

Competing Risks Analysis of MLB Draft Data

Poster · September 2023
DOI: 10.13140/RG.2.2.35021.4636

CITATIONS
0

1 author:

Eric Gerber
Northwestern University
187 PUBLICATIONS 1,819 CITATIONS
[VIEW PROFILE](#)

READS
538

COMPETING RISKS ANALYSIS OF MLB DRAFT DATA

Eric Gerber
Northeastern University



Introduction

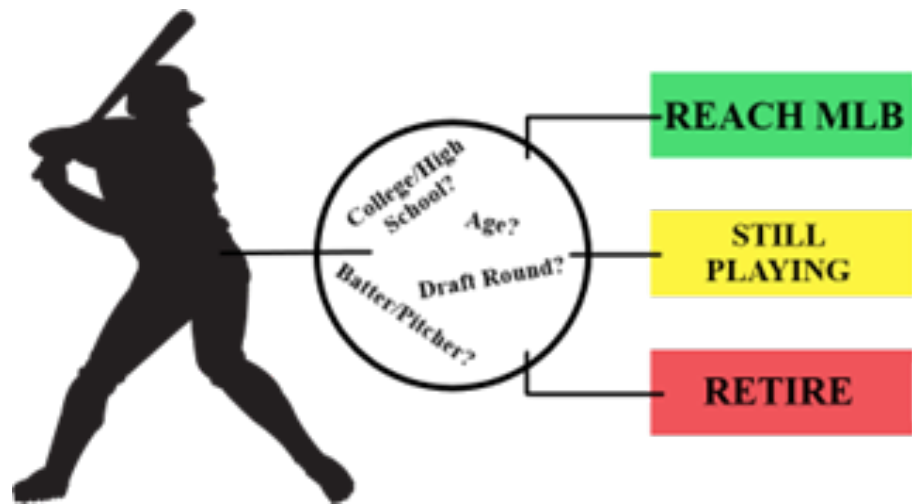
Baseball is unique in the major US sports in that nearly every player who is drafted will spend significant time in the minor leagues (MiLB) before reaching the major league (MLB). Beginning in 2021, the MLB draft was cut in half from 40 rounds to 20, yet still most will spend years in MiLB and retire before making it to the big leagues. Statistical analyses that can quantify the likelihood a player reaches MLB, and in what time frame, are of immense use to the players, their agents, and even the teams themselves.

Methods and Data

Recent studies have done an excellent job of evaluating how performance impacts becoming a successful MLB player [1] and predicted income [6]. Our goal is to assess how draft day factors, independent of performance, impact the time it takes draftees to either reach MLB or retire without doing so.

Competing risks analysis is a subset of survival analysis. Survival analysis deals in analyzing the expected time until an event occurs and the risk associated with that event occurring at a given time, while allowing for data observations for which the event was not observed, or has not occurred yet (called censored observations).

Competing risks analysis allows for multiple, competing, events to occur. In baseball, drafted players may reach the MLB (event 1), retire without reaching the majors (event 2), or still be playing in the MiLB (censored).



Competing risks analysis will allow us to assess what factors impact the time to these events and give better context to players as they decide to pursue a career in baseball.

Our data consist of all signed draftees from 2012 to 2016 ($N = 4573$). Since some of these draftees are still playing in MiLB as of 2022, we employ a competing risks approach to the analysis. There are various ways of fitting competing risks models [2, 3]. Analysis was conducted via the **cmprsk** [4] package in **R**.

