 — Nov 2023
ASCO GU 2024 — Feb 2024

Abbreviations: AACR, American Association for Cancer Research; ASCO, American Society of Clinical Oncology; ESMO, European Society for Medical Oncology; GU, Genitourinary Cancers Symposium; ISPOR, The Professional Society for Health Economics and Outcomes Research; QCS, Quality Care Symposium.

Screening results

- At the screening stage, the review time was more than 10 times shorter with LiveSTART™ compared with human review, with a total time of 135 minutes for LiveSTART™ and 1,440 minutes for the human reviewer, and a mean of 17 minutes and 180 minutes per congress, respectively (**Table 2**).

Table 2. AI versus human screening times

Conference	Abstracts	Screening time (minutes)	
		AI	Human
AACR 2023	1,085	13	100
ASCO 2023	5,710	35	526
ASCO GU 2023	736	12	67
ASCO QCS 2023	581	11	53
ESMO 2023	2,184	19	201
ISPOR EU 2023	2,332	17	214
ISPOR 2023	2,334	16	214
ASCO GU 2024	706	12	65
TOTAL	**15,668**	**135**	**1,440**

Abbreviations: AACR, American Association for Cancer Research; AI, artificial intelligence; ASCO, American Society of Clinical Oncology; ESMO, European Society for Medical Oncology; EU, Europe; GU, Genitourinary Cancers Symposium; ISPOR, Professional Society for Health Economics and Outcomes Research; QCS, Quality Care Symposium.

- The mean specificity and accuracy of LiveSTART™ versus human screening were 93% and 92%, respectively (**Figure 3**).

Figure 3. Specificity and accuracy of LiveSTART™ review versus human review

Specificity
ASCO GU 2023: 87%; AACR 2023: 96%; ISPOR 2023: 98%; ASCO 2023: 91%; ESMO 2023: 93%; ASCO QCS 2023: 91%; ISPOR Europe 2023: 98%; ASCO GU 2024: 89%

Accuracy
ASCO GU 2023: 85%; AACR 2023: 96%; ISPOR 2023: 97%; ASCO 2023: 91%; ESMO 2023: 92%; ASCO QCS 2023: 91%; ISPOR Europe 2023: 98%; ASCO GU 2024: 85%

Abbreviations: AACR, American Association for Cancer Research; AI, artificial intelligence; ASCO, American Society of Clinical Oncology; ESMO, European Society for Medical Oncology; GU, Genitourinary; ISPOR, The Professional Society for Health Economics and Outcomes Research; QCS, Quality Care Symposium.

Results (cont.)

Extraction results

- Following human screening, 427 abstracts were included.

