

Estimating the Striking Accuracy of UFC Fighters

Russel Luber

Introduction

Mixed Martial Arts (MMA) is a combat sport built on controlled aggression and technical skill. Under the [Unified Rules](#), athletes may strike with punches, kicks, elbows, and knees, and may grapple using throws, takedowns, submissions, and ground control. The [Ultimate Fighting Championship](#) (UFC), now more than three decades old, is the sport's premier organization, staging events viewed by millions around the world.

MMA performance rests on two broad pillars: striking and grappling. Striking focuses on inflicting damage from a distance or in close range, while grappling emphasizes controlling an opponent on the ground and threatening submissions. Both elements contribute to winning fights, but they do so in different and measurable ways.

Research Questions

This project investigates two central questions about performance in UFC competition:

1. Which in-fight performance differential metric is the most predictive of winning?
2. What is each fighter's *latent* probability of landing a significant strike, after accounting for gender, weight class, and fight-to-fight randomness?

Understanding which advantages translate most reliably into victory helps clarify what matters strategically inside the cage. Just as importantly, striking accuracy is a core determinant of offensive success: every significant strike thrown has the potential to meaningfully damage an opponent. Estimating each fighter's latent ability to land these strikes provides insight into persistent skill differences that shape competitive outcomes.

Data Collection

The data was collected from the [UFC Stats](#) website was webscraped using the `rvest` package. The scripts for my webscraping pipeline can be found [here](#). I made use of the `tidyverse` collection of packages to do data cleaning and exploratory data analysis.

Methodology

To answer my research questions, I fit two separate statistical models: a predictive logistic regression model for win probability (RQ1) and a hierarchical Bayesian binomial model for significant strike accuracy (RQ2). The second model is motivated by results from the first.

The first research question focuses on understanding which performance metrics are most predictive of winning a UFC fight. In particular, I chose four metrics: significant strike differential, takedown differential, knockdown differential, and control time differential.

For each fight, I computed differential statistics by subtracting one fighter's performance from the other's. For example, if we have a fight between Fighter A and Fighter B, then Fighter A's significant strike differential would be:

$$\text{Fighter A's sig. strike differential} = \text{sig. strikes landed by A} - \text{sig. strikes landed by B}$$

Fighter B's significant strike differential would be the other way around. And the same goes for the other performance metrics of knockdown differentials, takedown differentials, and control time differentials.

All predictors were then standardized so that regression coefficients reflect the effect of a one-standard deviation increase in each differential metric.

I fit a **binomial logistic regression model** where the outcome is whether the fighter won (1) or lost (0). The model estimates how each standardized differential contributes to the log-odds of winning.

The second research question shifts from predicting wins to estimating latent fighter skill: each fighter underlying probability of landing a significant strike, after adjusting for structural factors and random variation.

Observed accuracy varies across fighters, weight classes, genders, and individual fights, and is often noisy due to differences in opponent, pacing, and fight context. To separate these sources of variation, I fit a **hierarchical Bayesian binomial logistic regression model** that treats significant strikes landed out of attempts as binomial outcomes.

The model includes:

- Population-level (fixed) effects for gender and weight class, which capture systematic differences in strike accuracy across divisions.
- Fighter-level random effects, which estimate each fighter’s deviation from their division and gender-adjusted baseline accuracy. These represent inherent striking ability.
- Fight-level random effects, shared by both fighters in a given fight, capturing contextual variation such as stylistic matchups, altitude, or fight pace.

Together, these components allow the model to estimate each fighter’s “typical” strike-landing probability in an average fight, while accounting for both structural factors and fight-to-fight randomness. Weakly informative priors are used to stabilize estimates, especially for fighters with limited data.

Data Summaries

The dataset used in this analysis contains 16,402 observations, where each row represents a single fighter in a single UFC bout. Because each fight contributes two rows (one per competitor in the fight), the dataset corresponds to 8,201 unique fights.

Each observation includes detailed information about the fighter’s performance, identity, and context within that bout. Key variables include:

Variable Name	Description
<code>fight_id</code>	The unique identifier for a UFC fight.
<code>fighter</code>	The name of the <code>fighter</code> of interest.
<code>opponent</code>	The name of the <code>fighter</code> ’s opponent.
<code>weight_class</code>	The <code>weight class</code> that the fight is taking place at. For example “welterweight” i.e. fighters who weigh over 155 lbs and up to 170 lbs during the fight.
<code>gender</code>	Gender of the <code>fighter</code> .
<code>sig_strikes_landed</code>	The number of <i>significant</i> strikes that the fighter landed <i>on</i> the <code>opponent</code> .
<code>sig_strikes_thrown</code>	The number of significant strikes that the fighter threw at the <code>opponent</code> .
<code>sig_strikes_landed_by_opp</code>	The number of <i>significant</i> strikes landed <i>by</i> the <code>opponent</code> on the <code>fighter</code> .
<code>sig_strike_diff_z</code>	Z scored measure of striking advantage: how many standard deviations above or below average a fighter’s significant strike differential was in a given fight.

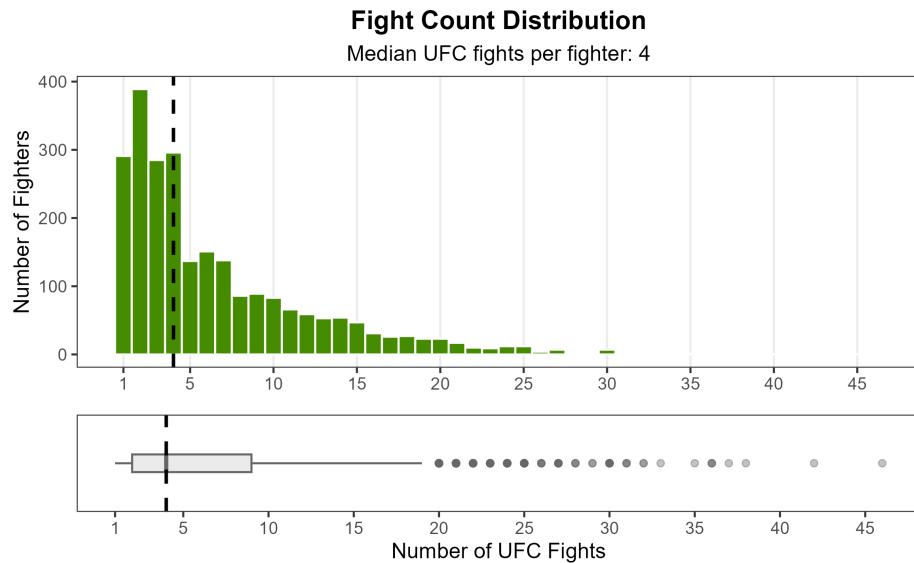
Across all observations, the data include 2,451 unique fighters, covering UFC athletes across all major divisions.