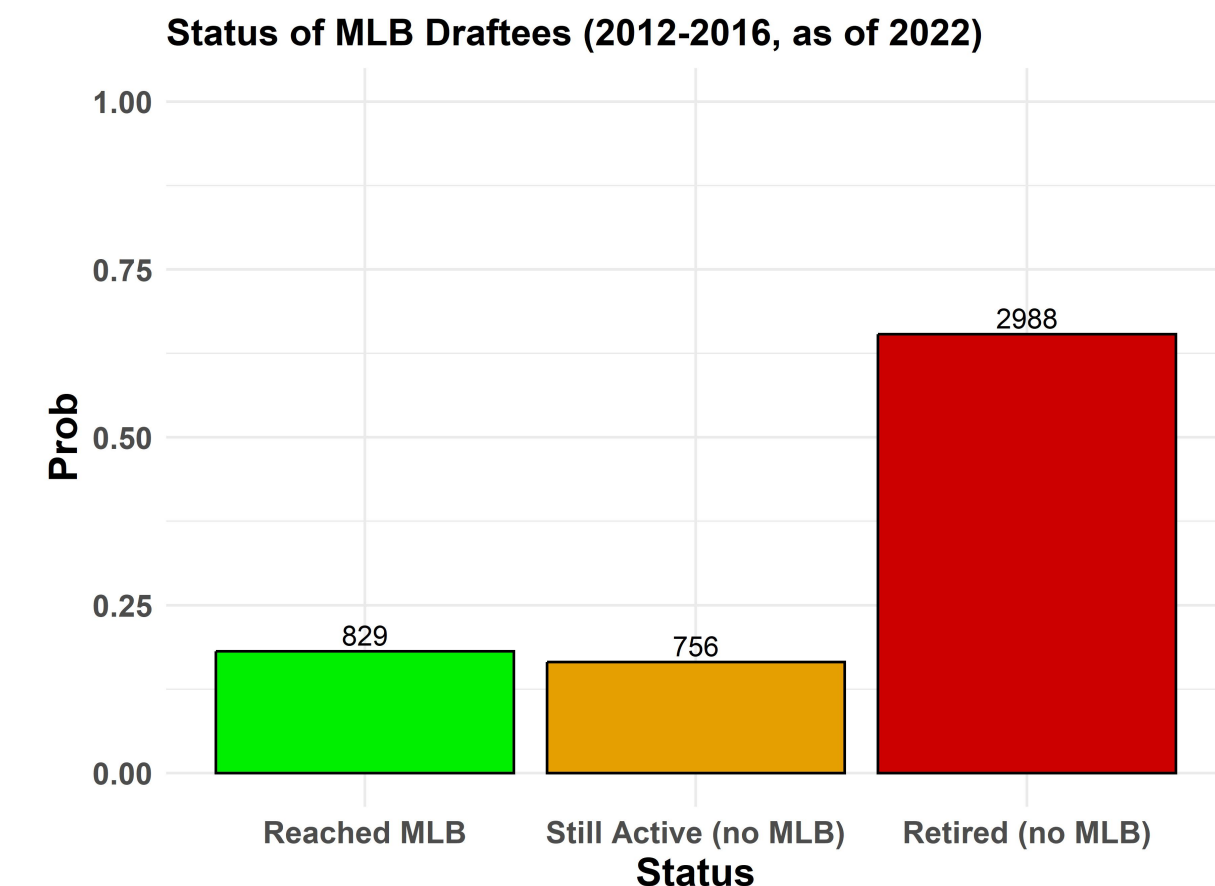


Fig. 1: Most players retire without reaching MLB.

