Discovering the effect of hero choice on the outcome of a Dota 2 game

Noyan Eren Toksoy
Supervisor(s): Jesse Krijthe, Rickard Karlsson
EEMCS, Delft University of Technology, The Netherlands

June 19, 2022

A Dissertation Submitted to EEMCS faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering

Abstract

Dota 2 is one of the most popular MOBA (Multiplayer Online Battle Arena) games being played today. A Dota 2 match is played by two teams of 5 players. The main goal of the game is to destroy the opposing team's Ancient tower, the team that manages to do so, wins the game. An essential part of a match is the hero selection phase before it starts. There are different ways to select heroes in different game modes, but the game mode that is used for this research is the Captain's mode where each team is assigned a captain and the captains take turns picking and banning heroes. The main question that this research aims to answer is: "What is the effect of the Pudge hero being picked in a team on the outcome of a Dota 2 game?" Causal inference, a discipline that is concerned with discovering causal relationships using data analysis under certain assumptions about the data is used in this research to measure this effect. More specifically, the g-formula is the causal inference method of choice for this research. The data that is used for this research is gathered through the OpenDota API. After running the correctly formatted data through the g-formula implementation, the effect of the Pudge hero being picked in a team on the game outcome was estimated as -0.2848%. Meaning that, on average, there is a 0.2848% less chance of a team winning the game, if the Pudge hero is in that team.

1 Introduction

Dota 2 is one of the most popular MOBA (Multiplayer online battle arena) games being played today [1]. A Dota 2 match is played between two teams of 5 players, namely the Dire and the Radiant. The teams are formed through hero selection from a pool of 123 unique heroes [2]. After each player has picked a hero, the game starts. The main goal of the game is to destroy the other team's Ancient tower, which is the main tower located at each team's base. The team that destroys the opposing team's Ancient tower wins the game. Besides the casual players of the game, there is a vast community of professional players who play this game competitively and millions of dollars can be at stake. Naturally, there exists a significant effort around the community to make certain predictions on the critical processes of the game, such as its outcome, whether the Radiant or the Dire team is more likely to win, using public data on the game. One example of such research is the work of Yang, Qin and Lei, where a real-time match result predictor was implemented using player data collected prior to the match, as well as in-game events [3]. Logistic regression was used as their prediction model to predict the game outcome.

Causal inference is a discipline that is concerned with discovering causal relationships between variables using data and its analysis to estimate the effect of one variable, commonly called the treatment, on another, usually the outcome [4]. This analysis is conducted under certain assumptions on the data used, which will be discussed in more detail in section 2.2. The process of predicting the outcome of a Dota 2 game is a complicated task because of the nature of the game and the number of factors involved. This research is not focused on predicting the outcome of games, however the aforementioned factors also play an important role when the causal effect of a hero choice on the game outcome is investigated. As an example, consider MMR (Matchmaking rating), a measure of performance of a player. This is a factor that can affect both the selection of the Pudge hero, the treatment, and the outcome. Because for example, Pudge is one of the rather beginner-friendly heroes, which means that the MMR can affect its selection. It can also affect the game outcome because beginners with low MMR are more likely to make mistakes and concede a loss in the game. It is not a trivial task to dissociate such factors when estimating the effect of the treatment. These factors are classified as the confounding factors in causal inference literature [5]. These factors are known to affect both the control/treatment groups and the outcome. To briefly explain the concept of control and treatment groups, consider a simple example where drug A's effect on curing illness X is investigated. To design an experiment to measure the treatment effect, a population would be randomly assigned to two groups: control and treatment, without the participants being aware of their group. The treatment group would be assigned to take drug A, which is also called the intervention, where the control group would be given a placebo. This way the effectiveness of drug A can be measured by comparing the number of cured people in the treatment group vs. the control group. A possible confounding factor could arise when this effect can not be measured through an actual experiment but only through data analysis. An example confounding factor would then be if a person has a certain existing medical condition that makes him/her more likely to be assigned drug A, while also affecting the possibility for illness X to be cured.

The specific research question that will be addressed in this research is: "What is the effect of the Pudge hero being picked in a team on the outcome of a Dota 2 game?" The main reason behind the choice of hero for this research is the fact that Pudge is the most popular hero of all time in terms of its choice-rate and it also has a relatively high win-rate [6]. The hero choice is one of the most important phases of a Dota 2 game because each hero has unique powers and strategies that are known to work well with them. However, there are numerous confounding factors involved in this causal relationship such that, if controlled for, (e.g. only selecting games where the skill rate is above a certain number) when trying to measure the treatment effect (Pudge hero being picked in a team), can possibly lead to selection bias due to samples being excluded from the data [7]. Not controlling for these factors is also not a good solution because then there is a bias term introduced by the difference between the treated and the untreated when estimating the treatment effect (lack of exchangeability), so it is necessary to correct for them [5]. The concept of a confounding factor and mathematical notations used to represent them will be further discussed in section 2.

One way of correcting these confounding factors is using the g-formula [5], which is also the main methodology used in this research that will be further discussed in section 2 and 3. Currently, there is a lack of open-source code that uses the g-formula, therefore this method is not as popular as some of the other methods used for causal inference, despite the fact that it was discovered in the 80s [8]. An important motive of this research is to provide an open-source, transparent research to showcase the effectiveness of g-formula for the correction of confounding factors, so that accurate measurements can be made on the treatment effect. The main research question can be broken down into the following sub-questions such that when all of the answers to them are combined, the research question itself can be answered:

1. What is the causal diagram with the relevant factors?
2. What are the confounding factors that are involved in the causal relationship?
3. How can these confounding factors be corrected for when applying causal analysis?
4. How should the data be separated (Control/Treatment groups) to conduct the causal analysis?

