

Temporal Patterns of Immune-Related Adverse Events Enable Phenotype-Guided Risk and Timing Prediction Using OMOP-Based Real-World Data

Liyuan Gong¹, Harry Reyes Nieva², Jeong Yun Yang¹, Peter Shahinnian¹, Nitish Aswani¹, Alexander Wei¹, Chin Hur¹

¹Department of Medicine, Columbia University, New York, NY, USA; ²Department of Biomedical Informatics, Columbia University, New York, NY, USA

Background

Immune checkpoint inhibitors (ICIs) have improved outcomes across multiple cancers, but immune-related adverse events (irAEs) remain difficult to anticipate because they vary by organ system, severity, and time of onset.^{1,2}

Clinically significant irAEs may require systemic immunosuppression, treatment interruption, or discontinuation. Existing risk models often treat irAEs as a binary outcome and focus mainly on early events, limiting their ability to guide long-term monitoring. For the OHDSI community, irAE prediction is an important real-world evidence use case because it requires standardized cohort construction, reproducible phenotype definitions, temporal outcome anchoring, and clinically interpretable prediction using observational health data.^{3,4}

We developed an OMOP-based machine learning framework to characterize early and late moderate-to-severe irAEs among cancer patients receiving ICIs. Our goal was to determine whether baseline clinical phenotypes are associated with irAE timing and whether a staged prediction approach can support both risk stratification and time-resolved monitoring.

Methods

We conducted a retrospective cohort study using structured electronic health record data from the All of Us Research Program, harmonized to the Observational Medical Outcomes Partnership common data model. Adult patients with documented exposure to anti-PD-1, anti-PD-L1, anti-CTLA-4, or related ICI agents were identified using standardized RxNorm concepts. The index date was the first recorded ICI administration. Patients were required to have at least 12 months of baseline observation and up to 24 months of follow-up after ICI initiation. The final analytic cohort included 500 patients.

irAE outcomes were defined using anchor-based OMOP phenotypes combining organ-specific inflammatory diagnosis codes with temporally proximate clinical management, including systemic corticosteroids, immunosuppressive therapy, hormone replacement therapy, or ICI discontinuation. Moderate-to-severe irAEs were defined as Grade ≥ 2 events. Early and late irAEs were defined based on time from ICI initiation to irAE onset, with late events occurring after 12 months.

Baseline features were extracted from the 12 months before ICI initiation and included demographics, body mass index, cancer type, ICI regimen, comorbidities, prior chemotherapy, baseline medications, and laboratory values. We used UMAP followed by Leiden clustering to identify baseline clinical phenotypes among patients with moderate-to-severe irAEs. We then trained binary models to predict moderate-to-severe irAEs within 6, 12, and 24 months. XGBoost was selected as the primary model based on nested 5-fold cross-validation. Decision curve analysis was used to identify clinically relevant risk thresholds, and patients exceeding these thresholds were evaluated using random survival forests to estimate time to irAE onset.

Results

Among 500 ICI-treated patients, 217 had no identified irAE, 180 had Grade 1 irAEs, 95 had Grade 2 irAEs, and 8 had Grade ≥ 3 irAEs within 24 months. The cumulative incidence of moderate-to-severe irAEs increased over time, reaching approximately 21-22% by 24 months.

Organ-specific analyses showed substantial heterogeneity in timing, with some irAEs occurring predominantly earlier and others appearing later during follow-up.

Unsupervised clustering identified four baseline clinical phenotype groups among patients with moderate-to-severe irAEs. These clusters differed in the distribution of early and late irAE onset. One cluster had a higher proportion of late-onset irAEs, while another consisted exclusively of early-onset cases. Cluster membership was significantly associated with irAE timing (chi-square $p=0.005$; Cramer's $V=0.35$), suggesting that baseline clinical phenotypes capture meaningful temporal heterogeneity in irAE development.

Multivariable analyses also suggested different predictors for early versus late irAEs. Prior chemotherapy and thyroid disease were strongly associated with early-onset moderate-to-severe irAEs, while renal cancer was more strongly associated with late-onset events. For binary prediction, XGBoost showed the strongest overall performance across tested models. Discrimination was higher for 6- and 12-month outcomes and declined for 24-month prediction, supporting the hypothesis that late-onset irAEs are harder to predict using baseline features alone. Decision curve analysis showed positive net benefit across clinically relevant threshold probabilities. In time-to-event analysis, phenotype-informed random survival forests improved stratification, increasing the concordance index from approximately 0.69 to 0.73.

Table 1 Summary of core findings and relevance to the OHDSI community.

Analytic component	Key finding	OHDSI relevance
Cohort and phenotype	500 ICI-treated patients; Grade ≥ 2 irAEs reached approximately 21-22% by 24 months	Standardized OMOP cohort construction and anchor-based outcome definition
Temporal heterogeneity	Four phenotype clusters associated with early vs late irAE timing ($p=0.005$)	Demonstrates timing-aware adverse event phenotyping
Prediction and timing	XGBoost supported binary risk stratification; phenotype-informed RSF improved C-index from ~ 0.69 to ~ 0.73	Actionable staged prediction framework for monitoring and network validation

Conclusion

Moderate-to-severe irAEs show meaningful temporal heterogeneity that can be characterized using OMOP-based real-world EHR data. Baseline clinical phenotypes were associated with early versus late irAE onset, and a staged machine learning framework combining binary risk prediction, decision curve-guided thresholds, and phenotype-informed survival modeling improved time-resolved risk stratification. For the OHDSI community, this work highlights a reproducible analytic use case for standardized irAE phenotyping, temporal outcome definition, and clinically interpretable risk prediction using observational data. Future work will focus on external validation across additional OMOP datasets and refinement of phenotype definitions to support broader network studies of ICI toxicity.

References

1. Michot JM, Bigenwald C, Champiat S, Collins M, Carbone F, Postel-Vinay S, et al. Immune-related adverse events with immune checkpoint blockade: a comprehensive review. *Eur J Cancer*. 2016;54:139-148.
2. Postow MA, Sidlow R, Hellmann MD. Immune-related adverse events associated with immune checkpoint blockade. *N Engl J Med*. 2018;378:158-168.

3. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform.* 2015;216:574-578.
4. All of Us Research Program Investigators. The All of Us Research Program. *N Engl J Med.* 2019;381:668-676.