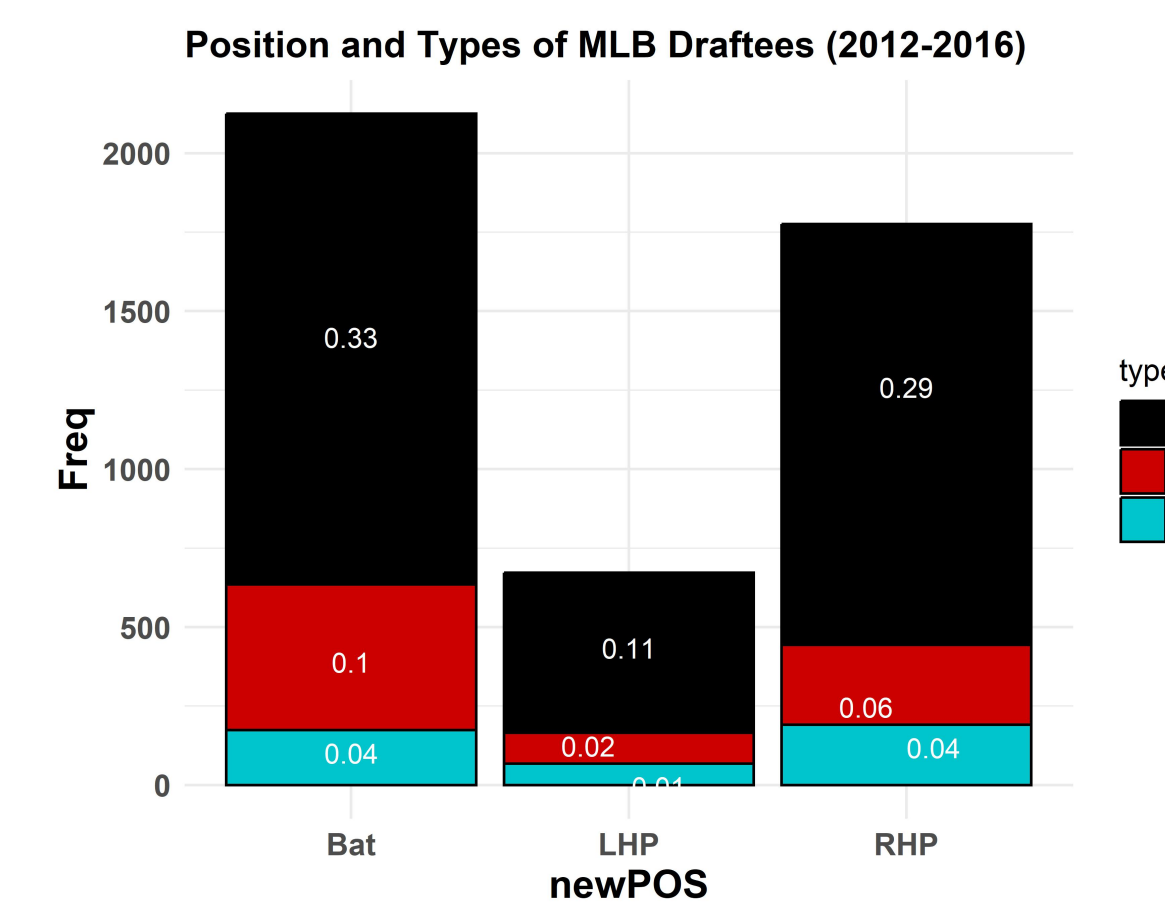


Fig. 2: Positions and types of drafted players.

Results

Basic competing risks [5] provides a framework for testing differences in survival for finite groups, but not regression with multiple covariates. Since other factors may impact these results, the proportional subdistribution hazard model [3] allows accounting for other covariates. After model selection, significant predictors include: overall draft position (pick #), signing bonus (in millions), position (batter, left-handed pitcher, right-handed pitcher), type, and interaction between overall pick and bonus.

