

SIGNIFICANCE

Statistics have been part of the discourse surrounding sports from the beginning, and sports continue to provide interesting problems solved via statistical analysis. Each year, Major League Baseball teams draft players out of high school or college. Can we use data to help these players, their agents and the teams that draft them understand their chances on draft day? Statistical analysis can give insight into the futures of hundreds of young athletes every year.

Competing Risks in Baseball

A UNIQUE SPORT:

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). Even then, less than half of players drafted will one day make it to 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. Further, until 2022 MiLB teams were not required to provide housing for their players, and most MiLB players still earn only poverty-level wages. 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.

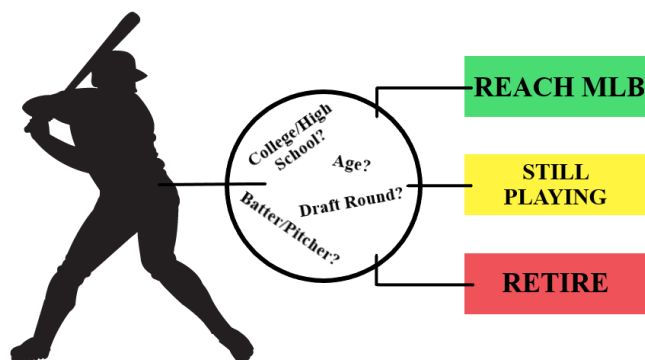
MODELING TIME TO COMPETING EVENTS:

Competing risks analysis is a subset of survival analysis; a statistical technique used most often in medical settings. 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.

WHAT OUR DATA LOOK LIKE:

All MLB draft data from 2012 to 2016 ($N = 4573$). Included are the years the player reached the majors, retired, or if they are still playing, as well as demographics including position, type of player (high school or college), signing bonus, and pick number, among others.



Competing risks analysis is used to understand the time until a drafted baseball player either reaches, or retires before reaching, the major leagues and how different factors impact these competing events.

RESULTS WITH AN IMPACT:

Several important takeaways reveal themselves from the analysis. Consider one notable example of first-round batter selections, assuming a \$2 million signing bonus: 4-year college players are more likely than high schoolers to reach MLB and to retire in each year after being drafted. The relative risk (college risk/high school risk) is higher for reaching MLB than retiring across all years. What this means is that talented, first-round college batters' futures in baseball are not only decided sooner than high school batters (for better or worse), but their chances relative to high school batters are better for reaching MLB. This suggests that even if they are one of the top talents available, high school batters may be risking quite a bit by signing early instead of attending college. By contrast, this relationship of the relative risks switches by year 7 for left-handed pitchers, suggesting more hope for high school lefties.

While each player should be considered on an individual basis, trends such as this can help players better understand where their chances stand, and provide context for a life-changing decision.