American Journal of Epidemiology, 2025, **00**, 1–8
https://doi.org/10.1093/aje/kwaf120
Advance access publication date June 6, 2025
Breakthroughs in Epidemiology

A causal and replication analysis of claims that jet lag affects team sport performance

Matthew S. Tenan*[,1], Ali R. Rezai[1], Andrew D. Vigotsky[2]

[1]Rockefeller Neuroscience Institute, West Virginia University, Morgantown, WV, United States
[2]Northwestern University, Evanston, IL, United States
*Corresponding author: Matthew S. Tenan, Rockefeller Neuroscience Institute, 33 Medical Center Drive, P.O. Box 9301, Morgantown, WV 26506 (matthew.tenan@hsc.wvu.edu)

Abstract

It is broadly accepted that jet lag impacts human performance in sport; however, this has never been validated within a causal inference framework. Here, we used the potential outcomes framework to determine if jet lag, conditional on game time, causes collegiate football teams to under- or overperform expectations. We also attempted to replicate seminal noncausal analyses in professional football. Nine collegiate football seasons (2013-2022, omitting 2020), for a total of 6245 games were analyzed. A generalized additive model with penalized splines and independence weights was used to evaluate interactive causal effects of hours gained/lost in travel and game time on the probability of away teams beating the Las Vegas spread. Furthermore, we used our collegiate cohort to exactly replicate two previous studies purporting to show that NFL teams are adversely affected by jet lag. Jet lag's effects were compatible with the null hypothesis in both the causal analysis ($P = .133$) and noncausal replications. Practically, if the effects of jet lag are unclear, then it is also unclear whether interventions to treat jet lag in elite sport are warranted for teams crossing fewer than 3 time zones if the primary goal is to optimize winning.

Key words: causal inference; jet lag; reproducibility; team performance; football; circadian disorder.

Introduction

Jet lag is a subtype of circadian sleep disorders[1] induced by a temporary misalignment between an individual's circadian clock and the destination's required sleep/wake schedule.[2] Jet-lagged individuals experience sleep disturbances, daytime fatigue, and impaired concentration[1,3] that depend on travel direction (westward vs eastward) and the number of time zones crossed. It is broadly accepted that jet lag impairs individual human performance, which has been extrapolated to team sports in Position Stands.[4-10] Despite the broad acceptance that jet lag influences team performance, this has not been rigorously demonstrated. Such a demonstration can have important economic (betting, team spending), biomedical (player health and injuries), and cultural (effects of winning) implications.

The causal effect of jet lag on team sport performance remains unclear. In the United States, domestic travel, where flights traverse ≤3 time zones, provides the most common opportunity for athletes (and non-athletes) to experience jet lag. Therefore, regular domestic travel could constitute an important source of performance variation and, thus, influence the probability of winning. Yet, the evidence is mixed. Small lab-based experiments have generally not supported the notion that jet lag impacts surrogate measures of human performance, such as sprinting, or perceptual measures, such as fatigue and sleep quality.[11-13] In contrast, "big data" observational studies suggest jet lag impairs team performance.[14-20] However, these observational studies' methodological shortcomings preclude causal inferences since they are not strongly informed by theory (ie, a putative causal structure). Instead, they employ univariable analyses,[16-19] infer causality from multivariable correlations,[14,15,18,19] and omit data, such as major sports markets in the middle of the continental United States.[16,18-20] These methodological choices may partly explain the heterogeneous and discordant conclusions across the observational literature. For instance, studies have now found jet lag-attributed decrements in performance *only* occur for westward travel (football,[18,20] basketball,[16,17] hockey[17]) or *only* occur for eastward travel (baseball,[14,19] basketball[15]), with some studies indicating that this effect *only* occurs in evening games.[17,18,20] In this work, we provide a principled answer regarding the causal effect of jet lag or circadian misalignment, using the same treatment heuristic of directional travel and game time, on team performance by operating within the potential outcomes framework.