5. Which evaluation method should be used to measure effect?

This paper is structured as follows. The causal inference methodology of choice, as well as the assumptions about the data will be introduced in section 2. Section 3 contains detailed explanations of the tasks involved with this research project such as the data gathering process and the g-formula implementation. Section 4 is dedicated to showcase the results obtained. Section 5 is a discussion on responsible research. Section 6 serves as a discussion and interpretation of the obtained results. Concluding remarks are made in section 7 and section 8 briefly discusses the limitations and the possible future work.

2 Methodology

In this section, the main causal inference method that is used to answer the research question will be discussed in detail. The motivations for its use, existing research on the methodology, advantages vs. disadvantages and also the necessary assumptions will be discussed.

2.1 G-formula

Due to the existence of the confounding factors, the effect of the treatment can not be isolated from the effect that the confounding factor introduces. This results in confounding bias that needs to be corrected for when estimating the average treatment effect [5]. To better illustrate this, mathematical notations from causal inference literature can be used. The term treatment (T) refers to the intervention whose effect is being estimated, in the case of this research, it is the selection of the Pudge hero. In causal notation it is a binary variable, $T \in \{0, 1\}$ shows whether the treatment is applied or not. Y represents the outcome but it is usually used together with the treatment variable t, in the form $Y^{t=\{0,1\}}$, where $Y^{t=0}$ refers to the outcome that would have been observed, if the treatment was not applied and similarly, $Y^{t=1}$ refers to the outcome that would have been observed, if the treatment was applied [5]. To observe the effect of the treatment, $E[Y^{t=1}] - E[Y^{t=0}]$ can be calculated, where $E[x]$ refers to the mean or the expected value of x [5]. In real-word scenarios this direct calculation is impossible because once the treatment is applied, it is not possible to reverse it, meaning that if $E[Y^{t=1}]$ is known, $E[Y^{t=0}]$ becomes counterfactual. $E[Y|T = 1] - E[Y|T = 0]$, which refers to association between the treatment and the outcome can be used to calculate the causal effect but unfortunately, in the presence of confounding factors, unconditional exchangeability does not hold, meaning that the control and the treatment groups are essentially quite different in terms of important characteristics prior to the application of the treatment due to some existing confounding factors [5]. This means that there is an additional bias term introduced by the confounding factors that needs to be corrected for. Figure 1 illustrates the visual representation of a confounding factor, inside a causal diagram. It can be seen that the confounding factor (L) is affecting both the treatment and the outcome.

Ideally, randomisation is the cure for taking care of this confounding bias [7]. However, it is not possible to conduct randomised controlled tests with observational data, which is the type of historical data that this research has access to. Luckily there are other existing methods that can correct for the confounding factors and also work with observational data. G-methods, where G stands for general, is one of the categories consisting of such methodologies like Inverse probability weighting, the parametric g-formula and g-estimation [5].

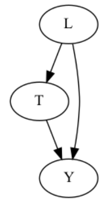


Figure 1: A simple causal diagram in which T is the treatment, Y is the outcome and L is the confounding factor. The arrows indicate the causal relationship, e.g L causes T and Y

The chosen method for this research is the parametric g-formula. A popular use case of the g-formula among causal inference methods is to correct for the time-varying confounding factors [9]. However it is not limited to it and can also be used to correct for non-time-varying confounding. The g-formula requires the following conditional estimation: $E[Y|T = t, L = l]$, this is then modelled by standardisation [10]. This allows for the Average Treatment Effect, $E[Y^{t=1}] - E[Y^{t=0}]$ to be calculated. Here is the full g-formula that will make it possible to calculate the Average Treatment Effect [5]:

$$E[Y^t] = \sum_l E[Y|T = t, L = l] \times \Pr(L = l)^1$$

The estimate is obtained by a fully-saturated parametric regression. Parametric regression was chosen over non-parametric regression because non-parametric regression is prone to overfitting the data and the curse of dimensionality [11]. For the parametric regression, an assumption has to be made that the data follows a certain probability distribution [10]. A second choice was made regarding the saturation level of the model. Saturation level refers to the number of the interactions between the treatment and the confounding factors included in the regression model [12]. As more and more interactions are included, the regression will fit the data better. As assumptions related to the relationships between the parameters are added, the model will become less saturated [13]. For this research, such assumptions were not deemed relevant and therefore a fully-saturated regression model was used.

2.2 Assumptions

There are certain assumptions that need to be made about the data and the factors involved in order to ensure that there is no bias term when the g-formula is applied [14]. These assumptions are:

- (Conditional) exchangeability: This assumption can be formulated as $Y^t \perp\!\!\!\perp T|L$. Which means that the entities belonging to the control and treatment groups are exchangeable given the confounding factors. Meaning that if the treatment was applied to the control group with the same confounding values, the outcome would have been the same. This is a critical assumption because otherwise the control and treatment groups would not be comparable. This is approximately true for this experiment because the confounding factors included captures the distinguishing information between players. Unfortunately, it is not possible to test and assure that this assumption

[1] Pr(x) refers to the probability of x

is correct with observational data, even when a lot of confounding factors are involved [5]. Therefore this assumption is deemed as approximately true for this experiment.

- Positivity: The positivity assumption suggests that the probability of being assigned to either control or the treatment group must be bigger than zero, positive. This assumption holds for this experiment because in the game modes chosen, Pudge hero is an available hero pick. Therefore there is a positive probability that it can be picked, and therefore receiving treatment. The remaining teams, where the Pudge hero was not selected, are the control group by design. This is also proven when the collected data is observed, there are teams from both groups.

