

Player Tracking Facilitates Valid Causal Inference: The Average Treatment Effect of Defender Proximity on Scoring

Matthew S. Tenan PhD, ATC, FACSM

Ali R. Rezai MD

Rockefeller Neuroscience Institute, West Virginia University, USA

Paper Track: Basketball

Paper ID: 891922

Abstract

Probably the hottest topic in sport analytics is the development of prediction models. However, stakeholders often incorrectly assume that changing the input to a prediction model in the real-world causes a change in the desired outcome. This is a subtle but important fallacy. While considerable work has been put forward to see what retrospectively “describes” an effective player/team or prospectively “predicts” performance or injury, considerably less analytical work has been put into determining “what CAUSES performance to increase” or “what CAUSES soft-tissue injuries to decrease”, the latter of which is important to actually change the outcome of games and seasons. The goal of this current work is to show within a Causal Inference framework how valid causal conclusions can be made from high-dimensional player tracking data in basketball. It is suggested that this is a critical time regarding data availability, analytical methods development, and personnel expertise to start expecting valid causal conclusions to be drawn from data in sport.

1. Introduction

The need for quality defensive play is nearly never in doubt; indeed, the oft-repeated quote is that “Offense sells tickets, but defense wins championships.” The extent to which this axiom is true in basketball has been hotly debated.(García et al. 2013) Typical basketball outcomes indicative of positive defensive performance are often defined by simplistic metrics such as blocks, defensive rebounds, turnover ratio, and opponent field goal percentage, amongst many others. Previous research has focused on descriptive classification of these simplistic metrics and indicated both field goal percentage and defensive rebounds are characteristic of winning teams.(García et al. 2013; Milanović et al. 2016; A. Gómez et al. 2017; Çene 2018) Importantly, simplistic measures of defensive performance that are reported in box scores do not account for other contextual factors, such as flow of game over time, location of game, opponent field goal percentage prior to game, player location on the court, and others. More complex models for prediction problems have been developed with an application focused on betting markets(Song and Shi 2020; Song et al. 2020) or simple game win-loss.(Loeffelholz et al. 2009; Caliwag et al. 2018) However, neither simplistic descriptive studies nor complex prediction models are helpful in demonstrating to coaching staff what strategies or style of game-play actually cause a team to win or cause a player to make a shot in a given scenario.

The advent and widespread proliferation of machine learning prediction algorithms in sport may mislead coaches, clinicians and front-office staff into thinking that changing some aspect of what

goes into a prediction model can actually “cause” an effect in a real-world outcome. This fundamentally misconstrues the purpose and validity of prediction models. While machine learning metrics such as Variable Importance and Shapley Values are helpful in understanding how a machine learning model is internalizing and weighing the input data, it does not tell you that prospectively changing a data input would cause a change in performance or outcome for the player/team. The latter requires the use of Causal Inference methodologies.

The field of Causal Inference is dominated by two theoretical approaches, the Rubin Causal Model (Rubin 1974) based on the Potential Outcomes framework and the Structural Causal Model based on do-calculus (Pearl 1995). Both of these frameworks have their ardent adherents and can be mathematically equivalent in some instances (Malinsky et al. 2019); however, the approach used in this manuscript is the Potential Outcomes framework because 1) it is theoretically grounded in approximating the randomized control trial, which is a recognized gold-standard for causal study design; 2) it is nearly a century old, with a wealth of biomedical literature dealing in nuanced approaches to various problems; and 3) it explicitly handles nested relationships. The ability to handle nested relationships in causal inference is fundamental to many instances in both sport and medicine where important information may only be available (or unobservably encoded) in an aggregate form, such as hospital-wide resources or a basketball team’s style of play (Weber et al. 2018). Judah Pearl, the most well-known proponent of the Structural Causal Model, has stated that he does not believe these nested structures represent causal relationships (Pearl 2019); however, extensive biomedical research in this area demonstrates that failing to model these relationships results in incorrect estimands in randomized-control trials (Candlish et al. 2018; Moerbeek and Schie 2019; Vorland et al. 2021).

