# Football injuries in the bundesliga
# More sick days through Covid?

**Lyndon Sehne**
Matrikelnummer 3922334
lyndon.sehne@student.uni-tuebingen.de

## Abstract

Although VfB Fans strongly believe they are only so bad because of their injured players, an analysis of injuries in the Bundesliga cannot be directly linked to success or failure. Some evidence was found to imply Covid19 might cause longer term injuries, but the increase in match density did not increase the total number of injuries. Further research into Long Covid and Covids (mental) health implications for athletes is needed. All code and data for this project can be found on Github.

### 1 Data Overview

#### 1.1 Fussballverletzungen.com

Fabian Siegel who I found on Twitter compiled a data set of all injuries in the Bundesliga since 2009. Apparently he has been collecting this data himself and uses it to post graphs about different aspects concerning this data.
I mainly used the length of injuries, neglecting other factors such as the type of injury. Sadly there was no factor for age.
A problem with my analysis I would not like to overlook with other data is that injuries often occurred beyond just one season. For practical reasons I dismissed this, but it clearly violates the premise of my assumption that the data between the seasons is independent, and also the possibility arises that injury days could be counted several times (in different seasons).

#### 1.2 openligadb.de

As mentioned in a tutorial openligadb (an open source Bundesliga statistics site) provided the basis for combining my main data set to the seasons league results.
To get a more coherent data set I only analyzed the 10 teams that have remained in the Bundesliga since 2014.

### 2 Analysis

#### 2.1 Injuries during and before Covid19

I would have liked to conduct a one sided t-test to test if Covid had increased injuries. For this I would have tested both for the average length of an injury and for the amount of injuries per season (comparing the averaged seasons 10/11, ..., 18/19 with season 20/21, as 19/20 was when Covid started and 21/22 is still ongoing). This seemed very promising when I first compared season 20/21 to the averaged seasons. But then I looked at the numbers for the other seasons (Figure 1, season abbreviated to first year, ie season 10/11 is called season 10).

Figure 1: Supposed injury data per season

| Season | mean injury length | sum days |
|---|---|---|
| 10 | 14.62 | 541 |
| 11 | 10.75 | 344 |
| 12 | 13.16 | 329 |
| 13 | 17.62 | 969 |
| 14 | 20.84 | 32884 |
| 15 | 18.41 | 27014 |
| 16 | 17.74 | 29385 |
| 17 | 16.88 | 27872 |
| 18 | 19.79 | 33948 |
| 19 | 20.21 | 30685 |
| 20 | 20.79 | 27344 |
| 21 | 13.73 | 11177 |

The obvious problem is the discrepancy of the summed up injury days before and after season 14/15, as before they are under 1000 days and after around 30000. After this insight it is easily visible that the summed days for season 20/21 are not increased, compared to 14/15 - 18/19.
As for the injury length: conducting a one-tailed Welch 2-sample t-test showed a significantly higher injury length for the season 20/21 compared to the seasons 14/15 - 18/19. To visualize the mean injury length I created a box plot of all used seasons and the log injury time (Figure 2). Without using the log the box plots would be squished and thus not useful. For context the highest outlier is at 721 days. Due to the high outliers the mean (diamond in Figure 2) is always above the median of the box plot.

