Customization of a Large Language Model Approach to Capture PSA and Imaging Derived Real-World Progression Events in Prostate Cancer

RWD123

Scan to learn more

Kelly Magee, MS, FNP-BC[1]; Patrick Ward, PhD, MPH[2]; Wanjing Chen, MS[2]; Eunice Hankinson, MSN, FNP-C[1]; Aaron Dolor, PhD[1]
[1]Flatiron Health, New York, NY, USA; [2]Flatiron Health, Durham, NC, USA

Background

- Progression is a critical endpoint in cancer outcomes research and accurate identification of cancer progression events from electronic health records (EHRs) can support the development of high-quality real-world evidence (RWE) and enable promising novel uses such as predicting disease trajectory and assessing treatment effectiveness
- We have previously demonstrated the performance of a large language model (LLM)-based approach to capture real-world progression (rwP) in solid tumors, including prostate cancer (PC), from clinician documentation of worsening disease.[1] However, progression of PC often involve changes in prostate-specific antigen (PSA) and may be reflected differently in routine EHR documentation compared to other solid tumor cancer types
- We adapted our existing solid tumor approach to incorporate PSA-based progression events and evaluated its performance using elements of the VALID Framework[2]

Methods

- **Data source:** This study used the US-based, longitudinal Flatiron Health Research Database (FHRD), an EHR-derived, deidentified database, comprising patient-level data originating from >220 US cancer clinics (>1,600 sites of care; primarily community oncology settings) and curated via technology-enabled abstraction and artificial intelligence-based extraction methods, including NLP and LLMs.[3] The FHRD contains data for over 370,000 patients with PC
- **Setting:** This study focused on patients receiving treatment for metastatic hormone-sensitive prostate cancer (mHSPC) or metastatic castration-resistance prostate cancer (mCRPC)
- **Variable:** Progression events were generated using a LLM designed to capture clinician-documentation of worsening disease and augmented by progression events derived from changes in PSA values. PSA values were captured from both structured and unstructured EHR data and a PSA-based progression event was defined as a PSA increase of ≥25% between consecutive PSA results (within 180 days of one another) followed by an additional PSA result in 21-90 days that showed further increase[4]
- **Statistical analysis:**
  - We assessed the completeness of rwP events overall and by source evidence type among patients with a metastatic diagnosis date after January 1, 2011 and with mHSPC receiving first-line therapy (1L) and mCRPC receiving 1L, second-line (2L), or third-line therapy (3L). Source evidence types included events generated from clinician-documentation and events generated only from PSA changes. We also estimated the proportion of events where the clinician cited radiology as a source of their determination for progression
  - We evaluated the relationship between first rwP event following line start and clinically relevant downstream events (treatment change or death) overall and by source evidence types
  - We also estimated real-world progression-free survival (rwPFS) by sites of metastases (SoM) to ensure clinical validity

Results

- The proportion of patients with ≥1 rwP event was 53% in 1L mHSPC (n = 71,936) and 73%, 77%, and 77% in 1L (n = 46,663), 2L (n = 26,254), and 3L (n = 14,382) mCRPC (**Table 1**)
- 49%-68% of patients experienced ≥1 clinician-documented event and 20%-36% experienced ≥1 PSA-derived only event. Radiology was cited as source evidence for progression determination by the clinician in 42%-62% of events (**Table 1**)
- The mean number of rwP events was similar across line settings, with a mean of 2.8-3.2 events per patient (**Table 1**)
- The percent of first rwP events associated with a downstream clinical event increased in later lines (39% in 1L mHSPC to 50% in 3L mCRPC) with higher rates among clinician-documented events compared with PSA-only events (41%-57% vs. 23%-33%) Events where radiology was cited as source evidence had similar rates of downstream events compared to the broader clinician-documented events group (45%-58%; **Table 1**) suggesting clinicians are more likely to change treatment in response to radiology-based progression vs. PSA only based progression
- When considering all rwP events regardless of source evidence type, consistent with the known prognostic impact of visceral vs. non-visceral metastases[5], PFS was shorter for patients with liver metastases compared with those with bone-only metastases (3.2 vs. 5.9 months in 1L mCRPC, 6.9 vs. 17.8 months in 1L mHSPC) (**Figures 1 & 2**)

