**Knees, Ankles, and Shoulders…oh my!**

***Note:*** *For this project, I will be assuming the identity of a data scientist working for a company that produces support braces (e.g. knee braces, ankle braces, etc).*

**Problem Statement:**

Our company, Best Braces USA, expends a great deal of capital each year on advertising. Our top three products (Knee Braces, Ankle Braces, and Shoulder Braces) account for 75% of that spend. As the various channels (TV, Radio, Social Media, Point of Sale, etc) to engage with different age groups become more diverse, we want to maximize our ad dollar efficiency by right-channeling those ads for the biggest consumers of each product. That is to say, for example, if we can find that 18-30 year olds have the biggest need for Knee Braces, then we may want to spend most of our Knee Brace ad dollars in the social media areas. The company as a whole, and our shareholders, will benefit from this more focused advertising. In particular, this will clearly be powerful information for our sales and advertising departments.

**What Data Can We Use?**

Every year, the U.S. Consumer Product Safety Commission tracks emergency rooms visits to approximately 100 hospitals. The complete year by year dataset is available [here](http://www.cpsc.gov/en/Research--Statistics/NEISS-Injury-Data/). For this project, we will focus on the 2023 data, located [here](https://docs.google.com/spreadsheets/d/1dji4d9bt6ebytPmXb5EufoGPoTwtYoC1/edit?usp=drive_link&ouid=108941037787073549843&rtpof=true&sd=true). This data contains, among other things, the age and sex of the patient, what their final diagnosis was, and what was the part of the body that was injured. We will take a subset of this data: Males with a diagnosis of Sprain/Strain to the Knee, Ankle, or Shoulder. This data is of particular interest to us since our braces are specifically designed to 1) prevent sprains/strains as well as 2) provide rehab support for those having suffered such an injury.

**How Will We Conduct the Research?**

We will start with the Null Hypothesis as “There is no statistical difference in the mean age of males who suffer sprains/strains to the knee, ankle, or shoulder”. The Alternative Hypothesis then becomes “There **is** a statistical difference in the mean age of males who suffer sprains/strains to the knee, ankle, or shoulder”. The dataset will be divided into the three effected body parts, and then we will use histograms and statistical features (e.g. skewness and kurtosis) to determine normality (or not). Based on results there, we will choose between pairwise t-tests, one-way ANOVA test, and or Kruskal-Wallis test to determine if the Null Hypothesis can be rejected or not. We will report using the 95% confidence interval and this a p-value of .05.