- Consistency: This assumption relates to the application of the treatment. The treatment and its application needs to be well defined because otherwise, it can be applied in different ways, possibly producing different causal effects [5]. In the case of this experiment, the assumption holds because there is no variation possible in the application of the treatment, the Pudge hero can not be picked in different ways.

- No measurement error: This assumption depends on the reliability of the collected data from the OpenDota API and also the game itself when the player data is measured. This API is a service of Steam, which is one of the largest video game online markets [15]. It is also commonly used for research on the game of Dota 2 [16]. Therefore there are valid reasons to assume that there are no measurement errors in the collected data.

- No model misspecification & No unmeasured confounding: These two assumptions are closely related to each other, in the sense that no model misspecification mainly depends on the absence of unmeasured confounding. No model misspecification is the assumption that the regression model that is used is correct with all the coefficients [17]. A major misspecification could possibly occur when there are confounding factors that were not measured and adjusted for in the regression. From the data that was made available, all confounding factors identified were measured and included in the regression. Again, there is no way to guarantee that there are no unmeasured confounding factors, due to the complex nature of the game of Dota 2 and the data that is made available.

3 The effect of hero choice on the game outcome

This section is dedicated to a discussion of the main ideas of the research. the ideal experiment, accomplished tasks, important details about the causal experiment design and also the implementation of the methodology.

3.1 The ideal experiment

The ideal experiment would be mainly using the power of randomization to mitigate the effects of the confounding factors and observational data. Random Dota 2 players would be selected to participate in the experiment. The players would be randomly assigned to teams of 5, with a balanced hero role distribution, which will be explained in detail in subsection 3.3.1. Team match-ups would be randomly made and the games would be carried out with the criteria that the Pudge hero can not be selected. The results of these games would be collected. Then one player of a team from each match-up would be randomly selected to pick the Pudge hero in the hero-selection phase for the next round of games. All other players would pick the same heroes as before and use the same strategies as before, so there is no knowledge carried over from the previous round. This makes sure that there is no bias introduced because the treatment and the control group are exactly the same before the treatment is applied. The matches would be carried out again with the same match-ups but with the treatment (picking the Pudge hero) applied and the results would be collected. After this it would be possible to calculate the average treatment effect because the counterfactual cases, what would have happened if the treatment was not applied, are available and arrive at a conclusion on the effect of picking the Pudge hero on the game outcome.

Unfortunately there are a number of limitations in a real-world scenario, like this research project, that does not allow the ideal experiment to be conducted. One of these limitations is the fact that the available data for this research is historical data collected from past games. It is impossible to make sure that the same group of players and team compositions play a game twice with the same hero picks and same strategies. Even if there were games played twice with the same people and the only difference is that one hero picks Pudge, it can not be concluded with 100% confidence that $\underbrace{\{E[Y^{t=0}|T=1] - E[Y^{t=0}|T=0]\}}_{BIAS} = 0$, meaning that the control and treatment groups are not exchangeable prior to the application of the treatment. This is because Dota 2 is a game where experience is built over time, and once a game is played with a group of opponents, their strategies and the proper countermeasures to apply are learned by the other team. So when the second game starts, the only difference will not be the selection of the Pudge hero but other factors as well, such as the skill level, strategy used and the like. This ideal experiment is therefore not possible to be conducted using real players but it is still useful to keep in mind as a thought experiment and to get a better understanding of the effect that is being measured.

3.2 Forming the causal diagram

A causal diagram is a directed acyclic graph (DAG) that shows the causal relationship between the factors involved [7]. Forming this diagram allows for the categorization of these factors such as confounding, co-variate, mediator and the like. The causal diagram that was formed for this research can be seen in Figure 2, which can be used to display the confounding factors involved. Some of these confounding factors will be explained in detail in the modelling of the confounding factors subsection.

To form this causal diagram, the confounding concepts were determined through the heuristic knowledge of the game as well as exploration of the API used to gather the data. This exploration of the API also allowed some of the immeasurable concepts to be transformed into measurable ones, because the information inside the end-points of the API define the limits of the information that can be obtained throughout the project. An example of this is, game knowledge demonstrates a certain skill level in the game, which can be quantitatively measured by MMR (Matchmaking rating) which is a collective measure of game performance that almost all players have. Therefore, this measure can be included in the data. Some of these confounding factors (e.g. Team balance) can not simply be measured by a single existing field in the data-set, therefore these factors need to be modelled through combining multiple fields of information. This will be discussed in-depth in subsection 3.3.1.

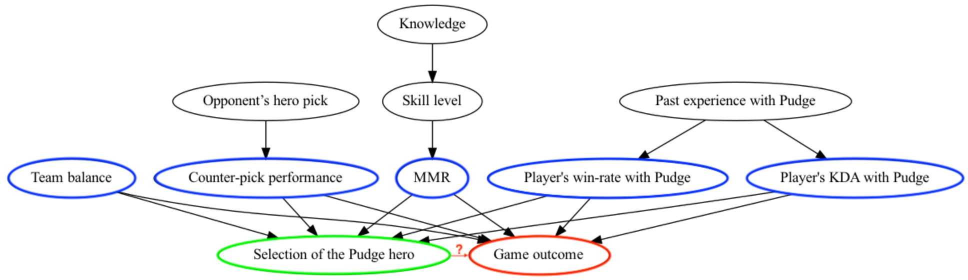


Figure 2: The causal diagram containing the relevant factors for this research (treatment (green), outcome (red) and the confounding factors (blue)). Player's win rate with Pudge confounding factor is a percentage of the number of matches each player won using Pudge divided by the total number of games with Pudge and KDA with Pudge refers to the (kills + assists)/deaths ratio of each player with Pudge.