The dates of the fights range from November 1, 2000 to November 15, 2025, corresponding roughly to the period from the adoption of the Unified Rules of MMA to the most recent scraped event. This provides a comprehensive historical dataset spanning 25 years of UFC competition, enabling both cross-sectional comparisons between fighters and longitudinal analyses of fight outcomes and striking performance.

Exploratory Data Analysis

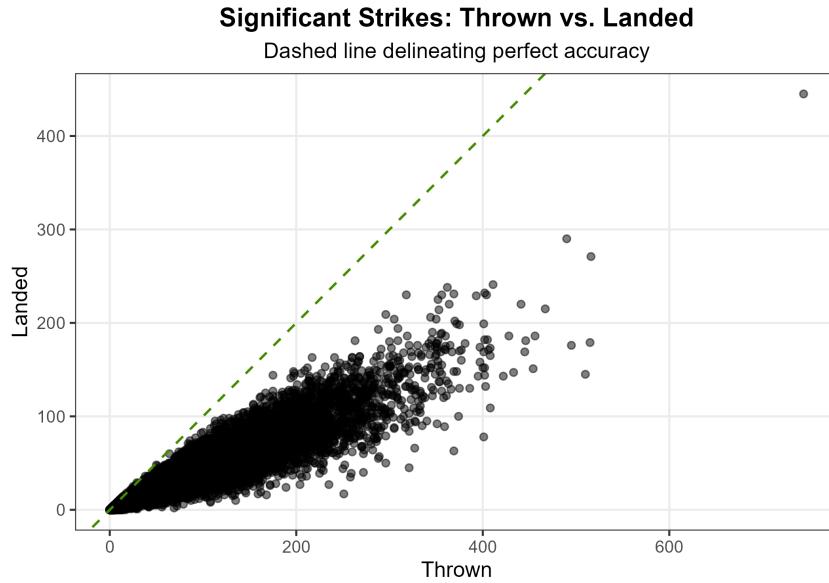
To motivate my analysis, I did some exploratory work and asked questions about the data to get a sense of it.

How many fights does each fighter have in the UFC?



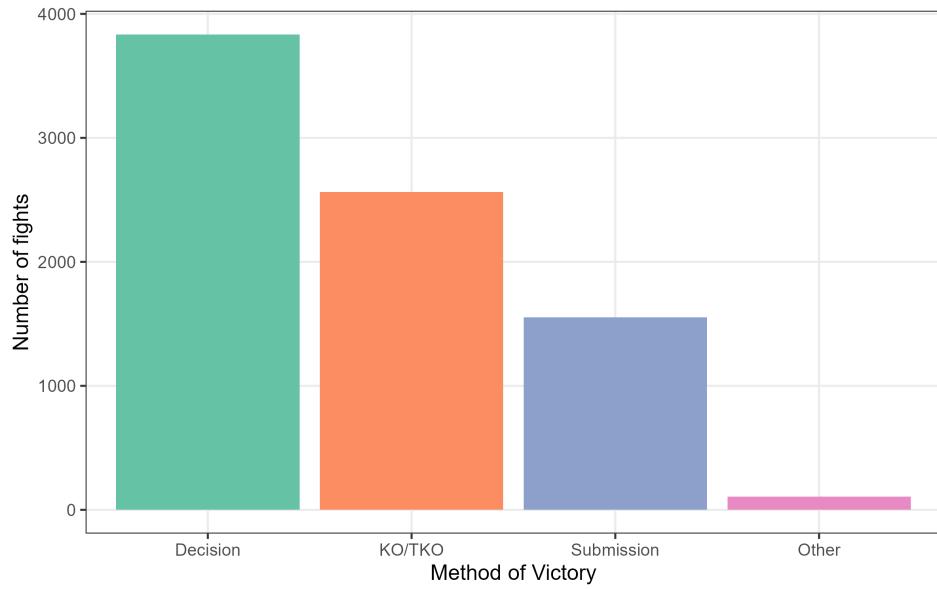
Most fighters have less than 5 fights on their record, with the most common number of fights being 2. It's an extremely right-skewed distribution. Almost power law-like. Many fighters enter the UFC but only a few accumulate long careers. The median number of fights in the UFC population is 4, meaning that 50% of fighters have fewer than 4 fights and 50% have more than 4 fights.

How many strikes do fighters throw versus how many do they actually land?



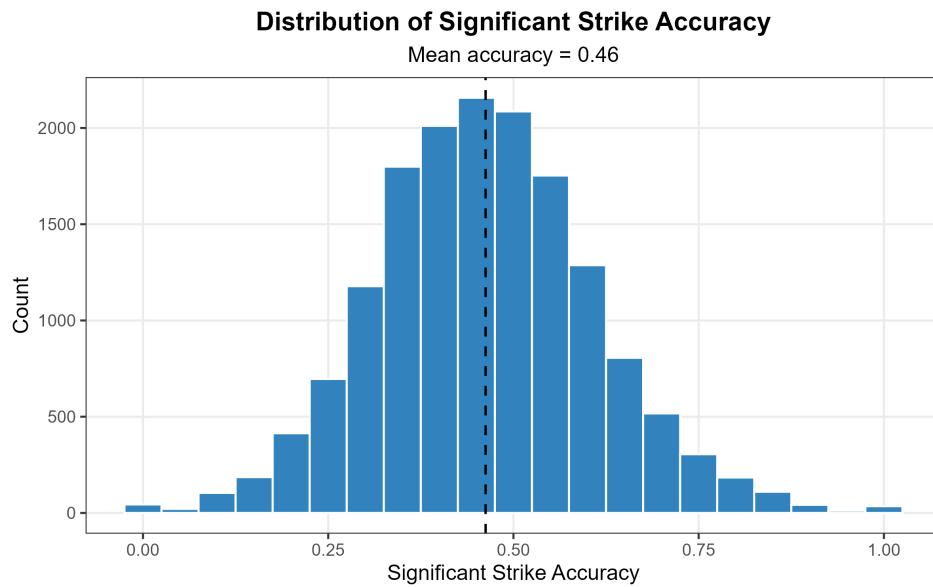
In this figure, we see how accurate UFC fighters are in general. Each dot is a fighter in a fight. The dashed line denotes what is possible i.e. You can't land more strikes than you throw. We see that the dots are well below the perfect accuracy line. This shows that fighters throw more than they land. And that's par for the course in MMA.

