The Paper: Telling the Story of Running Related Injuries

Predict 455/ Section 56
June 4, 2017

Baseer Ahmad
Tracy Valentine
Luis Alejandro Mesen
Jen Vroomen

Introduction

The Show is a visual exploration that examines running as a sport and associated injuries. The inspiration for this study originated with an article titled "Training Errors and Running Related Injuries: A Systematic Review". This article was written in 2012 by Nielsen, Buist, Sorensen, Lind, and Rasmussen. The Show is divided into three sections: an introduction, a discussion, and a conclusion. The introduction chapter will introduce the audience to the topic and provide a broad overview of running and related injuries. The discussion will identify and explore causal factors to running injury risk. Furthermore, this chapter will discuss study findings from the original source by Nielsen et al. Finally, the conclusion phase will provide a solution to avoid running injury.

The Show discusses and touches on several topics related to running injuries. The objective of The Show is to present information and research regarding running as a training potential, outline associated injuries, as well as provide a potential solution to avoid injury. Armed with this information, individuals who are either interested in running as a sport or who are working on a running related performance goal may be able to better avoid common injuries and therefore reach their full potential as a runner.

Methodology

Each visualization was initially created using a common process. The research was performed on the topic of running and running related injuries. This research initiated by first studying the article that was the original inspiration by Nielsen et al. Next a search was performed using relevant keywords on PubMed, Google Scholar, and the Northwestern Library. This research resulted in the collection of several tables and written data files that could be made into visualizations and would tell a complete story around the topic of running and running related injuries. Next, each set of numeric data was recorded in a common Excel spreadsheet. Each team member developed the visualizations by pulling a data set into the desired program and using best practices to develop the most effective graphic to represent the data at hand.

The introduction included seven visualizations. The first visualization provides a graphic showing common running surfaces (Taunton J et al, 2013). The larger the circle in the graphic, the more frequently this surface type is used by runners. The second graphic is a bar graph that shows how often runners train in a given week (Taunton J et al, 2013). Both of these graphics were made in Tableau.

The third visualization is a geospatial map created in R that uses a blue to coral color gradient to represent the percentage of a given state's population that have completed and ultra-distance running race. This map compares the density of ultra runners in a given state. The data for this visualization comes from a website that records the number of ultra-distance finishes in each state (Ultra Distance Magazine, 2014).

Visualizations four and five use data from the 1993 Auckland Citibank Marathon (Satterthwaite P et al, 1999). Both of these visualizations were created in Tableau. The graphics show the distributions of age groups and genders respectively that participated in the marathon. The graphic displaying age groups is a color coded table while the graphic showing gender is a filled box graphic with two different colors representing the proportion of male to female athletes.

The sixth visualization provides an overview of the likelihood of injury for a wide variety of sports (Pons et al., 2010). The data for this visualization was numeric and the program used to design the visualization was Tableau. A blue to red color gradient was used to portray different likelihood of injury (as a percent) for a given sport.

The last visualization in the Introduction chapter is a word cloud created from the various definitions of a "running injury" as written by 31 different articles (Nielsen et al, 2012). The text from these definitions was saved in a .txt file and the words "injury", "definition", and "running" were removed. Next, the .txt file was imported into RStudio and a word cloud was created based on the frequency of a given word. Only words that were used more than two times were included in the word cloud and the limit of number words in the cloud was set to 150.

The Discussion includes twelve visualizations. These visualizations were further divided into two subchapters: 'A Case Study on Running Injuries', and 'Defining A Predictive Injury Model'.

The first visualization in the Discussion chapter shows the most common injury types and the associated incident of injury as a percentage (Christinsen S, 2016). This visualization was created in Tableau and employed differently sized circles to show the likelihood of injury for the given injury and used shades of blue to group injuries by similar likelihood of injury percentage. The groupings represent high, moderate, low moderate, and low likelihood groups. The next seven visualizations in the Discussion chapter examine specific injuries and their association with either gender, age group, or running experience. Consequently, these graphics make up

the case study subchapter. The title and order of these visuals are as follows: “Quadricep Injury by Age Group”, “Calf Injury by Age Group”, “Calf Injury by Gender”, “Knee Injury based on Marathon Experience”, “Hamstring Injuries based on Gender”, “Hamstring Injuries Based on Marathon Experience”, and “Hip Injuries Based on Gender”. All of these visualizations are bar charts created in Tableau using data from the 1999 paper by Satterthwaite et al.