Disclosures: This study was sponsored by Flatiron Health, Inc.—an independent member of the Roche Group. During the study period, KM, PW, WC, EH, and AD reported employment with Flatiron Health, Inc. and stock ownership in Roche. Data first presented at ISPOR 2026 in Philadelphia, PA, US on May 19, 2026.
Contact information: Kelly Magee (kmagee@flatiron.com)

Table 1. Completeness and downstream events by source evidence combination

Pts with Index line of therapy (n)	rwP evidence type	Percent of patients	Mean of rwP events	Percent of rwP with downstream event
mHSPC 1L (71,936)	Overall	52.9	3.0	38.8
	PSA Only	19.6	1.7	22.7
	Clinician-documented	49.4	2.1	41.2
	Clinician cited radiology	41.8	2.1	44.8
mCRPC 1L (46,663)	Overall	73.4	3.2	42.7
	PSA Only	35.2	1.6	25.8
	Clinician-documented	66.4	2.1	47.4
	Clinician cited radiology	59.0	2.1	50.2
mCRPC 2L (26,254)	Overall	76.7	3.0	45.8
	PSA Only	36.0	1.5	28.3
	Clinician-documented	68.4	2.1	51.5
	Clinician cited radiology	62.3	2.1	53.4
mCRPC 3L (14,382)	Overall	77.0	2.8	50.4
	PSA Only	33.7	1.4	32.5
	Clinician-documented	67.9	2.0	56.6
	Clinician cited radiology	62.4	2.0	57.7

1L, first-line; 2L, second-line; 3L, third-line; mCRPC, metastatic castration-resistance prostate cancer; mHSPC, metastatic hormone sensitive prostate cancer; PSA, prostate-specific antigen; rwP, real-world progression; rwPFS, real-world progression-free survival.

Figure 1. rwPFS by sites of metastases in 1L mHSPC

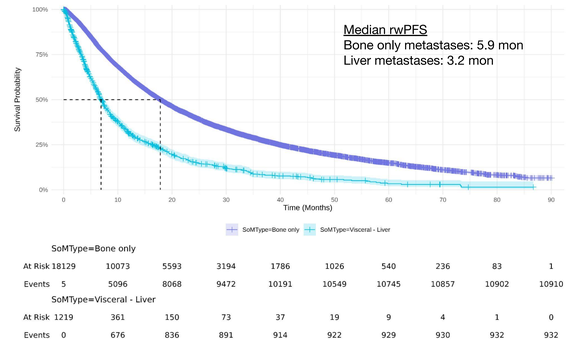

Figure 2. rwPFS by sites of metastases in 1L mCRPC

This study establishes the **performance and validation of an LLM-based approach to capturing prostate cancer progression data** from EHRs; supporting the efficient and reliable generation of valuable outcome data across large cohorts

Conclusions and Future Directions

- The combination of clinician-documented rwP events and PSA-derived rwP events generates complete and reliable progression data for patients with PC
- Future work exploring the relationship between assessment of progression in the real world vs in the clinical trial setting among patients with PC would be valuable in contextualizing rwPFS findings
- Large scale progression data may support novel applications including predicting treatment effectiveness and developing tooling to support clinician decision making

References

1. Cohen AB et al. Abstract B006: Using large language models for scalable extraction of real-world progression events across multiple cancer types. *Clin Cancer Res.* July 2025; 31(13_Supplement):B006. doi:10.1158/1557-3265.AIMACHINE-B006
2. Estevez M et al. Ensuring reliability of curated electronic health record–derived data: The Validation of Accuracy for Large Language Model–/Machine Learning–Extracted Information and Data (VALID) Framework. *JCO Clin Cancer Inform.* 2026 Mar;10:e2500215.doi:10.1200/CCI-25-00215
3. Flatiron Health. Database Characterization Guide. Flatiron.com. Published March 18, 2025. Accessed April 14, 2026. https://flatiron.com/database-characterization
4. Scher HI et al. Trial design and objectives for castration-resistant prostate cancer: Updated recommendations from the Prostate Cancer Clinical Trials Working Group 3. *J Clin Oncol.* 2016;34(12):1402-1418. doi:10.1200/JCO.2015.64.2702
5. Moul JW. Prostate. In: Amin MB, et al., eds. AJCC Cancer Staging Manual. 8th ed. Springer; 2017:663-672