The Hot Hand in Call of Duty Search and Destroy

A Preprint

Tony ElHabr

anthonyelhabr@gmail.com

Andrew ElHabr
School of Industrial and Systems Engineering (ISyE)
Georgia Institute of Technology
Atlanta, Georgia, USA
andrewelhabr@gatech.edu

August 21, 2022

Abstract

Our research investigates patterns in round win percentages in professional Search and Destroy (SnD) matches of the popular first-person shooter game Call of Duty (CoD). First, we find evidence supprting the hypothesis that round win probability can be modeled as a constant across the series, although not at the naive 50%. Second, we examine post-streak round win probability, given the series outcome. We find no evidence that post-streak win rate is significantly different than expected for combinations of streak length, series length, and series winner, suggesting that the perception of momentum at the team level is just that, perception.

Keywords esports · Call of Duty · hot hand · runs test

1 Introduction

1.1 Call of Duty Search and Destroy

Call of Duty (CoD), first released in 2003, is one of the most popular first-person shooter (FPS) video game franchises of all-time. The most popular mode in the competitive scene is "Search and Destroy" (SnD).[1] SnD is a one-sided game mode in which one team, the offensive side, tries to destroy one of two designated bomb sites on the map.

In professional CoD SnD, the two teams take turns playing offense and defense every round. They must win six rounds to win the series.[2] A round can end in one of five ways:

1. One team eliminates all members of the other team prior to a bomb plant. (Eliminating team wins.)
2. The offensive team eliminates all members of the defensive team after a bomb plant.[3] (Offense wins.)

---

[1]SnD bears resemblance to "Bomb Defusal" in Counter-Strike and "Plant/Defuse" in Valorant, two other FPS games played in more popular professional leagues.

[2]A maximum of 11 even rounds can be played. There is no "sudden death" or "win by two" rule like there are for SnD equivalent in professional Counter-Strike and Valorant matches.

[3]

- The bomb can be picked up by any member of the offensive team.
- The bomb carrier is not obstructed at all by carrying the bomb (i.e. movement is the same, weapon usage is the same).
- The defense does not get any visual indication for who is carrying the bomb.
- A bomb plant takes five seconds. The timer resets if the player stops planting site prior to completing it.
- A bomb defuse takes seven seconds. The timer resets if the player "drops" the bomb.
- The bomb takes 45 seconds to defuse after being planted.

3. The defensive team defuses the bomb after a bomb plant.[4] (Defense wins.)
4. The offensive team does not make a plant by the time the round timer ends. (Defense wins.)

We adopt the terminology "series" to refer to what CoD SnD players typically call a "match", so as to emulate the terminology of playoff series in professional leagues like the National Basketball Association, National Hockey League, and Major League Baseball. A "game" or a "match" in such leagues is analogous to a "round" of CoD SnD. Further, it is analogous to a trial in statistical analysis of Bernoulli events.

SnD is not the only game type played in competitive CoD. Teams play in a head-to-head, best-of-five (first-to-three wins) format, where SnD is always played as the second and fifth game mode (if a fifth variant is needed). The best-of-five across game types could also be called a "series", but since we analyze only the SnD game types, we refer to SnD games as series (of rounds).

Anecdotally, most observers swear by the existence of momentum in CoD SnD series, to the extent that vernacular has been developed to describe such phenomenon. People say that teams "ice up" when the team wins several consecutive rounds to come back to win a series, often culminating with a series-clinching round win.[5] Further, viewers have come to embrace the "5-3" phenomenon, where teams win three consecutive rounds after facing a 5-3 deficit to win 6-5. There is even a term for the rare 0-5 comeback—a "full sail".

1.2 Data

CoD has roughly gone through three eras of professional gaming: (1) Major League Gaming (MLG) tournaments prior to 2016; (2) the CoD World League (CWL), initiated in 2016; and (3) the 12-franchise CoD League (CDL), operating since 2020. The CDL has completed three year-long "seasons" as of August 2022.[6]

The data set consists of all SnD matches played in major tournaments and qualifiers during the CDL era, totaling 7,792 rounds across 852 series.[7] Data was collected in spreadsheets by community member "IOUTurtle".[8]

The observed offensive round win percentage across all rounds is $\tau_O = 47.8\%$.[9] Table 1 shows round win percentages by series "state" (i.e. the number of round wins by each team prior to an upcoming round). Offensive round win rate is not quite constant, although never veers more than 10% from this overall average.

Table 1: Offensive round win rates for the upcoming round, given both the offensive and defensive team's prior number of round wins. Numbers in parentheses are sample sizes.