- The total extraction time of the 427 records was 68 minutes for LiveRef™ and 1,025 minutes for the human reviewer.
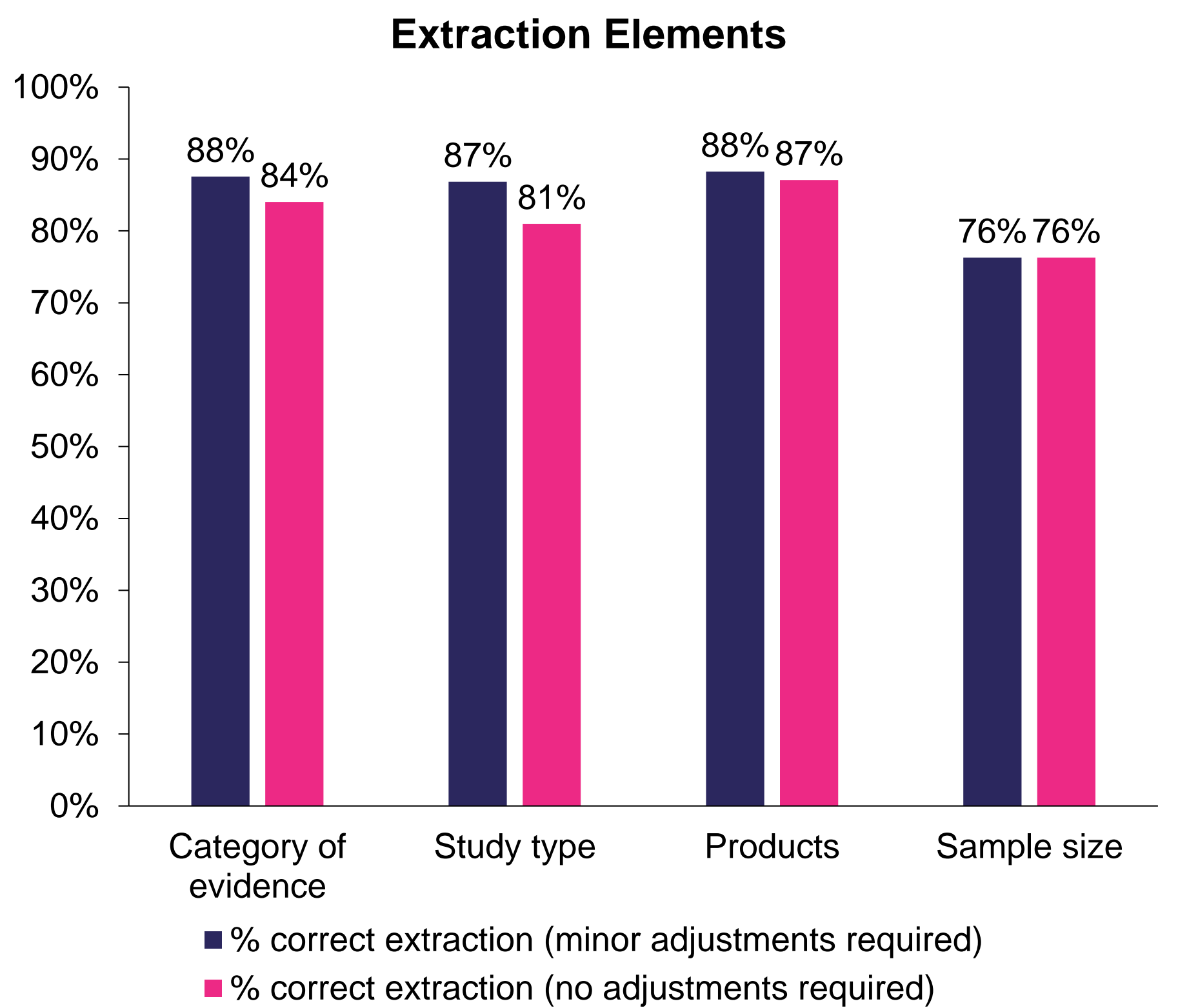
Table 2. AI versus human extraction times

Conference	Abstracts	Extraction time (minutes)	
		AI	Human
AACR 2023	12	10	10
ASCO 2023	93	9	120
ASCO GU 2023	117	9	180
ASCO QCS 2023	6	9	15
ESMO 2023	60	8	60
ISPOR Europe 2023	10	7	10
ISPOR 2023	19	7	30
ASCO GU 2024	110	9	150
TOTAL	**427**	**68**	**575**

Abbreviations: AACR, American Association for Cancer Research; AI, artificial intelligence; ASCO, American Society of Clinical Oncology; ESMO, European Society for Medical Oncology; GU, Genitourinary Cancers Symposium; ISPOR, Professional Society for Health Economics and Outcomes Research; QCS, Quality Care Symposium.

- The extraction accuracy with minor adjustments of LiveRef™ was 88% for evidence category, 87% for study type, 88% for interventions, and 76% for sample size. (**Figure 4**).

Figure 4. AI-enabled extraction accuracy

Extraction Elements

	% correct extraction (minor adjustments required)	% correct extraction (no adjustments required)
Category of evidence	88%	84%
Study type	87%	81%
Products	88%	87%
Sample size	76%	76%

- The evidence categories included clinical trials, real-world evidence (RWE), economic, and health-related quality of life (HRQOL).

- Among identified studies, 143 were clinical trials, 224 were RWE studies, 22 were economic studies, and 23 were HRQOL studies.

- Overall, new results from phase 2 or 3 studies of treatments for MPC were captured from 98 abstracts.

Conclusions

- The AI models, LiveSTART™ and LiveRef™, were effectively used for frequent literature monitoring and performed study screening and extraction with high specificity and accuracy within a notably shorter timeframe compared with human review.

- LiveRef™ efficiently extracted key study characteristics, providing early insights from a large volume of abstracts released periodically throughout the year.

References

1. Cacciamani GE, et al. Nat Med. 2023;29:14–5. **2.** Hamel C, et al. BMC Med Res Methodol. 2021;21:285. **3.** National Institute for Health and Care Excellence. Use of AI in evidence generation: NICE position statement 1.0 (2024). Retrieved from https://www.nice.org.uk/about/what-we-do/our-research-work/use-of-ai-in-evidence-generation--nice-position-statement.