

Statistical Baseball Research Bibliography
Charlie Pavitt

The goal of this essay is to introduce the 2022 version of the Statistical Baseball Research Bibliography and explain its use. The Bibliography is the result of a comprehensive survey of published baseball literature. Along with many books, it includes articles in baseball journals (most notably the *Baseball Analyst*, *Baseball Research Journal*, and *By the Numbers*), academic sports journals (for example, *Journal of Sports Economics*, *Journal of Sports Analytics*, and *Journal of Quantitative Analysis in Sports*), and other academic journals (for example, *Operations Research*, *American Journal of Sports Medicine*, and *The American Statistician*). I am currently in the process of adding online material, most notably those for Baseball Prospectus and FanGraphs along with unique ones such as those for the annual MIT Sloan Sports Conference. A lot of online material is here, but there is still a lot left to add.

Articles have been included in the Bibliography if they meet the following criteria:

1 - They have been intended to make a contribution to our knowledge about baseball as a statistical science. This does not mean that the article must include statistical analysis. Many worthy articles have made theoretical or critical contributions without performing statistical analyses.

2 - While articles that present methods for evaluating or ranking teams or players are included if they make a meaningful contribution, articles that do nothing more than evaluate or rate teams or players are not included. Books that appear to have been intended to do little more than exploit the popular market for books on baseball statistics and make no contributions to our knowledge have not been included and will not be added in the future.

3 - Books must either have been published in conventional markets (either academic or trade) or by SABR. Self-published works will only be included if they have made an unusual contribution (such as Cook's *Percentage Baseball and the Computer* and the early James *Baseball Abstracts*).

Each entry has eleven columns of information. These are as follows:

Column A - Last name of first author.

Column B - First name and middle initial of first author.

Columns C and D present a code system identifying the content of the article. These provide the organizational structure of the Bibliography. If an article includes more than one clearly different content area, it will be entered under each of them. In the case of books, it will be entered under each content area covered. The code system consists of two hierarchically organized levels, respectively called the macrocode, and microcode. Each of these codes is symbolized by one or two capital letters. To begin, each article is categorized within a general subject area. This general subject area is indicated by the article's macrocode, which can be found in Column C. Each general subject area is divided into more specific content areas. Each

specific content area is indicated by the article's microcode, which can be found in Column D. It is important to remember that the same microcode may symbolize a different category for different general areas. For example, the microcode S indicates Sample Size within the macrocode category Batting Evaluation, Starter/reliever within the macrocode category Pitching Issues, and Selection within the macrocode category Managing.

Column E - The title of the article or book. Due to space limitations, the title may be shortened or paraphrased.

Column F- If journal article, title of journal, If book, name of publisher, If contributed chapter in book edited or mostly written by someone else, editor/author listed here. Look for listing of book under editor or primary author for title/publisher/year. If webpost, URL.

Column G - If journal article, volume or issue journal. If book, location of publisher.

Column H- Date of publication.

Column I- Pages that article is on. If article/book includes more than one subject area, pages will be specific to the discussion of the subject area if it is clearly distinct from the others.

Column J- Comments. If article is multi-authored, coauthors will generally be listed here. If article is part of debate, extension of earlier article, etc., other article(s) in series will be cited. When Column E consists of book title, title of book chapter may be listed here.

I update the Statistical Baseball Research Bibliography periodically with both new and old material that I find. Therefore, I would be interested in seeing any statistically-based articles anyone is familiar with not in the present version of the Bibliography and considering them for inclusion in future versions if they meet the criteria listed above.

Macrocode	Microcode	Comments
Base Running	<ul style="list-style-type: none"> Advancement on Infield balls in play Advancement on Outfield balls in play reaching Base on Error other Base running Issues Base runner Speed Pinch Running base Running Evaluation base Stealing Evaluation base Stealing Strategy 	Evaluation and strategy
Batting Evaluation	<ul style="list-style-type: none"> Ability Ballpark adjustments performance Consistency run Expectancy Fundamental metrics a team of a Given player Hitf/x and pitchf/x ranKing replacement Level comparison among Methods Other Predicting Performance Reliability Sample Size Top-down bottom-Up 	<ul style="list-style-type: none"> All nonsituational aspects of offense Measured overall and for specific skills In batters Type of evaluation method BA, OBA, SLG, RBI Type of evaluation method Type of evaluation method Methods for measuring who is best Type of evaluation method Miscellaneous evaluation methods Of measures And estimating performance Type of evaluation method Type of evaluation method
Batting Issues	<ul style="list-style-type: none"> Age/experience Clutch performance enhancement Drugs reaching on Errors Great feat odds Historical changes K - strikeouts bLack/white/Latin piNchhitting Protection hot/cold Streaks Transactions silVer slugger Walks 	<ul style="list-style-type: none"> Changes in performance over career Does it exist? If so, how to measure it Impact Differences among players DiMaggio streak and others Differences in performance As compared to being in lineup Does it exist? Do they exist? Effect of changing team on individual performance Predicting who wins Impact on offense; measuring batting eye