High-resolution player tracking systems such as ShotTracker at the collegiate-level and Second Spectrum at the NBA-level allow for far more granular detail about the context under which events occur in a game, such as a defensive rebound or a made shot. This three-dimensional position data appears to be under-utilized when it comes to deriving causal insights that can be directly applied to basketball training, nutrition or game strategy. The ability to define a causal model for on-court performance, whatever the desired outcome metric may be, can be used to ascertain if a certain supplement or recovery modality is having a real, causal effect on performance. This can both enhance team/individual performance as well as save franchises/universities hundreds of thousands of dollars a year on interventions which have no real causal effects. The goal of this present work is more modest: demonstrate how the Potential Outcomes framework can be combined with player tracking in collegiate basketball to determine the Average Treatment Effect of defender proximity on scoring likelihood as a function of shot distance.

2. Methods

2.1. Player Tracking Data

Using the ShotTracker positioning system, 15,835 shots with sufficient data were obtained from games in the Big 12 and Mountain West Men’s Basketball leagues for the 2020 and 2021 seasons. The ShotTracker system is made up of three components with sensors located within the ball, on players, and anchors located within the stadium’s rafters. Additionally, the ShotTracker system is calibrated to each stadium’s court, providing millimeter-level resolution in three-dimensions. All data was accessed via the ShotTracker Application Programming Interface (API); this analysis was approved by the West Virginia University’s Institutional Review Board (Protocol #2205569473) and the West Virginia University Athletics Department.

2.2. Tracking Data Preprocessing

All recorded shots from the 2020 and 2021 seasons for the within-conference games were extracted from the ShotTracker API. Within the system, all data points have UNIX time but some observations are missing game time or shot clock time. It was possible to interpolate missing game time data via linear interpolation using adjacent shots with recorded game times and accompanying UNIX time. Shot clock time was unable to be recovered with any validity and was discarded as a data source. Some games had overtime indicated, whereas other games lasting longer than 40 minutes did not. A satisfactory resolution to this issue was not able to be reached so all games were truncated at 40 minutes and overtime periods removed from analysis.

The ShotTracker API was also leveraged to obtain general player shooting statistics for the player prior to each game. In this way, there was no information leakage since a player's 2- and 3-point Field Goal percentages going into a given game were aligned with any shots taken during that game. For each recorded shot, the on-court player lineup was obtained. It was found that 9% of shots had less than 10 players in their ShotTracker lineup. This incomplete data was removed from further analysis. For shots with complete lineups, the location of each player on the court was obtained and the proximity from the shooter for each defender was calculated via the Pythagorean theorem.

2.3. Causal Inference Analysis

The Potential Outcomes framework defines estimands in randomized experiments as functions of potential outcomes for units. For each unit under a dichotomous set of treatments, there are an array of outcomes a unit could have under treatment 1 or treatment 2. This framework can be aligned with Heisenberg's uncertainty principle where we can state that it is not physically possible to know the outcome of BOTH treatment 1 and treatment 2 on a unit with absolute certainty as they cannot be observed at the same time, even if all initial conditions are pre-specified. However, if we can make the Stable Unit Treatment Value Assumption, or SUTVA, in a randomized experiment/trial, the Average Treatment Effects of an intervention can be discerned.

SUTVA has two parts:

1. For each unit there is only one form of treatment and one form of control treatment (in a dichotomous treatment paradigm)
2. There is no interference among units or each unit's outcome remains the same no matter what treatment other units receive

SUTVA is explicitly addressed in the study design of randomized control trials and the "randomized" aspect of that study design removes any systematic bias in the assignment of a unit to a particular treatment by balancing the treatment groups based upon covariates that are perceived to be meaningful. For observational studies leveraging the Potential Outcomes framework, SUTVA must be reasonable and it is important to balance treatment assignment on meaningful covariates. (Holland 1986)

In the context of the current analysis, our "treatment" is not a binary, but rather the continuous "proximity to closest defender" and so meeting SUTVA requires there to be only one "form" of defender proximity to shot taken. Pilot work by our group indicated that the defender's 'angle' to the shot had had no meaningful effect on scoring probability and this was the only potential "other form" of treatment considered possible. The second aspect of SUTVA for the current work would

assume that defender proximity to one shot does not interfere with the proximity of a defender to another shot, an assumption which seems reasonable at face-value.

Having satisfied SUTVA, we assessed if each “unit” or shot could have been randomly assigned to any defender proximity (i.e. treatment balance). Since shot’s are taken by players of different skill levels, players are defended by other players of different styles of play, and teams have different styles of play both on offense and defense which lead to more “open shots” or more “contested shots”, this treatment balance cannot simply be assumed.