Three visualizations are included in the ‘Defining a Predictive Injury Model’ subchapter of the Discussion. The first graphic in this subchapter is titled “The Relationship between Miles Per Week and Running Related Injury” and is a comparative bar graph that compares this relationship between men and women (Nielsen et al, 2012). The next visualization is titled “Relationship between Duration of Workout and Running Related Injury” and is a bar graph with a fitted line showing how the incident of injury changes with the duration of a workout (Nielsen et al, 2012). Lastly, the final graphic in this subchapter is titled “Proposal of using Acute: Chronic Workload as a Solution for Assessing Injury Risk”. This graphic is a box plot that shows the average risk of high chronic load and low chronic load as a function of acute load (Gabbett T, 2015). Each of these visualizations was created in Tableau.

The conclusion chapter discusses using the ratio metric of acute to chronic load as a way to avoid the running injuries discussed in this presentation. This chapter contains two visualizations that present relevant data on the topic of acute versus chronic loads. The first of these is a shaded region plot that shows that an athlete will want to maintain an acute to chronic load ratio between .8 and 1.3 in order to avoid injury (Windt J and Gabbett TJ, 2017). This plot was created in R. The second visualization in the conclusion section is a heat map that details the likelihood of injury as a function of acute to chronic load ratios (Blanch P and Gabbett TJ, 2016). This map was developed in Excel.

Transition slides were used to divide The Show into the Introduction, Discussion, and Conclusion chapters respectively. These slides serve to keep the presentation flowing in a logical way so that the information is easier to follow. A series of stock photos were used to enhance the transition slides. There is a section in the References chapter that lists the sources for these photos. The Show was put together using impress.js. A free CSS theme was downloaded from https://templated.com/ and modified to align with the specific needs of the

presentation. Each of the visualizations outlined above was saved as a PNG file and submitted as a slide in The Show. GitHub was employed for versioning the source control (link: https://github.com/lamesen/TheShow).

Discussion

Bar charts have been employed for nine of the visualizations utilized in The Show. These charts are generally the preferred choice for data that is comparative in nature because they allow easy comparison of values across categories along the same scale. For the bar charts used in The Show, runners have been split into subgroups based on gender, age, running frequency, and marathon experience. Two clustered bar charts were also created for 'Knee Injury Based on Marathon Experience' and 'The Relationship Between Miles Per Week and Running Related Injury.' Clustered bar charts allow us to introduce an extra dimension of comparison by breaking our data down even further into additional subcategories. Color is used here to meaningfully distinguish between these subgroups.

Bubble charts were used for 'Surface Type For Running' and 'Incident Percentage of Running Injuries.' These are used to effectively make comparisons across different categories with an added dimension of proportionality represented by the size of each circle or bubble. In the case of the 'Surface Type For Running' visual, color is utilized to further distinguish the various surface types represented in the plot. With the second bubble chart, color is used in tandem with incident percentage—darker colors are associated with more frequent occurrences of injury while lighter shades are associated with less common injuries. The size of the bubbles makes it easier to compare the relative frequency of these injuries.

Choropleth maps allow us to compare relative values across distinct geographical areas. These values are then color coded based on a scale that enables us to evaluate the density within specific regions and variation over the map as a whole. The visualization 'Density of Ultra-Distance Runners by State' depicts the percentage of a state's population that has completed one or more ultra-endurance running event. The use of contrasting colors here makes it easy to distinguish between states with a low percentage of ultra-distance runners and those with a higher concentration of ultra-distance runners, while the shade of these colors provides a better understanding of these percentages.

Word clouds are used to display the frequency with which words occur in textual data. The graphic 'Defining Running Injury: A Systematic Review' shows the most common words used when referencing running injuries and how experts traditionally define it.

Heat maps are utilized to compare values from two different categories to understand the strength of their relationship. The heat map for 'Injury Likelihood Based on Acute and Chronic Workloads' aligns chronic workload on the y-axis and acute workload along the x-axis. Scenario analysis can then be performed using the percentages of both chronic and acute workload to determine how a sudden change in acute workload would affect the likelihood of injury. A three-color scale (green-yellow-red) is effectively used to convey an increasing chance of injury. Green is considered generally safe while red is considered extremely dangerous.