Using a causal framework is particularly difficult in team sports since several sources of variation are not easily quantifiable at scale. To assess performance, it is imperative to standardize the away team's probability of winning a game, particularly based on information known before the game begins. This is not straightforward to estimate in practice, as whether a team wins or loses a game is impacted by many measured and *unmeasured* variables. For instance, player injury, lineup changes, coaching changes, coaching strategy/tactics, home field advantage, winning streaks, team records, and several other factors likely affect performance. One may account for a team's expectation of winning a game by benchmarking their performance

Received: September 18, 2024. **Accepted:** May 12, 2025

against the Las Vegas spread.[18,20] Direct standardization of "winning", in our current examination of jet lag, also accounts for scheduling dynamics inherent to college football that could act as a direct confounder to an analysis of jet lag on performance. Although the general schedule (ie, which teams play at which location) is set 8-12 months in advance, the exact kickoff time for each game is typically only set 6-12 days prior to that game occurring. Television networks largely decide kickoff times to maximize viewership and advertising revenue in concert with their contractual agreements laid out at the conference-level. Therefore, the expected game "quality" and performance expectations directly cause kickoff times to change. The "wisdom of the crowd" Las Vegas spread (Appendix S1), while imperfect, appears to be a reasonable method of determining if a team under- or overperforms expectations by directly standardizing across games, implicitly accounting for a range of unmeasured confounders, thereby satisfying the conditional unconfoundedness assumption for causal inference.

In this study, we attempted to conceptually replicate previous work but increase the level of rigor by using an explicit causal inference framework to determine how much jet lag, conditional on game time, causes a collegiate football team to underperform or overperform relative to expectations across a 9-year period (ie, whether they beat the Las Vegas betting spread). Furthermore, we directly attempted to replicate previous studies indicating that westward travel decreases National Football League (NFL) team performance in the evenings.[17,18] By using both a principled causal inference framework and the naïve analyses employed by other seminal papers in the field, we could evaluate the causal effects of jet lag and kickoff time on collegiate football team performance compared to other scenarios.

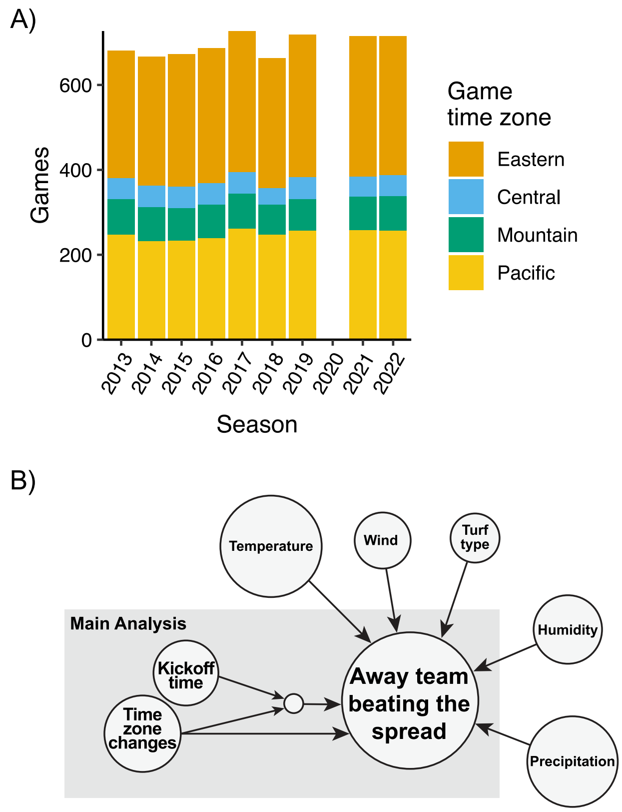

Figure 1. Summary of NCAA football dataset and our proposed directed acyclic graph (DAG) of the data-generating mechanism. (A) Distribution of games played from 2013 to 2022, along with the time zones in which those games were played (ie, home team time zone). A large majority of games were played in the Eastern and Pacific time zones. (B) DAG of variables contributing to the probability of the away team beating the spread. Environmental factors, such as temperature, wind, turf type, humidity, and precipitation will affect the performance of the away team, especially if they are not used to the conditions. We are specifically interested in the role of time zone changes on performance. However, time zone changes do not exist in isolation—their effects may be moderated by kickoff time. For instance, a West Coast team that travels to the East Coast may be more susceptible to performance decrements if the game is played earlier in the day than later in the day.

Methods

Throughout the causal real-world evidence study, we follow Pearl's 4-step process for identifying causal relationships (define, assume, identify, and estimate)[21] plus a final fifth step of "inference", which we believe he underemphasizes. Our "Causal Structure" section focuses on defining the estimand, explicitly stating our assumptions, and determining that the estimand is identifiable (steps 1-3). Estimation and inference (steps 4 and 5) are covered in subsequent sections ("Multivariate Weights" and "Causal Inference via Randomization Inference"), which are complex due to the multivariate treatment/exposure and a desire to not assume linearity of the treatment effects.