2.3.1. Propensity Score & Validation

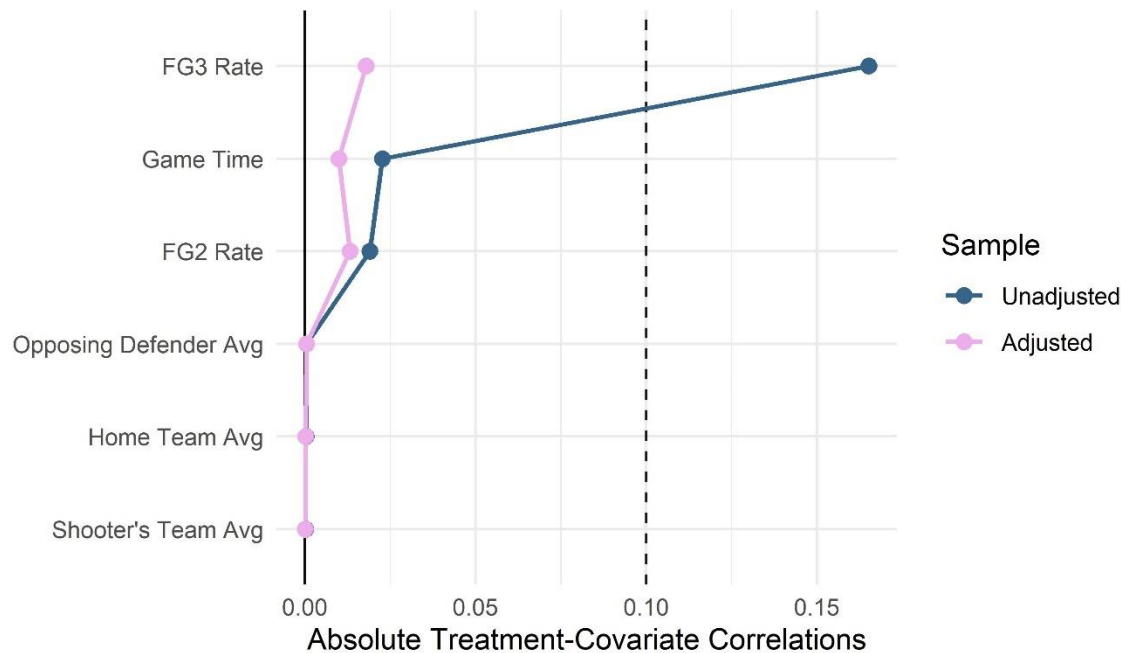
Current work requires that, for each observed shot, the defender’s proximity could have been randomly assigned or ‘balanced’.(Fong et al. 2018) We achieved this balance via a non-linear generalized additive mixed model propensity score where shots are nested within players and the shooter’s team, the team playing at home, the player’s 2-point and 3-point percentages prior to the game, game time when shot is taken, and the player guarding the shooter are used to predict the defender’s proximity. This propensity score seeks to capture information relating to how “tight” a shooter may be played both by modeling their shooting history, the team they play for (and thus play-calling/style of play), and the player defending the shot. The benefit of creating the propensity score via a generalized additive mixed model is that random effects (modeling the variance at the shooter-level) as well as non-linearities can be effectively incorporated. Nonlinearities in continuous predictors for the propensity score (game time when shot is taken, player’s 2-point percentage, and player’s 3-point percentage) were modeled with thin plate splines and 30 dimensions of the basis. The validity of the generalized additive mixed model was verified by examining the histogram of residuals (to assess normality) and plotting the residuals vs predicted values (to assess constant variance).

Most classical propensity score methodologies employ a logistic or binary outcome model, creating a model-predicted probability between 0 and 1 that is directly interpretable as a propensity score weight.(Westreich et al. 2010; Wyss et al. 2014) The most common approach to continuous treatment propensity scores is through the use of stability weights. Continuous case propensity scores are stabilized with a numerator other than 1 because unstabilized scores will have infinite variance. In the multilevel continuous treatment case, the stabilizing numerator is created either through a cluster-level Gaussian density or a marginal Gaussian density.(Schuler et al. 2016)

After the propensity score is stabilized, covariate balance is checked.(Austin 2019) Covariate balance is assessed to verify that each unit/shot is balanced across the treatments (i.e. defender proximity) and provide a level of evidence that assignment of this treatment could have been random. For continuous treatments, the propensity score weights are assessed by examining the correlation between the baseline covariate and the treatment, the rationale being that the baseline covariate should be statistically independent from the treatment and a threshold of 0.1 has been shown in simulations to result in minimal confounding.(Zhu et al. 2015) After reviewing the balance statistics for both the cluster-stabilized and margin-stabilized propensity scores, it was determined that the margin-stabilized propensity scores produced the best and most effective balancing of baseline covariates across treatments (defender proximities). The balance statistics for the margin-stabilized propensity score weights can be seen in Figure 1, presented in a so-called ‘Love Plot’ format.