Side-by-side boxplots have the same advantages of a single boxplot, while simultaneously allowing us to compare the center measurement, range and variance of our data. The visual 'Proposal of Using Acute: Chronic Load As a Solution for Assessing Injury Risk' provides a better understanding of the relationship between acute and chronic workloads. Acute load is divided into seven categories, while chronic load is split between high and low conditions. This visual lets us see the general upward trend in injury risk as acute load is increased while at the same time, giving us the error rate of these estimations. Color is used here to depict the severity of injury risk.

Conclusion

In conclusion, The Show was constructed using best data visualization best practices and conveys the research performed regarding running and running related injuries. We also recognize that there is room for improvement in the future. For one, we leaned heavily on the use of bar graphs. As we learn about additional visualization types going forward, perhaps we could try to experiment with other charts or graphs that would present this information in new and refreshing ways. Additionally, as a group, most of us were using these visualization tools for the very first time. Consequently, many of our visualizations are relatively simple. This is not necessarily a bad thing as we still tried our best to follow good data visualization practices with all of them. However, as we continue to hone our skills we would like to produce more dynamic or interactive visuals that would allow us to convey an even more engaging story.

References

Blanch P, Gabbett TJ (2016). Has the athlete trained enough to return to play safely? The acute:chronic workload ratio permits clinicians to quantify a player's risk of subsequent injury.*British Journal of Sports Medicine,* 50:471-475.

Christensen S (2016). Epidemiology of Running Injuries. Complete Track & Field Blog. http://completetrackandfield.com/epidemiology-of-running-injuries/

Gabbett T (2015). The training-injury prevention paradox: should athletes be training smarter and harder? *British Journal of Sports Medicine*. doi:10.1136/bjsports-2015-095788.

Nielsen RO, Buist I, Sorensen H, Lind M, Rasmussen S (2012). Training Errors and Running Related Injuries: A Systematic Review. *The International Journal of Sports Physical Therapy*, 7(1)58-71.

Oja P, Titze S, Kokko S, Kujala U, Heinonen A, Kelly P, Koski P, Foster C (2015). Health benefits of different sport disciplines for adults: systematic review of observational and intervention studies with meta-analysis. *British Journal of Sports Medicine*, 49:434-440.

Pons- Villanueva J, Segui-Gomez M, Martinez-Gonzalez MA (2010). Risk of Injury according to participation in specific physical activities: a 6-year follow-up of 14,356 participants of the SUN cohort. *International Journal of Epidemiology*, 39(2):580-7.

Satterthwaite P, Norton R, Larmer P, Robinson E. (1999). Risk factors for injuries and other health problems sustained in a marathon. *British Journal of Sports Medicine,* 33(1):22-26.

Taunton JE, Ryan MB, Clement DB, et al (2003). A prospective study of running injuries: the Vancouver Sun Run "In Training" clinics. *British Journal of Sports Medicine,* 37:239-244.

UltraRunning Magazine (2014). Which State has the most UltraRunning finishers as a percentage of population? https://ultrarunning.com/featured/which-state-has-the-mostultrarunning-finishers-as-percentage-of-population/

US Department of Health (2000). CDC Recommendations Regarding Selected Conditions Affecting Women's Health, 31(49): 23.

Windt J, Gabbett TJ (2017). How do training and competition workloads relate to injury? The workload—injury aetiology model. *British Journal of Sports Medicine,* 51:428-435.

Stock Photo Sources

http://www.stuff.co.nz/life-style/well-good/inspire-me/84844189/nz-finally-catching-up-to-world-with-marathon-fundraising

http://www.stuff.co.nz/life-style/well-good/inspire-me/84844189/nz-finally-catching-up-to-world-with-marathon-fundraising

http://www.runnersworld.com/sites/runnersworld.com/files/styles/slideshow-desktop/public/oldrunner_davide_gambino.jpg?itok=mBbLPZ6p

http://www.runnersworld.com/sites/runnersworld.com/files/styles/article_main_image_2200px/public/articles/2017/02/charlie_westrip_1.jpg?itok=8Jj0QwcT

http://naturalrunningcenter.com/wp-content/uploads/2012/09/shutterstock_102336898.jpg

http://i.dailymail.co.uk/i/pix/2014/07/28/article-2708628-2011DA0C00000578-524_634x434.jpg

http://en.alalam.ir/news/1850908

Jin J (2014). JAMA Patient Page: Running Injuries, 312(2):202