Defense round wins	Offense round wins					
	0	1	2	3	4	5
0	47.8% (852)	46.6% (408)	43.1% (216)	43.5% (115)	43.3% (67)	40.5% (37)
1	48.6% (444)	49.3% (418)	51.5% (309)	43.4% (205)	43.3% (120)	39.4% (99)
2	52.8% (218)	48.9% (305)	48.9% (315)	46.6% (262)	48.7% (189)	42.1% (133)
3	54.5% (123)	46.0% (200)	49.6% (250)	45.6% (248)	44.4% (214)	44.8% (174)
4	56.9% (65)	54.5% (145)	47.2% (193)	44.7% (228)	55.2% (221)	50.5% (208)
5	47.4% (38)	49.4% (83)	47.1% (136)	50.9% (175)	45.2% (177)	46.0% (202)

---

[4]Often the defensive team will try to eliminate all team members prior to making the defuse, but in some cases, they may try to "ninja" defuse.

[5]This vernacular is not specific to CoD SnD in specific, and is not only used in the contexts of streaks.

[6]CoD is fairly unique compared to other esports in that it runs on an annual lifecycle (released coming in the late fall), where a new game is published every year under the same title. Each new game bears resemblance to past ones, often introducing relatively small variations ("improvements") to graphics, game modes, and other facets of gameplay. During the CDL era, the games released have been Modern Warfare (2020), Cold War (2021) and Vanguard (2022).

[7]288 of the series occur in major tournaments, which are considered to be more "competitive" than the qualifiers since they are played in person (COVID-permitting), whereas the qualifiers are played online.

[8]Raw data: `https://linktr.ee/CDLArchive`. Author: `https://twitter.com/IOUTurtle`. Processed data can be found at `https://github.com/tonyelhabr/fps_round_games/blob/master/data/cod_rounds.csv`.

[9]Offensive round win percentage has been nearly constant across the three games during the CDL era: 1. 47.2% in MW (2020) 2. 47.9% in Cold War (2021) 3. 48.1% in Vanguard (2022)

On the extremes, 6-0 sweeps make up 4.7% of all series, while 6-5 series make up 23.7 of series.

At tournaments, higher-seeded ("better") teams choose whether they want to start on offense in the SnD series played as the second game or the fifth game, and their opponent is assigned to play offense in the first round of the SnD slot not chosen.[10][11] ("Competitive Settings | Call of Duty League," n.d.) With online qualifiers, teams all play each other twice throughout the season. Each team gets to play the higher-seeded role once in their matchups, despite their win-loss records. Overall, we might say that bias due to stronger teams playing defense (slightly advantageous) more often is very minimal.

2 Literature review

There have been a handful of studies of the distribution of games played in a series of a professional sport. Most assume a constant probability ϕ of a given team winning a game in the series, regardless of the series state. Mosteller (1952) observed that the American League had dominated the National League in Major League Baseball's (MLB) World Series matchups, implying that games should not be modeled with a constant (null) $\phi_0 = 0.5$. Mosteller proposed three approaches for identifying the optimal constant probability value of the stronger team in the World Series, finding $\phi \approx 0.65$ in each case: (1) solving for ϕ from the observed average number of games won by the loser of the series, which he called the "method of moments" approach; (2) maximizing the likelihood that the sample would have been drawn from a population in which the probability of a team winning a game is constant across the series (i.e. maximum likelihood), and (3) minimizing the chi-square goodness of fit statistic χ^2 for ϕ.

Chance (2020) re-examines the constant probability notion in Major League Baseball's (MLB) World Series (1923–2018), the National Basketball Association's (NBA) Finals (1951–2018), and the National Hockey League's (NHL) Stanley Cup (1939–2018). Chance finds strong evidence against the null hypothesis of $\phi_0 = 0.5$ in the MLB and NHL championship series when applying Mosteller's first and second methods. Chance goes on to outline a conditional probability framework (likelihood of winning a game given the series state) which can exactly explain the distribution of the number of games played.

Momentum, one of most discussed topics in sports analytics, goes hand-in-hand with a discussion of the nature of series outcomes.[12] Two opposing fallacies are observed in the context of momentum: the "gambler's fallacy" (negative recency) and the "hot hand fallacy" (positive recency). Per Ayton et al. (2004), negative recency is "the belief that, for random events, runs of a particular outcome ... will be balanced by a tendency for the opposite outcome", while positive recency is the expectation of observing future results that match recent results.