Figure 1. Margin Stabilized Propensity Score Love Plot

The baseline covariates are listed on the y-axis and the x-axis is the correlation between Defender Proximity and the covariate. The Unadjusted line (blue) is the correlation between the Defender Proximity and the covariate prior to adjustment with the propensity score and the Adjusted line (pink) is after application of the propensity score. The dashed line at 0.10 is the threshold for covariate balance, which is met after propensity score weighting. The Average Correlations are shown for the Opposing Defender, Home Team Name and Shooter's Team Name both for display readability and to maintain anonymity compliance per ShotTracker agreements, but all Adjusted individual correlations were below 0.1.



2.3.2. Final Model & Analysis

The formal analysis for causal inference was then made with a player-nested logistic regression for shot make/miss as dependent variable. The shot's distance from hoop center, defender proximity to shooter and the interaction effect of those data points served as independent variables. The observations were weighted with stabilized propensity scores, and robust sandwich estimates for standard error calculations.(Lumley and Scott 2017) Multicollinearity was checked to insure that the Variable Inflation Factors were not greater than 5.

2.3.3. Analytical Software and Open Code

All analyses and API calls were made using the R programming language.(R Core Team 2022) In addition to base R, the following add-on packages were leveraged for data retrieval, management, analysis and/or visualization: httr,(Wickham 2022a) httr2,(Wickham 2022b) jsonlite,(Ooms 2014) glue,(Hester and Bryan 2022) data.table,(Dowle and Srinivasan 2022) lubridate,(Grolemund and Wickham 2011) plyr,(Wickham 2011) dplyr,(Wickham et al. 2022) stringr,(Wickham 2022c) purr,(Henry and Wickham 2022) ggplot2,(Wickham 2010) tidyr,(Wickham and Girlich 2022)

rlist,(Ren 2021) tibble,(Muller and Wickham 2022) mgcv,(Wood 2011) gratia,(Simpson 2022) WeighIt,(Greifer 2022a) cobalt,(Greifer 2022b) survey,(Lumley 2010) interactions,(Long 2019) and RColorBrewer.(Neuwirth 2022) All code and pseudodata for the final causal inference analysis is available at the author's GitHub site: https://github.com/TenanATC/SSAC_2023

