ChatGPT

Plan for At-Bat Simulation Using 2025 Cubs Statistics

Available Statistics from Baseball Reference

Hitting (Batter) Stats: The 2025 Chicago Cubs team page provides standard batting stats for each player. Key metrics include:

- **Games (G)** and **Plate Appearances (PA)** – total opportunities for the batter.
- **At-Bats (AB)** – official at-bats (excludes walks, hit-by-pitch, etc.).
- **Hits (H)** – total hits, broken down into **Doubles (2B)**, **Triples (3B)**, and **Home Runs (HR)**.
- **Runs (R)** and **Runs Batted In (RBI)** – scoring and runs driven in.
- **Stolen Bases (SB)** and **Caught Stealing (CS)** – base-running stats (not directly used in at-bat outcome, but available).
- **Walks (BB)** and **Strikeouts (SO)** – counts of plate appearances ending in a walk or strikeout.
- **Batting Average (BA)**, **On-Base Percentage (OBP)**, **Slugging Percentage (SLG)**, and **On-Base Plus Slugging (OPS)** – rate stats summarizing hitting performance.
- *Advanced ratios:* From these, we can derive **Strikeout Rate (K%)** and **Walk Rate (BB%)**, which are the percentage of plate appearances that result in a strikeout or walk, respectively [1]. We can also compute **Batting Average on Balls In Play (BABIP)** for each batter, which measures how often balls they hit into play (excluding HR) fall for hits [2]. *(BABIP formula: (H – HR) ÷ (AB – HR – SO + SF)* [3] *.)*

Pitching Stats: The team page also lists comprehensive pitching data for each pitcher:

- **Wins (W), Losses (L), Win-Loss %**, and **Earned Run Average (ERA)** – summary outcomes (used for context, not directly in our at-bat model).
- **Games (G), Games Started (GS), Complete Games (CG), Shutouts (SHO), Saves (SV)** – pitching appearances (not directly used in single at-bat simulation except to identify starters vs. relievers).
- **Innings Pitched (IP)** – total innings (outs recorded) by the pitcher.
- **Hits Allowed (H)** – total hits conceded.
- **Runs (R) and Earned Runs (ER) Allowed** – total and earned runs given up.
- **Home Runs Allowed (HR)** – crucial for simulating long-ball outcomes.
- **Walks (BB)** and **Strikeouts (SO)** – pitching outcomes count. From these we can derive **strikeout percentage (K%)** and **walk percentage (BB%)** for pitchers (i.e. strikeouts or walks per batter faced) [1] [4]. Often Baseball Reference provides K/9 and BB/9 as well, which can be converted to these rates.
- **Batting Average Against (BAA)** or opponent OBP/SLG – how opponents hit off this pitcher (useful as an aggregate measure of hittability).
- **WHIP** (Walks + Hits per Innings Pitched) – baserunners allowed per inning (a general indicator of how often the pitcher allows any runner).
- *Advanced:* **Strikeout-to-Walk ratio (K/BB)** may be listed, and stats like **ERA+** or **FIP** might appear but are more for evaluation than direct simulation input.

Fielding Stats: The team fielding section provides both team totals and individual player fielding data:

- **Defensive Efficiency (DefEff)** – the percentage of balls in play that the team's defense converts into outs [5]. (This is essentially 1 – BABIP allowed, a key measure of overall team fielding effectiveness.)
- **Innings (Inn)** – defensive innings played (often given in outs, e.g. 3 outs = 1 inning).
- **Total Chances (Ch)** – opportunities to make an out (putouts + assists + errors).
- **Putouts (PO)** and **Assists (A)** – how many outs were recorded by fielders directly or via assists.
- **Errors (E)** – number of mistakes allowing a runner to reach base.
- **Fielding Percentage (Fld%)** – the rate of plays successfully made = (PO+A) ÷ Ch.
- **Double Plays (DP)** – number of double plays turned.
- **Advanced metrics: Total Zone Runs (Rtot)** and **Defensive Runs Saved (Rdrs)** may be listed, indicating fielders' run-saving value above average. These can indicate the quality of defense beyond simple errors. (For example, a high Rdrs or Rtot means fielders convert more balls into outs than average.)

These stats from Baseball Reference will feed our simulation. In particular, batter and pitcher **K, BB, HR, and hit statistics** (and derived percentages) will drive outcome probabilities, and fielding stats (like DefEff and errors) will adjust the chances of balls in play becoming hits or outs.

