

How HS Replay Deals with their Meta Analysis

Yesterday I had a discussion with a group of HS players on the differences between using “tracker” side data vs using “opponent” side data. We concluded that they could lead to different results. For those interested, the file is here:

<https://docs.google.com/document/d/1ihc-pu1iOqpPEnfqsf-T2TJuohI0BSVU3XlqKZS1wOI/edit?usp=sharing>

Two aspects were clear:

Aspect 1. HS replay uses “Tracker” win rates in their “deck” tab, but seems to be pooling that with “opponent” win rates in their Meta Analysis. We also concluded that they do not explain how they do this.

Aspect 2. VS uses a simple average of “opponent” and “tracker” win rates. This is documented in their FAQ.

Personal Claim 1: using “tracker” win rates exclusively is crazy if what you care about is trying to represent potential win rates for the “average” player (or for a random player in the distribution of skills). But this is part of the other file.

It is obvious that, in order to use “opponent” win rates, you need an algorithm to determine what “archetype” the opponent is playing. In some cases, it could be super clear: the opponent plays rogue and on turn 1 plays the quest. Yeah! “Quest Rogue”. In other cases, it is hard to tell unless several key cards are played. In some games, it may be impossible to tell two archetypes apart. That’s a limitation we have to accept and live with.

Personal Claim 2. VS deals with this aspect a lot better than HS replay does.

Let’s analyze my claim in the context of warlock. There are two archetype of warlock that have several cards in common: Control Warlock and Cube Warlock. However, their respective matchups are quite different. Let’s first look at what HS replay does (from their META TAB).

 Warlock			
Archetype ▲	Winrate ▲▼	Popularity ▲▼	Games ▲▼
 Control Warlock	49.9%	1.3%	200,000
 Cube Warlock	56.5%	8.9%	1,400,000
 Even Warlock	53.7%	0.4%	60,000
 Zoo Warlock	53.3%	1.1%	180,000
 Other	49.2%	3.8%	600,000

What do we see? First, lots of player playing Cube. It makes sense as it seems to be a ``high`` win rate deck. Some people playing Zoo too. But wait! 600,000 are coming from “Other” Warlock!..... That is, the second most popular Warlock deck in HS Replay is an archetype that.....yes, you have never seen on ladder.

Here is the problem. HS replay attempts to distinguish between Control Warlock and Cube Warlock. But oftentimes, you have games that go like this for the Warlock: Turn 1: Kobol, Turn 2: tap, Turn 3: tap, Turn 4: Hellfire, Turn 5: Lackey, Turn 6: (Lackey was silenced and you are dead) concede.

If you ever played Warlock, you will know that many games go like that. Now, if that is ALL you get to see, which archetype was that? Was it Cube lock? Was it Control Warlock? Well, we have no idea because both decks have all those cards in common. Since we can’t tell, let’s just call them “Other Warlock”. That doesn’t look problematic....except, it is because we are not doing this ``at random``. What we just did was to massively remove games that Control Warlock and Cube Lock commonly lose and tossed them into “Other Warlock”. If we only remove games like that, then that only inflates the win rate for Cube and Control. (You could also imagine situations where turn 5 is lackey coin dark pack into voidlord and your opponent conceding which will lead to the opposing bias).

VS has consistently refused to separate Cube from Control Warlock and reports combined win rates (this, in turn, is not ideal given that we really want to see these two archetypes as “different”, but at least they acknowledge and understand the problem).

Now, some players will say ``but HS replay has much more data``. No questions about that. But if the statistical analysis is done improperly, the conclusions you may draw may be quite quite poor. This common misconception is in vogue these days with “Big Data” everywhere. Well, Big Data doesn’t solve “data” issues unfortunately; something that many people fail to understand.

Warlock is not the only class that suffers from the problem above but is provides the simplest illustration. You could look into other classes. Take Paladin for example.

 Paladin			
Archetype ▲	Winrate ▲▼	Popularity ▲▼	Games ▲▼
 Aggro Paladin	53.8%	0.3%	41,000
 Even Paladin	59.8%	3.2%	510,000
 Murloc Paladin	59.0%	0.3%	39,000
 Odd Paladin	58.8%	8.2%	1,300,000
 Secret Paladin	55.8%	0.6%	100,000
 Other	48.2%	5.1%	810,000

Again, how could it be that 5% of the players on ladder are playing “Other Paladin”, when even, odd, secret, murloc, and aggro are already accounted for!!?? If you play enough HS, ask yourself: When was the last time you run into a Paladin that did not belong into any of the above five archetypes?

Remark 1: The cases of Paladin and Warlock are conceptually different. In the case of Warlock we should convince ourselves that there are many games where it’s impossible to tell apart Control from Cube. That’s a limitation of using “opponent” information. In the case of Paladin, I can only conclude that the algorithm used by HS Replay is not doing a great job as Paladin Archetypes are not that hard to segregate. Two **very** different issues leading to similar problems.

Summary: HS replay's recognition algorithm dumps around 20% of the games into the trash introducing bias for the archetypes they do report. These games that are dumped do not count for matchups, yet they are *not random occurrences*. In Statistics and Econometrics this is called “selection” and when you have “selection”, even if you had infinite data, you do not consistently estimate the parameters you care about.

Disclaimer: the goal of this note is not to complain about HS Replay or VS. They provide valuable resources for the HS community. The goal is to help players understand the differences and start a discussion that, eventually, may lead to “better” data analysis.