Hypotheses for Research Questions

1) Does simply being deemed as **high risk** on the soft tissue watch list **predict HSI incidence**? Even though being deemed high risk is based upon many known risk factors, we believe simply being deemed as high risk on the soft tissue watch list is not completely reliable to predict HSI incidence. It can definitely be a strong predictor of HSI incidents. As mentioned in much research, there are not just controllable but also uncontrollable factors at play, making it more difficult to predict HSI incidence. We believe that being deemed high risk, along with other factors, can significantly contribute to predicting HSI incidence.

Also wary of a high false positive rate in the high-risk category. We believe, on the big picture, with a focus on validating all risk categories, low-high is more beneficial than honing in on purely the high risk category for better overall and individual predictability.

2) If so, how do the components of the **watch list predict subsequent HSI injury risk**?
A history of previous ACL or HSI injuries, a greater percentage distance away from the positional norm in the weaker direction, a greater percent strength imbalance, and strength trend towards weaker will all indicate a higher subsequent HSI injury risk.

3) If the components of the **watch list are predictive of HSI**, can we establish **better thresholds** as being flagged?
If the components of the watch list are predictive of HSI, then we believe we can establish better thresholds for HSI.

OTHER QUESTIONS TO KEEP IN MIND

1. How does what we track compare to what these studies suggest? Is there a predictor we're missing?

2. If our predictors are accurate, is it possible to determine which factors predict which degree? (i.e. poor hydration predicts a strain while imbalance predicts a tear)

3. Remember that with physiological data, the data from the day of can be as important as the days leading up to it. This can be in terms of data the week of the injury (accumulated playerload, average hydration, average soreness if available) as well as trends over time (does athlete imbalance worsen leading up to the injury? how about force proportional to body weight). When considering fatigue as a factor it's helpful to think about the season phase. Pre-season training is taxing with little recovery, and the 1st and 2nd game of the season usually come with a few injuries as athletes aren't used to the stimulus yet. Fatigue might be greater later in the season due to cumulative load for 14 weeks.

4. Do "bye" weeks (weeks without a game) have any effect on HSI?
-time series data working with games and weeks not in data (aka bye weeks)
5. Is there a difference in HSI rates among starting vs non-starting athletes?
-do we have data pn starters vs non starters?

Roadmap

Documentation Style: finished products of work uploaded as a PDF
Timeline: follow to best ability

<u>Week 1</u>
Tuesday
- orientation + setting up workspaces

Wednesday
- find research literature to help our understanding + summarize findings from these sources

Thursday
- finish up research + give hypotheses on research questions + look over/sort through datasets, make sense of all predictors, etc.

<u>Week 2</u>
Monday
- meeting 10 am + github tutorial + continue familiarizing ourselves with the data

Tuesday
- week 1 review Q's due + literature review + brainstorm how to determine accuracy of risk report + submit project roadmap and literature review to week 1 deliverables

Wednesday
- start cleaning and organizing data + transform variables if needed

Thursday
- meeting 10 am + continue data wrangling + explore distributions of individual predictors for better understanding + simple pots of relationships

<u>Week 3</u>
Monday
- meeting 10 am + analyze the data to determine if players deemed high risk experienced injury on incident report

Tuesday
- reflect on hypothesis for question 1 + analyze the incident report for all athletes on the soft tissue watchlist

Wednesday
- continue to analyze the accuracy of risk assessments for athletes based on incident report

Thursday
- meeting 10 am + look into how previous injury status correlates to occurence of new hamstring injuries

<u>Week 4</u>
Monday
- meeting 10 am + look into how strength statistics correlate to incidence of hamstring injuries

Tuesday
- continue to look into the strength statistics + experiment with other statistics such as sprinting statistics

Wednesday
- analyze how changes to the thresholds of risk impact accuracy

Thursday
- meeting 10 am + continue to experiment with threshold changes + reflect on hypothesis for question 2 + peer/performance review

Week 5
Monday
- meeting 10 am + begin creating powerpoint presentation + begin creating draft of the written report + create meaningful visualizations for the data

Tuesday
- continue creating powerpoint presentation + continue drafting written report

Wednesday
- finish powerpoint presentation + finish written report

Thursday
- meeting 10 am + commit final changes to repository + push README replacement markdown file + present to stakeholders

Rough Outline Given

1. Background and research (1-2 weeks)
   a. Sport Science staff will provide initial research sources, data sets, and questions to interns
   b. Interns will build a roadmap for the rest of their project (including hypothesis, documentation style and modality, timelines, etc.)
   c. Interns will create a brief report (mini-literature review) (2-3 pages) on initial research sources found and their relation to answering the question
   d. Interns will receive feedback from Sports Science staff on viability of sources (we can provide different or additional sources for them to move on)
   e. First peer review/ performance review

2. Data analysis (3 weeks)
   a. Interns will then start to transform the dataset to start analysis as they see fit
   b. Bi-weekly check ins with Sports Science staff for guidance, Q&A, general housekeeping
   c. Interns will also have opportunities to gain exposure to sports science jobs (i.e. data entry, sports performance technology demos, meeting with other departmental staff members)
   d. Interns will create a logical data workflow with reproducible steps
   e. Each day of work (whatever schedule is set) will see at least 1 meaningful notebook commit to the team repository per Intern f. i. ii. iii. iv. Second peer review/ performance review

3. Conclusion Presentation (1-2 weeks)
   a. Interns will create meaningful visualizations of the data
   b. Interns will present their findings and explain statistical significance
        i.   Sampling methods are expected (if applicable)
        ii. Possible bias identification
              1. How do we limit these biases in the future?
        iii. Scalability
              1. What happens if we add additional variables?
              2. Can we use this for other sports? The whole athletic department?
        iv. CUSS
              1. Center
              2. Unusual features
              3. Shape
              4. Spread
              5. BE SPECIFIC!
   c. Interns will create a professional presentation and present to Sports Science Staff, their fellow interns, and 1 original stakeholder (i.e. AT/ S&C coach/ Sport Coach who presented question)
   d. Final peer/ performance review
   e. Clear and actionable set of principles