By what method do fighters win?



Most fighters win by judges' decision. Going to the score cards is the most common way to win in the UFC. Knocking your opponent out is the second most frequent method of victory, followed by submitting your opponent. The "other" category here includes winning because the opponent was disqualified and winning because the on-site doctor determines that the opponent cannot continue.

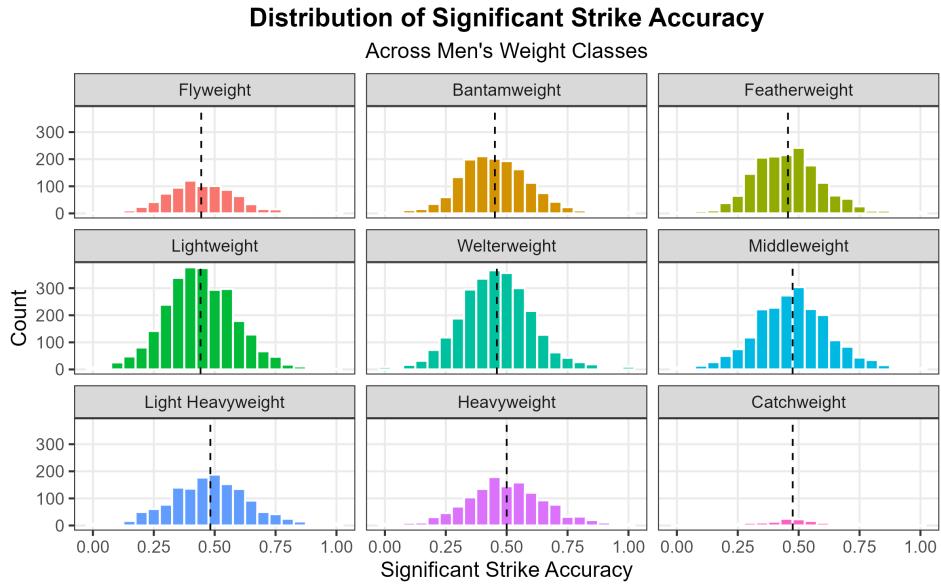
What proportion of significant strikes do UFC fighters land?



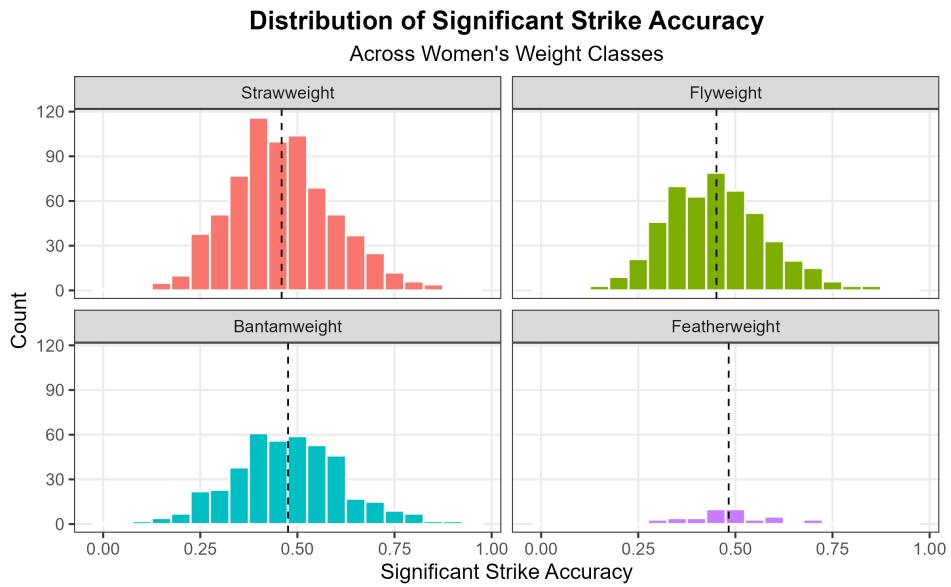
The distribution of significant strike accuracy for the whole UFC population looks approximately Normal. There are 0.00 and 1.00 accuracies because there are some fights where a fighter never threw a strike and other fights where a fighter threw one or more strikes and they all landed. I enforced a minimum significant strikes thrown condition and subset the dataset (< 4% of the data filtered out) to curtail this issue but it persists.

How does significant strike accuracy differ across gender and weight class?

The natural question to ask after looking at the whole population is to look at sub-populations.



The mean significant strike accuracy across men's weight divisions hovers around 44-50%. Just like the UFC population as a whole, the distribution of significant strike accuracy by weight class is approximately Normal. In terms of average accuracy, the heavyweight division seems to be the most accurate at 0.5 while the lightweight division is on the low end at 0.44.



For women, the mean significant strike accuracy varies between 46-48%. The featherweight division's mean accuracy is 0.48, while the flyweight division's mean accuracy is 0.46. It's

worth noting that women's MMA is a relatively recent addition to UFC, which started only in 2013.

With hundreds to thousands of fighters in each weight class and gender, we observe some differences in the typical levels of significant strike accuracy across groups. While the distributions within each group appear roughly normal, their centers shift subtly from one weight class to another. This indicates that gender and weight class introduce structure into the data that's important to keep in mind when examining fighter accuracy.

Research Question 1

Which in-fight performance differential metric is the most predictive of winning?