Influence of Each Stat on At-Bat Outcomes

Batter's Statistics Influence: A batter's individual stats set the baseline for what happens during their plate appearances. Key influences include:

- **Strikeout Rate (K%)** – If a batter strikes out frequently, it increases the probability that any given at-bat ends in a strikeout. For example, a batter with a 30% K% has a high chance to fan in our simulation. This batter-specific tendency will weigh into the matchup outcome.
- **Walk Rate (BB%)** – Batters with good plate discipline (high BB%) draw more walks, raising the at-bat's chance to result in a base on balls. A batter with, say, a 12% walk rate contributes significantly to walk outcomes in the model.
- **Power/Hitting Profile** – Batter stats like HR count (or HR% of PA) determine the chance of a home run in an at-bat. Similarly, the distribution of the batter's other hits (singles, doubles, triples) influences the probability of each hit type when the ball is put in play. For instance, if 25% of a batter's hits are doubles, we'll reflect a higher chance of a double given a hit. High slugging or ISO (isolated power) indicates more extra-base hits, affecting the outcome mix (more doubles/triples/HRs versus singles).
- **Batting Average on Balls in Play (BABIP)** – A batter's BABIP measures how often their balls in play fall for hits. A consistently high BABIP batter tends to hit balls harder or in ways that avoid fielders. This will elevate the chance that when *this* batter puts the ball in play (i.e. doesn't strike out or homer), the result is a hit rather than an out. In our model, the batter's BABIP contributes to the hit probability after making contact.

Pitcher's Statistics Influence: The opposing pitcher's stats likewise affect the at-bat's outcome probabilities, often in an opposing way to the batter's tendencies:

- **Strikeout and Walk Rates (K%, BB%)** – A pitcher's ability to strike batters out or avoid walks heavily dictates those outcomes. Pitchers largely control strikeout and walk frequency [4]. A pitcher with a very high K% will raise the likelihood of a strikeout in the matchup, even against a contact-oriented hitter, while a pitcher with a low K% gives the batter more chance to put the ball in play. Similarly, a pitcher with a low BB% will suppress walk outcomes for even a patient hitter. These pitcher rates will be combined with the batter's rates to get a matchup-specific K or BB probability.

- **Home Run Rate (HR/PA or HR/9)** – Pitchers have tendencies to allow home runs. A pitcher who gives up many home runs (e.g. a high HR/9) increases the probability that a given ball in play or plate appearance results in a homer for the batter. Conversely, a pitcher who rarely allows homers (maybe due to keeping the ball on the ground or in a big park) will reduce the batter's chance of homering. We will incorporate the pitcher's HR rate along with the batter's power to determine the overall HR chance in an at-bat.
- **Batting Average Against (BAA) and BABIP allowed** – If the pitcher has a low BAA or BABIP allowed, it indicates they are good at limiting hits on balls in play (whether due to skill or the defense behind them). This will **lower** the probability that a ball in play becomes a hit. For instance, a pitcher with an opponents' BABIP of .250 (very low) will often turn balls in play into outs (either by inducing weak contact or via good defense). We factor this in when a batter makes contact – such a pitcher will tilt the odds toward an out.
- **Control and Other Factors** – Other stats like WHIP or opponent OBP summarize how many baserunners the pitcher allows. A low WHIP pitcher typically has low hit and walk rates, reinforcing the above effects (fewer hits/walks per inning means more outs – many via Ks or weak contact). While we won't use WHIP directly, it correlates with the probabilities we derive from K, BB, and BABIP.

Fielding (Defense) Influence: The defense behind the pitcher can turn would-be hits into outs (or vice versa), affecting the outcome of balls in play:

- **Defensive Efficiency (Team DefEff)** – This metric directly measures the percentage of balls in play that are converted into outs by the defense [5]. A high DefEff means the team is excellent at fielding (fewer balls drop in for hits). In our simulation, if a ball is put in play (not a HR), a higher DefEff will reduce the probability of that ball becoming a hit. For example, if the league-average BABIP is ~.300 (DefEff ~.700) and the Cubs' defense has DefEff .708, the Cubs turn slightly more balls into outs than average. We will adjust hit/out chances on contact to reflect this – essentially scaling down the batter/pitcher expected BABIP by the defense's performance. A strong defense "steals" some hits and turns them into outs.
- **Errors and Fielding Percentage** – Errors (E) and Fld% indicate how often fielders mishandle balls. While rare on a per-play basis, errors do allow batters to reach base on what would have been outs. We can incorporate a small probability of reaching on error for balls in play, based on the defense's error rate. For instance, if the team has 80 errors in 162 games, that's roughly 0.5 errors per game, or about a 0.2–0.3% chance per ball in play turning into an error-induced reach. In the simulation, after determining an out, we could give a slight chance that the out is actually an error (meaning the batter is safe). Better fielding teams (higher Fld%) have lower error probabilities, so their opponents less often reach on errors.
- **Individual Fielding Metrics** – If we wanted to be more granular, we could use positions and metrics like Rtot or Rdrs for specific fielders. For example, if a particular outfielder has a very high Rdrs (meaning they get to more balls), balls hit toward them might have a smaller hit probability. However, for simplicity at the at-bat level (without simulating exact ball trajectory), we'll mostly use the team-level defensive efficiency. We assume the overall defense quality uniformly affects balls in play. (In future, one could map batted-ball types to specific fielders and use individual fielding stats, but that's beyond our current scope.)

In summary, **a strikeout or walk outcome is primarily influenced by the batter and pitcher's K%/BB%** [1] **, a home run outcome by the batter's power and pitcher's HR tendency, and any ball in play outcome by the batter's hitting skill, pitcher's contact management, and the defense's ability to convert plays into outs** [6] . A good defense will tilt borderline outcomes toward outs, whereas a weak defense might allow more hits on contact. These factors will all come together in our model of an at-bat.

Combining Batter, Pitcher, and Fielding Stats into a Probabilistic Model

To simulate an at-bat, we need to combine the batter's and pitcher's tendencies (with a defensive adjustment) to calculate the probability of each possible outcome. We will use the following procedure to build a probabilistic model for a single plate appearance:

1. **Identify Outcome Categories:** We will categorize at-bat results into a set of mutually exclusive outcomes: Strikeout (K), Walk (BB), Hit by Pitch (HBP), and Ball in Play (which can further result in an out or one of several hit types). Home runs (HR) will be treated as a separate category from other hits, since a HR is not a fieldable ball in play (and does not depend on defense once it happens). Thus, our top-level outcomes to calculate probabilities for are: **K, BB, HBP, HR, and Ball in Play (in-play contact that is not a HR)**. For simplicity, HBP can be grouped with walks as "free passes" if its probability is very low, or we can assign a small separate probability using batter and pitcher HBP stats (if available). Initially, we might ignore HBP or use league-average HBP% since it's relatively rare.

2. **Calculate Three True Outcomes First:** The "three true outcomes" – strikeouts, walks, and home runs – are largely batter vs. pitcher driven (fielders don't influence these). We combine batter and pitcher stats for each:

3. **Strikeout Probability (P_K):** Take the batter's K% (e.g. batter strikes out in 20% of PAs) and the pitcher's K% (e.g. pitcher strikes out 25% of batters faced). We need a method to combine these into one probability for this matchup. A simple approach is to take a weighted average or midpoint between the two percentages, perhaps giving a bit more weight to the pitcher (since pitchers have significant control over strikeouts) [4] . For example, we might set P_K = 0.5*(batter K% + pitcher K%). If league-average K% is considered, we could refine by scaling relative to average: for instance, if the batter is 1.2× league K rate and the pitcher is 1.1× league K rate, then the matchup might be ~1.2×1.1 = 1.32× league K rate. But in the interest of simplicity, an average or mild weight approach is sufficient. The key is that if either party is extreme (high or low), the combined probability reflects that (e.g. a high-K batter facing a high-K pitcher yields a very high strikeout chance).
4. **Walk Probability (P_BB):** Similarly, combine the batter's walk rate and pitcher's walk rate. If a batter walks 10% of the time and the pitcher walks 8% of batters, we might set the matchup walk probability around the average (9% in this case, possibly weighted toward the pitcher's 8% because pitchers largely control walks). Again, a simple average or weighted average (e.g. 60% pitcher, 40% batter) can be used. This ensures a wild pitcher and a patient hitter yield a high walk chance, whereas a control pitcher and free-swinging batter yield a low walk chance.

