

INTRODUCTION *In the Dugout*

Never did I expect that writing this book would lead me into the San Diego Padres dugout. On July 28, 1997, however, two hours before that evening's game against the Philadelphia Phillies, there I was! Tony Gwynn had just returned his Louisville Slugger to the bat rack after batting practice.

"Tony!—I'm Michael Schell," I called out. Tony Gwynn, the 7-time batting champion from the San Diego Padres, turned toward me and replied, "Soooo—you're the guy!"

A month earlier I had sent a press release to the media relations people at the Padres saying that Gwynn was on the verge of clinching his standing as the best pure hitter in baseball history. A week earlier I had sent him a congratulatory note after he got the clinching hit. Two days before, sportswriter Wayne Lockwood presented my findings in the *San Diego Union-Tribune*.

"I'll talk with you in a minute," Gwynn added as he headed to chat with some early-bird fans that he knew.

I waited expectantly and a little nervously. Shaun O'Neill, a sports reporter for the *North County Times* told me that Gwynn was a very unassuming ballplayer who would downplay what I was going to say but would listen intently.

"You're putting pressure on me!" Gwynn moaned jokingly as he approached.

He was hitting .391 and the media had been hounding him for weeks about the possibility of becoming the first player in 56 years to hit .400 for the season. That week, in fact, Gwynn was on the cover of both *Sports Illustrated* and *The Sporting News*.

"You've got no pressure from me," I countered, "you've already done it!"

He sat down beside me and said, "Show me what you found."

That's what I plan to do in this book—show the reader what I've

found by developing a method to compare players across baseball history, from the first pitch in 1876 to the present day.

The Tyranny of Traditional Top Hitters Lists

Most baseball encyclopedias and many almanacs have lists of top lifetime hitters or single season batting champions. Young fans memorize the names—Ty Cobb, Rogers Hornsby, Ted Williams, Babe Ruth—and often their lifetime batting averages as well (.366, .358, .344, .342, respectively). These players have become mythological heroes of the game.

There is some sadness, though, among fans today since our favorite active players—Tony Gwynn, Mike Piazza, Ken Griffey Jr. are—hopelessly out of the top positions. Gwynn would rank 16th, Piazza 21st and Griffey fails to make the top 100. Moreover, Hall of Famers Hank Aaron and Willie Mays and all-time hits leader Pete Rose are off the list, while a host of relatively unknown players like Bibb Falk, Cecil Travis, Rip Radcliff, and Elmer Smith are on it. How can this be?

Knowing how extraordinary these current and recent players are, we become mystified by those on the list. How did they do it? Why were they so much better? The punch line, which is the subject of this book, is that they are *not* so much better. We fans have been misled by the averages. It is the unfortunate fact of life that fair appraisals of anything rarely come without effort.

Grandparents may tell you about how they bought a house for \$15,000. Did they also tell you, however, that they only earned \$6,000 per year then, too? Simply defined statistics, like *batting average* (which equals hits divided by at bats), may be fine to make comparisons between ballplayers playing in the same year in the same ballparks against the same pitchers. But why should these averages be used at all to compare a player who played at night in a domed stadium with astroturf with another player who played only day games in the open air on natural grass?

The question of who the greatest hitters are is a subject of considerable interest to baseball fans. It is a source of argument between father and son, between Dodger and Yankee fan, between the pure hitter fan and the slugger fan. The good news is that we can reasonably answer this question, when it is clearly posed. This is the legacy pro-

vided by baseball, which has a wealth of statistical data over a hundred-year period. The bad news is that the answer is not easily found in the baseball encyclopedias and almanacs. It is the aim of this book to identify the 100 greatest hitters, by applying four adjustments to the standard batting average.

“Best Hitter” Defined

What does the phrase “best hitter” mean? Hitting is composed of many things. For example, Tony Gwynn is excellent at getting hits but relatively few of his hits are home runs. On the other hand, Mark McGwire is only average in getting hits but they go a country mile when he does connect! So which one is the “better” hitter?