To answer this research question, I fit a logistic regression model predicting the probability that a fighter wins the bout as a function of four standardized in-fight performance differentials: significant strikes, knockdowns, takedowns, and control time. Each predictor is z-scored, so the coefficients represent the change in log-odds of winning associated with a one standard deviation in that metric.

```
# Logistic regression model
perf_diff_model <- glm(
  res_win ~ sig_strike_diff_z + kd_diff_z + td_diff_z + ctrl_time_diff_z,
  data = win_df,
  family = binomial()
)

summary(perf_diff_model)
```

Call:

```
glm(formula = res_win ~ sig_strike_diff_z + kd_diff_z + td_diff_z +
  ctrl_time_diff_z, family = binomial(), data = win_df)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.774e-15	2.351e-02	0.00	1
sig_strike_diff_z	2.050e+00	4.284e-02	47.84	<2e-16 ***
kd_diff_z	1.321e+00	3.495e-02	37.80	<2e-16 ***
td_diff_z	4.514e-01	3.839e-02	11.76	<2e-16 ***
ctrl_time_diff_z	3.052e-03	1.648e-04	18.51	<2e-16 ***

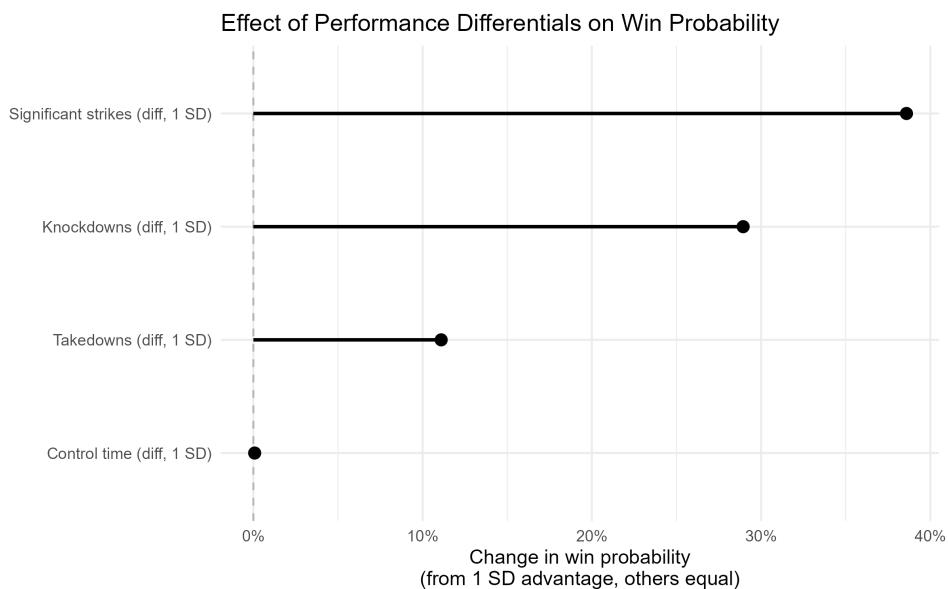
```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 22342 on 16115 degrees of freedom  
Residual deviance: 11425 on 16111 degrees of freedom  
AIC: 11435
```