5. **Home Run Probability (P_HR):** We combine the batter's HR rate (HR per PA) and the pitcher's HR rate (HR allowed per PA). If a batter homers in 5% of PAs and the pitcher allows 3% HR/PA, the matchup HR probability might be around the lower of the two (since a pitcher can avoid some HR by

style, but a power hitter can still muscle some out). We could use an average or slightly favor the lower value (to be conservative) – e.g. here maybe ~4%. Another method is to base it on contact: first determine how often the ball is hit in the air by the batter vs pitcher (if such info were available) and then the HR/FB rates – but since we only have overall rates, a direct blend is fine.

6. **Ball-in-Play Probability:** Next, we determine the probability that the at-bat results in a **ball in play (BIP)** that is not a home run. This essentially is 1 minus the probability of the "true outcomes" we computed (K, BB, HR, plus HBP if included). For example, if P_K + P_BB + P_HR = 40% of the outcomes, then the remaining ~60% will be balls hit into play (this includes singles, doubles, triples, or outs in the field). We might denote this P_BIP. We expect P_BIP to roughly equal (1 – K% – BB% – HR%) of the combined matchup. Notably, high strikeout matchups will have a lower P_BIP (fewer balls put in play), while contact-oriented matchups have higher P_BIP.

7. **Outcome of Balls in Play – Hit vs Out:** When a ball is put in play, we need to decide if it becomes a **hit or an out**. This is where batter, pitcher, and fielders intersect strongly. We will use a combination of the batter's batting skill (BABIP), the pitcher's typical BABIP allowed, and the defense's efficiency. For the matchup's **expected BABIP** (on contact):

8. Start with the batter's season BABIP (say .310, meaning 31% of his in-play balls go for hits) and the pitcher's BABIP allowed (say .280, meaning hitters average 28% on balls in play against him). The defense behind the pitcher is partly why the pitcher's BABIP is what it is [6] . We can average these (.295 in this example) as a baseline chance that a ball in play off this batter vs pitcher is a hit.
9. Adjust for the defense: If the Cubs' team defensive efficiency (DefEff) is, for example, .708 (meaning 70.8% of balls are turned into outs, implying opponents' BABIP ~.292), and league-average DefEff is .700 (BABIP .300), then the Cubs defense is about 2% better than average at converting balls to outs. We would reduce the baseline hit probability accordingly when the Cubs are fielding. In our example, baseline .295 BABIP vs a .708 DefEff defense might drop to ~.285 hit probability (rough estimate), because the defense will steal some hits. Conversely, if we were simulating the Cubs batting *with* an opposing defense that's worse than average (say DefEff .680), we'd increase the hit probability. In formula terms: `P_hit_on_contact = Combined_BABIP * (Opponent_DefEff_adjustment)`, where Opponent_DefEff_adjustment is a factor less than 1 for above-average defenses and >1 for below-average.

10. Also adjust for extreme fielders if desired: for instance, if a particular elite fielder is known (high Rdrs), certain types of contact might have extra reduction. But in our at-bat level model, we stick to the team aggregate.
    Thus, we get **P_hit_given_BIP** (probability a ball in play becomes a hit). Then the chance a ball in play is an out is simply **1 – P_hit_given_BIP**, minus a tiny fraction for errors (explained next).

11. **Error Chance:** A small portion of balls in play result in the batter reaching on error instead of a clean hit or out. We can incorporate this as follows: based on the team's error rate, say the Cubs made 73 errors in 162 games (just as an illustration). That's roughly 0.45 errors per game. Given an average of about 25–30 balls in play per game per team, the chance of an error on any one ball in play is about 1.5–2%. We will include an **error outcome** with a probability ~1–2% of all balls in play (adjusted by the specific defense – a team with fewer errors would have <1% perhaps). When an error occurs, we treat it as the batter reaching base (like an "extra" outcome aside from hit). To implement: for a ball in play, we first decide if it's an error (with probability based on E/Chances). If not an error, then use

the P_hit_given_BIP for hit vs out. This way, a great fielding team with high Fld% (few errors) will almost never give an error outcome, whereas a poor defensive team might occasionally allow an extra baserunner via error.

12. **Hit Type Distribution:** If an at-bat is determined to be a hit (specifically a non-HR hit), we then decide what kind of hit: single, double, or triple. We will use the batter's statistical distribution of hit types for this. For example, from the stats we know how many of each hit type the batter has. If the batter has 100 hits on the season, with 70 singles, 20 doubles, 5 triples, 5 homers (just an example distribution), then conditional on a **hit (that is not a HR)**, the probabilities would be 70/95 ≈ 73.7% single, 20/95 ≈ 21.1% double, and 5/95 ≈ 5.3% triple. (We exclude HR in that conditional mix because HRs we handled as a separate outcome before the BIP stage.) We will apply those percentages to decide the hit type. Pitchers' tendencies could also factor in (e.g. if a pitcher is known for yielding more extra-base hits, we might subtly adjust these, but such detail is beyond the basic data on BRef; we assume the batter's profile dominates). Note that a **home run** outcome was already handled earlier (as P_HR), so when we reach this step we are only dealing with singles/doubles/triples.