There are many different baseball statistics. Batting average and slugging average both combine singles, doubles, triples, home runs, and outs into single measures. Batting average is computed by totaling the different kinds of hits and dividing by the number of at bats, while slugging average totals the number of bases that you reach on the hits before being divided by the total number of at bats. However, both of them ignore factors like the walk average, number of RBIs and less well-measured things like hit-and-run or clutch hitting ability. There are other ballplayer abilities as well, such as run scoring, base stealing, and fielding. Bill James, with his *Runs Created* formulae, and John Thorn and Pete Palmer, with their *Total Player Rating*, have combined batting, fielding, and stealing data into a single rating. Both composite statistics are useful and interesting.

The statistics of Bill James and Thorn and Palmer seem to be searching for the best *players*. Statistics that combine various hitting events, which may include weighting of the values of singles, doubles, triples, and home runs (and possibly walks, strikeouts, or other batting events) are searching for the best *batters*. The search in this book is for the best *hitters*, that is, the players with the best chance to get a hit in a given at bat. Thus, we will use the preeminent baseball statistic, the *batting average*. However, we will adjust this average for each year in baseball history based on the ease with which hits could be attained and the player’s home ballpark. This leads to batting averages that are *relative* to the league batting averages. Consequently, the talent pool of the league must also be considered. Also, at bats late in the careers of

the longest playing stars will not be included, since batting ability clearly wanes then. Because these adjustments are needed to *level the playing field*, standard batting average lists do not properly order the top hitters.

Minimum Requirements of Qualifying Players

In order to determine the 100 greatest hitters of all time there must be a minimum playing time. I have chosen 4000 at bats as the minimum. One could choose 4000 plate appearances, which includes both at bats and walks and a few other minor events, so as not to penalize individuals who walked frequently. However, since the focus here is on the ability to hit, not the ability to get on base, a minimum of at bats is used. A player who plays full-time will get 400–600 at bats per year, so 4000 at bats represents 7 to 10 years of full-time play.

This minimum — 4000 at bats — is close to those used by the three major contemporary baseball encyclopedias. *The Baseball Encyclopedia* uses the same definition. *The Sports Encyclopedia: Baseball* also requires a 10-year career. *Total Baseball* requires appearances in 1000 games. Thus, *Total Baseball's* top 100 list includes 5 players who have fewer than 4000 at bats, including Bob Fothergill, who had a mere 3265. The others are John McGraw, Mike Donlin, Dale Mitchell, and Taffy Wright.

It is particularly regrettable that John McGraw is left out. This outstanding, turn-of-the-century third baseman is in the Hall of Fame as the manager who skippered many successful New York Giant teams. He was a very good hitter and one of the best players of his day at drawing a walk. Had he not been so good at walking, he would easily have gotten the extra 76 at bats to qualify. He will be included when on base average is considered later in the book.

I also require that the player be retired or have at least 8000 at bats, if he is still active. (This additional requirement will be explained further in the chapter on late career declines.) Thus 10 active players (as of spring 1998) are included: Harold Baines, Wade Boggs, Joe Carter, Chili Davis, Gary Gaetti, Tony Gwynn, Rickey Henderson, Paul Molitor, Tim Lincecum, and Cal Ripken Jr.

Through 1995, 8259 players have played major league baseball, excluding the Negro Leagues. However, only 836 players — who will be

called *qualifying players*—qualify for consideration, by having had at least 4000 at bats and being retired or having had at least 8000 at bats, if still active. Thus, the top 100 hitters are members of an elite group—comprising only 1.2% of all major league batters and 12% of qualifying players.

Unfortunately, records from the Negro Leagues are incomplete although they are now being compiled and included in baseball encyclopedias. As a result, they are not ranked in this book.

Batting Average Data Sources

There is no universally accepted list of top hitters. Besides the minimum eligibility criteria, reference books differ on how many hits and at bats each player actually had. Fortunately, this problem is largely limited to players from the 19th and early part of the 20th century. For example, Cap Anson, who played from 1876 to 1897 batted $.329 = (3000 / 9108)$ according to *The Baseball Encyclopedia, Tenth Edition*, 1996, $.329 = (2995 / 9101)$ according to *Total Baseball, Fifth Edition*, 1997, and $.333 = (3022 / 9067)$ according to *The Sports Encyclopedia: Baseball, Seventeenth Edition*, 1997.