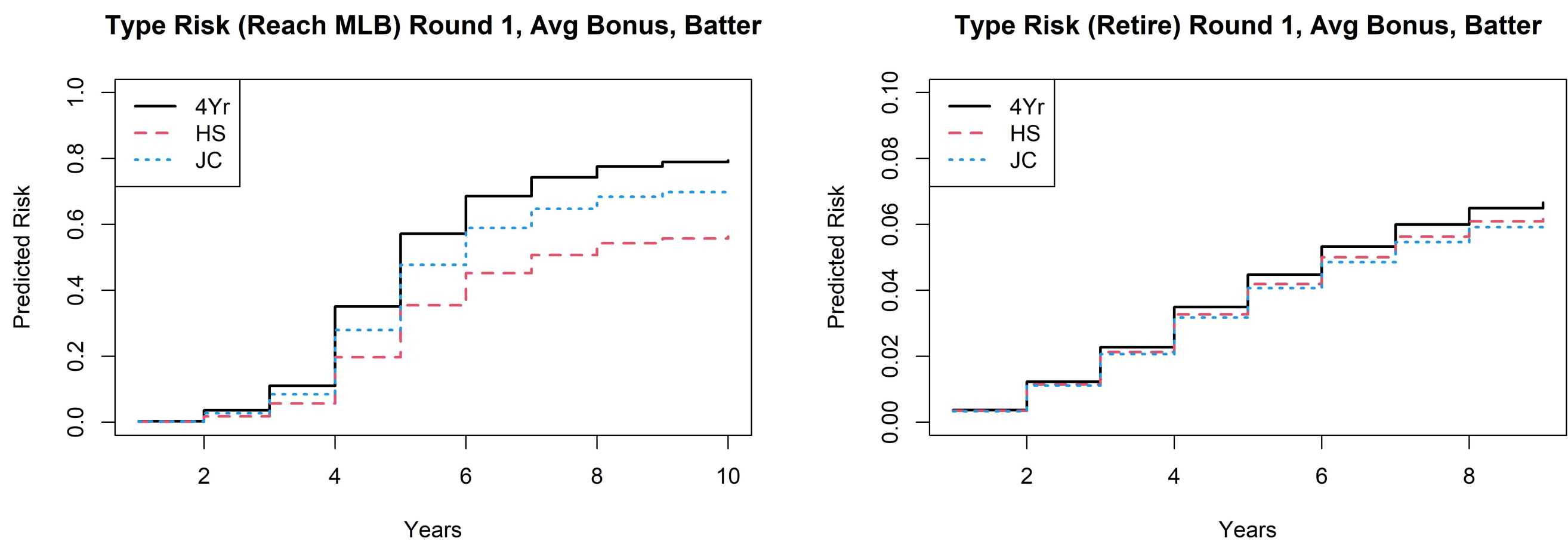


Fig. 3: Predicted risks of Reaching MLB and Retiring for first round (average bonus) Batters

