# Long-Term Impact of Superstars on Colleagues[*]

**Masaya Nishihata[a]**

### Abstract

This study examines the impact of superstars on their colleagues' opportunities and career trajectories using data from Major League Baseball (MLB). While superstars are often entrusted with key responsibilities due to their exceptional abilities, their presence may unintentionally limit teammates' chances to gain experience, potentially hindering long-term skill development and career progression. To address potential endogeneity, we employ coarsened exact matching (CEM) and leverage exogenous variation from superstar injuries. The results reveal that superstars significantly reduce their teammates' playing opportunities, and players who share a position with a superstar face a higher likelihood of exiting MLB, both in the short and long run. These findings have broader implications for workplace dynamics, suggesting that organizations' myopic reliance on a small number of highly talented individuals for short-term efficiency may have unintended consequences, limiting the growth and career prospects of those around them. Such short-term optimization can generate unintended long-term costs by discouraging human capital accumulation and undermining overall team sustainability.

Keywords: superstar, human capital, long-term career
JEL classification: J24, J44, L83

---

[*] This work was supported by JST SPRING, Grant Number JPMJSP2123. This paper has no conflicts of interest to disclose.
[a] Graduate School of Economics, Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan (e-mail: nishihata.masaya@gmail.com).

## 1. Introduction

Worker productivity is not uniform; even small differences in productivity can lead to disparities in task allocation, with certain tasks concentrating on a few highly talented individuals. Rosen (1981) highlights a demand structure that prioritizes quality over quantity, using examples from musicians, artists, and athletes, and theoretically explains that top performers dominate markets through technological advancements, such as the internet, and earn huge incomes. In the context of workplace task allocation, while the scale of this concentration may be less extreme, assigning key tasks to a few talented individuals could limit their colleagues' opportunities to gain experience, potentially hindering human capital accumulation and long-term career development. Understanding these negative externalities among colleagues is essential for optimizing task allocation and worker assignments.

This study uses data from Major League Baseball (MLB) to examine the impact of superstars on their colleagues. Specifically, we first investigate whether the presence of a superstar on the same team and in the same position reduces plate appearances by other players during a season. Next, we analyze the long-term effects of superstars on their colleagues by examining how variations in the presence of a superstar during a player's rookie year influence their career length.

## 2. Data

### 2.1. Data source

We use data from the Lahman Baseball Database, which contains comprehensive pitching, hitting, and fielding statistics for MLB players dating back to 1871, as well as All-Star Game results. Our analysis focuses on batters, excluding designated hitters (DH), since their position allows us to define substitution relationships with superstars. We supplement these data with records of player placements and activations on the Injured List (IL; formerly Disabled List, DL) from Baseball Prospectus, a baseball analytics platform that maintains player cards and transaction logs documenting IL/DL stints.

### 2.2. Definition of superstars

Based on an extensive literature review, Call et al. (2015) define a star employee as one whose (a) performance, (b) visibility, and (c) relevant social capital are disproportionately high and sustained. According to this definition, high visibility without corresponding competence or a "one-hit wonder" excelling in only a single season does not qualify as a star. Therefore, we define a superstar as a player who has been selected for the All-Star Game at least twice.

All-Star Game selection reflects a player's performance, and requiring at least two selections balances two risks: misclassifying players with insufficient merit as superstars and failing to recognize genuinely deserving players. A higher threshold for the number of All-Star selections would reduce misclassification risk but might exclude legitimate superstars, while a lower threshold could inflate the number of superstars by including marginal cases. For this study, we adopt the criterion of at least two All-Star selections to achieve this balance effectively.

## 3. Empirical framework

### 3.1. Superstar effect on game appearances

First, we assume conditional independence and implement coarsened exact matching (CEM). The rich detail in MLB data, including individual attributes and performance metrics, allows us to compare players in similar situations, differing only in the presence of a superstar in the same team and position. CEM involves coarsening continuous variables into bins and creating strata based on these coarsened variables and categorical variables. Only observations within strata that are exactly matched and exhibit variation in the treatment (presence of a superstar) are retained for analysis, with weights assigned to ensure that the distribution of the control group aligns with that of the treatment group.

To implement CEM, it is essential to properly assess the similarity between players. We use the variables employed in the calculation of similarity scores introduced by James (1994) to match comparable players. The similarity score starts at 1000 points and is reduced based on standardized differences between two players in terms of game appearances, at-bats, runs scored, hits, doubles, triples, home runs, runs batted in, walks, strikeouts, stolen bases, batting average, and slugging percentage. Additionally, a positional adjustment is applied based on the defensive positions played. By using these variables, the similarity score captures not only overall performance but also player type, distinguishing, for example, between power hitters and leadoff hitters. Following this methodology, we match players using the cumulative values of these variables up to the previous season, along with batting handedness (both-, left-, or right-handed), ensuring that players with comparable career profiles are grouped together. Continuous variables are coarsened using Sturges' rule, which is commonly applied to determine bin width for histogram-based data grouping.