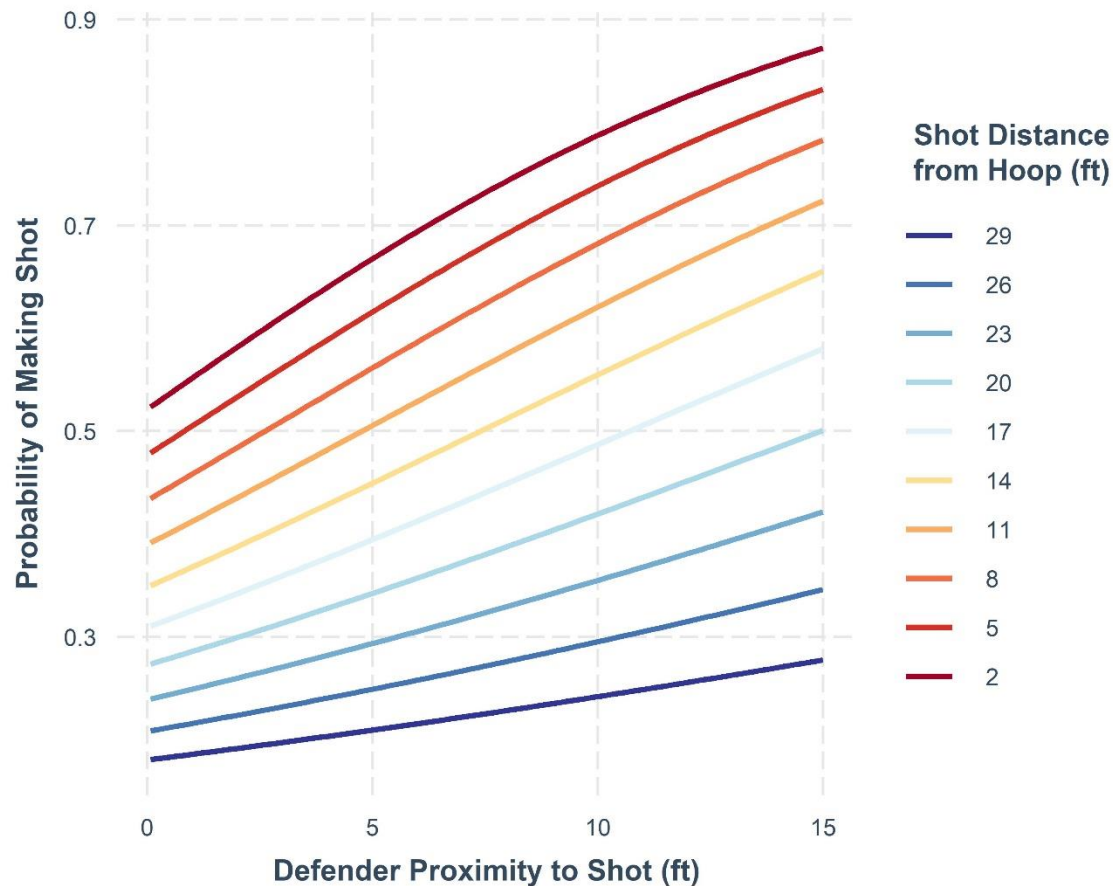
3. Results

There was a significant interaction effect ($p = 0.002$), indicating that defender proximity to shot causes a lower likelihood of making a basket and that the proximity of the defender has stronger causal effects as shot distance decreases. Figure 2 shows the Average Treatment Effect of defender proximity on shot probabilities for different shot distances.

The Average Treatment Effect is the difference in outcomes between units (shots) assigned to different treatments (proximities to defender). This analytically demonstrates the causal effect of how close a defender is to a shooter and how this impacts the likelihood of making the shot; furthermore, we demonstrate that this causal effect is differential, more profound, and non-linear as shots are taken closer to the basket. Unsurprisingly, a wide-open shot taken close to the hoop has a high likelihood of success, but this likelihood can be substantially decreased even by a defender being 5 feet away.

Figure 2. Causal Interaction Analysis

The Average Treatment Effect of Defender Proximity (in feet, x-axis) on probability of making a shot (y-axis). This effect is examined at various distances a shot may be taken from the hoop.



3.1. Causal Analysis in Basketball

The causal inference work in the current manuscript is unlikely to surprise coaches, players or front-office personnel. Of course, more closely guarded shots are less successful and it seems natural that the highest probability shots closest to the hoop are made far less likely if a “rim defender” or shot-blocker is nearby. However, to our knowledge, this is the first analytical report detailing out the causal nature of this relationship. This type of work would not have been possible without some sort of continuous player tracking information, which has only recently become widely available in basketball. Furthermore, while the Potential Outcomes framework has existed in some form since the 1970’s, (Rubin 1974) only recently have methods been devised and validated which can handle complex study designs necessary to handle most applications in sport. The current analysis leverages multilevel models in both the propensity score calculation and the final Causal Inference model, and nonlinear penalized splines are leveraged in the design of the propensity score. While both of those statistical tools have been around since the 1990’s, (Raudenbush 1988; Hastie 1990) only in the last two decades have they become widely used, analytically feasible with a personal computer, and taught in graduate school curriculums. Therefore, we may be at the start of a critical phase for causal inference in sport where there is technical data available, widely validated statistical methods, and personnel educated to employ these methods.

Future analytical work with player tracking data could answer causal inference questions that have a greater impact, for example:

- Does a zone defense cause there to be lower quality shots?
- Does instantaneous player acceleration with the ball cause more fouls?
- Does player height cause them to get more rebounds (or does wingspan, jump height, etc.)?

However, even more interesting causal inference questions can be answered when combining multiple data sources such as player tracking, demographics, nutrition, sports medicine, strength & conditioning, and sport psychology. For example:

- Does using a nutritional supplement cause increases in late-game defensive or offensive performance?
- Does prophylactic ankle taping/bracing cause a decrease in ankle injuries?
- Does prophylactic ankle taping/bracing cause a decrease in angle injuries as a function of decreasing player acceleration?
- Does pre-game hydration cause a decrease in cramping during games?
- Does pre-game hydration cause an increase in late-game performance?
- Does visuo-motor training cause an increase in in-game reaction time to passes (i.e. decreased 'bobbles')

None of the above questions are necessarily challenging given the right data sources and Causal Inference methods, whether the analyst is employing propensity scores, directed acyclic graphs, regression discontinuity, or difference-in-differences. The key is having both the “right data” and the “right people” who can deploy these analytic techniques and explain them with all applicable caveats and assumptions to appropriate stakeholders.