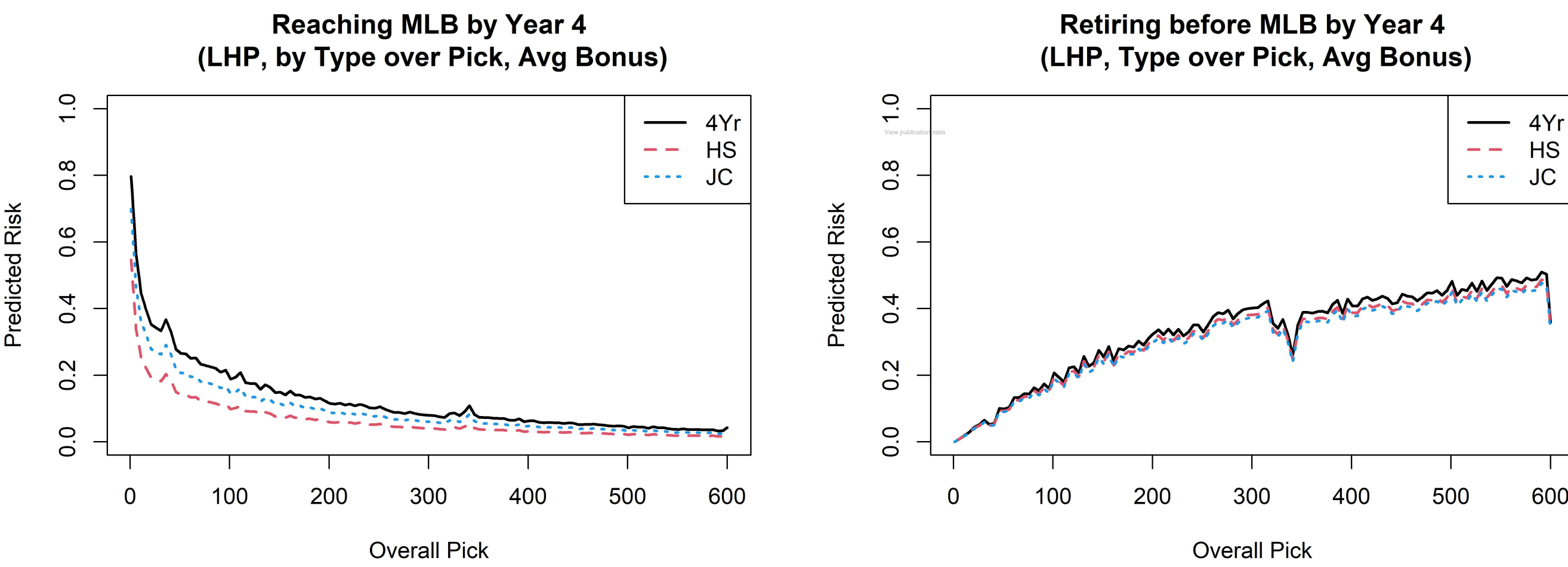


Fig. 4: Predicted risks of Reaching MLB and Retiring across picks (average bonus) for LHP