Total Baseball received much of its information from sabermetricians (members of the Society for American Baseball Research—SABR), who carefully researched box scores for many of these discrepancies. Thorn and Palmer described some of the corrections that they made to “official” major league records in the “Errors and Controversies” section of the introduction of the Player Register for *Total Baseball*, 1989. Since I believe that it provides the most accurate numbers available, *Total Baseball* will be the principal data source for hits and at bats for players in this book.

The Traditional Top 100 Hitters List

Table I gives the traditional list of top 100 hitters. This is *not* my list of top 100 hitters. However, this list is useful as a starting point. A look at the traditional list shows that 48 hitters were in their prime during the 1920s and 1930s! On the other hand, only 8 players on the list were in their prime on or after 1960.

Table I The Traditional List of Top 100 Hitters

Rank	Player	Average	Mid Yr	Top 5
1	Ty Cobb	.366	1917	16
2	Rogers Hornsby	.358	1926	12
3	Joe Jackson	.356	1914	7
4	Ed Delahanty	.346	1896	9
5	Ted Williams	.344	1950	12
6	Billy Hamilton	.344	1895	7
7	Tris Speaker	.344	1918	<u>13</u>
8	Dan Brouthers	.342	1892	<u>11</u>
9	Babe Ruth	.342	1925	<u>8</u>
10	Harry Heilmann	.342	1923	<u>8</u>
11	Pete Browning	.341	1888	8
12	Willie Keeler	.341	1901	10
13	Bill Terry	.341	1930	<u>7</u>
14	George Sisler	.340	1923	7
15	Lou Gehrig	.340	1931	<u>9</u>
16	Tony Gwynn	.340		13
17	Jesse Burkett	.338	1898	7
18	Nap Lajoie	.338	1906	7
19	Riggs Stephenson	.336	1928	2
20	Al Simmons	.334	1934	7
21	Paul Waner	.333	1935	8
22	Eddie Collins	.333	1918	<u>11</u>
23	Wade Boggs	.331		<u>11</u>
24	Stan Musial	.331	1952	16
25	Sam Thompson	.331	1896	<u>5</u>
26	Heinie Manush	.330	1931	<u>6</u>
27	Cap Anson	.329	1887	10
28	Rod Carew	.328	1976	12
29	Honus Wagner	.327	1907	13
30	Tip O'Neill	.326	1888	5
31	Jimmie Foxx	.325	1935	7
32	Earle Combs	.325	1930	1
33	Joe DiMaggio	.325	1944	6
34	Babe Herman	.324	1936	2
35	Hugh Duffy	.324	1897	<u>3</u>

(continued)

Table I The Traditional List of Top 100 Hitters (*continued*)

Rank	Player	Average	Mid Yr	Top 5
36	Joe Medwick	.324	1940	6
37	Edd Roush	.323	1922	7
38	Sam Rice	.322	1925	0
39	Ross Youngs	.322	1922	3
40	Kiki Cuyler	.321	1930	3
41	Charlie Gehringer	.320	1933	5
42	Chuck Klein	.320	1936	4
43	Pie Traynor	.320	1929	1
44	Mickey Cochrane	.320	1931	2
45	Ken Williams	.319	1922	2
46	Kirby Puckett	.318	1989	5
47	Earl Averill	.318	1935	1
48	Arky Vaughan	.318	1940	4
49	Roberto Clemente	.317	1964	10
50	Chick Hafey	.317	1931	1
51	Joe Kelley	.317	1900	1
52	Zach Wheat	.317	1918	6
53	Roger Connor	.317	1889	6
54	Lloyd Waner	.316	1936	1
55	Frankie Frisch	.316	1928	2
56	Goose Goslin	.316	1930	2
57	George Van Haltren	.316	1895	1
58	Bibb Falk	.314	1926	1
59	Cecil Travis	.314	1940	3
60	Hank Greenberg	.313	1939	2
61	Jack Fournier	.313	1920	4
62	Elmer Flick	.313	1904	5
63	Bill Dickey	.313	1937	0
64	Johnny Mize	.312	1945	6
65	Joe Sewell	.312	1927	0
66	Fred Clarke	.312	1905	2
67	Barney McCosky	.312	1946	2
68	Bing Miller	.312	1929	2
69	Hugh Jennings	.311	1905	1
70	Fred Lindstrom	.311	1930	2