3.3 Data gathering

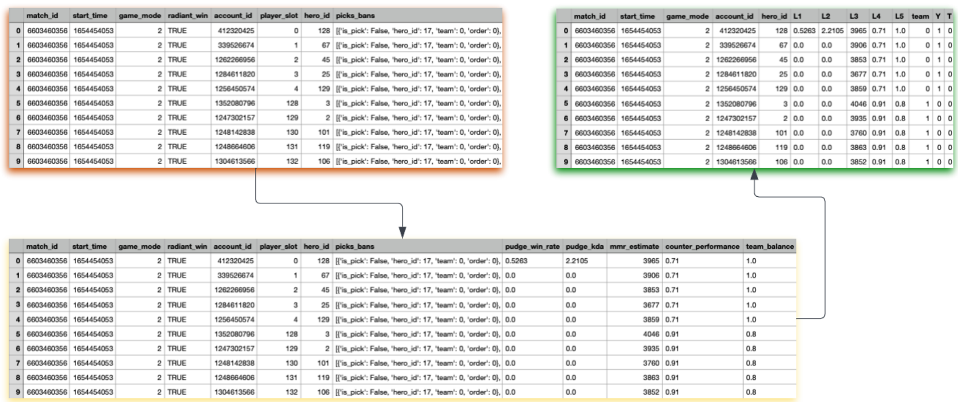
The match data necessary to apply causal inference methods were obtained using the OpenDota API [18] and Dotabuff [6]. The data was gathered through a collection of automation scripts using Python which gathered recent match information starting from the latest matches and going back in the historical data.[2] The explorer endpoint of the API was used to query the database using SQL queries. The query consisted of selecting as many relevant fields as possible to calculate the confounding factors, however some fields were not made accessible by the API for the SQL queries. Such issues encountered with the data will be further discussed in section 8. Using this method, 10100 unique matches were gathered, dating back to 16/11/2021. The majority of these games were from the Captain's mode, which is another limitation of the API. Captain's mode is a game mode where there is a captain selected for each team and the captains take turns selecting heroes for their team [2]. This game mode is often used in professional matches and tournaments and it is a good fit for this research because this allows for an important confounding factor, counter-pick performance to be estimated. After gathering these matches, additional API calls were made per each match and player to gather all the necessary data. This data was then passed to the modelling methods to model some of the more complex confounding factors and obtain a numerical value. These matches were then formatted to be fed into the g-formula algorithm. Figure 3 illustrates the step-by-step data change to arrive at the final formatted data.

3.3.1 Modelling the confounding factors

The counter-pick performance and the team balance confounding factors need to be modelled in order to represent them as numerical values in the data-set.

The counter-pick performance is about how the team responds to the opponent's hero picks throughout the game. Counter-picking is a very important strategy in a Dota 2 game because for each hero, there are heroes that are known to perform better or worse against them [6]. This makes the counter-picking strategy an essential part of the game. Each team can be assessed for their counter-picking performance by looking at the draft order of a

[2]The full implementation of the data-gathering scripts can be found at `https://github.com/noyant/research-project-dota2`



	match_id	start_time	game_mode	radiant_win	account_id	player_slot	hero_id	picks_bans
0	6603460356	1654454053	2	TRUE	412320425	0	128	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},
1	6603460356	1654454053	2	TRUE	339526674	1	67	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},
2	6603460356	1654454053	2	TRUE	1262266956	2	45	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},
3	6603460356	1654454053	2	TRUE	1284611820	3	25	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},
4	6603460356	1654454053	2	TRUE	1256450574	4	129	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},
5	6603460356	1654454053	2	TRUE	1352080796	128	3	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},
6	6603460356	1654454053	2	TRUE	1247302157	129	2	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},
7	6603460356	1654454053	2	TRUE	1248142838	130	101	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},
8	6603460356	1654454053	2	TRUE	1248664606	131	119	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},
9	6603460356	1654454053	2	TRUE	1304613566	132	106	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},

	match_id	start_time	game_mode	account_id	hero_id	L1	L2	L3	L4	L5	team	Y	T
0	6603460356	1654454053	2	412320425	128	0.5263	2.2105	3965	0.71	1.0	0	1	0
1	6603460356	1654454053	2	339526674	67	0.0	0.0	3906	0.71	1.0	0	1	0
2	6603460356	1654454053	2	1262266956	45	0.0	0.0	3853	0.71	1.0	0	1	0
3	6603460356	1654454053	2	1284611820	25	0.0	0.0	3677	0.71	1.0	0	1	0
4	6603460356	1654454053	2	1256450574	129	0.0	0.0	3859	0.71	1.0	0	1	0
5	6603460356	1654454053	2	1352080796	3	0.0	0.0	4046	0.91	0.8	1	0	0
6	6603460356	1654454053	2	1247302157	2	0.0	0.0	3935	0.91	0.8	1	0	0
7	6603460356	1654454053	2	1248142838	101	0.0	0.0	3760	0.91	0.8	1	0	0
8	6603460356	1654454053	2	1248664606	119	0.0	0.0	3863	0.91	0.8	1	0	0
9	6603460356	1654454053	2	1304613566	106	0.0	0.0	3852	0.91	0.8	1	0	0