Data sources and preprocessing

The data for this work arose from multiple sources: Pro Football Focus (PFF), Wikipedia, Google Maps, gridMET, collegefootballdata.com, and the "lutz" R package.[22] First, the National Collegiate Athletics Association (NCAA) Football teams in PFF were matched with the physical locations of each institution from Wikipedia.[23] The location of each institution was then converted into a latitude and longitude via the Google Maps application programming interface (API) and used to obtain the time zone of each location via the "lutz" R package. The PFF API was then queried for all game information from the 2013-2022 NCAA football seasons. The PFF game information included the stadium at which the game was played and the latitude/longitude coordinates, which were subsequently cross-indexed to obtain the time zone for each stadium, accounting for locations that do not observe daylight savings time. The stadium latitude, longitude, and game date were then matched with associated meteorological data from the gridMET database.[24] Finally, each game was matched with the historical consensus Vegas spread data from collegefootballdata.com via team names and game scores for each season.[25] In our current analysis, we used the "consensus line" for the day before each game was played, which should best represent the team's expected performance.

This process rendered 6245 complete games for analysis (Figure 1A). This publicly available data does not require Institutional Review Board review or approval (45 CFR Part 46). All game times were standardized to the US Eastern Time Zone and rounded to their nearest hour on a 24-hour clock, which decreased computational costs and mitigated model nonconvergence. The time difference between the away team's time zone and the time zone in which the game was played was calculated. Observations where the home team played at a stadium located in a different time zone than their institution were removed. All games where either the stadium's location or the away team's institutional location was not in the continental United States were removed. As weather data is not pertinent when games are played inside, wind and precipitation weather variables for inside games were set to 0, humidity was randomly sampled from a bounded 45-65 uniform distribution, and temperature was randomly sampled from a bounded 70-75 uniform distribution. Finally, binary variables for away team winning the game and away team beating the Vegas spread were created where winning and beating the spread were coded as 1 and losing and not beating the spread were coded as 0 for each game.

Causal structure

The goal of the current work was to ascertain if the estimand for the directed acyclic graph in Figure 1B (main analysis) was more likely to occur than random chance when many potential outcomes were simulated (see following Randomization Inference section).

$$\partial\tau\,(\mathbf{t},\mathbf{x}) = \mathbb{E}\,[\nabla_t Y\,(\mathbf{t}) \mid X = \mathbf{x}]$$

We examined the above estimand equation, which is the marginal conditional average treatment effect (CATE). This is the instantaneous rate of change (∂) of the average treatment effect (τ) for each hour gained/lost in travel (t), conditional on kickoff time (x). This is estimated from the expected value ($\mathbb{E}$) of the partial derivative (∇_t) of the away team's performance ($Y(\cdot)$) as a function of hours gained/lost in travel on that game's performance ($Y(t)$), conditional on the kickoff time for that game (X). The hours gained/lost in travel, conditional on kickoff time, has been used by previous larger scale studies as a proxy for jet lag or circadian misalignment, in lieu of direct measures of an individual's sleep-wake cycle.[14-19] The term "jet lag" is used throughout the manuscript both for brevity and to be consistent with previous literature but readers are encouraged to keep the actual estimand in mind when interpreting the totality of our analytical results and extensions to future work.

The primary outcome variable of interest, the away team beating the Las Vegas spread, encodes a lot of underlying information, such as player personnel, coaching, home-field advantage, and other expectations about team play and capabilities. One thing the pregame consensus Vegas spread cannot fully account for is weather patterns that impact gameplay during the game as this is only fully known *post hoc* (Figure 1B). Additionally, our directed acyclic graph (DAG) assumes that the Vegas spread is not capturing causal information regarding hours gained/lost from time zone changes, the conditional effect of kickoff time on time zone change, and turf type (field turf, artificial turf, or real grass). As it is vitally important to our analysis and conclusions that the Vegas spread does not already capture causal information on the impact of jet lag in our cohort, a secondary analysis was performed to confirm the Vegas spread did not reliably capture salient information concerning jet lag, as captured by hours gained/lost from time zone changes with effect modification of kickoff time (P = .448; Appendix S2 and Figure S1). The effect modification of kickoff time is shown on the DAG per the conventions proposed by Attia *et al.*[26]