Studying both player streaks and team streaks in basketball, in both observational and controlled settings. Gilovich et al. (1985), henceforth GVT, do not find evidence for the hot hand phenomenon. However, Miller and Sanjurjo (2018), henceforth MS, provide a framework for quantifying streak selection bias, which effectively works in the manner posited by the gambler's fallacy. Specifically, MS say that a "bias exists in a common measure of the conditional dependence of present outcomes on streaks of past outcomes in sequential data" implying that, under i.i.d. conditions, "the proportion of successes among the outcomes that immediately follow a streak of consecutive successes is expected to be strictly less than the underlying (conditional) probability of success". When applying streak selection bias to GVT's data, MS come to the opposite conclusions as GVT.

Other research has borrowed a common approach from the field of quality control, looking at unlikely sequences of events with the Wald-Wolfowitz runs test (Peel and Clauset 2015), (Steeger, Dulin, and Gonzalez 2021). Peel and Clauset find no evidence for unlikely sequences of scoring events in the NHL, College Football (CFB), and National Football League (NFL), but do in the NBA. As a check on their entropy approach to momentum identification, Steeger et al. found three NHL teams who sequence of wins in the 2018-2019 regular season violated the Wald-Wolfowitz null hypothesis, although these teams did not match the teams identified with their primary approach.

---

[10]"Hardpoint" has been played as the first and fourth games in a matchup; the variant for the third game was Domination in the 2020 season, and Control in the 2021 and 2022 seasons.

[11]Anecdotally, the higher-seeded teams tend to choose to play defense in the first round of the SnD series played as the second game, although this choice is not consistent across teams, or with the same team over time.

[12]We often use use "streaks" and momentum interchangeably, but as (Steeger, Dulin, and Gonzalez 2021) note, momentum implies dependence between events, whereas streaking does not.

2.1 Our contribution

Despite the plethora of existing research on games played in a series and momentum in sports, these topics have yet to be investigated heavily, if at all, in esports. Work has been done to examine intra-round win probability in other FPS titles such as Counter-Strike (Xenopoulos, Freeman, and Silva 2022) and Valorant (DeRover 2021), both of which are round-based like CoD SnD. However, considering Counter-Strike and Valorant specifically, research on round-level trends seems non-existant, perhaps for one of many reasons:

1. Both have economic aspects that can create clear advantages for one side in a given round, given how prior rounds played out. CoD has no such equivalent, except for perhaps "kill streaks", which infrequently play a role.
2. Teams play either offense or defense for many consecutive rounds. (The former has a 14-14-1 format and the latter uses a 12-12-1 format.) On the other hand, teams in CoD SnD rotate roles every round, analogous to a 1-1-1-1-1-1-1 format for home advantage in best-of-seven series for professional sports like the MLB, NBA, and NHL.[13] While theoretically one might be able to account for any kind of format, such as a 5-5-1, the rotation of team sides every round is convenient for convincing ourselves that rounds could reasonably be modeled as i.i.d. Bernoulli trials.
3. Both Counter-Strike and Valorant have overtime rules—team must win by two rounds—which can make end-of-series sequences difficult to model. CoD SnD does not have overtime rules.

To our knowledge, there is no existing public statistical research on CoD SnD format, beyond descriptive analysis on social media. While intra-round trends may be more directly applicable to teams looking for an advantage on their competition, broader investigation on a common concept like the hot-hand phenomenon in a new field—particularly one that is less subjected to factors that may be difficult to control for, e.g. weather—should be useful as a precedent or supplement for future researchers.

3 Methodology

3.1 Distribution of rounds played

In a best-of-s format, assuming a constant round win probability ϕ ($0 \leq \phi \leq 1$) for one team[14], the probability that a series ends in r rounds ($r \leq s$), $\hat{\Phi}(r)$ is

$$m = \frac{s+1}{2} \tag{1}$$

$$\hat{\Phi}(r) = \frac{(r-1)!}{(m-1)!(s-r)!}(\phi^m(1-\phi)^{r-m} + \phi^{r-m}(1-\phi)^m) \tag{2}$$

$s = 11$ rounds for CoD SnD.[15] For example, assuming $\phi = 0.5$, the probability of a series ending in 9 rounds in CoD is

$$\hat{\Phi}(9) = \frac{(9-1)!}{(6-1)!(11-9)!}(0.5^6(1-0.5)^{9-6} + 0.5^{9-6}(1-0.5)^6) = 56(0.5^9 + 0.5^9) = 0.21875.$$

To evaluate the constant round win probability null hypothesis—that is, that the expected and observed round win rates, $\hat{\Phi}(r)$ and $\Phi(r)$ respectively, are equal to one another—we can compute the chi-square goodness of fit statistic

$$\chi^2 = \sum^{r \in R} \frac{(\Phi(r) - \hat{\Phi}(r))^2}{\hat{\Phi}(r)}. \tag{3}$$

---

[13]1-1-1-1-1-1-1 is not used today in these leagues, but it was at least once in each league.