	match_id	start_time	game_mode	radiant_win	account_id	player_slot	hero_id	picks_bans	pudge_win_rate	pudge_kda	mmr_estimate	counter_performance	team_balance
0	6603460356	1654454053	2	TRUE	412320425	0	128	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},	0.5263	2.2105	3965	0.71	1.0
1	6603460356	1654454053	2	TRUE	339526674	1	67	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},	0.0	0.0	3906	0.71	1.0
2	6603460356	1654454053	2	TRUE	1262266956	2	45	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},	0.0	0.0	3853	0.71	1.0
3	6603460356	1654454053	2	TRUE	1284611820	3	25	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},	0.0	0.0	3677	0.71	1.0
4	6603460356	1654454053	2	TRUE	1256450574	4	129	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},	0.0	0.0	3859	0.71	1.0
5	6603460356	1654454053	2	TRUE	1352080796	128	3	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},	0.0	0.0	4046	0.91	0.8
6	6603460356	1654454053	2	TRUE	1247302157	129	2	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},	0.0	0.0	3935	0.91	0.8
7	6603460356	1654454053	2	TRUE	1248142838	130	101	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},	0.0	0.0	3760	0.91	0.8
8	6603460356	1654454053	2	TRUE	1248664606	131	119	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},	0.0	0.0	3863	0.91	0.8
9	6603460356	1654454053	2	TRUE	1304613566	132	106	[{'is_pick': False, 'hero_id': 17, 'team': 0, 'order': 0},	0.0	0.0	3852	0.91	0.8

Figure 3: The flow of the collected data throughout the data gathering process. The flow is from orange shadow (top left), where raw data is collected from the API to yellow shadow (bottom) in which additional API calls are made per player to form the confounding factor columns, to green shadow (top right) where the data is re-formatted to be used for the g-formula algorithm.

game in which the heroes that are picked after the opponent's hero selection can be seen. This can then be compared to the known hero counter-picks to assess the performance of the team in terms of the hero selection against the opponent's heroes. The hero counter-pick information was obtained from Dotabuff [6]. In the website, there is information on the possible counter-picks for each hero, sorted by their "disadvantage" percentage. The disadvantage metric goes beyond just using win rate against other heroes. "It is calculated by establishing their win rates both in and outside of the match-up and comparing the difference against a base win rate." [6]. The disadvantage score of each hero against each hero (123 rows and 122 columns) is collected and stored in a JSON file to be used for calculations of the counter-pick performance. For each team, its counter-pick performance is calculated by summing up the disadvantage that is introduced to the opponent in each counter-pick until the selection of the Pudge hero, because counter-picks after Pudge is selected no longer affects the treatment. This value is then assigned to each member of that team.

Team balance refers to how balanced a team composition is, in terms of the number of carry and support roles in a team. There are a number of official roles for heroes in Dota 2, these are: Carry, Support, Nuker, Initiator, Disabler, Durable, Escape, Pusher, Jungler [2]. With the main, distinctive roles being carry and support. Carry heroes are usually ones with more offensive power and useful later on in the game, whereas the support heroes are usually concerned with helping and boosting other team members [2]. It is suggested by the existing game guides from Dota experts to have 2 supports and 3 carry role heroes for a balanced team composition [19]. Therefore, this carry and support distribution was selected to represent the ideal case of a team composition. In the "/heroes" endpoint of the API, there are a number of roles assigned to each hero. For each hero, its carry-support percentage was calculated by assigning 1.0 if the hero contained the carry role, 0.5 if it carried both carry and support or none of them, and 0.0 if the hero contained the support role. With these numerical values, the ideal team should have an average carry-support percentage of $\frac{3*1.0+2*0.0}{5} = 0.6$. For each team, its balance up to the selection of the Pudge hero was calculated by $1 - |\frac{\sum_{i=0}^{n} S_i}{n} - 0.6|$ where S_i is the function that gives the carry-support percentage of a hero. This way, a team with a balanced team composition (0.6 carry-support percentage) will have a high score of 1, whereas the less balanced ones will have lower scores. The obtained value was assigned to each player of the team.

3.4 G-formula implementation

For the implementation of the g-formula, a fully-saturated parametric regression was used to estimate the conditional probabilities which is then used by standardisation. This implementation was provided by zEpid library's TimeFixedGFormula [20]. Here is the saturated regression model that was used, where the variables L1 to L5 are the confounding factors and the ":" operator allows the possible interactions between the treatment and the confounding factors to be included:

$$Y = T + \sum_{i=1}^{5} L_i + T : L_i$$

This model was used to first fit the expected outcomes for the treated, $T = 1$, and then the same process was applied to the control group, $T = 0$. The difference between these values resulted in the Average Treatment Effect ($E[Y^{t=1}] - E[Y^{t=0}]$) in terms of a percentage increase/decrease in the win-rate. There is also another important step to estimating the causal effect which is analysing the 95% confidence interval. To calculate the confidence interval, bootstrapping technique was used to re-sample the data 2000 times. For each new sample, the Average Treatment Effect (ATE) was calculated and appended to a list. The standard error (SE) of this list was calculated to be used in the calculation of the confidence interval. Here is the formula used for finding the 95% confidence interval, where 1.96 is the corresponding z-value for a 95% confidence interval:

$$[ATE - 1.96 * SE,\ ATE + 1.96 * SE]$$

$$SE = \frac{\sigma}{\sqrt{n}},\ \text{n = Sample size}$$

4 Experimental Setup and Results

This section is dedicated to a detailed description about how the hypothetical experiment with observational data was conducted and the results obtained.

4.1 Experiment with the observational data

After the data was collected with necessary information from the confounding factors, its statistical distributions were analysed, which served both as a sanity check and also provided insight on the population of the collected data. Then, the data needed to be formatted, where the control and treatment groups were also formed, before actually calculating the Average Treatment Effect using the g-formula.