(*continued*)

Table I The Traditional List of Top 100 Hitters (*continued*)

Rank	Player	Average	Mid Yr	Top 5
71	Jackie Robinson	.311	1952	<u>4</u>
72	Baby Doll Jacobson	.311	1921	0
73	Rip Radcliff	.311	1939	1
74	Ginger Beaumont	.311	1905	<u>3</u>
75	Mike Tiernan	.311	1893	3
76	Denny Lyons	.310	1891	2
77	Luke Appling	.310	1940	3
78	Irish Meusel	.310	1921	1
79	Elmer Smith	.310	1894	1
80	Bobby Veach	.310	1919	2
81	Jim O'Rourke	.310	1890	5
82	John Stone	.310	1933	0
83	Jim Bottomley	.310	1930	2
84	Sam Crawford	.309	1908	8
85	Bob Meusel	.309	1925	0
86	Jack Tobin	.309	1921	1
87	Spud Davis	.308	1937	0
88	Richie Ashburn	.308	1955	4
89	King Kelly	.308	1884	6
90	Jake Beckley	.308	1898	3
91	Paul Molitor	.308		6
92	Stuffy McInnis	.308	1918	0
93	Don Mattingly	.307	1989	<u>4</u>
94	Joe Vosmik	.307	1937	<u>1</u>
95	Frank Baker	.307	1915	3
96	George Burns	.307	1922	3
97	Matty Alou	.307	1967	<u>4</u>
98	Hack Wilson	.307	1929	0
99	Johnny Pesky	.307	1948	3
100	George Kell	.306	1950	<u>7</u>

Notes:

Mid Yr = year that the player attained half of his career at bats (blank for active players).

Top 5 = number of years among the league's top 5 hitters.

Underline = one batting title; bold = two or more titles.

As the book develops, I will drop and add players to the list, changing it after each adjustment. The ranking of players who remain on the list will also change, sometimes rather dramatically. After all adjustments have been made, only 18 of the 50 hitters from 1920–39 will remain. On the other hand, the number of players whose prime was 1960 or later will jump from 8 to 39.

A Note About Statistical Methods

This book makes use of statistical methods. They are presented as simply as possible and it is my intention that readers with no prior statistical knowledge will still enjoy this book and understand the basic ideas.

The use of statistics is the way that science often judges whether or not an idea that somebody dreams up is supported by the data. When the evidence from data is overwhelming, no statistics are needed. It is only in close cases that statistical analysis is needed. Just as a microscope enhances the seeing power of the eye, so statistics allows us to obtain convincing evidence about some question with less data than is required to have *overwhelming* evidence.

I am a professional statistician in a cancer research center. The center seeks knowledge on how to prevent and best treat cancer. Since it is important to make discoveries in cancer as soon as possible, statisticians are part of the team. Why shouldn't we use the same tools scientists do so that we can make additional discoveries about baseball?

Statistical methods or concepts used in this book have been placed in boxes labeled *Technical Notes*. These may be skipped without interrupting the flow of the book. Use of the statistical method or concept will usually immediately follow the box. Readers who are not interested in these details can look at the interpretation that I provide. Once a statistical method is introduced, it may be used several times in the book. A list of the Technical Notes is given at the beginning of the book for easy reference.

Organization of the Book

The book is divided into two parts: Methods and Findings. Chapters 1–5 comprise the Methods section. In chapter 1, the basic characteris-

tics of the players eligible for consideration in the top 100 are described. In chapters 2–5 the four adjustments to batting averages will be introduced and applied. Chapters 6–12 comprise the Findings section. In chapter 6 the fully adjusted top 100 hitters are identified. Later chapters deal with top hitters by position, top single-season batting averages, the best batting teams of all-time, ballpark effects for the 20th-century stadiums, top players for on base percentage, who should really be in the Hall of Fame, where today’s hitters would place among the top 100 hitters, and a wrap-up. Readers who can’t wait for the top 100 list can jump to chapter 6.