Underdog Achievement and Randomness in Team Ball Sports

Thaksheel Alleck*, *Roman Mitchell*, *Ori Remen*, *Tommaso Giovannelli*, *Luis Nunes Vicente

LEHIGH UNIVERSITY — INDUSTRIAL AND SYSTEMS ENGINEERING ISE

Industrial and Systems Engineering Department

Research Project within the OutreachISE Program

Research Goal 1: Find the **team ball sport** with the highest **underdog achievement** (likelihood of weaker teams winning against stronger ones).

Research Goal 2: Investigate how **underdog achievement** is influenced by **randomness factors** that affect match outcomes in team ball sports.

Data Collection: Match Scores and Team Rankings

We created a **match score dataset** by scraping Wikipedia pages using Python for 12 team ball sports.

For each edition of a competition:

- We computed a **team ranking** based on victories, losses, and ties.

- We aggregated past team rankings into a **weighted ranking**.

Team Ball Sport	Major International Competition
Basketball	Summer Olympic Games
Cricket	ICC Men's Cricket World Cup
Field Hockey	Men's FIH Hockey World Cup
Futsal	FIFA Futsal World Cup
Handball	Summer Olympic Games
Ice Hockey	Winter Olympic Games
Lacrosse	World Lacrosse Men's World Cup
Roller Hockey	World Skate Roller Hockey World Cup
Rugby	Rugby World Cup
Soccer	FIFA World Cup
Volleyball	FIVB Volleyball Men's World Cup
Water Polo	FINA Men's Water Polo World Cup

How to Identify Weak Teams?

Given a match between teams T1 and T2, T1 is a **weak team** if

$$R(T1) \leq R(T2) - \textbf{Threshold}$$

where R(·) denotes the position of a team in the **weighted ranking**. We refer to |R(T1) - R(T2)| as the rank difference between T1 and T2. For each sport, we set the **Threshold** to the median of the corresponding rank difference distribution.

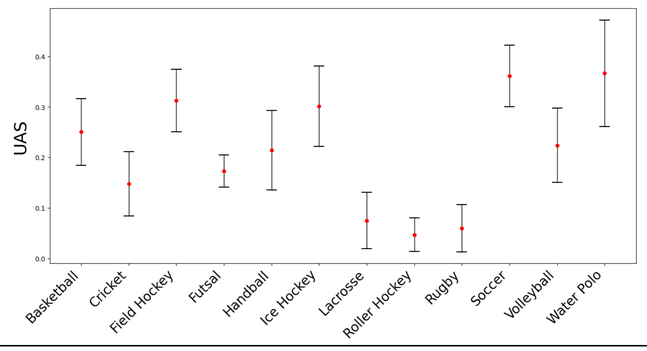
How to Quantify Underdog Achievement?

For each team ball sport, based on the **weighted ranking**, we computed an **underdog achievement score**.

$$\textbf{UAS}_e = \frac{\text{Number of victories or draws by a weak team in edition } e}{\text{Number of matches with a weak team in edition } e}$$

$$\textbf{UAS} = \frac{1}{\text{Number of editions}} \sum_{e=1}^{\text{Number of editions}} (\textbf{UAS}_e)$$

95%-confidence intervals for the **UAS** value for each sport.



Soccer, **water polo**, **field hockey**, and **ice hockey** have the highest **UAS**, while **lacrosse**, **roller hockey**, and **rugby** have the lowest.

Explaining Underdog Achievement with Our Randomness Model

We quantified average **randomness factor values** for each sport, resulting in a dataset containing 12 rows (one per sport) and 14 columns (one per factor). We perform a principal component analysis (PCA) and a correlation analysis to gain insights into the relationship between **UAS** and **randomness factors**.

PCA computes principal components (PCs), linear combinations of column values that preserve original dataset variability. PC1 and PC2 explain 56% of the variability.