Multivariate weights

Current evidence in the field of causal inference indicates that variables unrelated to the exposure but related to the outcome should always be included in a propensity model as it decreases the variance of the exposure effect without a change in bias.[27] Therefore, it was necessary to include all weather variables and turf types with causal paths to the outcome only. To adjust for these variables, we derived multivariate energy-balancing weights according to the method by Huling *et al.*, which uses energy distance covariance to mitigate the dependence between the treatment and competing exposures.[28,29] Briefly, the distance covariance optimal independence weight (DCOW) is useful for both continuous[28] and binary treatments[29] in that it does not rely on correctly specifying the conditional density, a common requirement for inverse probability weighting methods, and does not require hyperparameter tuning or choosing which moments to decorrelate, as is common with many methods for directly estimating propensity weights (eg, covariate balancing propensity score or entropy balancing weights). In our empirical testing, we found the DCOW method to far outperform other weighting methods both in terms of not yielding extreme weights and minimizing the correlation between covariates and exposure.

In the context of the current experiment, we want it equally likely that each game could have theoretically been "assigned" to any level of the effect modification "treatment" (kickoff time and hours gained/lost from time zone changes, see Figure 1B).[30] In addition to explicitly having the covariates in Figure 1B, the away team was also added to the energy-balancing procedure. This addition makes theoretical sense as a team like the University of North Carolina-Chapel Hill on the East Coast will have time zone changes ranging from 0 to +3 as they travel from east to west, and a team such as the University of Texas-Austin will have time zone changes ranging from −1 to +2, depending on if they travel to the East Coast or toward the West Coast. Similarly, many East Coast teams are likely to have a higher probability of playing earlier in absolute time than West Coast teams as they will play relatively more games in the Eastern Time Zone (eg, the University of Southern California would not play a home game at noon Eastern Time, as this is 9 AM Pacific Time). Balancing the multivariate treatments should help account for these known structural effects in the data.

Prior to weighting, the distance covariance was 1.030 and the final weights had a multivariate DCOW of 0.011, where a DCOW of "0" indicates complete independence. After creating the energy balancing weights, the univariable covariate balance was checked to verify that each covariate was balanced across the effect modification treatment by examining the Pearson correlation in both the weighted and unweighted samples.[31] The goal of the univariable balance diagnostics was to confirm that the correlation between the covariates and effect modification treatment decreased after weighting and minimally was below the widely used 0.1 threshold.[32,33] The plot of adjusted and unadjusted univariable correlation with the effect modification treatment can be seen in Figure 2.

Causal inference via randomization inference

A generalized additive model (GAM) enables us to model the nonlinear interactive relationship of hours gained or lost in travel and kickoff time (exposure in the days preceding the game) on the binary outcome, beating the Las Vegas spread. Our exposure assumes that hours gained or lost in travel is a proxy for jet lag, and jet lag's effects are moderated by kickoff time. We used a smoothed effect for hours gained/lost with 5-dimension bases and a tensor product interaction smooth for the effects of hours gained/lost and kickoff time with 7-dimension bases. In doing so, we assumed that history dependence was negligible, such that so-called "repeated exposures" (eg, an East Coast team playing on the West Coast multiple times) do not have cumulative effects since teams typically travel back to their own time zone between games and have ample time to readjust to their home time zone before traveling again. We originally assumed that this model would require a nested structure, but away team-level intraclass correlation coefficients (ICCs) were negligible (<0.001),[34] suggesting this additional complexity, which introduced considerable

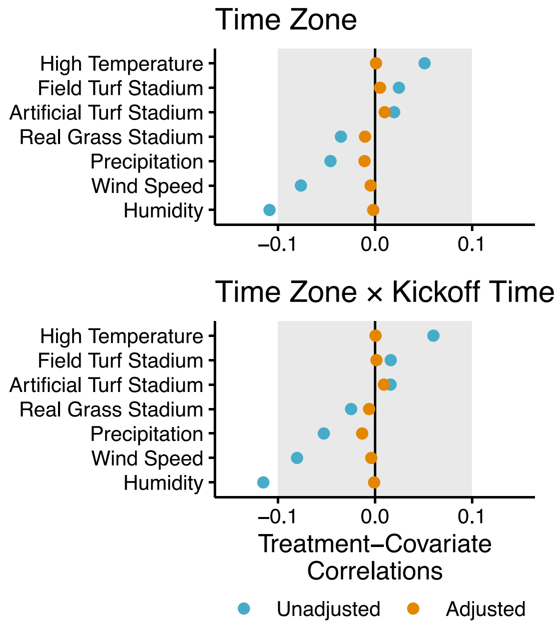

Figure 2. Love plot showing that the energy-balancing weight adjustment balances the treatments across covariates by attenuating the treatment-covariate correlations. These weights were used in our subsequent causal analysis that was fit using a GAM.