```
Number of Fisher Scoring iterations: 6
```

The model fits the data well. The residual deviance decreases substantially relative to the null deviance (22,342 to 11,425), indicating that the four differentials collectively explain a large amount of the variability in fight outcomes. All predictors are highly statistically significant.



This plot shows how much a fighter's probability of winning is expected to increase when they have a one standard deviation advantage in each performance metric, assuming all other metrics are equal.

A one standard deviation advantage in significant strikes produces the largest effect, increasing the predicted chance of winning by roughly 40 percentage points. This means that in an otherwise even fight, outperforming an opponent by this margin in striking volume shifts the expected outcome from essentially 50–50 to close to a 90 percent win probability.

A one standard deviation advantage in knockdowns increases win probability by about 30 percentage points, indicating that knockdowns are also highly consequential but still secondary to overall striking differential.

Takedown differential has a more moderate effect, raising win probability by approximately 10 percentage points for a one standard deviation advantage.

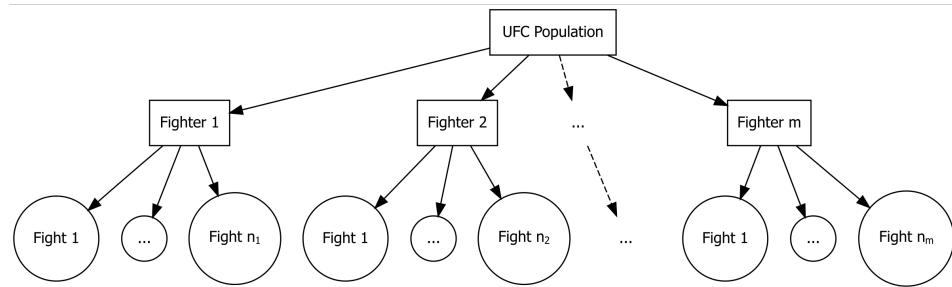
Finally, a one standard deviation advantage in control time has almost no effect on predicted win probability once the other metrics are accounted for, highlighting its limited independent predictive value in this particular model.

Taken together, the results from RQ1 show that significant striking differential is the strongest in-fight predictor of winning, far outweighing the effects of knockdowns, takedowns, or control time. This naturally raises a deeper question about where these striking advantages come from in the first place. Given its connection to winning, striking accuracy is an important target of estimation for each fighter.

Research Question 2

What is each fighter's *latent* probability of landing a significant strike, after accounting for gender, weight class, and fight-to-fight randomness?

Before building the model to estimate a fighter's probability of landing a significant strike, it's important to understand how the data is structured.



Each fighter i has n_i fights. As we've seen, the median UFC fighter has about 4 fights in their career. But there's a lot variation in the number of fights fighters have fought in.

To estimate each fighter's underlying probability of landing a significant strike, we need a model that respects the structure of UFC data. Fighters belong to a broader UFC population, but they vary widely in how many fights they have, who they face, and the conditions under which their performances are observed. Some have long careers with dozens of bouts, while others fight only once or twice. Because of this uneven data structure, treating each fighter independently would overfit the data for fighters with very few observations, while fully pooling fighters together would obscure meaningful skill differences.

A **Bayesian multilevel model** provides a principled middle ground through **partial pooling**. In this framework, each fighter's latent striking ability is estimated while simultaneously borrowing strength from the overall distribution of abilities in the UFC population. Fighters with many fights have estimates driven mostly by their own data, while fighters with few fights are stabilized by the population-level prior. This approach accounts for gender and weight class differences, incorporates fight-to-fight randomness, and yields coherent, shrinkage-adjusted estimates of each fighter's true striking accuracy.

Model

Population-level parameters

Intercept

$$\alpha \sim \mathcal{N}(\text{logit}(0.46), 1)$$

Gender effect

$$\beta_{\text{gender}} \sim \mathcal{N}(0, 0.3)$$

Weight-class effects

For each non-baseline weight class j :

$$\beta_{\text{wc},j} \sim \mathcal{N}(0, 0.3)$$

Fighter-level random effects (varying intercepts)

For each fighter $f = 1, \dots, F$:

$$u_f \sim \mathcal{N}(0, \sigma_{\text{fighter}}^2)$$

Fight-level random effects (varying intercepts)

For each fight $k = 1, \dots, K$:

$$v_k \sim \mathcal{N}(0, \sigma_{\text{fight}}^2)$$

Hyperpriors

$$\sigma_{\text{fighter}} \sim \mathcal{N}^+(0, 0.4)$$

$$\sigma_{\text{fight}} \sim \mathcal{N}^+(0, 0.4)$$

Likelihood

For each observation i :

- y_i = significant strikes landed
- n_i = significant strikes thrown (attempted)
- $\text{fighter}(i), \text{fight}(i)$ = fighter and fight indices
- $\text{gender}(i), \text{wc}(i)$ = covariates

Linear predictor

$$\ell_i = \alpha + \beta_{\text{gender}(i)} + \beta_{\text{wc}}(i) + u_{\text{fighter}(i)} + v_{\text{fight}(i)}$$

Strike landing probability

$$p_i = \text{logit}^{-1}(\ell_i)$$

Observed data model

$$y_i \mid n_i, p_i \sim \text{Binomial}(n_i, p_i)$$

Likelihood function

$$\Pr(\mathbf{y} \mid \alpha, \beta, u, v, n) = \prod_{i=1}^N \text{Binomial}(y_i \mid n_i, \text{logit}^{-1}(\ell_i))$$

Posterior distribution

Let

$$\Theta = (\alpha, \beta, u, v, \sigma_{\text{fighter}}, \sigma_{\text{fight}})$$

be the full set of parameters.

Given the data

$$\mathcal{D} = \{y_i, n_i, \text{gender}(i), \text{weight class}(i), \text{fighter}(i), \text{fight}(i)\}_{i=1}^N$$

Bayes' rule states that

$$\Pr(\Theta | \mathcal{D}) \propto \Pr(\Theta) \times \Pr(\mathcal{D} | \Theta)$$

Written out fully:

$$\underbrace{\Pr(\Theta | \mathcal{D})}_{\text{Posterior}} \propto \underbrace{\Pr(\alpha) \Pr(\beta) \Pr(\sigma_{\text{fighter}}) \Pr(\sigma_{\text{fight}})}_{\text{Priors}} \left[\prod_{f=1}^F \Pr(u_f | \sigma_{\text{fighter}}) \right] \left[\prod_{k=1}^K \Pr(v_k | \sigma_{\text{fight}}) \right] \times \underbrace{\left[\prod_{i=1}^N \text{Binomial}(y_i | n_i, \text{logit}^{-1}(\alpha + \beta_{\text{gender}(f)} + \beta_{\text{wc}(f)} + u_f)) \right]}_{\text{Likelihood}}$$

Posterior fighter-level accuracy

To obtain a fighter's latent probability of landing a significant strike (removing fight-level randomness), we have:

$$\theta_f = \text{logit}^{-1}(\alpha + \beta_{\text{gender}(f)} + \beta_{\text{wc}(f)} + u_f)$$

Generative Story

The model assumes that every fighter has an underlying tendency to land significant strikes, and every fight has its own quirks that make landing either easier or harder than usual. We begin by imagining a baseline level of strike landing log odds for the baseline category fighter, which in this model is a male Bantamweight fighter in an average fight. Gender and weight class then shift this baseline up or down in systematic ways, holding the other predictor constant.

In addition to these population-level shifts, each fighter receives their own deviation that captures how much more or less likely they are to land strikes compared to fighters of the same gender and weight class. Each fight also receives its own deviation that reflects the unpredictable circumstances of that particular matchup, such as style interactions, pacing, or contextual conditions.

For each fighter in each fight, these components are added together on the log odds scale and transformed into a probability that a strike attempt will land in that specific context. Finally, the observed number of landed strikes is generated by drawing from a binomial distribution using this probability and the number of strikes attempted. Bayesian inference then uses the observed data to update our beliefs about fighter-level tendencies, fight-level contextual variation, and the overall influence of gender and weight class on strike landing probability.

Summary of the Generative Story

1. Draw a global baseline log odds of landing a significant strike, representing the baseline category fighter (male Bantamweight) in an average fight.
2. Draw systematic effects for gender and weight class, each representing how those predictors move up or down, holding the other predictor constant.
3. Draw fighter-specific log-odds deviations, which represent how much each fighter tends to land strikes more or less often than expected based on gender and weight class alone.
4. Draw fight-specific log-odds deviations, which represent the contextual quirks of each fight that make landing strikes easier or harder for everyone in that fight.
5. For each fighter-fight observation:
 - Sum the baseline, fixed effects, fighter deviation, and fight deviation to create the linear predictor ℓ_i .
 - Transform ℓ_i to a significant strike landing probability $p_i = \text{logit}^{-1}(\ell_i)$.
 - Generate the observed significant strikes landed y_i from a Binomial distribution with parameters n_i (significant strike attempts) and p_i .

This defines the full joint distribution over all parameters and data. Bayesian inference conditions this generative story on the observed data to obtain the posterior distribution.

Model Building

```
# Model
formula_acc <- bf(
  sig_strikes_landed | trials(sig_strikes_thrown) ~
    1 + gender + weight_class +
    (1 | fighter_id) +
    (1 | fight_id)
)

# Priors
mu0 <- qlogis(0.46) # baseline mean on logit scale
```

```

priors_acc <- c(
  # Intercept
  set_prior(paste0("normal(", mu0, ", 1)"), class = "Intercept"),
  
  # Gender, weight class effects (fixed effects)
  set_prior("normal(0, 0.3)", class = "b"),
  
  # Fighter variation
  set_prior("normal(0, 0.4)", class = "sd", group = "fighter_id"),
  
  # Fight variation
  set_prior("normal(0, 0.4)", class = "sd", group = "fight_id")
)

```

Prior Predictive Simulation

Before fitting the model, I ran a prior predictive simulation to check whether the chosen priors generate realistic UFC-like striking accuracies. This ensures the priors are weakly informative but still produce plausible ranges of fighter ability and fight-to-fight variation. Prior predictive checks help confirm that the model is sensible *before* it sees real data.

```

fit_prior_acc <- brm(
  formula = formula_acc,
  data = df,
  family = binomial(),
  prior = priors_c,
  sample_prior = "only",
  iter = 1000,
  chains = 2,
  cores = 2,
  backend = getOption("brms.backend"),
  seed = 1738
)

```

Loading required package: rstan

Loading required package: StanHeaders

rstan version 2.32.7 (Stan version 2.32.2)

```

For execution on a local, multicore CPU with excess RAM we recommend calling
options(mc.cores = parallel::detectCores()).