A detailed statistical summary (mean, standard deviation, min, max, value) of the obtained data can be found in Appendix A. But as a quick way to get insight into the population of the data, the distribution of the relevant columns can be seen in Figure 4. To highlight some of the interesting patterns in the data, it can be observed that the MMR estimate of the players are mostly located around a very high score of around 5000 with two peaks around 3900 and 6200. This means that most of these players are from the "Divine" rank which consists of the top 5% of all ranked players in the game of Dota 2 [21]. This is an important fact to keep in mind when analysing the results of this research because MMR, skill level, is naturally a very important factor of the game. Another interesting observation is that the counter-pick performance closely resembles a normal distribution, centred around 1.24. Lastly, it can be seen that the team balance metric is quite high for all teams involved, with the lowest value being around 0.6. One possible explanation of this distribution is the fact that the overall MMR of the population being quite high, so they would be expected to pick relatively balanced teams.

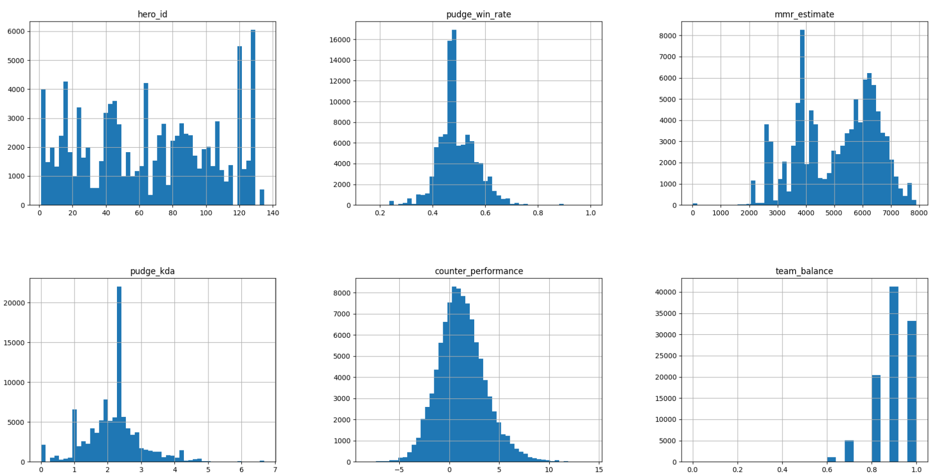


Figure 4: The histogram plots of the relevant columns from the collected data. The graph with label hero_id displays the distribution of the heroes picked through all games, ordered by the hero id on the x-axis.

The first step for formatting the data was to divide the data into control and treatment groups. The treatment group consisted of teams in which the Pudge hero was selected. The control group was the remaining teams. After this, the outcome (Y) column was formed based on the outcome of the game for each team and also the confounding columns were renamed (L1...L5). This reformatted data was fed to the g-formula algorithm to retrieve the Average Treatment Effect value.

4.2 Results

The result obtained from the g-formula algorithm, with different adjustments can be seen in Table 1. The Average Treatment Effect (ATE) was calculated as -0.2848% when all of the confounding factors were included in the regression model, as can be seen in subsection 3.3. Whereas the ATE without any adjustment for the confounding factors, meaning that the regression model only contains T, was calculated as -5.4006%. This means that, before adjusting for the confounding factors, the teams in which the Pudge hero was selected, had a 5.4006% less chance of winning the game compared to the teams that did not select Pudge. After adjusting for the confounding factors, this chance increased from -5.4006% to -0.2848%.

It can be seen in Table 2 that the combined effect of all the confounding factors is 5.1158%. This is a positive effect which caused the increase in the ATE when the confounding factors are adjusted. The individual effect of each confounding factor was also analysed to discover the factors that play a significant role. This effect was calculated by: $ATE(x) - ATE(y)$, $x =$ Adjusted for single confounding factor, $y =$ No adjustment. Most notably, MMR is the confounding factor with the highest effect of 4.5007%, whereas the counter-pick performance resulted in the least difference of 0.0057%.

	Average Treatment Effect (%)
Adjusted for all confounding factors	-0.2848
Without adjusting for the confounding factors	-5.4006

Table 1: A table with the obtained results from the causal inference method of choice for this research, g-formula.

	Confounding Effect (%)
All confounding factors	5.1158
Pudge win rate	0.8691
Pudge KDA	0.6127
MMR	4.5007
Counter-pick performance	0.0057
Team balance	0.0194

Table 2: A table with the confounding effect of each factor, as well as the combined effect of all confounding factors. The effect is calculated by subtracting the ATE of modelling each factor individually in the g-formula from the ATE without the confounding factors.

After these results were obtained, the bootstrapping method was applied with 2000 resampling iterations to obtain the confidence interval. Figure 6 illustrates the confidence interval through a probability density function of ATEs obtained from all iterations. 95% confidence interval is between -3.3936% and 2.8240%, which also contains the calculated ATE of -0.2848. The mean value of the ATEs obtained through bootstrapping is -0.3691%. The difference between this mean and the ATE is 0.0843%.

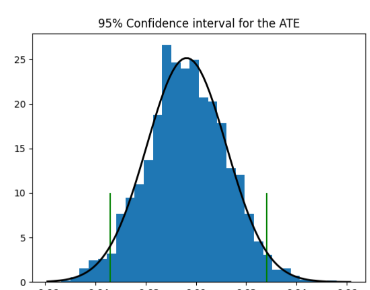


Figure 5: The probability density function of the ATE list obtained by bootstrapping 2000 times. The green vertical lines indicate the start and end of the 95% confidence interval (-0.033936, 0.028240).

5 Responsible Research

This section is dedicated to a discussion on the ethical aspects of this research as well as its reproducibility. This research's main source of data is gathered through OpenDota API, which is an open-source platform where extensive information on the game of Dota 2 and its players can be obtained. Dota 2 is a game that can be downloaded through Steam, an online video-game market [15]. The data about the game and players are collected by Steam, and then it is accessed by the OpenDota API through Steam's WebAPI. All players playing the game have the option to not share their data with Steam by disabling the "Expose Public Match Data" option in the settings [18]. Which means that the data used throughout this research is consented by holders of the data, the players. However, it is still necessary to make sure that there is no abusive usage of the data, that could violate the user's privacy, because not all users may be aware of the option to disable data sharing. The data collected throughout this research was limited to the relevant factors of the game mentioned throughout the report, which do not contain any personal information that is subject to abuse or misuse.