[14]If $\phi > 0.5$, we can say that this is the stronger team.

[15]ϕ and other symbols are selected so as to reserve symbols like p for usage in other contexts without causing confusion for the reader. Upper-case symbol, e.g. Φ, are consistently used in this paper to represent proportions, while lower-case symbols are used to represent Bernoulli trials success probability. "Hats", e.g. $\hat{\Phi}$, are used to convey expectations, while bare symbols symbols convey observational data.

in which $R = [6, 7, 8, 9, 10, 11]$ for CoD SnD.

3.2 Momentum

3.2.1 MS post-streak probability

Let us now consider round win rate immediately following a streak of k wins, given that the series lasts r rounds. Removing our knowledge of a streak, we might model the Bernoulli round win probability as

$$p_0(\text{win}|r) = p_0^{+|r} = \begin{cases} \frac{m}{r} & \text{team wins series} \\ \frac{r-m}{r} & \text{team loses series} \end{cases} \tag{4}$$

using m from Equation 1.

As shown by MS with their Theorem 1, we should expect the proportion of rounds wins immediately following a streak of k rounds wins for a series lasting r rounds, $\hat{P}_{MS}^{+|k,r}$, to be strictly less than $p^{+|r}$.[16] Unfortunately, there does not exist a closed form representation for the expected value of $\hat{P}_{MS}^{+|k,r}$ for $k > 1$. Nonetheless, one may recursively build a collection of dictionaries to compute the expectation analytically, or, more conveniently, one may run simulations to estimate the expected value, as We choose to do.[17]

Although it would be convenient to emulate the hypothesis testing for difference of proportions performed by MS (and GVT), our context is fundamentally different from that of MS, who focus on longitudinal data in controlled settings.[18] The number of trials is fixed in their experimental designs, but in CoD SnD, the number of rounds played is determined as a function of the max number of possible rounds (s) and whether or not the team wins the series. The Bernoulli trial success probability, i.e. a round win in CoD SnD, is not independent of the opponent. Consequently, a statistical test of the difference in $\hat{P}^{+|k,r}$ and $\hat{P}^{-|kr}$, as performed by MS to evaluate their hypothesis regarding post-streak success rates, is not completely appropriate, although useful as a reference.

3.2.2 "Notional" post-streak probability

One can consider another form of the expected proportion of rounds won immediately after a streak of k round wins in a best-of-s series given the length of the series (r rounds), the "notional" proportion $\hat{P}_0^{+|k,r}$. The Bernoulli round win probability underlying $\hat{P}_0^{+|k,r}$ is

$$p_0(\text{win}|k,r) = p_0^{+|k,r} = \begin{cases} \frac{m-k}{r-k} & \text{team wins series} \\ \frac{s-m-k}{r-k} & \text{team loses series} \end{cases} \tag{5}$$

using m from Equation 1.

We can perform a binomial test to evaluate the null hypothesis $\omega = \omega_0$ for the observed probability of success ω and a user-specified ω_0, where $0 \leq \omega_0 \leq 1$.[19] If there are r^+ observed successes in a sample of r trials and we expect that there should be $r * \omega_0$, the probability of arriving at this expected number of successes is

$$\Pr(r^w) = \binom{r}{r^+} \omega^{r^+} (1-\omega)^{r-r^+}. \tag{6}$$

Treating the notional proportion $p_0^{+|k,r}$ as the null ω_0 and plugging in the observed proportion $P^{+|k,r}$ for ω in Equation 6 (treating proportions as probabilities), we can evaluate the null hypothesis that the observed probability is equal to the notional probability. If we can reject this null hypothesis, then we can consider team momentum, represented by post-streak success, plausible.

---

[16]See Appendix E.1 of MS for the proof.

[17]We have adapted the code from Vafa (2017), which implements MS's framework (2018).

[18]GVT also performs statistical tests on shots from players in live games, i.e. "observational" data, but they note that their findings are likely affected player shot selection in the face of defensive strategy by the opposing team.

[19]ω and other symbols are chosen so as to not conflict with symbols already used elsewhere for in other equations, either generic or for CoD SnD specifically.

One can perform the same binomial test for the MS's streak-selection-adjusted proportion, $\hat{P}_{MS}^{+|k,r}$. FOOTY: Given the caveats mentioned before, the results should be heeded with caution.