	Hit by pitch size	Who gets hit, who does the hitting Relation with performance
Batting Strategy	At bat Bunt hits Days off sacrifice Flies Hit and run Lineup Order Pinch hitting Sacrifice hits platooning swing decisions	Strategies other than swing decisions Bunts not scored as sacrifices Impact on performance Are they strategic? When does it help? Who plays? Impact on performance What type of player is best? When do they help? When does it help? And plate discipline
Fielding Evaluation	Catching Double play Errors Field/x-based methods Gold Glove Infielding Miscellaneous methods Outfielding Range factor Team Zone rating (and progeny)	Methods for evaluating Impact on overall defense Significance of For evaluating Predicting who wins Methods for evaluating For evaluating Methods for evaluating Relevant discussion Methods for evaluating Methods for evaluating
Fielding Strategy	Back pick Fielding position Infield positioning Outfield positioning Pitch out and Pick offs fielding Substitutions intentional Walks	What we know Where should they play? When does it help? When does it help? Does it help? What effects do they have? When do they help?
Game	Big bang theory Game length Odds of winning Runs per Game Runs per Inning	Validity

General	Introductions to baseball research	No macro or microcodes.
Inning	Event type Leadoff hitter Odds of scoring/winning Performance Run expectancy	Does it matter? Impact of In various base-out situations In various base-out situations In various base-out situations
injury	other Arm injuries Concussion Elbow Forearm General Hand and wrist Illness Leg Predicting injury Shoulder Trunk	Incidence rates and performance outcomes
Managing	Coaches Evaluation Play call challenges Selection Tenure Turnover	Evaluation Methods for measuring how good How successful? Characteristics affecting it Characteristics affecting it Characteristics affecting it
Overall player evaluation	All-star Hall of fame miscellaneous Issues Most valuable player Other methods Ranking win Shares Total player rating War	Total performance Predicting who gets chosen Predicting who gets in Predicting who wins Methods for measuring who is best Method for Methods for measuring
Pitching Evaluation	Ability Dips Era based pitchF/x replacement Level	Measured overall and for specific skills Issues concerning its validity Type of evaluation method Type of evaluation method Type of evaluation method

post Mccracken Methods	Type of evaluation method
Other methods	
Predicting Performance	Type of method
Quality start	Type of method
Run expectancy tables	Type of evaluation method
Transplanted offensive methods	Type of evaluation method
relieVers	Type of evaluation method
Winning average	Attempts to salvage

Pitching Issues

Age/experience	Its effects
Breakdowns	Groundball/flyball, power/finesse, starter/reliever
Clutch	Does it exist? If so, how to measure it
performance enhancing Drugs	Impact
consistEncy	In performance
Historical changes	
Interdependence	Among pitchers
K – Pitcher hitting	
bLack/white/Latin	Differences in performance
No hitters	How likely? Who is likely to throw one?
wOrkload	Analysis of ideal load, effects of layoffs and overwork
Run support	Impact
Transactions.	Effect of changing team on performance
Walks	Their implications
cY Young award	Predicting who wins
siZe	Relation with performance

Pitching Strategy

Hit by pitches	As strategy, and other reasons for it
location/type/velocity Interactions	
pitch Location	Impact
Other strategy	
Reliever usage patterns	
pitch Sequencing	Impact
pitch Type	Impact
starter Usage patterns	
pitch Velocity	Impact

Situational

Batter/pitcher matchups	Can one “own’ the other?
Count	Impact of balls vs. strikes
Doubleheaders	
Fly ball/ground ball	Impact on performance
Home/away	Impact on performance
late Inning pressure situations	Impact on performance

Left/right	Implications of platoon differentials
day/Night	Impact on performance
Opposition	Implications of who team is playing
ball Park	Impact on performance
Runners in scoring position	Impact on performance
Season	Tendencies from month to month
Turf/grass	Impact on performance

Team Issues

Analytics	Influence of
Birth date	
the Coase-rottenberg theorem	
the Draft	Position and future performance
Ethnicity	Stacking
Front office evaluation	How good?
birth Location	Country, section of U.S.
Minor league development	Major league career and performance
Pay and performance	
Roster design	Transactions, salary, player turnover, etc.
diSability	Impact

Team Performance

winning Average	Relevant issues
competitive Balance	
Components	Relative impact of hitting, pitching etc.
run Differential	And winning percentage
Ethnicity	Impact
miscellaneous Factors related to team performance	
Hot/cold streaks	
Matchups	Which team wins
Other	
the Plexiglass Principle	Regression to the mean
Ranking	Standings
run Scoring	Attempts to model
World series and playoffs	What makes winners?

Umpire

Accuracy	In pitch calling
Bias	In pitch calling
Changes over time	In pitch calling
Ethnic bias	Does it exist?
Other	On topics other than pitch calling
Roboump	Accuracy and impact