13. **Putting it Together:** Now we have all pieces to form a complete probability distribution for a matchup. For a given batter vs pitcher (with a known defense behind), we determine:

14. P(K), P(BB), P(HBP) – from step 2, combined rates.
15. P(HR) – from step 2, combined rate.
16. P(BIP) – = 1 – [P(K)+P(BB)+P(HR)+P(HBP)].
17. Given BIP: P(hit) and P(out) (with P(hit) possibly split into P(single/double/triple) per step 6, and P(out) including a small slice of error probability turning into a reached base).
    For example, imagine a matchup yields: 20% K, 8% BB, 2% HBP, 3% HR. That sums to 33%, leaving 67% balls in play. Now suppose on those balls in play, we calculate a 30% chance of a hit (due to batter/pitcher/fielders) and 70% out. Further, of the 30% hits: 24% are singles, 5% doubles, 1% triples (and HR were separate). And maybe a 1% error chance on any ball. We would then have final probabilities like: 20% strikeout, 8% walk, 2% HBP, 3% home run, and among the 67% BIP: about 20% *overall* hits (which corresponds to 16% singles, 3.3% doubles, 0.7% triples) and ~1% reach on error, ~46% regular outs. (Those would sum to 100%.)

The simulation engine will use these probabilities to decide each at-bat's outcome. It's essentially a weighted random choice among: K, BB, HBP, 1B, 2B, 3B, HR, Out, Error (error could be considered a type of out event where batter reaches base). We ensure the weights reflect the stats-driven probabilities.

Random Outcome Selection and Real-World Variance

Once we have the probabilities for each outcome in an at-bat, the simulation will **use randomness to select an actual result**. This ensures the simulation is non-deterministic and that over many at-bats the frequency of outcomes will reflect the probabilities, much like real baseball where probabilities hold on average but any given plate appearance is uncertain.

Random Selection Method: For each at-bat, the program will generate a uniform random number between 0 and 1 (or a random integer range, any equivalent method). We will then map that number to an outcome based on the cumulative probability distribution. For example, suppose our outcome probabilities (in decimal) for a particular matchup are:

- Strikeout = 0.20
- Walk = 0.08
- HBP = 0.02
- Home Run = 0.03
- Single = 0.16
- Double = 0.033
- Triple = 0.007
- Reach on Error = 0.01
- Out = 0.46 (note: out includes all non-error outs in play)

We would establish cumulative thresholds: 0.00–0.20 = Strikeout, 0.20–0.28 = Walk, 0.28–0.30 = HBP, 0.30–0.33 = HR, 0.33–0.493 = Single, 0.493–0.526 = Double, 0.526–0.533 = Triple, 0.533–0.543 = Error, 0.543–1.000 = Out. The sim draws a random number and finds where it falls [7]. For instance, a random number 0.25 would land in the Walk range; 0.85 would be an out. This technique is standard for Monte Carlo simulations of categorical events [7] – essentially we're "rolling a weighted die" for each at-bat outcome.

Mirroring Real-World Variance: Because each outcome is chosen randomly, the simulation will naturally produce variance akin to real games. A batter with a 30% true strikeout chance won't strike out exactly 3 times in 10 in every simulation run – sometimes they might strike out 5 times in a row, sometimes not at all in a game. This random fluctuation mimics how, in reality, a .300 hitter can still go 0-for-4 in a given game or a low-power hitter might run into a homer unexpectedly. Over a large number of at-bats or repeated simulations, the results will converge toward the input probabilities (law of large numbers), but any single game can have unexpected swings. We can enhance the realism of variance by ensuring the random number generation is truly uniform and by running many simulations to see distribution of outcomes.

To incorporate a **touch of realistic statistical variance** beyond pure randomness, we could optionally introduce slight game-to-game variability in the probabilities themselves (for example, treat the batter's true probabilities as a mean and sample from a beta distribution each game to simulate "hot" or "cold" streaks). However, that is a more advanced feature; at minimum, the randomness per at-bat provides plenty of variance. The system thus produces a different sequence of outcomes every time it's run, while still reflecting the overall stats.