Physical Environment Factors

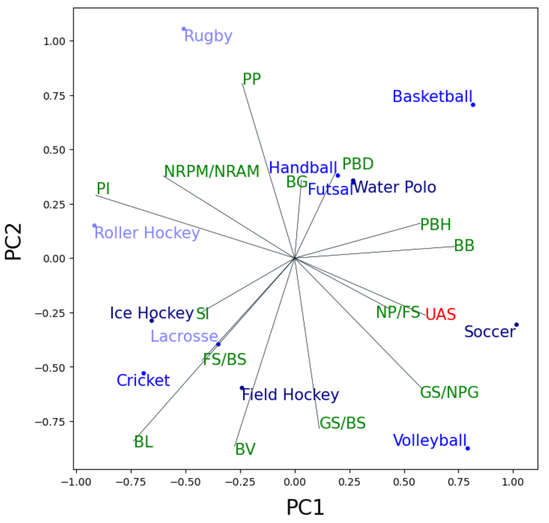
BL	Ball lightness
BV	Ball velocity
FS/BS	Field size/Ball size
GS/BS	Goal size/Ball size
BG	Ball geometry
BB	Ball bounciness

Player Factors

PP	Player powerfulness
PBH	Player ball handling
PBD	Player ball dispossession
PI	Player inexperience

Team Factors

NP/FS	Number of players/Field size
GS/NPG	Goal size / # players who can effectively defend the goal
SI	Scoring infrequency
NRAM/NRPM	# rules about movement / # rules that prevent movement



Observation 1. In **soccer**, **PBH**, **BB**, and **GS/NPG** values are high due to players using various body parts, a highly bouncy ball, and one player defending the goal.

Observation 2. For **hockey sports**, main randomness factors include **PI**, **SI**, **FS/BS**, **BL**, and **BV**. Players retire young, scoring frequency is lower than basketball, ball size is small (resulting in high **FS/BS** values), ball weight is light, and ball velocity is high.

Observation 3. For **water polo**, main randomness factors are **PBD**, due to low play time, and **PP**. Similar conclusions apply to **handball**, **futsal**, and **basketball**.

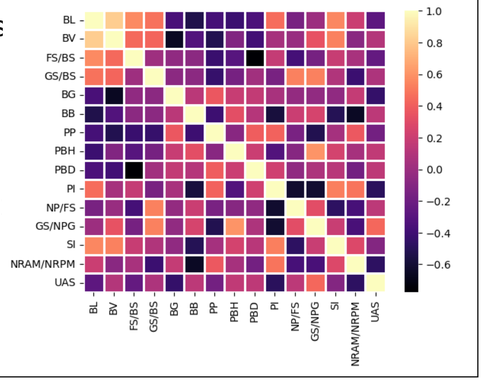
Observation 4. For **rugby**, key randomness factors include **PP**, **BG**, **PI**, and **PBD**, each at maximum values. **BG**'s high value stems from the unique shape of rugby balls, increasing match outcome randomness. **PI** is high due to players retiring young. **PBD**'s high value results from minimal play time and many players.

Engineering Impact.
- Design training strategies in sports to address randomness.
- Improve the profitability of the gaming industry (e.g., betting markets).
- Enhance sports performance analysis.

Future Work.
- Analyze professional leagues and collegiate competitions.
- Analyze team non-ball sports.
- Investigate the impact of referee errors in seasonal competitions and home vs away matches in seasonal competitions.

The heatmap on the right illustrates the Pearson correlation coefficient between each pair of factors, including **UAS**.

The factors with the highest impact on **randomness** are those with a positive correlation with **UAS**, i.e., **GS/NPG**, **NP/FS**, **PBD**, **PBH**, and **BB**.



Paper 1: "Match score dataset for team ball sports", to be submitted to Data in Brief.
Paper 2: "Why is soccer so popular: Understanding underdog achievement and randomness in team ball sports", to be submitted to Journal of Sports Sciences.