For the matched sample, we estimate the following model using linear regression (OLS):

$$PA_{igt} = \beta_1 WithStar_{igt} + \gamma OPS_{igt} + f(Age_{it}) + X_{igt}\theta + \mu_i + \nu_{gt} + \varepsilon_{igt}, \quad (1)$$

where, $PA_{igt}$ represents the number of plate appearances by player $i$ on team $g$ during season $t$. $WithStar_{igt}$ is a binary variable indicating whether a superstar is in the same team and position, $OPS_{igt}$ is the OPS, capturing batting ability, $f(Age_{it})$ includes indicator variables for age, $X_{igt}$ is control variables such as within-season transfer of player $i$ and the superstar, $\mu_i$ represents individual fixed effects, $\nu_{gt}$ denotes dyad fixed effects for team and season, $\varepsilon_{igt}$ is the error term. The coefficient of interest, $\beta_1$, captures the impact of the presence of superstars on appearances. To address potential endogeneity, we control for these variables in the regression in addition to using CEM. $WithStar_{igt}$ is assumed to be exogenous under the conditional independence assumption.

Second, we present an estimation strategy leveraging exogenous variation that closely resembles a natural experiment as a robustness check. For the sample where $WithStar_{igt} = 1$, we estimate the following model, using the periods during which the superstar was on the injured list as the explanatory variable:

$$PA_{ijgt} = \beta_2 StarIL_{ijgt} + \gamma OPS_{ijgt} + f(Age_{it}) + POS_{igt}\delta + X_{ijgt}\theta + \eta_{ij} + \nu_{gt} + \varepsilon_{igt}, \quad (2)$$

where, $StarIL_{ijgt}$ indicates the periods during which the superstar $j$, who shares the same team and position as player $i$, was on the injured list. $POS_{igt}$ represents the main defensive position. The term $\eta_{ij}$ represents dyad fixed effects for player $i$ and superstar $j$. The coefficient of interest, $\beta_2$, captures the effects of the superstar's absence. Consistent with expectations, $\beta_2$ is anticipated to have the opposite sign to $\beta_1$.

### 3.2. Superstar effect on career duration

In this study, we employ the Cox proportional hazards model (Cox, 1972) to analyze the survival times of individuals, which, in this context, refers to the career lengths of MLB players. The Cox model is a semi-parametric approach widely used in survival analysis. It models the hazard rate at t—the instantaneous risk of an event (exit from MLB) occurring—as a function of covariates while allowing for a non-parametric baseline hazard function.

The model is specified as follows:

$$\lambda(t|\boldsymbol{X}) = \lambda_0(t)\exp(\beta_3 WithStar_{igt} + \beta_4 WithStar_{it_0} + POS_{igt}\delta + \phi(Age_{it}) + X_{igt}\theta + \omega_g), \quad (3)$$

where $\lambda(t|\boldsymbol{X})$ is the hazard function at time $t$ conditional on the covariates, representing the instantaneous exit rate of exiting MLB during the marginal time interval $[t, t+1)$, given that no exit has occurred before $t$. The term $\lambda_0(t)$ represents the baseline hazard function, which accounts for the risk of exit conditional on all covariates being zero.

The key covariates include $WithStar_{it_0}$, a binary variable indicating whether a superstar was on the same team and in the same position as player $i$ during their rookie year. Unlike $WithStar_{igt}$, which reflects whether a player shares a position with a superstar in a given season, this variable captures the potential long-term impact of early career experiences. $\phi(Age_{it})$ accounts for the player's age and its squared term. $\omega_g$ represents team fixed effects. Unlike previous models, this analysis does not control for $OPS_{igt}$, as doing so could lead to an underestimation of cumulative effects. If early career experiences limit a player's OPS growth, controlling for OPS would obscure the long-term influence of initial exposure to a superstar. The coefficients of interest are $\beta_3$ and $\beta_4$. $\beta_3$ captures the impact of having a superstar teammate in the same position during a given season, while $\beta_4$ reflects the long-term effect of sharing a position with a superstar during a player's rookie year. A positive $\beta_4$ suggests that such an early-career experience increases the hazard rate, implying a higher probability of a shorter MLB career.