computational demand, was unnecessary. Although various probability weighting methods are firmly established for linear versions of clustered or time-series data,[35,36] none can handle splines to account for nonlinear relationships like GAMs. To our knowledge, only 2 popular software packages implement weighting in GAMs: Stata and R.[37,38] Both software packages implement frequency weights, which will produce correct parameter estimates but invalid SEs, precluding valid interval coverage and frequentist hypothesis testing rates. Standard errors that consider the unique properties of probability weights have not yet been developed for GAMs.

In many scenarios, including linear regression, bootstrapping would be a valid solution for extracting appropriate SEs and CIs; however, it has been previously demonstrated that bootstrapping is not pointwise valid with GAMs and could lead to misleading inferences.[39,40] After an extensive survey of the literature and computational software, we concluded that short of deriving new estimators, randomization inference was the only viable solution.

Randomization inference of *F*-statistics was performed to facilitate causal inferences.[41] This involved fitting an intercept-only "null model" to predict the outcome measure (away team beating the spread) while including the independence weights. We then performed an analysis of variance (ANOVA) by comparing the null model with the full model (including the smoothed effects) to obtain an *F*-statistic. While this *F*-statistic is not valid for inference based on a standard *F*-distribution, it can be contrasted to a distribution of *F*-statistics generated from all potential outcomes (or a sufficiently large sample of potential outcomes) to determine how a model fit using our real data compares to a model fit using randomly generated potential outcome data.[41]

To perform randomization inference, we need to build many datasets of treatments (ie, noise treatments) while maintaining the current positivity status. Positivity is the concept that all games should be equally exposed to each given treatment—something our study, by definition, violates (termed structural positivity).[42] Our noise treatments should be theoretically possible and not suggest that the North Carolina Tar Heels can play an away game where they lose 2 hours due to travel, which would imply they were playing somewhere in the Atlantic Ocean. The kickoff time positivity requirement is neither fully random nor fully structural. A school's local time, location of their conference (eg, Atlantic Coast Conference, Southeastern Conference, etc.), and the school makeup of that conference naturally create a differential probability of kickoff times (standardized to the Eastern Time Zone) that a team is likely to play; therefore, the historical kickoff times for away teams in our dataset were used to probabilistically assign a random kickoff time for each game.

Five thousand noise treatment datasets, each with 6245 observations, were simulated. These 5000 datasets were analyzed with the same GAM model used with the real data, and then an ANOVA was performed between that full noise model and the null noise model to render a "noise *F*-statistic". The 5000 noise *F*-statistics were then compared with the observed *F*-statistic. The quantile of the observed *F*-statistic within the distribution of noise *F*-statistics was used to determine the compatibility of our data with the null hypothesis—ie, a *P*-value.

Noncausal replication of previous studies

Smith *et al*. previously examined the number of points by which a West Coast NFL team beat the spread, and we exactly replicated their analysis.[18] Briefly, the outcome variable was the score of the West Coast team minus the score of the East Coast team, plus the point spread. In this way, the resulting value is positive if the West Coast team beat the point spread and zero or negative if the West Coast team did not beat the point spread. This was then compared to zero using a 1-sample *t*-test. We also performed the above analysis only for evening games, which Smith *et al*. defined as games starting after 8 PM Eastern Time.[18]

Roy and Forest[17] are less explicit in their analytical plans than Smith *et al*.[18] Their primary analysis appears to be a linear probability model, where the binary (0 (loss) and 1 (win)) dependent variable is modeled as a linear function of the number of time zones crossed, only for events occurring after 5:30 PM Eastern Time.[17] Notably, the numbering of time zones in the Roy and Forest analysis is the inverse of what is used in our primary causal analysis (ie, −3 in Roy and Forest corresponds to Westward travel of 3 time zones, whereas −3 corresponds to Eastward travel of 3 time zones in our causal analysis). For the purposes of replication, we modified our data to correspond with Roy and Forest.

Results

The analysis includes 6245 games and played by 131 unique teams, across the Eastern (2818), Central (2193), Mountain (594), and Pacific (640) time zones (Figure 3A). Eastward-traveling teams tended to start games relatively earlier, while westward-traveling teams tended to start games relatively later (Figure 3B). There was no strikingly clear bias for teams beating the Vegas spread, but such an inference could be considered "naïve" since it does not consider many other factors that affect performance. Our DAG provides a framework for our causal modeling.