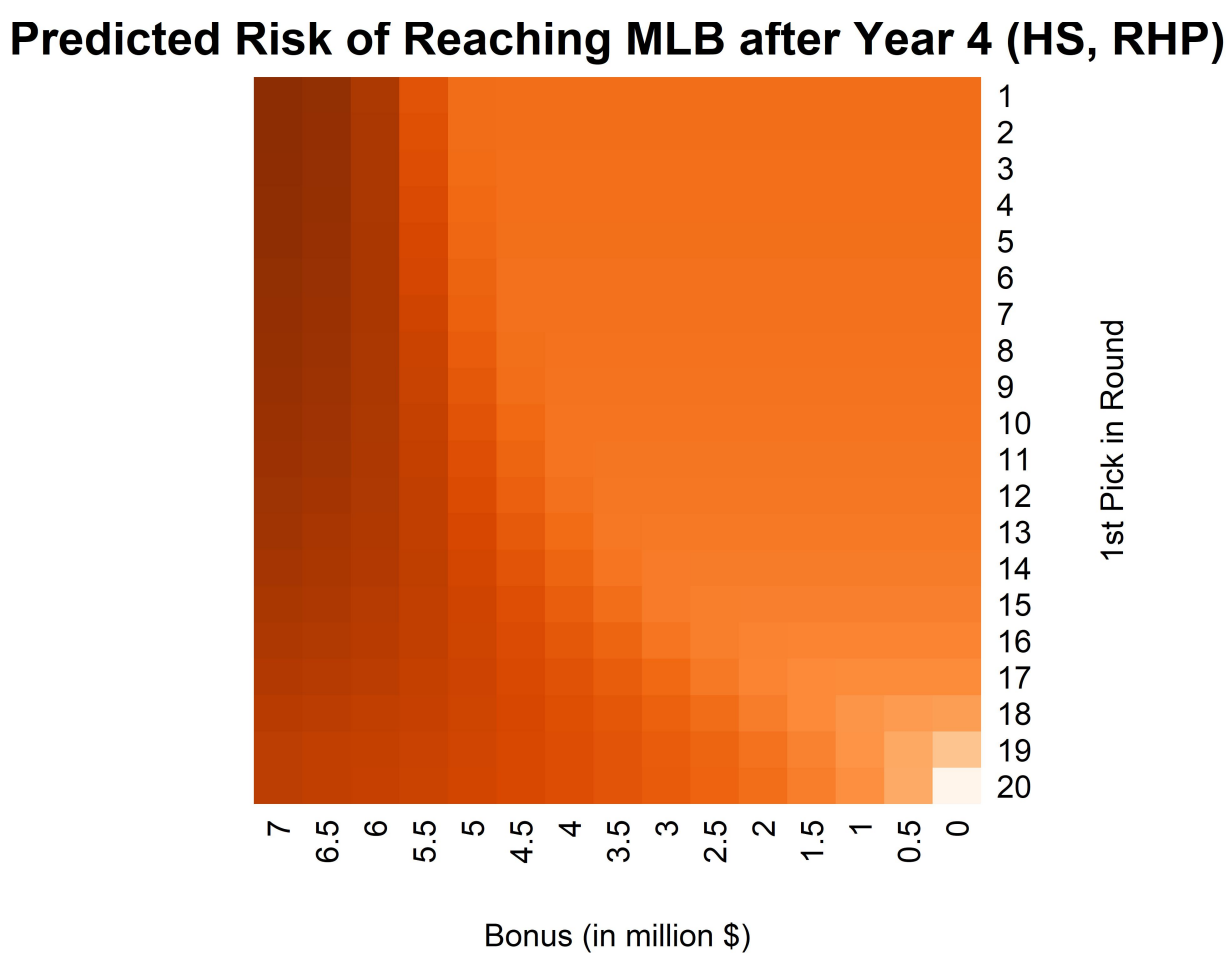


Fig. 5: Predicted risks of Reaching MLB for HS RHP across Round and Bonus

Other notable results:

- Leaving all else fixed, 1st round 4Yr batters have approx. 78% better chances on average (between 40% and 100% over time) of reaching MLB than similar HS draftees, while the chances of retiring remain only approx. 6% higher for 4Yr players.
- The coefficients of the model suggest RHP have 24% better odds of reaching MLB than batters, and LHP even higher at 39% better odds. Batters are also 15% more likely to retire early than pitchers. This suggests that LHP are in higher demand from MLB teams and have longer staying power.
- These results reflect some of those from [6] that suggest college players careers are decided more quickly than high school players, and that draft position plays an important role on future success.

Discussion

Each player should be considered on an individual basis. Trends such as those found in this research are useful tools in providing some context for young players on their chances of achieving their dream, but do not account for specific talent, performance, or personal situations.

You can also try out the model with our R Shiny App (pending update):