To avoid recompilation of unchanged Stan programs, we recommend calling
rstan_options(auto_write = TRUE)

For within-chain threading using `reduce_sum()` or `map_rect()` Stan functions,
change `threads_per_chain` option:
rstan_options(threads_per_chain = 1)

```

Do not specify '`-march=native`' in '`LOCAL_CPPFLAGS`' or a `Makevars` file

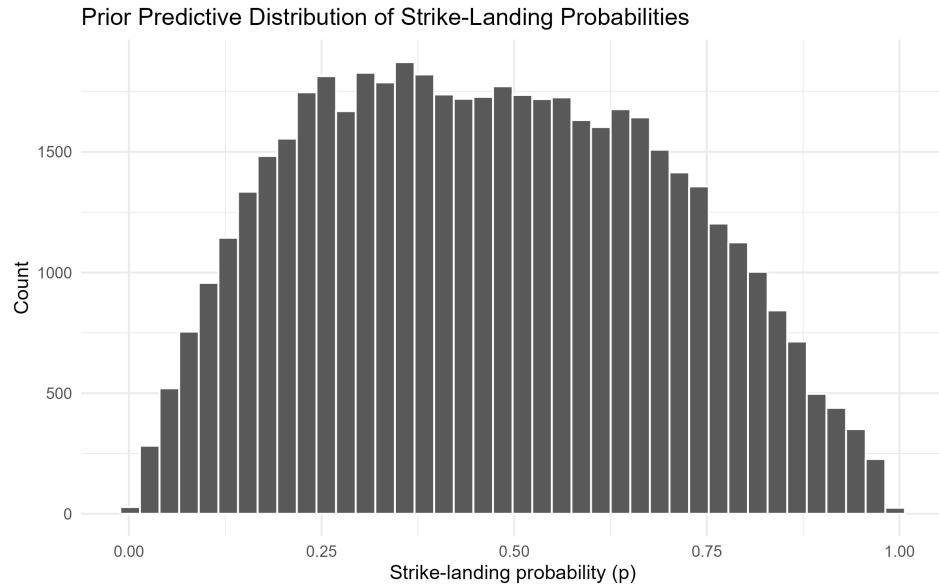
Attaching package: 'rstan'

The following objects are masked from 'package:posterior':

`ess_bulk`, `ess_tail`

The following object is masked from 'package:tidyর':

`extract`



The prior predictive distribution of strike-landing probabilities is broad and centered around moderate accuracy levels. Most simulated probabilities fall between roughly 0.15 and 0.80,

with the density peaking around 0.35-0.55 and tapering toward the extremes near 0 and 1. This indicates that the priors allow a wide range of plausible striking abilities while still concentrating mass in a realistic region for UFC fighters, avoiding implausibly high or low accuracy values.

Computational Implementation

The model was fit using `brms`, which compiles the multilevel binomial logistic model into Stan and performs posterior inference via Hamiltonian Monte Carlo. I used four chains with 2,000 iterations each (1,000 warmup), along with a slightly increased `adapt_delta` to reduce divergences.

```
fit_acc_model <- brm(
  formula = formula_acc,
  data = df,
  family = binomial(),
  prior = priors_acc,
  sample_prior = "yes",    # include prior draws in the fit object
  iter = 2000,
  warmup = 1000,
  chains = 4,
  cores = 4,
  backend = getOption("brms.backend"),
  seed = 1738,
  control = list(adapt_delta = 0.95)    # reduces divergences in logistic models
)
```

Sampling completed successfully, providing stable posterior draws for all parameters.

Fitting and Diagnostics

```
summary(fit_acc_model)
```

```
Family: binomial
Links: mu = logit
Formula: sig_strikes_landed | trials(sig_strikes_thrown) ~ 1 + gender + weight_class + (1 | :
Data: df (Number of observations: 16344)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
      total post-warmup draws = 4000
```

Multilevel Hyperparameters:

	Estimate	Est.Error	l-95%	CI	u-95%	CI	Rhat	Bulk_ESS	Tail_ESS
~fight_id (Number of levels: 8199)									
sd(Intercept)	0.41	0.00	0.41	0.42	1.01	897	1462		
~fighter_id (Number of levels: 2448)									
sd(Intercept)	0.39	0.01	0.38	0.41	1.00	616	1043		

Regression Coefficients:

	Estimate	Est.Error	l-95%	CI	u-95%	CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	-0.33	0.02	-0.38	-0.28	1.01	366			
genderWomen	0.20	0.04	0.13	0.27	1.02	358			
weight_classCatchweight	0.10	0.06	-0.02	0.21	1.00	456			
weight_classFeatherweight	0.04	0.03	-0.02	0.09	1.01	433			
weight_classFlyweight	-0.07	0.03	-0.13	-0.02	1.01	537			
weight_classHeavyweight	0.31	0.04	0.24	0.38	1.01	473			
weight_classLightHeavyweight	0.16	0.04	0.09	0.22	1.01	492			
weight_classLightweight	0.01	0.03	-0.04	0.07	1.01	458			
weight_classMiddleweight	0.11	0.03	0.05	0.18	1.01	459			
weight_classStrawweight	-0.13	0.04	-0.22	-0.04	1.01	438			
weight_classWelterweight	0.09	0.03	0.04	0.15	1.01	454			
									Tail_ESS
Intercept			650						
genderWomen			902						
weight_classCatchweight			970						
weight_classFeatherweight			860						
weight_classFlyweight			1129						
weight_classHeavyweight			1118						
weight_classLightHeavyweight			875						
weight_classLightweight			777						
weight_classMiddleweight			864						
weight_classStrawweight			944						
weight_classWelterweight			914						

Draws were sampled using `sample(hmc)`. For each parameter, Bulk_ESS and Tail_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

The fitted model indicates meaningful variability in strike-landing ability both across fighters and across fights. The fighter-level standard deviation (≈ 0.39 on the log-odds scale) reflects

substantial underlying differences in striking accuracy between athletes, even after accounting for gender and weight class. The fight-level standard deviation (≈ 0.41) is similarly large, suggesting that contextual factors unique to each bout (such as pace, matchup dynamics, or environmental conditions) also contribute significantly to how easy or difficult it is to land strikes in a given fight.

The population intercept corresponds to the baseline category (male bantamweight fighters in an average fight), and the fixed effects show that gender and most weight classes produce modest shifts from this baseline. In particular, women exhibit slightly higher strike-landing log-odds than men, while weight-class effects cluster tightly around zero, indicating that once fighter identity is accounted for, weight class alone does not strongly predict accuracy. All parameters show excellent convergence diagnostics ($\hat{R} \approx 1$ and high ESS), indicating that the posterior samples are stable and reliable for inference.

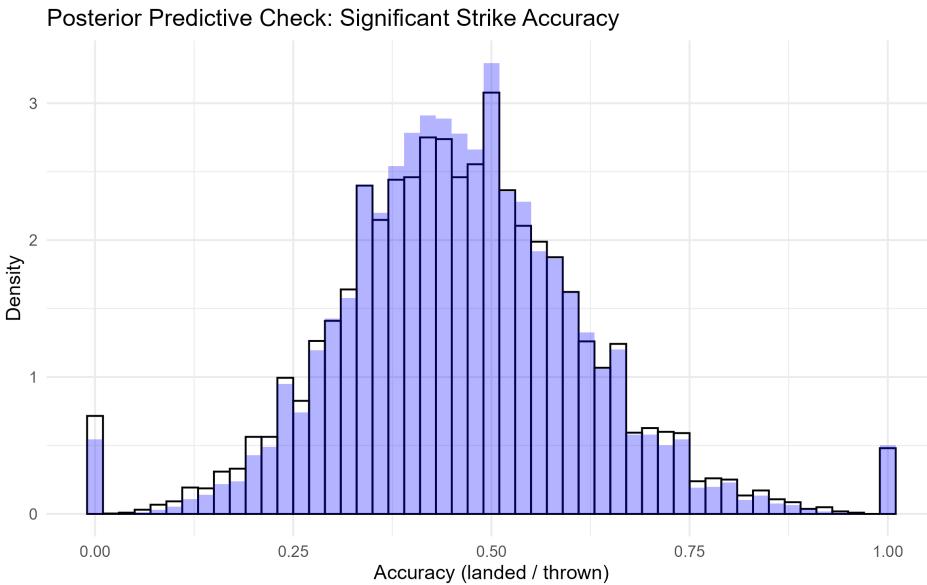
Model Evaluation

One way to evaluate the model is to answer the question:

Does the model generate simulated data that look like the real UFC data?

In other words, I performed a posterior predictive check by comparing the observed strike-landing outcomes to data simulated from the posterior predictive distribution.