## 4. Results

Table 1 presents the regression results examining the impact of superstars on plate appearances. Column (1) shows the results of Equation (1) using the unmatched sample, where the coefficient of $WithStar_{igt}$ is -42.438. This indicates that the presence of a superstar on the same team and in the same position reduces a teammate's plate appearances by 42.438 per season. Column (2) presents the results of Equation (1) after applying CEM to address confounding factors. The coefficient of $WithStar_{igt}$ is -29.244 even after mitigating potential endogeneity. Column (3) reports the results of Equation (2), leveraging exogenous variation as a robustness check. The coefficient of $StarIL_{ijgt}$ is 0.149, meaning that for each day a superstar is on the injured list, their teammates gain an additional 0.149 plate appearances. Since a season is approximately six months long, if a superstar is absent for an entire season (about 180 days), this effect accumulates to approximately 26.82 additional plate appearances.

Table 1: Superstar effects on teammates' plate appearances

| | (1) | (2) | (3) |
|---|---|---|---|
| *Dependent variable:* | $PA_{ijgt}$ | $PA_{ijgt}$ | $PA_{ijgt}$ |
| $WithStar_{igt}$ | -42.438*** | -29.244*** | |
| | (4.238) | (5.460) | |
| $StarIL_{ijgt}$ | | | 0.149** |
| | | | (0.075) |
| $OPS_{igt}, f(Age_{it}), X_{igt}$ | ✓ | ✓ | ✓ |
| $POS_{igt}$ | ✓ | | ✓ |
| $\mu_i$ | ✓ | ✓ | |
| $\eta_{ij}$ | | | ✓ |
| $\nu_{gt}$ | ✓ | ✓ | ✓ |
| Sample | whole | matched | w/ stars |
| Observations | 14022 | 5673 | 6742 |
| $R^2$ | 0.655 | 0.737 | 0.764 |
| Adjusted $R^2$ | 0.568 | 0.578 | 0.642 |
| RMSE | 122.50 | 100.20 | 101.25 |

*Notes*: Robust standard errors with clustering at the individual level are in parentheses. *** and ** denote significance at the 1% and 5% level, respectively.

Table 2 presents the results from the Cox proportional hazards model, analyzing the impact of superstars on the career length of MLB players. The coefficient of $WithStar_{igt}$ in Column (1) is 0.172, indicating that sharing a position with a superstar in a given season is associated with an increased hazard rate. This suggests that players competing with a superstar for playing time face a higher likelihood of exiting MLB in the short term. More notably, the coefficient of $WithStar_{it_0}$ in Column (2) is 0.123 and remains significant in Column (3), even after controlling for the presence of a current-season superstar. This finding reveals that early-career exposure to a superstar increases the hazard rate in the long run, independent of short-term superstar effects. In other words, even after accounting for the possibility that early-career exposure to a superstar increases the likelihood of sharing a position with a superstar in later seasons, players who shared a position with a superstar during their rookie year are more likely to experience shorter MLB careers. Overall, these results suggest that both short-term and long-term superstar effects shape a player's career trajectory, with immediate competition limiting opportunities and early-career experiences exerting a lasting influence on career length.

Table 2: Superstar effects on teammates' career duration

| | (1) | (2) | (3) |
|---|---|---|---|
| *Dependent variable:* | $\lambda(t\|\boldsymbol{X})$ | $\lambda(t\|\boldsymbol{X})$ | $\lambda(t\|\boldsymbol{X})$ |
| $WithStar_{igt}$ | 0.172*** | | 0.161*** |
| | (0.054) | | (0.054) |
| $WithStar_{it_0}$ | | 0.123** | 0.111** |
| | | (0.053) | (0.053) |
| $POS_{igt}, \phi(Age_{it}), X_{igt}$ | ✓ | ✓ | ✓ |
| $\omega_g$ | ✓ | ✓ | ✓ |
| Observations | 11137 | 11137 | 11137 |
| AIC | 22056.8 | 22061.3 | 22052.6 |
| BIC | 22364.1 | 22368.6 | 22367.2 |
| RMSE | 0.40 | 0.40 | 0.40 |

*Notes*: Robust standard errors with clustering at the individual level are in parentheses. *** and ** denote significance at the 1% and 5% level, respectively.

Reference

Call, M. L., Nyberg, A. J., and Thatcher, S., 2015. Stargazing: an integrative conceptual review, theoretical reconciliation, and extension for star employee research. Journal of Applied Psychology 100(3), 623.

James, B., 1994. The Politics of Glory: How Baseball's Hall of Fame Really Works. Macmillan.

Rosen, S., 1981. The economics of superstars. American Economic Review 71(5), 845-858.