For implementation, most programming languages have random number libraries. We will use those to compare against our cumulative probability cutoffs and select outcomes. This Monte Carlo approach is straightforward and efficient for simulating entire games play-by-play. In pseudocode, for each atBat(): generate random 0-1, if <P_K then "Strikeout"; else if <P_K+P_BB then "Walk"; ... and so on through all outcome categories.

[7] illustrates this approach, where a function uses a random draw and checks it against cumulative probabilities for each outcome category to decide the result of the at-bat. By following the distribution we set (derived from the stats), the random selection ensures outcomes occur with the right frequencies on average, but in a given game they can deviate – just like real baseball where probabilities only guarantee long-run trends.

Limitations and Future Improvements

While this simulation design uses a rich set of Baseball-Reference stats and captures many important aspects, there are some limitations and assumptions to note:

- **No Pitch-by-Pitch or Count Dynamics:** We model each at-bat as a single probabilistic event with fixed probabilities. We do not simulate individual pitches or changing probabilities based on count (e.g., a 0-2 count makes a strikeout more likely in reality, but our model doesn't account for that). This simplification means we ignore strategic nuances like pitch sequencing, but it keeps the model simple and focused on the overall stats.

- **Context-agnostic Outcomes:** The probabilities are generally static for a given batter-pitcher matchup. We aren't adjusting for game context such as clutch situations, weather, or park dimensions. Baseball Reference stats are season averages mostly, so we assume those averages hold constant. In real games, a batter might face a tiring pitcher in later innings or benefit from a wind blowing out, but our simulation doesn't handle those dynamic context changes yet.

- **Lineups and Fatigue:** We assumed fixed lineups and one pitcher per team for the whole game (no substitutions). In reality, pitchers tire and relievers with different stats come in, and pinch hitters may be used. Our current model effectively simulates a starting pitcher going a full game against a set lineup. This is a simplification intended by the user's constraints. In future extensions, we'd introduce pitcher stamina and a bullpen – e.g., after a certain number of innings or pitches, switch to a reliever (with his own stats affecting probabilities) – and allow lineup substitutions. The framework is extensible: we could carry stamina for pitchers and reduce their effectiveness (e.g., their K% drops, BB% rises as they fatigue) or swap in relievers with fresh stats. Likewise, pinch hitters or defensive replacements could be slotted in, altering the batter or fielder stats mid-game. For now, none of that is modeled, which makes the simulation less realistic for a full 9-inning game but much simpler to implement.

- **Fielding Simplifications:** We used team-level defensive efficiency and a generic error rate. This means every ball in play faces the same odds of becoming a hit or error, regardless of *which* fielder it might be hit towards or the type of batted ball. Realistically, a grounder to a Gold Glove infielder is more likely to be an out than a grounder to a poor fielder, and fly balls vs. ground balls have different out probabilities. Our model doesn't distinguish these scenarios because we don't simulate the batted ball type or direction. If needed, we could incorporate splits (e.g., using ground ball vs fly ball tendencies if available, and fielders' strengths by position), but Baseball Reference doesn't provide detailed batted-ball type stats on the team page. So a caveat is that our use of DefEff and error% is an average that might gloss over some situational detail. Still, DefEff is a solid overall measure of defense quality [5], so it should reasonably approximate the impact of fielders in aggregate.

- **No Base Running or Strategy:** Our at-bat simulation ends with the outcome (e.g., single or out). We haven't detailed how runners advance on hits or outs. In a full game simulation, we'd need rules for advancing runners (e.g., how a single might move a runner from first to third, chances of tagging up on fly outs, etc.). The prompt limited scope to individual at-bats, so our design stops at determining the outcome of each at-bat (who got on base or who got out). To turn this into a full game, we'd implement base-runner advancement and scoring logic after each at-bat outcome. This could be done with simple rules (everyone advances one base on a single, two on a double, etc., as was done in other basic simulators [8] [9]), or with probabilistic base-running stats. For now, we note this omission and plan to address it in the future when we simulate full gameplay.