```
Warning: The `x` argument of `as_tibble.matrix()` must have unique column names if
`.name_repair` is omitted as of tibble 2.0.0.
i Using compatibility `$.name_repair`.
```



The posterior predictive check on the accuracy scale shows that the model reproduces the overall distribution of strike-landing accuracy very well. The simulated accuracies (purple) closely overlap the observed accuracies (black outline) across the entire range: the center of the distribution around 40–60 percent, the spread toward both lower and higher accuracies, and even the small mass near 0 and 1 that occurs when fighters throw very few strikes. This strong agreement indicates that the hierarchical binomial model captures both the central tendency and variability of per-fight striking accuracy in the UFC data.

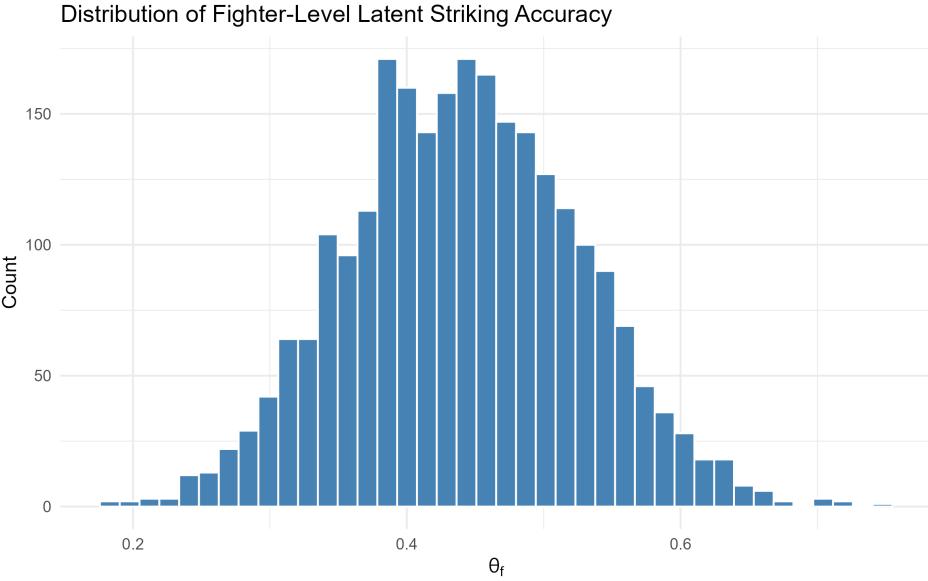
Fighter Accuracy Estimation

To answer the second research question, I use the fitted multilevel model to estimate each fighter's underlying probability of landing a significant strike, independent of fight-to-fight randomness.

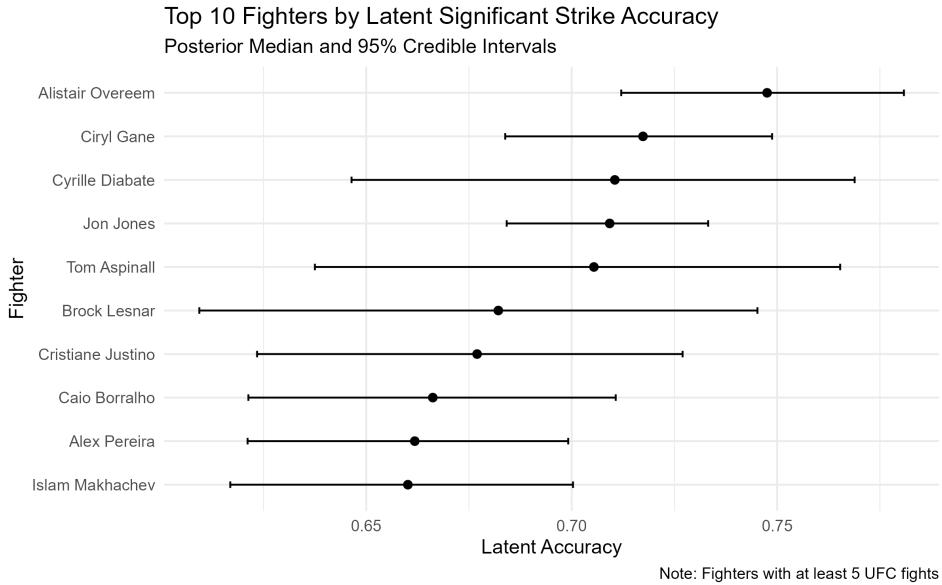
In this framework, every fighter has a latent accuracy parameter, θ_f , which represents their true strike-landing ability after accounting for gender, weight class, and partial pooling toward the population mean. Mathematically, θ_f is obtained by combining the population intercept, the appropriate fixed effects, and the fighter's random effect, then transforming the result through the logistic function. The posterior distribution of θ_f therefore provides a probabilistic estimate of each fighter's striking accuracy in an “average” UFC fight.

$$\theta_f = \text{logit}^{-1}(\alpha + \beta_{\text{gender}(f)} + \beta_{\text{wc}(f)} + u_f)$$