Another important aspect of a research is its reproducibility. A reproducible research would mean that the validity of the research can be verified by other researchers and that it can also be extended for a larger research or a variation of the existing research. Throughout this research, special attention was paid to the reproducibility aspects. The data gathering process is explained in detail in subsection 3.2 and also in the code repository. This is also the case for other aspects of the research like the data analysis techniques used for the g-formula implementation. The code repository is well documented, containing most of these details that can allow for this research to be reproduced. As long as the open-source OpenDota API is operational this research can be reproduced and/or extended.

6 Discussion

When analysing the Average Treatment Effect obtained from the g-formula algorithm, it is important to check its potential impact on the game, e.g. what would it mean for a hero to have a 5% less chance of winning. Because the percentage of effect is very much dependent on the nature of the research, where small numbers can still have significant implications. In the case of this research and Dota 2, when the hero stats for the last 12 months are analysed, it can be seen that the heroes' win rate percentages range from 43.33% to 53.70 [6]. This means that quite small numbers can still make an impact when the effect of a hero selection on the game outcome is analysed.

The Average Treatment Effect obtained before correcting for the confounding factors was -5.4006%, which is quite a significant effect because having almost 5.4% less chance of winning the game with a certain hero would be a significant disadvantage. Considering such a change would move the hero with the highest win rate to a lower mid position in the rankings. After adjusting for the confounding factors, this effect got much smaller in magnitude to just -0.2848%. From the analysis, it can be observed that the confounding factor with the highest effect is MMR. One possible explanation of this behaviour could be that the players with relatively low MMR might have selected the Pudge hero and then proceeded to lose their games, which would explain the significantly bigger negative effect prior to correction. Pudge is also the most popular hero of all-time in the game, so it can also be seen as one of the beginner-friendly heroes in the game, which could make it more desirable for the players with relatively low MMR.

Perhaps the most critical criteria for arriving at definite conclusions in the area of causal inference and observational studies is making sure that the assumptions made about the data hold, or are approximately true [5]. The results obtained through causal inference are as valid as the assumptions that are made about the data [5]. In section 2 of this paper, efforts were made towards reasoning about why the necessary assumptions for the causal inference method of choice holds for this research. However, as also stated in the same section, for some of the assumptions, like no model misspecification and no unmeasured confounding, it is not possible to test whether it holds with 100% confidence. All that can be done is reasoning as much as possible to show that it approximately holds. This is also a common problem in the field of causal inference, when a certain method requires assumptions that are not empirically testable [5].

7 Conclusion

The main objective of this research is to discover the effect of the Pudge hero being picked in a team on the game outcome for that team in the game of Dota 2. There were 5 sub-questions that needed to be answered such that the answers to those questions would allow the research question itself to be answered as well. The answers to these questions were presented throughout this paper in their relevant sections.

To discover this effect and have precise measurements for it, a causal inference technique known as the g-formula was used. Causal inference is about discovering causal relationships between variables using data analysis. In the case of this research, the variables were the selection of the Pudge hero, the treatment, and the game outcome. A common methodology used to discover the treatment effect is by calculating the Average Treatment Effect. This is possible thanks to the g-formula implementation, which estimates the counterfactual outcomes that can not be known otherwise, e.g. what would have happened if a team had not selected the Pudge hero. This way, a numerical comparison can be made between the actual outcome and the counterfactual ones to arrive at an estimated effect of the treatment. For the causal scenario of this research, the effect of the Pudge hero being picked in a team on the game outcome for that team was estimated to be -0.2848%. This is a relatively small effect that would not necessarily result in a significant advantage or disadvantage on the game but an interesting observation made from the results obtained was that this effect was -5.4006% prior to the correction of the confounding factors[3], which is a significant effect that would suggest that the hero is possibly disadvantageous.

In conclusion, g-formula is a causal inference technique that can be used to correct for the confounding factors when trying to estimate the treatment effect. The measured effect for the causal scenario of this research was -0.2848%. This measure is as accurate as the validity of the assumptions made about the data, which can be found in section 2.

[3]Factors known to affect both the treatment and the outcome. They can be corrected for using the g-formula. More information can be found in the Introduction section.

8 Limitations and Future Work

Throughout this research, there were a number of important limitations encountered and there is certainly room for improvement in possible future iterations. The majority of these limitations are from the data gathering process, related to the OpenDota API that is used. It is quite difficult to retrieve a large number of matches played at once through the API because for most of the endpoints, there is a limit of maximum 100 results per call. In the initial efforts to collect data for this research, an endpoint was used to get matches less than a certain match_id, while lowering the match_id in each iteration to get new games each time. This method did work very nicely and resulted in a dataset of 300k games that only went back 5 days in history. However, upon close inspection of the data, a lot of the matches did not contain all the information about the players. In fact, there were only about 300 games with all player information out of the 300k dataset. This forced a change to the method of gathering data, because this was not feasible considering the huge number of API calls being made, 10 per game to check account information so 3 million calls just to get 300 games. A possible solution was to use another endpoint, that allowed the user to query the database with SQL queries. This method did work and it resulted in the final dataset used for the research but a compromise had to be made in terms of the number of games collected and their recency in order to assure accuracy.