Randomization inference revealed that the conditional effect of jet lag on the away team's probability of beating the Las Vegas spread was compatible with the null hypothesis of "no effect" (P = .133, Figure 4A). In addition, the findings of our replication attempts based on previously reported NFL data were also highly compatible with the null hypothesis of "no effect" of jet lag on winning. Using Smith *et al*.'s exact data reduction and analysis method,[18] our data in a collegiate population show a

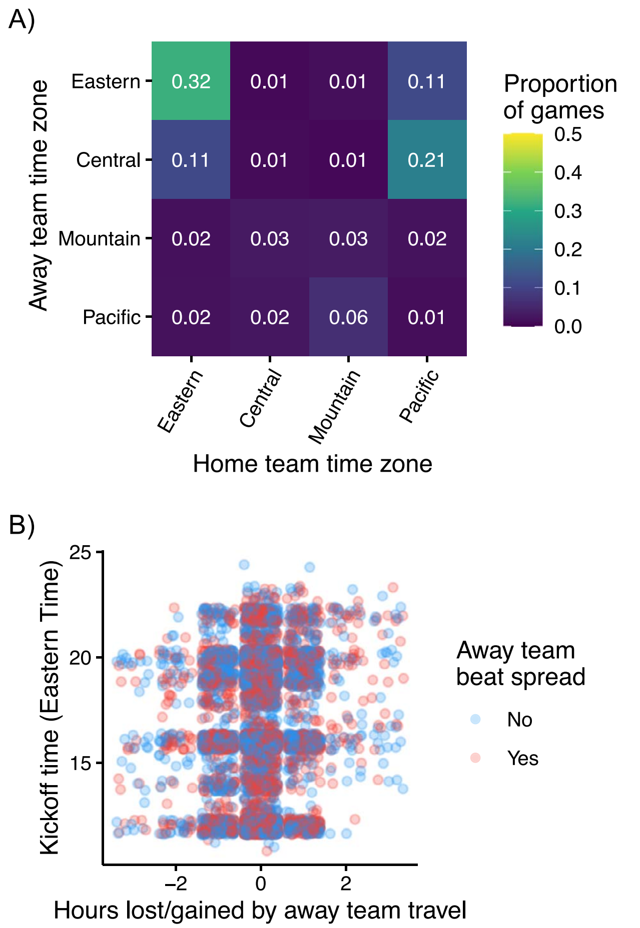

Figure 3. (A) Joint distribution of home and away team time zones. Most of the games played in the Eastern time zone involve away teams from the East Coast, whereas most of the games played in the Pacific time zone involve teams from Eastern and Central time zones.
(B) Distribution of hours lost/gained by the away team, kickoff time, and whether the away team beat the Vegas spread. Note that the data are sparse with eastward travel and late start times, in addition to westward travel and early start times.

