**CIBMTR - Equity in Post-HCT Survival Predictions Analysis Report**

**Executive Summary**

This report presents an analysis of the Center for International Blood and Marrow Transplant Research (CIBMTR) dataset aimed at developing predictive models for event-free survival following hematopoietic cell transplantation (HCT). The analysis focuses on ensuring equitable predictions across diverse racial groups, addressing potential disparities in survival outcomes. Multiple machine learning approaches were explored, with a gradient boosting model incorporating racial interaction features demonstrating the best performance while maintaining fairness across demographic groups.

**Data Description**

The dataset consists of 59 variables related to hematopoietic stem cell transplantation (HSCT), encompassing a wide range of recipient and donor characteristics:

* **Demographic information**: Age, sex, race/ethnicity (with equal representation across White, Asian, African-American, Native American, Pacific Islander, and multiracial groups)
* **Medical characteristics**: Disease type, disease risk index, comorbidity scores, cytogenetic scores
* **Treatment details**: Graft type, conditioning intensity, HLA matching, GVHD prophylaxis
* **Outcome measures**: Event-free survival (efs) and time to event (efs\_time)

The data was synthetically generated using SurvivalGAN methodology, trained on a large cohort of real CIBMTR data while preserving important relationships between variables and survival outcomes.

Initial analysis revealed significant variations in survival outcomes across racial groups:

* Median survival ranged from 7.98 months for White patients to 16.07 months for patients identifying as more than one race
* The survival disparity ratio (maximum/minimum) was approximately 2.01
* Event rates (proportion of patients experiencing an event) varied from 46.6% to 62.6% across racial groups

**Main Objectives**

The primary objectives of this analysis were:

1. Develop accurate predictive models for event-free survival following hematopoietic cell transplantation
2. Ensure equitable performance across different racial groups, minimizing disparities in prediction accuracy
3. Identify key factors influencing post-transplant survival outcomes
4. Evaluate whether the incorporation of racial interaction features improves model fairness and overall performance

**Modeling Approach**

**Survival Analysis Framework**

This analysis employed survival analysis techniques appropriate for time-to-event data with censoring. Multiple modeling approaches were explored:

1. **XGBoost with racial interaction features**: Gradient boosting incorporating specific interaction terms between race and key clinical predictors
2. **Hyperparameter-tuned XGBoost**: Optimized using randomized search over a comprehensive parameter space
3. **CatBoost**: An alternative gradient boosting implementation with inherent handling of categorical features

The Nelson-Aalen estimator was used to handle censored observations, transforming the time-to-event data into a continuous outcome variable suitable for regression models. The concordance index (C-index) was used as the primary evaluation metric, measuring the model's ability to correctly rank survival times.

**Feature Engineering**

Several specialized feature sets were created to address the racial equity objective:

1. **Race-specific interaction features**: Explicit interactions between racial categories and key clinical predictors (age, disease risk, comorbidity score, etc.)
2. **Demographic features**: Age groups, donor-recipient matching characteristics, and race-disease combinations
3. **One-hot encoded categorical variables**: Ensuring proper representation of categorical features

**Cross-Validation Strategy**

A stratified cross-validation approach was implemented to ensure consistent representation of racial groups across training and validation sets, providing more reliable estimates of performance across demographic subgroups.

**Key Findings**

**Model Performance**

The XGBoost model with racial interaction features demonstrated the best overall performance:

* Overall C-index: 0.644 ± 0.004
* RMSE: 0.244 ± 0.001

Performance by racial group showed relatively consistent results across demographics:

* African American: C-index 0.632 ± 0.013
* Asian: C-index 0.654 ± 0.009
* White: C-index 0.646 ± 0.006
* Native American: C-index 0.638 ± 0.008
* Pacific Islander: C-index 0.633 ± 0.006
* More than one race: C-index 0.653 ± 0.008

The hyperparameter tuning process identified an optimal XGBoost configuration with a C-index of 0.648, representing a modest improvement over the baseline model.

**Key Predictive Factors**

Analysis of feature importance revealed the following as the most influential predictors of post-transplant survival:

1. **Conditioning intensity**: The strongest predictor, reflecting the impact of preparative regimen intensity on outcomes
2. **Disease risk index (DRI)**: A composite measure of disease status and risk
3. **Comorbidity score**: Capturing pre-existing medical conditions
4. **Karnofsky performance score**: Measuring functional status
5. **Year of HCT**: Suggesting temporal trends in transplant outcomes
6. **Age-comorbidity interaction**: Highlighting the compounding effect of age and comorbidities
7. **Cytogenetic score**: Reflecting genetic risk factors

Notably, while race was not among the top predictors, the inclusion of race-specific interaction features improved the model's equitable performance across demographic groups.

**Disparities Analysis**

The analysis confirmed significant racial disparities in survival outcomes:

* Patients identifying as "More than one race" showed the highest median survival (16.07 months)
* White patients had the lowest median survival (7.98 months)
* The survival disparity ratio was 2.01, indicating the highest survival rate was twice the lowest

Despite these outcome disparities, the model maintained similar predictive performance across racial groups, with C-indices ranging from 0.632 to 0.654, suggesting fair predictive capability regardless of racial background.

**Limitations and Future Directions**

**Current Model Limitations**

Several limitations of the current approach should be noted:

1. **Synthetic data constraints**: While the dataset preserves important relationships, it may not capture all nuances present in real-world transplant data
2. **Limited external validation**: The model has not been validated on external datasets to confirm generalizability
3. **Black-box interpretability**: The gradient boosting models provide feature importance but lack the full interpretability of traditional statistical survival models
4. **Time-varying effects**: The current approach does not account for time-varying coefficients or effects that may change over the course of follow-up
5. **Small test set**: The final test dataset was extremely small (3 samples), limiting robust evaluation of the final model

**Future Directions**

To address these limitations and further advance this work:

1. **Incorporate traditional survival models**: Complement machine learning approaches with Cox proportional hazards models for improved interpretability
2. **Develop time-varying coefficient models**: Explore methods that can capture changing effects over the follow-up period
3. **External validation**: Test model performance on independent datasets from different transplant centers
4. **Causal inference methods**: Explore causal inference techniques to better understand the mechanisms underlying racial disparities
5. **Ensemble approaches**: Combine multiple survival models to potentially improve predictive performance
6. **Interactive decision support tools**: Develop clinician-facing tools that provide personalized survival predictions while highlighting uncertainty
7. **Incorporate post-transplant variables**: Explore the incorporation of post-transplant events as time-dependent covariates to improve prediction of subsequent outcomes

**Conclusion**

This analysis successfully developed a prediction model for post-HCT survival that maintains equitable performance across racial groups. By incorporating specific interactions between race and clinical factors, the model achieves good predictive accuracy while minimizing performance disparities. The identified key predictors align with clinical understanding of transplant outcomes, with conditioning intensity, disease risk, and comorbidity burden emerging as particularly important factors.

The analysis confirms significant disparities in actual survival outcomes across racial groups, highlighting the importance of addressing equity considerations in transplant care. While predictive models cannot eliminate these disparities, they can help ensure that prognostic information is equally accurate across demographic groups, supporting more personalized and equitable decision-making.

Future work should focus on external validation, enhanced interpretability, and the development of clinical decision support tools that can translate these findings into improved patient care and outcomes.