Evaluating Model Performance for Between-Country Survival Transportability

MSR169

Scan to learn more

Mohamed S. Ali, PharmD, SM[1]; Harlan Pittell, PhD[1]; Elsie Horne, PhD[2]; Philani Mpofu, PhD[1]; Qianyi Zhang, MS[1]; Blythe Adamson, PhD, MPH[3]
[1]Flatiron Health, New York, NY, USA; [2]Flatiron Health UK, London, United Kingdom; [3]Formerly Flatiron Health, New York, NY, USA

Introduction

- Transportability analyses can evaluate whether real-world evidence from one country can inform decisions in another. One way of assessing the plausibility of transporting overall survival (OS) evidence is outcome regression modeling. Such models are susceptible to model misspecification. It is, therefore, a good practice to assess model performance using internal validation within the data used to estimate the model
- The primary objective was to demonstrate a model that ensures internal validity when tested in an independent sample, before applying modeled OS for transportability. We analyzed a US cohort of patients with multiple myeloma (MM) as a case study

Methods

- **Data source:** The US-based, longitudinal Flatiron Health Research Database, an electronic health record-derived, deidentified database, comprising patient-level data originated from ~280 US cancer clinics (~800 sites of care; primarily community oncology settings) and curated via technology-enabled abstraction[1]
- **Setting and study population:** We included adult patients diagnosed with MM between January 1, 2015 and December 31, 2021. Patients were followed from initiation of first-line therapy (index date) until death or censoring at last recorded activity. Patients were required to have available baseline covariates (age, sex, treatment exposures, and stem-cell transplantation) and follow-up information sufficient to define OS
- **Main outcome measures:** The primary outcome was OS, defined as the time from first-line treatment initiation to death. Patients without a recorded death event were censored at their last known clinical activity date
- **Statistical analysis:** To evaluate model performance, the cohort was randomly split into a training set (70%) and an independent test set (30%). In the training set, OS was modeled using pooled logistic regression (PLR) on person-time data, including baseline covariates (age, sex, treatment exposures, and stem-cell transplantation) as well as time and covariate-time interaction terms to allow for non-proportional hazards. The fitted PLR model was applied to the test set to generate covariate level survival probabilities, which were marginalized over the covariate distribution of the test cohort. Model performance in the test set was evaluated using:
    - Discrimination, assessed using Harrell's C-index
    - Calibration, assessed by comparing PLR-predicted survival to Kaplan–Meier (KM) estimates and quantified using the mean absolute difference (MAD) over 0–60 months

Result

- **Participants:** A total of 9937 patients with MM were included in the analytic cohort. The cohort was randomly split into a training set (n = 6955) and an independent test set (n = 2982). Baseline demographic and treatment patterns are reported between the two sets in **Table 1**
- **Model performance:** In the test set, the PLR model demonstrated good discrimination, with a Harrell's C-index of approximately 0.689 (**Table 2**). Model calibration was also strong. The PLR-predicted survival curve closely aligned with the KM estimate in the test cohort (**Figure 1**). The MAD between predicted and observed survival over 0–60 months was approximately 0.009 (0.9 percentage points) (**Table 2**)
- **Survival estimates:** At 60 months, OS was approximately 55–60% in the test cohort as shown in KM curve, with minimal differences between PLR-predicted and observed KM estimate survival, indicating good agreement across the follow-up period

Table 1. Baseline characteristics of the study population by training and test sets

Characteristic	Overall (*N* = 9937)	Training set (*n* = 6955)	Test set (*n* = 2982)
Age at 1L initiation in years, mean (SD)	68.7 (10.5)	68.7 (10.5)	68.7 (10.6)
Sex, No. (%)			
Female	4512 (45)	3135 (45)	1377 (46)
Male	5425 (55)	3820 (55)	1605 (54)
CD38 inhibitor use, No. (%)	2076 (21)	1492 (21)	584 (20)
Chemotherapy, No. (%)	1288 (13)	908 (13)	380 (13)
IMiD use, No. (%)	7066 (71)	4948 (71)	2118 (71)
Proteasome inhibitor use, No. (%)	7978 (80)	5579 (80)	2399 (80)
Stem-cell transplantation within 1 year of diagnosis, No. (%)	2289 (23)	1623 (23)	666 (22)

Abbreviations: 1L, first-line; CD38, cluster of differentiation 38; IMiD, immunomodulatory drug; No. (%), number (percentage); PI, proteasome inhibitor; SCT, stem-cell transplantation; SD, standard deviation

Result (continued)

Figure 1. Observed (test KM) vs predicted (test PLR) overall survival in the test set

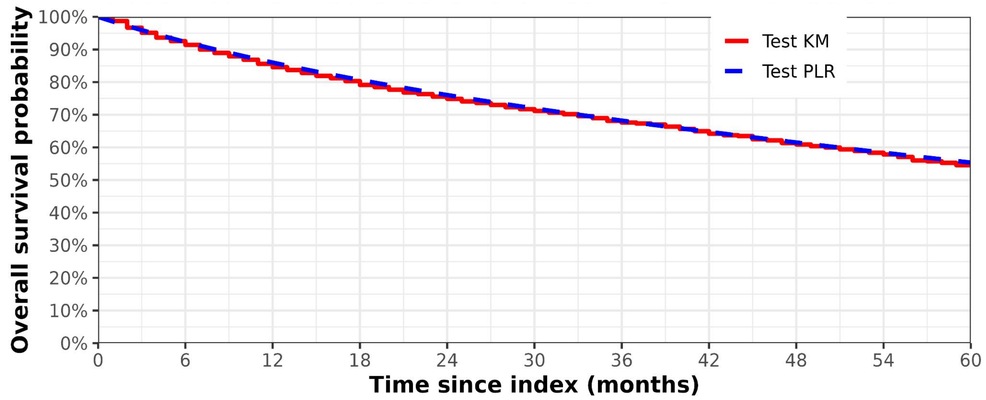

Table 2. Performance of the pooled logistic regression model in the test cohort

Metric	Value
C-index	0.689
Mean absolute difference (0–60 months)	0.009
Observed median OS (months)	68.90

Abbreviations: OS, overall survival

Conclusions/Main finding: The PLR model demonstrated good discrimination and close agreement between predicted and observed survival in the test cohort, indicating strong model performance. Model performance should be assessed before using modeled OS in transportability analyses to avoid inaccurately attributing country differences to model mispecification in the model development step

Acknowledgments: Vanessa Acquaah (Flatiron Health) for project management and Darren Johnson (Flatiron Health) for publication management
Disclosures: This study was sponsored by Flatiron Health, Inc.—an independent member of the Roche Group. During the study period, MSA, HP, EH, PM, QZ, and BA reported employment with Flatiron Health, Inc. and stock ownership in Roche. Data first presented at ISPOR in Philadelphia, PA, US on 17-20 May, 2026
Contact information: flatiron.forum@flatiron.com

References

1. Flatiron Health. Database Characterization Guide. Flatiron.com. Published March 18, 2025. Accessed April 15, 2026. https://flatiron.com/database-characterization