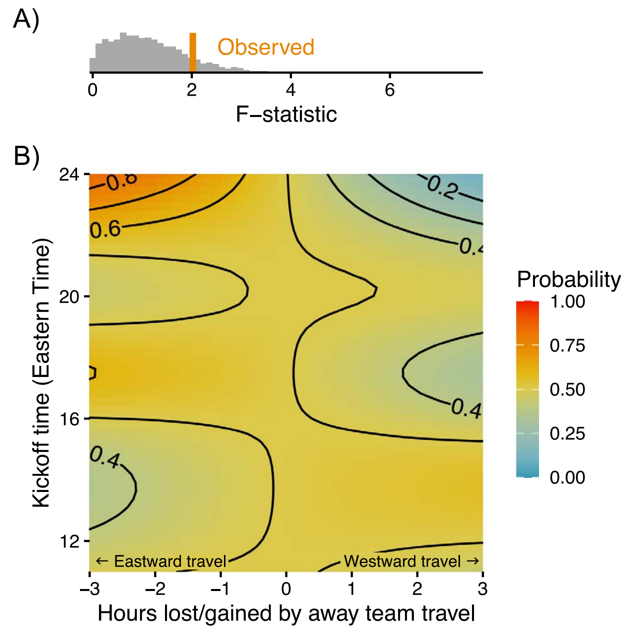

Figure 4. Time zone and kickoff time could not reliably identify jet lag's effects on team performance. (A) Randomization inference analysis for the causal effect of jet lag and kickoff time on beating the Vegas spread. *F*-statistic for observed effects of jet lag and kickoff time (solid vertical line) contrasted against the distribution of *F*-statistics for 5000 potential outcomes when kickoff time is randomized proportionally to existing data and hours gained/lost is fully random but consistent with positivity assumption. This resulted in a *P*-value of 0.133. (B) Surface plot of observed point estimates for the causal effect of jet lag (hours lost/gained in travel) and kickoff time on the probability of beating the Vegas spread, fit using a GAM with independence weights. Point estimate probabilities are denoted by the black lines. The left-hand side of the surface plot illustrates eastward-traveling West Coast teams (ie, "losing hours" in time zone changes) and the right-hand side of the plot represents westward-traveling East Coast teams (ie, "gaining hours" due to changing time zones). All kickoff times have been standardized to the Eastern Time Zone. The randomization inference helps put the surface plot into perspective, where some of the point estimates appear to have appreciable magnitudes (eg, 0.4), which may be at least partially attributable to data sparsity and poor estimates.

negligible effect that is highly compatible with the null hypothesis ($\hat{\beta} = -0.081 \pm 1.327$, $t = -0.061$, $P = .952$). Our findings were consistent even after we subset to include only evening games ($\hat{\beta} = -1.111 \pm 2.204$, $t = -0.504$, $P = .617$). Similarly, while Roy and Forest[17] demonstrated a large and positive relationship between jet lag and the probability of winning ($\hat{\beta} = 5.5$, $P = .049$, $R^2 = 0.113$), the same analysis in our cohort resulted in a minute negative relationship ($\hat{\beta} = -0.023$, $P = .056$, $R^2 = 0.001$). The low *P*-values in both datasets are at least partially attributable to pseudoreplication that inflates the degrees of freedom and decreases the *P*-values calculated from these models. Our follow-on analyses from replicating Roy and Forest's method indicated that failure to account for team-level correlations in winning was a meaningful flaw in their approach (ICC = 0.103).

Discussion

Empirical data are incompatible with a causal effect of jet lag

If jet lag adversely impacted performance, one would expect the probability of beating the spread to be systematically lower as hours are gained or lost in travel, particularly in the evening. However, our causal inference model produced a pattern of point estimates visually inconsistent with this jet lag hypothesis (Figure 4B) and compatible with the null hypothesis ($P = .133$). Although some of the point estimates are appreciable, they are likely imprecisely estimated and compatible with the null hypothesis of "no effect". Indeed, an exploratory analysis of point estimates using randomization inference and a step-down maxT correction for family-wise error rates concluded that the lowest *P*-value of all 98 predictions was 0.10 and compatible with the null hypothesis—about as likely as a fair coin landing on heads 3 times in a row. Indeed, if one takes the point estimates in Figure 4B at face value, the top-left corner would agree with the previous findings that West Coast teams playing on the East Coast beat the spread more often.[18] Despite the direction of individual estimates, they are compatible with the null hypothesis and have a high level of uncertainty. Given that 6245 games worth of data yielded such imprecise estimates that were compatible with the null hypothesis, it at least raises questions concerning the reliability, homogeneity, and practical significance of jet lag's effects.

Attempts to replicate naïve analyses for noncausal effects of jet lag

Our causal inference results may be discordant with previous findings due to (a) differences in datasets and/or (b) differences in analyses. To understand the extent to which our results may be specific to our causal inference analysis, we replicated the analysis of Smith *et al.*,[18] who studied the number of points by which West Coast teams beat the spread across 5 decades of NFL data. They found that West Coast teams beat the spread by more points than one would expect from random chance, a finding we could not reproduce with 9 years of collegiate data.

Additionally, Roy and Forest[17] analyzed *all* evening NFL games over 5 years to examine win-loss records, not limited to just East Coast versus West Coast games; this analysis allows for more uncontrolled confounding but also for the exploration of jet lag or circadian misalignment effects across multiple time zones. Their linear probability model used the number of time zones traveled and the direction to predict the probability of winning. Their analysis unveiled an appreciable time zone effect. In our collegiate football population, however, we did not observe similar results using the same analysis. While both our studies found "statistically significant effects", they differ in magnitude and direction, and the inferential statistics fall prey to inflated degrees of freedom. Thus, our discordant findings are *at least* due to differences in datasets. Future work should evaluate how more rigorous analyses, such as the one performed in this current study, would affect the inferences drawn from other datasets.