We can further decompose Equation 5 by the round i ($i \leq r$) in which the streak of length k carries into.

$$p_0(\text{win}|k,r,i) = p_0^{+|k,r,i} = \{ \begin{array}{ll} \frac{m-k}{r-k}, & \text{team wins series}, i \neq r \\ 1, & \text{team wins series}, i = r \\ \frac{s-m-k}{r-k}, & \text{team loses series}, i \neq r \\ 0, & \text{team loses series}, i = r \end{array} \tag{7}$$

Again, one might apply a binomial test to evaluate the hypothesis that the expected proportions, both $\hat{P}_0^{+|k,r,i}$ and $\hat{P}_{MS}^{+|k,r,i}$, are equal to the observed proportion $P^{+|k,r}$.

3.2.3 Wald-Wolfowitz runs test

Stepping back from the one-sided perspective of a single team's round win probability when streaking, one can attempt to detect the hot hand phenomenon with the Wald-Wolfowitz runs test. Under the null hypothesis, the number of runs in a sequence of r trials, $\zeta(r)$, is a random variable that can take on values r^+ and r^- for success and failure respectively, with the following mean μ and variance σ^2.

$$\mu = \frac{2r^+r^-}{r} + 1, \sigma^2 = \frac{(\mu-1)(\mu-2)}{r-1} \tag{8}$$

To incorporate our findings regarding constant round win probability, We can specify that $\Pr(r^+) = \phi$ (and, conversely, that $\Pr(r^-) = 1 - \phi$).

One can subset the observed series sequences to those that violate the run test null hypothesis and perform a test of proportions, where the null is that the observed rate ψ^+ of such series is equal to the frequency rate ψ_0. The test statistic Z

$$Z = \frac{\psi^+ - \psi_0^+}{\sqrt{\psi_d^+(1-\psi_d^+)(\frac{1}{N^+} + \frac{1}{\hat{N}^+})}} \tag{9}$$

where $\psi_d^+ = \frac{r^+ - \hat{r}^+}{N^+ - \hat{N}^+}$. If we can reject the null hypothesis for many sequences, then there is implicit evidence in favor of the hot hand.

4 Results

First, We investigate the constant probability assumption and the distribution of rounds played in a series. Chance's (2020) work is closely related to mine, and, in fact, provides a guide for this investigation. Afterwards, We investigate post-streak win rates within the context of momentum.

4.1 Distribution of rounds played

Using Equation 3, We find that $\chi^2 = 16.0$ (p-value of 0.0068) for the naive constant round win probability $\phi_0 = 0.5$. Thus, We can comfortably reject the constant probability hypothesis for the null $\phi_0 = 0.5$, even at a confidence level of $\alpha = 0.01$.

Table 2 shows the expected series lasting r rounds, $\hat{\Phi}_0(r)$ (under the assumption $\phi_0 = 0.5$), for $s = 11$, along with the observed proportions, $\Phi(r)$, in CoD SnD.

Table 2: The expected proportion of series lasting r rounds ($\hat{\Phi}_0(r)$) in a best-of-11 format, where $r \in R = [6, 7, 8, 9, 10, 11]$ under the assumption that each team has a constant round win probability $\phi_0 = 0.5$. The observed frequencies for CoD SnD are shown as a count $N(r)$ and as a proportion $\Phi(r)$ of all series ($\sum^{r \in R} N(r)$).

r	$N(r)$	$\Phi(r)$	$\hat{\Phi}_0(r)$
6	40	4.7%	3.1%
7	101	11.9%	9.4%
8	141	16.5%	16.4%
9	185	21.7%	21.9%
10	183	21.5%	24.6%
11	202	23.7%	24.6%

Table 3 shows the alternate values for the constant round win probability that We find when applying the three methods suggested by Mosteller (1952). Each is approximately or equal to 0.575, and, when applying Equation 3, each results in a χ^2 value for which We cannot reject the constant probability null hypothesis.

Table 3: Alternate estimates of the constant probability (ϕ) for winning a given round in a CoD SnD, applying the three methods suggested by Mosteller (1952), in addition to the naive ($\phi_0 = 0.5$).

