

## ReadMe File

### Materials included

#### Replication code final.R

This code is used to generate all results in the paper including those found in the online appendix. All sections can be easily collapsed using the side-bar to provide an overview of all tables and figures in the paper. After running the “preamble” and “data edits” sections, each chunk of code can be run independently to obtain the results for each Table or Figure respectively.

#### annonymized.Rdata

This is the final data set used to generate the all the results in the paper using “Replication code final.R”.

### Data collection and compilation

#### www.nhl.com (2010-2019)

This data contains the player-game statistics such as the name of the player, penalty minutes obtained, points scored as well as the date of the game along with team names and which team was the home team. Collecting the data from the NHL for seasons 2010-2019 is a bit easier. Here, we simply use the Rvest and XML2 packages in R and collected all “event summaries” for all games, which contain all variables listed above. We use a paste function as above to loop over the games.

```
paste("http://www.nhl.com/scores/htmlreports/", i, "/ES02", gameid)
```

Here *i* refers to the season going from 20102011 to 20182019 and *gameid* is simply a number going from 1 to a maximum of 1271 (max number of games in a season). The gameid must be formatted such that the number of characters is always equal to four. This is achieved by pasting the required number of zeroes before the number in the sequence (i.e. 1 becomes 0001).

#### Injury data

The injury data was purchased from ManGamesLost and subsequently cleaned and merged with the NHL data.

### Data Dictionary

Here we list and explain all the variables in the anonymized.Rdata data file. The unit of observation is a player during a game with their team during a given season.

injid	Unique identifier for each injury
peer	Anonymized identifier for the injured peer

peerpos	Position of peer (1 = offense, 2 = defense, 3 = goalie)
season	Annonymized season identifier
teamname	Annonymized team identifier
playername	Annonymized player identifier
playerpos	Position of player (1 = offense, 2 = defense, 3 = goalie)
gamenum	Unique game identifier (per team-season)
toi	Time on ice of player
goalcount	Goals scored in game by given player
assistcount	Assists scored by player in game
pointcount	Points scored in game by given player (goals + assists)
shotcount	Shots taken by player in game
injdistrib	addgame previous addgame of team
injdistrib	addgame next addgame of team
addgame	First game in which peer was injured and did not play as a result
injdur	Duration of injury in games missed
endgame	Game in which injured player is available to play again
distprevret	addgame previous endgame of team
distnextret	addgame next endgame of team
time	gamenum - addgame
timecat	Categorical variable for whether the game is before, during or after an injury period