In a possible future iteration or an extension of this research, the aforementioned problems about the data gathering step can be solved by using a different gathering method, proposing a change to the OpenDota API or having access to an API key with unlimited calls. This would allow the analysis to be conducted on a much larger, more recent dataset, meaning that the results can be more reliable. Another possible extension, maybe also the most interesting one, would be to extend this work to other heroes and estimate their effect on the game outcome. This could provide more insight on whether these heroes are advantageous/disadvantageous by design or are there other factors that have a bigger effect on the outcome. If this analysis is conducted on a big enough number of heroes, a generalization can even be made on the hero selection in the game of Dota 2. A possible research question could be: "What is the effect of hero selection on the outcome of a Dota 2 game?"

References

[1] B. Wirtz, "20 best MOBA games right now (for 2022 and beyond)," https://www.gamedesigning.org/gaming/moba/, Mar. 2019, accessed: 2022-5-7.

[2] "Dota 2," https://dota2.fandom.com/wiki/Dota_2, accessed: 2022-6-8.

[3] Y. Yang, "G-computation in causal inference," https://towardsdatascience.com/g-computation-in-causal-inference-774099da3631, Jun. 2019, accessed: 2022-5-7.

[4] J. Hill and E. A. Stuart, "Causal inference: Overview," in *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier, 2015, pp. 255–260.

[5] M. A. Hernán and J. M. Robins, *Causal Inference: What If.* Chapman & Hall/CRC, 2020.

[6] "Heroes - highest win rate, all time," https://www.dotabuff.com/heroes/winning?date=all, accessed: 2022-5-7.

[7] "Causal inference for the brave and true — causal inference for the brave and true," https://matheusfacure.github.io/python-causality-handbook/landing-page.html, accessed: 2022-5-30.

[8] S. McGrath, V. Lin, Z. Zhang, L. C. Petito, R. W. Logan, M. A. Hernán, and J. G. Young, "GfoRmula: An R package for estimating the effects of sustained treatment strategies via the parametric g-formula," *Patterns (N Y)*, vol. 1, no. 3, p. 100008, Jun. 2020.

[9] A. I. Naimi, S. R. Cole, and E. H. Kennedy, "An introduction to G methods," *Int. J. Epidemiol.*, p. dyw323, Dec. 2016.

[10] M. J. Smith, M. A. Mansournia, C. Maringe, P. N. Zivich, S. R. Cole, C. Leyrat, A. Belot, B. Rachet, and M. A. Luque-Fernandez, "Introduction to computational causal inference using reproducible stata, r, and python code: A tutorial," *Stat. Med.*, vol. 41, no. 2, pp. 407–432, Jan. 2022.

[11] C. Ayuya, "Parametric versus non-parametric models," https://www.section.io/engineering-education/parametric-vs-nonparametric/, accessed: 2022-5-30.

[12] M. Blackwell, "Psc 504: Regression." [Online]. Available: https://www.mattblackwell.org/files/teaching/s09-regression.pdf

[13] H. Anders, "Causal inference sequence part ii: Graphical models." [Online]. Available: https://www.lesswrong.com/posts/e8Refdq74jJ7Lf3ao/causal-inference-sequence-part-ii-graphical-models

[14] S. L. Taubman, J. M. Robins, M. A. Mittleman, and M. A. Hernán, "Intervening on risk factors for coronary heart disease: an application of the parametric g-formula," *Int. J. Epidemiol.*, vol. 38, no. 6, pp. 1599–1611, Dec. 2009.

[15] Valve Corporation, "Steam," 2022, last accessed june 2022. [Online]. Available: https://store.steampowered.com/

[16] I. Porokhnenko, P. Polezhaev, and A. Shukhman, "Machine learning approaches to choose heroes in dota 2," in *2019 24th Conference of Open Innovations Association (FRUCT)*. IEEE, Apr. 2019.

[17] Stephanie, "Model misspecification," https://www.statisticshowto.com/model-misspecification/, Jan. 2017, accessed: 2022-6-13.

[18] "OpenDota - dota 2 statistics," https://www.opendota.com/, accessed: 2022-5-30.

[19] "Steam community :: Guide :: Basics of team composition and tactics," https://steamcommunity.com/sharedfiles/filedetails/?id=2395798524, accessed: 2022-6-8.

[20] "zepid.causal.gformula.TimeFixed.TimeFixedGFormula — zepid documentation," https://zepid.readthedocs.io/en/latest/Reference/generated/zepid.causal.gformula.TimeFixed.TimeFixedGFormula.html, accessed: 2022-6-14.

[21] V. Milella, "Dota seasonal rank distribution and medals - april 2022," Apr 2022. [Online]. Available: https://www.esportstales.com/dota-2/seasonal-rank-distribution-and-mmr-medals

A Statistical summary of the collected data

data_merged_captains_final_summary

	pudge_win_rate	pudge_kda	mmr_estimate	counter_performance	team_balance	T
count	101000.0	101000.0	101000.0	101000.0	101000.0	101000.0
mean	0.4930316227722773	2.167269850495049	5101.459752475247	1.242322277227723	0.8991891089108911	0.019158415841584158
std	0.07241883282746849	0.8483177768006291	1383.4884706005234	2.3840703018097202	0.09157434754085861	0.13708230009393949
min	0.15	0.0	4.0	-8.26	0.0	0.0
%25	0.4506	1.7188	3865.0	-0.33	0.8	0.0
%50	0.4787	2.2775	5387.0	1.11	0.9	0.0
%75	0.5378	2.5137	6239.0	2.66	1.0	0.0
max	1.0	6.6842	7902.0	14.17	1.0	1.0

Figure A.1: A table with a statistical summary of the data. Only the columns for which a statistical analysis is useful are included in the table. The "T" column refers to the binary variable of treatment assignment.