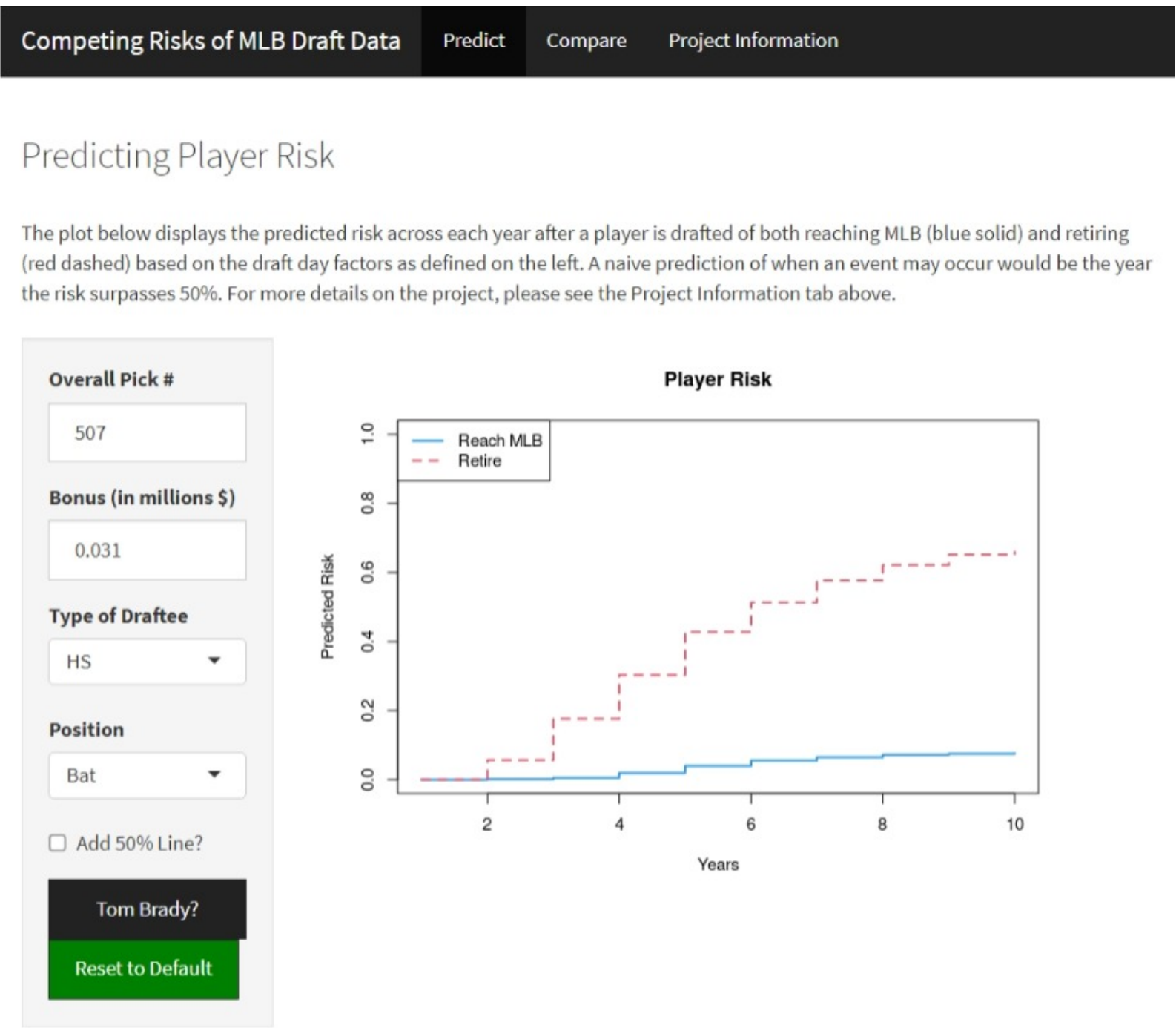


Fig. 6: The R Shiny App tells us Tom Brady made the right decision by playing football.



Acknowledgements

This research was funded in part through the Office of GRaSP at CSU Bakersfield as part of a faculty research funding program of the RCU. MLB draft data was manually collected from baseball-reference.com.

Students who have assisted with parts of this work include:

- Data collection: Nelson Guirado, Ryan Monahan
- EDA and basic survival analysis: Nikhil Bommarreddy
- App development and other data summaries: Kaamil Thobani, Vamshi Pagidi, Xi Chen, Yash Jayaprakash

References

- [1] G. Chandler and G. Stevens. “An exploratory study of minor league baseball statistics”. In: *Journal of Quantitative Analysis in Sports* 8.4 (2012), pp. 1–26.
- [2] S. Choi et al. “Weighted least-squares regression with competing risks data”. In: *Statistics in Medicine* 41.2 (2021), pp. 227–241. DOI: 10.1002/sim.9232.
- [3] J. P. Fine and R. J. Gray. “A Proportional Hazards Model for the Subdistribution of a Competing Risk”. In: *Journal of the American Statistical Association* 94.446 (1999), pp. 496–509.
- [4] Bob Gray. *cmprsk: Subdistribution Analysis of Competing Risks*. R package version 2.2-11. 2022. URL: <https://CRAN.R-project.org/package=cmprsk>.
- [5] R. J. Gray. “A class of K-sample tests for comparing the cumulative incidence of a competing risk”. In: *Annals of Statistics* 16.3 (1988), pp. 1141–1154.
- [6] N. D. Pifer et al. “Who Should Sign a Professional Baseball Contract? Quantifying the Financial Opportunity Costs of Major League Draftees”. In: *Journal of Sports Economics* 21.7 (2020), pp. 746–780.