Method	ϕ	χ^2 (p-value)
0. Naive	0.5000	16.0 (≤ 0.01)
1. Method of moments	0.5725	3.6 (0.6)
2. Maximum likelihood	0.5750	3.5 (0.62)
3. Minimum (χ^2)	0.5775	3.5 (0.62)

Table 4 shows the new $\hat{\Phi}(r)$ when re-applying Equation 2 for the maximum likelihood estimate $\phi_2 = 0.575$, resulting in a new set of expected proportions of series lasting r rounds $\hat{\Phi}_2(r)$.[20] We observe that $\hat{\Phi}_2(r)$ is larger than $\hat{\Phi}_0(r)$ for $r \in [6, 7]$, more closely matching $\Phi(r)$. $\hat{\Phi}_2(r)$ is also closer to the observed $\Phi(r)$ for $r \in [9, 10]$, although not for $r \in [8, 11]$.

Table 4: The observed proportion of series $\Phi(r)$ ending in r rounds in CoD SnD's best-of-11 format, compared to the expected proportion $\hat{\Phi}_0(r)$ the naive assumption $\phi_0 = 0.5$ and the expected proportion $\hat{\Phi}_2(r)$ under the maximum likelihood estimate $\phi_2 = 0.575$ for constant round win probability.

r	$\Phi(r)$	$\hat{\Phi}_0(r) = 0.5$	$\hat{\Phi}_2(r) = 0.575$
6	4.7%	3.1%	4.2%
7	11.9%	9.4%	11.2%
8	16.5%	16.4%	17.8%
9	21.7%	21.9%	21.8%
10	21.5%	24.6%	23.0%
11	23.7%	24.6%	22.0%

Observing that $\hat{\Phi}_2(r)$ reasonably matches $\Phi(r)$ (especially in comparison to $\hat{\Phi}_0(r)$), along with the null hypothesis rejection shown in Table 3, it is fair to conclude that the constant round win probability assumption can be valid in CoD SnD series with the appropriate choice of ϕ (≈ 0.575).

---

[20]The method of moments and minimum χ^2 estimates for ϕ are omitted simply because the results would be nearly identical to those for the maximum likelihood estimate of ϕ (since they are all ≈ 0.575).

4.2 Momentum

4.2.1 Post-streak probability

Given that people typically perceive streaks as beginning after the third success (or failure) at minimum (Carlson and Shu 2007), We focus on streaks of three round wins.[21] Table 5 compares the observed round win rate $P^{w,kr}$ following streaks of $k = 3$ round wins given the series outcome (r rounds and winner) with the notional and MS expected proportions, $\hat{P}_0^{+|k,r}$ and $\hat{P}_{MS}^{+|k,r}$ respectively.

Table 5: Given the round win streak $k = 3$, the length of the series (r rounds), and the series winner, the observed count of rounds wins, $r^{+|k=3,r}$, and proportion of round wins, $P^{+|k=3,r}$, among $N^{k=3,r}$ instances where a team could win after the streak ($P^{+|k=3,r} = \frac{r^{+|k=3,r}}{N^{k=3,r}}$). Additionally, the notional and MS expected proportions, $\hat{P}_0^{+|k=3,r}$ and $\hat{P}_{MS}^{+|k=3,r}$ respectively.

r	Win series?	$r^{+|k=3,r}$	$N^{k=3,r}$	$P^{+|k=3,r}$	$\hat{P}_0^{+|k=3,r}$	$\hat{P}_{MS}^{+|=3k,r}$
7	yes	156	209	74.6%	75.0%	75.7%
8	yes	130	209	62.2%	60.0%	61.9%
9	yes	100	193	51.8%	50.0%	52.7%
10	no	8	60	13.3%	14.3%	26.7%
10	yes	66	151	43.7%	42.9%	44.7%
11	no	31	129	24.0%	25.0%	30.8%
11	yes	60	150	40.0%	37.5%	38.8%

With the exception of $\hat{P}_{MS}^{+|k,r}$ when $r = 10$ and the eventual series loser is the team that wins after a streak of three round wins[22], all un-adjusted binomial test p-values are greater than the typical $\alpha = 0.05$ threshold, implying that we cannot reject the null hypothesis that the expected notional and MS post-streak round win rates are different than the observed proportion. Adjusting the p-values with the Benjamini & Yekutieli (BY) correction (2001), the null hypothesis cannot be rejected for any case.[23]

When performing the same tests for streaks of two, four, and five, there is no case in which We can reject the binomial null hypothesis for the expected notional probability $p_0^{+|k,r}$. The null hypothesis can only be rejected for the expected MS probability $p_{MS}^{+|k,r}$ when $k = 5, r = 7$ after applying the BY p-value correction.[24]

Now let us explicitly consider the round, i. As shown in Table 6, the notional win rate in the round i immediately following a streak of k round wins in a series lasting r rounds, $\hat{P}+|k, r, i_0$, is statistically different than the observed $P^{+|k,r,i}$ in several cases, notably when $r = i$ for each of $r \in [9, 10, 11]$.[25] Interestingly, the observed rate is greater than the expected rate in each case where we can reject the null hypothesis, whereas the observed rate is less than the expected rate in all other cases, as well in a strong majority of cases not shown.