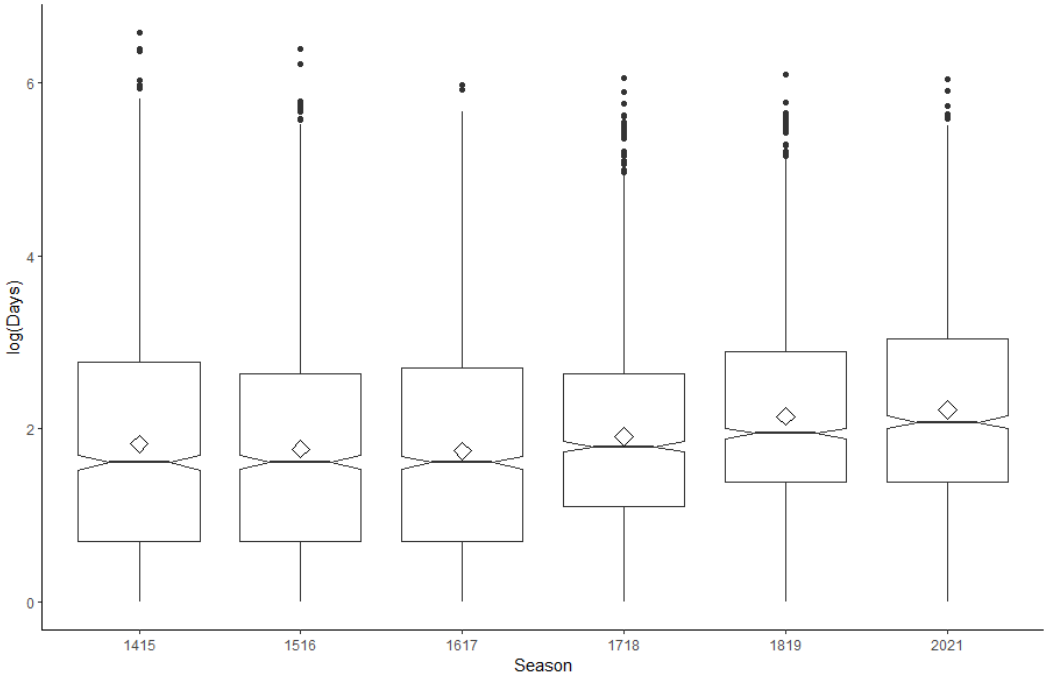


Figure 2: Box plots per season of the injury length in log(days).

### 2.1 Injuries and final points

Following the tip of looking into the correlation between the success of the teams and the magnitude of injury time, I found no connection between the two. However, it is very likely this is due to a missing factor in the analysis: team size. Without this it is not possible to calculate a base rate of injury impact. Thus, Figure 3 cannot deliver any insight other that there is no connection to be made with just these factors. Sadly there is no easily available source for team size, especially as it can fluctuate throughout the season (transfers, youth team players).

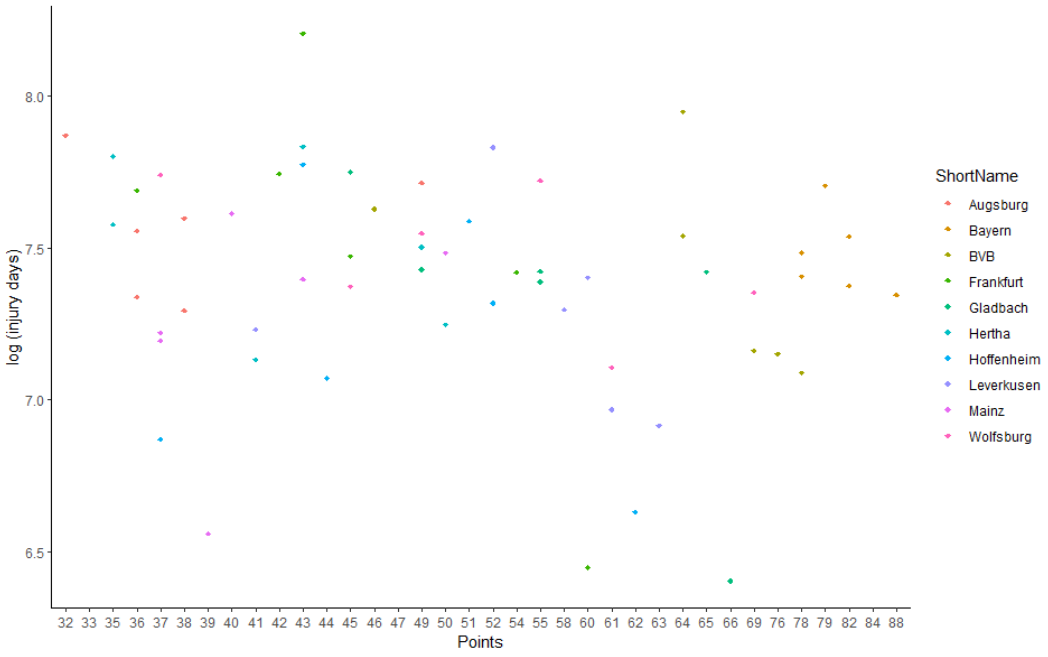


Figure 3: Scatter plot of points per season by injury length in log(days).

## 3 Discussion

Football fans love blaming bad luck, unfair referees and bad physical training coaches for any misfortune that befalls their club while not paying any attention to these matters in successful times. Despite far more exploratory analysis conducted beyond what I documented here (especially looking at VfB Stuttgart), it was not possible draw many conclusions from the used data. It is good to remember humans are amazing pattern recognizers, especially when motivated, and that this does not mean this is always true or provable.

The main motivation for looking into higher injury rates during Covid19 was because more games were being played in a shorter amount of time. This could not be demonstrated by the data. An increase in injury length for the 20/21 season was significant, however the conclusions from this are very limited (long covid?) and would need better research with more and more reliable data (La Liga, Premier League, Serie A, ...).

Also this whole data analysis would have greatly benefited from knowing the team sizes. An injury rate would have far greater weight on causality than summed up injury days.