4. Conclusion

Automated, high-resolution tracking of players in games, combined with modern statistical methods for Causal Inference, facilitate causal conclusions about game play in basketball. This current analysis places a causal and numerical framework around what is logically known; however, future analyses using player tracking and Causal Inference can explore more controversial topics on basketball strategy or performance. While much focus has recently been on prediction modeling in sport, Causal Inference is likely to have a higher return-on-investment when it comes to enhancing athlete and team performance.

References

- A. Gómez M, J. Ibáñez S, Parejo I, Furley P (2017) The use of classification and regression tree when classifying winning and losing basketball teams. *Kinesiology* 49.:47–56.
<https://doi.org/10.26582/k.49.1.9>
- Austin PC (2019) Assessing covariate balance when using the generalized propensity score with quantitative or continuous exposures. *Stat Methods Med Res* 28:1365–1377.
<https://doi.org/10.1177/0962280218756159>

- Caliwag JA, Aragon MCR, Castillo RE, Colantes EMS (2018) Predicting Basketball Results Using Cascading Algorithm. In: Proceedings of the 2018 International Conference on Information Science and System. Association for Computing Machinery, New York, NY, USA, pp 64–68
- Candlish J, Teare MD, Dimairo M, et al (2018) Appropriate statistical methods for analysing partially nested randomised controlled trials with continuous outcomes: a simulation study. *BMC Med Res Methodol* 18:105. <https://doi.org/10.1186/s12874-018-0559-x>
- Çene E (2018) What is the difference between a winning and a losing team: insights from Euroleague basketball. *Int J Perform Anal Sport* 18:55–68. <https://doi.org/10.1080/24748668.2018.1446234>
- Dowle M, Srinivasan A (2022) data.table: Extension of `data.frame`
- Fong C, Hazlett C, Imai K (2018) Covariate balancing propensity score for a continuous treatment: Application to the efficacy of political advertisements. *Ann Appl Stat* 12:156–177
- García J, Ibáñez SJ, De Santos RM, et al (2013) Identifying Basketball Performance Indicators in Regular Season and Playoff Games. *J Hum Kinet* 36:161–168. <https://doi.org/10.2478/hukin-2013-0016>
- Greifer N (2022a) WeightIt: Weighting for Covariate Balance in Observational Studies
- Greifer N (2022b) cobalt: Covariate Balance Tables and Plots
- Grolemund G, Wickham H (2011) Dates and Times Made Easy with lubridate. *J Stat Softw* 40:1–25. <https://doi.org/10.18637/jss.v040.i03>
- Hastie TJ (1990) Generalized Additive Models, 1st edition. Routledge, Boca Raton, Fla
- Henry L, Wickham H (2022) purrr: Functional Programming Tools
- Hester J, Bryan J (2022) glue: Interpreted String Literals
- Holland PW (1986) Statistics and Causal Inference. *J Am Stat Assoc* 81:945–960. <https://doi.org/10.2307/2289064>
- Loeffelholz B, Bednar E, Bauer KW (2009) Predicting NBA Games Using Neural Networks. *J Quant Anal Sports* 5:. <https://doi.org/10.2202/1559-0410.1156>
- Long J (2019) interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions
- Lumley T (2010) Complex Surveys: A Guide to Analysis Using R, 1st edition. Wiley, Hoboken, N.J
- Lumley T, Scott A (2017) Fitting Regression Models to Survey Data. *Stat Sci* 32:265–278. <https://doi.org/10.1214/16-STS605>
- Malinsky D, Shpitser I, Richardson T (2019) A Potential Outcomes Calculus for Identifying Conditional Path-Specific Effects. *Proc Mach Learn Res* 89:3080–3088