---

[21]Three happens to also be a reasonable number for series that last at maximum 11 rounds.

[22]This implies that the opponent, the eventual series winner, goes on to win the series 6-4 after going down 0-4. This is the only scenario in which one team wins four rounds consecutively but loses a best-of-11 series in 10 rounds.

[23]As noted before, the MS expectations may be unreliable for CoD SnD, so one is inclined to prefer the results of the notional binomial tests.

[24]At 36, the sample size for $k = 5, r = 7$ is not as large as most other combinations of k and r.

[25]The null can also be rejected for $r \in [7, 8]$, although these are not shown, and the sample sizes are smaller.

Table 6: Given the round i, the round win streak $k = 3$, the length of the series (r rounds), and the series winner, the observed count of rounds wins, $r^{+|k=3,r,i}$, and proportion of round wins, $P^{+|k=3,r,i}$, among $N^{k=3,r,i}$ instances where a team could win after the streak ($P^{+|k=3,r,i} = \frac{r^{+|k=3,r,i}}{N^{k=3,r,i}}$). Additionally, the notional proportion, $\hat{P}_0^{+|k=3,r,i}$, and the p-value of a binomial test comparing $\hat{P}_0^{+|k=3,r,i}$ to the observed proportion $P^{+|k=3,r,i}$. Table restricted to $r \in [10, 11]$ and $i < r$ and $N^{k=3,r,i} \geq 10$ for brevity.

r	i	Win series?	$r^{+|k=3,r,i}$	$N^{k=3,r,i}$	$P^{+|k=3,r,i}$	$\hat{P}_0^{+|k=3,r,i}$
10	7	no	2	13	15.4%	14.3%
10	7	yes	11	13	84.6%	42.9%
10	8	no	1	10	10.0%	14.3%
10	8	yes	9	10	90.0%	42.9%
11	5	no	9	13	69.2%	25.0%
11	5	yes	4	13	30.8%	37.5%
11	6	no	4	13	30.8%	25.0%
11	6	yes	9	13	69.2%	37.5%
11	7	no	3	10	30.0%	25.0%
11	7	yes	7	10	70.0%	37.5%
11	9	no	5	15	33.3%	25.0%
11	9	yes	10	15	66.7%	37.5%
11	10	no	5	10	50.0%	25.0%
11	10	yes	5	10	50.0%	37.5%

4.2.2 Wald-Wolfowitz runs test

In Table 7, the observed and expected count of sequences, $N^{\zeta r}$ and $\hat{N}^{\zeta r}$ respectively, are shown for sequences, $\zeta(r)$, for which one can reject the Wald-Wolfowitz null hypothesis at a confidence level of $\alpha = 0.05$, for $r \in [8, 9, 10]$. (In addition to these 13, there are 12 additional sequences for $r = 11$.) The relative frequency of the expected proportions are based on 10,000 simulations using the constant round probability $\phi_2 = 0.575$. A test for difference between the actual and expected proportions of all sequences indicates that the null hypothesis—that the two proportions are equal—cannot be rejected for any of the 25 significant sequences.

Table 7: Given the length of the series (r rounds), the sequence of round wins by the series winner $\zeta(r)$, where round wins r^w and losses r^- are indicated by 1 and 0 respectively, among sequences that reject the Wald-Wolfowitz null hypothesis for $r \in [8, 9, 10]$. The observed count, $N^{\zeta r}$, and proportion, $P^{\zeta r}$, of all possible sequences, as well as the expected count, $\hat{N}^{\zeta r}$, and proportion, $\hat{P}^{\zeta r}$. $\hat{N}^{\zeta r}$ is scaled to the observed number of series played, hence its non-integer value.