- **Statistical Assumptions and Small Sample Issues:** The model relies on season aggregate stats from Baseball Reference. If a player's stats are based on a small sample (e.g., a rookie with 50 PA), those percentages might not be very predictive. Likewise, the simulation assumes each at-bat is independent and the probabilities don't change – in reality, a pitcher might adjust how they pitch to a batter over the game, or a batter might learn and adjust. We aren't modeling any "adaptive" behavior or streakiness. All randomness is memoryless in our simulation. This is a common assumption for simplicity, but it's worth noting as a limitation.

- **Using BRef Data Only:** We constrained ourselves to stats available on Baseball Reference's team page. This forced a few proxies: for example, we used DefEff and simple BABIP rather than more granular Statcast measures (like exit velocity or catch probability) which could improve realism of ball-in-play outcomes. Also, we derived percentages (K%, BB%) from counts given on BRef – which is fine, but if BRef lacked something like a pitcher's exact BF (batters faced), we assumed BF ≈ (IP*3 + BB + HBP + etc.). Fortunately, most needed inputs (K, BB, HR, H, etc.) are indeed on BRef, so our model doesn't need outside data. The limitation is that BRef's stats are mostly retrospective averages – our simulation won't account for* why* those stats are what they are (e.g., pitch repertoire, match-up advantages) beyond their numeric influence.

In conclusion, this plan lays out a **comprehensive statistical simulation system** for one-at-bat outcomes using 2025 Cubs data. We listed all relevant stats from Baseball Reference and described how to use them to shape probabilities for strikeouts, walks, hits, etc., incorporating batter skill, pitcher skill, and team defense. By combining these factors and using random selection for each plate appearance, we can simulate a realistic distribution of outcomes in a game. The system is designed to be modular – we can plug in any team or player's stats (as long as we have their batting/pitching fielding data) and get a game simulation. With fixed lineups and one pitcher per side it's straightforward; and we've noted how one could extend it with substitutions, multiple pitchers, and more detailed nuances in the future [10].

Overall, this simulation will produce play-by-play results that mirror real baseball probabilities and variance, all grounded in the concrete stats from the Baseball Reference 2025 Cubs page. It provides a solid foundation for further refinement, ensuring that even at this simplified level, the **batter's, pitcher's, and fielder's contributions are all accounted for** in determining each at-bat's outcome.

Sources:

- Baseball Reference – 2025 Chicago Cubs team page (batting, pitching, fielding stats) [11] [5].
- FantasyPros Sabermetrics Guide – definitions of K% and BB% (strikeout and walk rates per PA) [1].
- RotoGrinders primer on K% – emphasizes calculation as strikeouts per batters faced [12].
- Fangraphs Sabermetrics Library – explanation of BABIP and its drivers (player quality, defense, luck) [6].
- Chicago Sun-Times – explanation of Defensive Efficiency (DefEff) as % of balls in play turned into outs [5].
- Jack Overby, "Basic Baseball Simulation" – example of using a random number to assign an outcome based on cumulative probabilities [7].

- Hardball Times, “10 Lessons from Creating a Baseball Simulator” – guidance on simulation structure and acknowledging which factors were included vs. simplified [13] [10].

---

[1] [4] Beginner’s Guide to Sabermetrics: K%, BB%, K%-BB% | FantasyPros
https://www.fantasypros.com/2015/05/beginners-guide-to-sabermetrics-k-bb-k-bb/

[2] [6] The Beginner’s Guide To Single-Season BABIP | Sabermetrics Library
https://library.fangraphs.com/the-beginners-guide-to-single-season-babip/

[3] Understanding the BABIP statistic - ESPN
https://www.espn.com/fantasy/baseball/flb/story?page=mlbdk2k12_understandingbabip

[5] Cubs' defense still ranks among best in baseball - Chicago Sun-Times
https://chicago.suntimes.com/2018/5/21/18316041/cubs-defense-still-ranks-among-best-in-baseball

[7] [8] [9] Basic Baseball Simulation. Are we in a simulation? Perhaps. If you… | by Jack Overby | Medium
https://jroverby92.medium.com/basic-baseball-simulation-afe3995e14b3

[10] [13] 10 Lessons I Learned from Creating a Baseball Simulator | The Hardball Times
https://tht.fangraphs.com/10-lessons-i-learned-from-creating-a-baseball-simulator/

[11] 2024 Chicago Cubs Statistics - Baseball-Reference.com
https://www.baseball-reference.com/teams/CHC/2024.shtml

[12] Strikeouts: The Most Important Pitching Stat
https://rotogrinders.com/fantasy/lessons/strikeouts-most-important-pitching-stat