- Milanović D, Štefan L, Škegro D (2016) Situational Efficiency Parameters of Successful and Unsuccessful Top Male Basketball Teams in the Olympic Tournament Games in London 2012. *Balt J Sport Health Sci* 1:. <https://doi.org/10.33607/bjshs.v1i100.44>
- Moerbeek M, Schie S van (2019) What are the statistical implications of treatment non-compliance in cluster randomized trials: A simulation study. *Stat Med* 38:5071–5084. <https://doi.org/10.1002/sim.8351>
- Muller K, Wickham H (2022) *tibble: Simple Data Frames*
- Neuwirth E (2022) *RColorBrewer: ColorBrewer Palettes*
- Ooms J (2014) The jsonlite package: A practical and consistent mapping between json data and r objects. *ArXiv Prepr ArXiv14032805*
- Pearl J (1995) Causal Diagrams for Empirical Research. *Biometrika* 82:669–688. <https://doi.org/10.2307/2337329>
- Pearl J (2019) *Causal Analysis in Theory and Practice*. <http://causality.cs.ucla.edu/blog/index.php/page/3/>. Accessed 28 Oct 2022
- R Core Team (2022) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Version 4.2.1. Vienna, Austria. URL <https://www.R-project.org/>
- Raudenbush SW (1988) Educational Applications of Hierarchical Linear Models: A Review. *J Educ Stat* 13:85–116. <https://doi.org/10.3102/10769986013002085>
- Ren K (2021) *rlist: A Toolbox for Non-Tabular Data Manipulation*
- Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 66:688–701. <https://doi.org/10.1037/h0037350>
- Schuler MS, Chu W, Coffman D (2016) Propensity score weighting for a continuous exposure with multilevel data. *Health Serv Outcomes Res Methodol* 16:271–292. <https://doi.org/10.1007/s10742-016-0157-5>
- Simpson G (2022) *gratia: Graceful ggplot-Based Graphics and Other Functions for GAMs Fitted using mgcv*
- Song K, Gao Y, Shi J (2020) Making real-time predictions for NBA basketball games by combining the historical data and bookmaker's betting line. *Phys Stat Mech Its Appl* 547:124411. <https://doi.org/10.1016/j.physa.2020.124411>
- Song K, Shi J (2020) A gamma process based in-play prediction model for National Basketball Association games. *Eur J Oper Res* 283:706–713. <https://doi.org/10.1016/j.ejor.2019.11.012>

- Vorland CJ, Brown AW, Dawson JA, et al (2021) Errors in the implementation, analysis, and reporting of randomization within obesity and nutrition research: a guide to their avoidance. *Int J Obes* 45:2335–2346. <https://doi.org/10.1038/s41366-021-00909-z>
- Weber S, Gelman A, Lee D, et al (2018) Bayesian aggregation of average data: An application in drug development. *Ann Appl Stat* 12:1583–1604. <https://doi.org/10.1214/17-AOAS1122>
- Westreich D, Lessler J, Funk MJ (2010) Propensity score estimation: machine learning and classification methods as alternatives to logistic regression. *J Clin Epidemiol* 63:826–833. <https://doi.org/10.1016/j.jclinepi.2009.11.020>
- Wickham H (2022a) *httr: Tools for Working with URLs and HTTP*
- Wickham H (2022b) *httr2: Perform HTTP Requests and Process the Responses*
- Wickham H (2011) The Split-Apply-Combine Strategy for Data Analysis. *J Stat Softw* 40:1–29. <https://doi.org/10.18637/jss.v040.i01>
- Wickham H (2022c) *stringr: Simple, Consistent Wrappers for Common String Operations*
- Wickham H (2010) *ggplot2: Elegant Graphics for Data Analysis*, 1st ed. 2009. Corr. 3rd printing 2010 edition. Springer, New York
- Wickham H, Francois R, Henry L, Muller K (2022) *dplyr: A Grammar of Data Manipulation*
- Wickham H, Girlich M (2022) *tidyr: Tidy Messy Data*
- Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R Stat Soc Ser B Stat Methodol* 73:3–36. <https://doi.org/10.1111/j.1467-9868.2010.00749.x>
- Wyss R, Ellis AR, Brookhart MA, et al (2014) The Role of Prediction Modeling in Propensity Score Estimation: An Evaluation of Logistic Regression, bCART, and the Covariate-Balancing Propensity Score. *Am J Epidemiol* 180:645–655. <https://doi.org/10.1093/aje/kwu181>
- Zhu Y, Coffman DL, Ghosh D (2015) A Boosting Algorithm for Estimating Generalized Propensity Scores with Continuous Treatments. *J Causal Inference* 3:25–40. <https://doi.org/10.1515/jci-2014-0022>