r	$\zeta(r)$	$N^{\zeta r}$	$P^{\zeta r}$	$\hat{N}^{\zeta r}$	$\hat{P}^{\zeta r}$
8	0-0-1-1-1-1-1-1	4	0.47%	7.67	0.90%
9	0-0-0-1-1-1-1-1-1	4	0.47%	3.75	0.44%
10	1-1-1-1-0-0-0-0-1-1	1	0.12%	2.39	0.28%
10	1-0-0-0-0-1-1-1-1-1	0	0.00%	2.04	0.24%
10	1-0-1-0-1-1-0-1-0-1	3	0.35%	2.04	0.24%
10	1-1-1-0-0-0-0-1-1-1	2	0.23%	2.04	0.24%
10	0-0-0-0-1-1-1-1-1-1	2	0.23%	1.79	0.21%
10	1-0-1-0-1-0-1-1-0-1	3	0.35%	1.70	0.20%
10	1-0-1-1-0-1-0-1-0-1	0	0.00%	1.70	0.20%
10	1-1-0-0-0-0-1-1-1-1	1	0.12%	1.19	0.14%
10	1-0-1-0-1-0-1-0-1-1	3	0.35%	1.11	0.13%

9

| 10 | 1-1-0-1-0-1-0-1-0-1 | 0 | 0.00% | 1.11 | 0.13% |
| 10 | 1-1-1-1-1-0-0-0-0-1 | 2 | 0.23% | 1.02 | 0.12% |

5 Discussion

For the sake of brevity, we did not delve into win rates and streaks split by offensive or defensive role. For example, it would be interesting to look at whether the round win rate is higher after a streaks of three round wins when the team starts the streak winning as the offensive team. This would mean that, in the round immediately following the streak, the streaking team would be playing defense, where teams are slightly more likely to win on on average. On the other hand, streaks of three starting with an offensive round win—theoretically, given $\tau_O = 47.7\%$ the expected frequency is $\tau_O^2 \times (1 - \tau_O) \approx 11.9\%$—are likely to occur less frequently than such streaks starting with a defensive win—$\tau_O \times (1 - \tau_O)^2 \approx 13.0\%$—so We would want to account for sample size differences.

Further regarding future work, We've reached out to the Twitter user "R11stats", who privately tracks in-round player engagements.[26] R11stats expressed intent on making the data public, which would allow for research into player-specific momentum.

Other talking points

- We probably should have expected no hot hand effect given the small "skill gap" in the CoD relative to other esports (andecdotal, but most professional esports player, including CoD players themselves, would say this).

References

Ayton, Peter, and Ilan Fischer. 2004. "The Hot Hand Fallacy and the Gambler's Fallacy: Two Faces of Subjective Randomness?" *Memory & Cognition* 32 (8): 13691378.

Benjamini, Yoav, and Daniel Yekutieli. 2001. "The Control of the False Discovery Rate in Multiple Testing Under Dependency." *Annals of Statistics*, 11651188.

Carlson, Kurt A., and Suzanne B. Shu. 2007. "The Rule of Three: How the Third Event Signals the Emergence of a Streak." *Organizational Behavior and Human Decision Processes* 104 (1): 113121.

Chance, Don. 2020. "Conditional Probability and the Length of a Championship Series in Baseball, Basketball, and Hockey." *Journal of Sports Analytics* 6 (2): 111–27. `https://doi.org/10.3233/JSA-200422`.

"Competitive Settings | Call of Duty League." n.d. `https://callofdutyleague.com/en-us/competitive-settings?tab=map-veto`.

DeRover, DeMars. 2021. "Round Win Probabilities Based on Who's Alive + Time." `https://www.reddit.com/r/VALORANT/comments/n3lpoo/round_win_probabilities_based_on_whos_alive_time/`.

Gilovich, Thomas, Robert Vallone, and Amos Tversky. 1985. "The Hot Hand in Basketball: On the Misperception of Random Sequences." *Cognitive Psychology* 17 (3): 295314.

Miller, Joshua B., and Adam Sanjurjo. 2018. "Surprised by the Hot Hand Fallacy? A Truth in the Law of Small Numbers." *Econometrica* 86 (6): 20192047.

Mosteller, Frederick. 1952. "The World Series Competition." *Journal of the American Statistical Association* 47 (259): 355380.

Peel, Leto, and Aaron Clauset. 2015. "Predicting Sports Scoring Dynamics with Restoration and Anti-Persistence." In, 339348. IEEE.

Steeger, Gregory M., Johnathon L. Dulin, and Gerardo O. Gonzalez. 2021. "Winning and Losing Streaks in the National Hockey League: Are Teams Experiencing Momentum or Are Games a Sequence of Random Events?" *Journal of Quantitative Analysis in Sports* 17 (3): 155170.

Vafa, Keyon. 2017. "Is the Hot Hand Fallacy a Fallacy?" `https://keyonvafa.github.io/hot-hand/`.

Xenopoulos, Peter, William Robert Freeman, and Claudio Silva. 2022. "Analyzing the Differences Between Professional and Amateur Esports Through Win Probability." In, 34183427.

---

[26] `https://twitter.com/R11stats`