```
Warning in left_join(., fighter_cov, by = "fighter_id"): Detected an unexpected many-to-many
i Row 38 of `x` matches multiple rows in `y`.
i Row 1936 of `y` matches multiple rows in `x`.
i If a many-to-many relationship is expected, set `relationship =
"many-to-many"` to silence this warning.
```

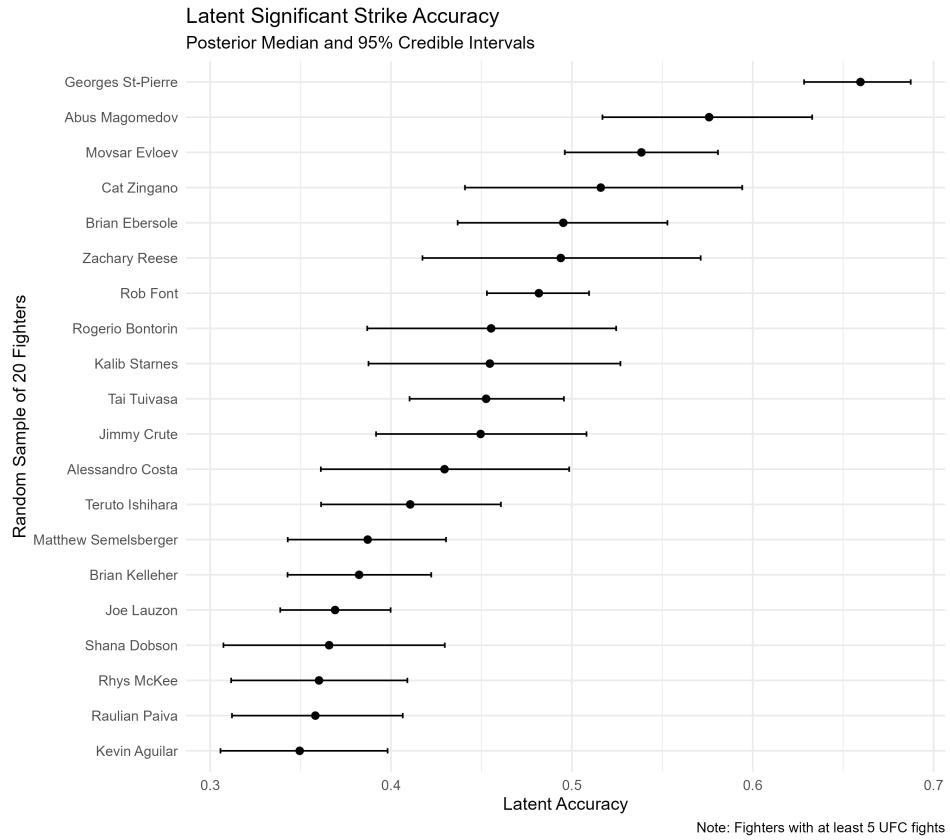


This histogram shows the posterior distribution of latent striking accuracy θ_f across all UFC fighters. The shape closely mirrors the empirical distribution of observed significant strike accuracy, indicating that the hierarchical model successfully recovers the underlying population pattern while smoothing out fight-to-fight noise. Most fighters fall between 0.40 and 0.50 accuracy with only a small minority displaying much higher or lower underlying precision.



This figure displays the ten fighters with the highest estimated latent striking accuracy among those with at least 5 UFC fights. Points denote posterior median accuracy and horizontal bars show 95 percent credible intervals, highlighting both the rank ordering and the uncertainty around each estimate. This plot emphasizes which fighters the model identifies as having the strongest underlying ability to land significant strikes.

For example, conditional on the model, there is a 95% probability that Alistair Overeem's significant strike accuracy falls between 0.71 and 0.78.



This figure shows posterior estimates of latent significant strike accuracy for a random subset of 20 UFC fighters who have recorded at least 5 bouts. Each point represents the fighter's posterior median accuracy, while the horizontal bar represents the 95 percent credible interval. This plot illustrates typical uncertainty levels across the roster and provides a representative look at the distribution of underlying striking skill in the population.

Conclusion

Across thousands of UFC bouts spanning twenty-five years, the predictive modeling in RQ1 showed that not all in-fight performance metrics contribute equally to determining who wins. Significant strike differential emerged as the dominant predictor, with each one-standard deviation advantage in landed-versus-absorbed significant strikes producing a large increase in the odds of winning. Knockdowns also contributed meaningfully, though less consistently due to their rarity, while takedown differential and control time had much weaker and less reliable associations. Overall, the model reveals that winning in modern UFC competition is driven far more by effective striking output than by positional or grappling control.

In RQ2, the Bayesian multilevel model uncovered a more fundamental layer of the sport: each fighter's underlying probability of landing a significant strike, separated from fight-to-fight randomness and adjusted for gender and weight class. The model identified meaningful variation across fighters, with some individuals showing substantially higher latent accuracy than others even after accounting for systematic differences across divisions. Posterior predictive checks confirmed that the model captures the broad distribution of observed striking accuracy, while partial pooling stabilized estimates for fighters with fewer bouts.

Taken together, these results provide both a descriptive and structural understanding of UFC performance. RQ1 isolates *what* in-fight metrics matter most for winning, clearly pointing to striking efficiency and output. RQ2 then reveals *why* fighters vary in those metrics, quantifying stable differences in underlying striking skill across the roster. The combination offers a coherent statistical picture of the sport: fights are won primarily through striking effectiveness, and fighters differ systematically in their ability to land strikes, even after adjusting for context and division.