Implications of the current work

College football has many scheduling similarities with the NFL in the United States: teams play 1 game roughly every 5-8 days with a "bye week" once or twice a season. This stands in stark contrast to basketball, soccer, or baseball where teams play multiple games per week and travel more frequently. As such, if a clear relationship between traversing time zones and either winning or the Vegas spread were to occur, American Football is probably the most quintessential sport to examine as they are less likely affected by consecutive travel or generalized fatigue. Our inability to (1) ascertain a clear causal relationship between jet lag and human performance and (2) replicate other noncausal analyses call into question previous beliefs that jet lag meaningfully and reliably impacts team performance in American football.

Although we are unable to discount the possibility that some teams in our dataset have attempted to employ jet lag mitigation techniques, our personal experience in collegiate athletics is that novel interventions are not implemented "at scale" (eg, it is far more tractable to implement a focused intervention in your 3 running backs than an entire team), not implemented systematically, and often quickly discarded when personnel changes occur or when staff become too busy to implement the focused interventions.

While there may be causal evidence to support the notion that phototherapy or melatonin administration alters chronobiology,[43-46] there is currently no causal evidence that these interventions improve performance in team sports and impact winning. In all cases examining jet lag and human performance, objective biomarkers of jet lag are eschewed, such as clock gene polymorphisms or melatonin phase-response curves,[47-49] in favor of assuming that traversing time zones induces jet lag or a circadian misalignment in individuals,[11-20] a potential source of person-level variation. Our study is the first to employ causal modeling on a large real-world dataset. Despite the size of our dataset, we were unable to ascertain jet lag's causal effects—this could be due to effect heterogeneity if any effects indeed exist or, perhaps more simply, that winning is just highly stochastic. Since the causal effects of jet lag themselves are so difficult to ascertain, we contend that it be incumbent on purveyors and proponents of jet lag therapies to causally demonstrate real-world efficacy and cost-effectiveness of these interventions on health or human performance and not rely on small, lab-based research or noncausal real-world correlations.

Conclusion

Using rigorous causal inference methods, we have shown that continental travel across time zones (ie, jet lag) has no discernable effect (relative to potential outcomes) on team performance in collegiate football. Both this causal result and the noncausal relational results contrast previous work on this topic, which decidedly concluded that jet lag had salient detrimental effects on performance. Notably, even when we replicated the analyses of seminal papers in this area, our findings were dramatically different from the findings in the original papers. To our knowledge, this is the first causal inference analysis on real-world data examining jet lag and also the first analysis of jet lag on team performance in intercollegiate athletes. Our results have direct applications to team performance in this amateur population and suggest that a higher level of analytical rigor, including the estimation of causal effects and their uncertainty, may be necessary to understand the jet lag dynamics in team athletics overall.

Acknowledgments

An earlier version of this work was presented at the 2024 American College of Sports Medicine Annual Meeting and a preprint was available at: https://doi.org/10.1101/2023.10.19.23296960.

Supplementary material

Supplementary material is available at *American Journal of Epidemiology* online.

Funding

None reported.

Conflict of interest

The authors declare no conflicts of interest.

Disclaimer

None.

Data availability

The final analysis was performed in the R programming language (version 4.3.1) and uses the following packages: mgcv, gratia, permuco, dplyr, WeightIt, cobalt, pbapply, lme4, ggplot2, ggpubr, stringr, stringdist, and independenceWeights. Fully reproducible code and data for the current analysis are available on the lead author's Github page: https://github.com/TenanATC/JetLag_AmJEpi.

Author contributions

M.S.T. (Conceptualization [equal], Data curation [lead], Formal analysis [lead], Investigation [lead], Methodology [equal], Project administration [equal], Visualization [supporting], Writing—original draft [equal], Writing—review & editing [equal]); A.R. (Conceptualization [supporting], Funding acquisition [lead], Investigation [supporting], Supervision [supporting], Writing—review & editing [supporting]); A.D.V. (Formal analysis [supporting], Investigation [supporting], Methodology [supporting], Validation [equal], Visualization [lead], Writing—review & editing [equal]).

References

1. *International Classification of Sleep Disorders, 3rd Edition*. American Academy of Sleep Medicine; 2014.
2. Eastman CI, Gazda CJ, Burgess HJ, et al. Advancing circadian rhythms before eastward flight: a strategy to prevent or reduce jet lag. *Sleep*. 2005;28(1):33-44. https://doi.org/10.1093/sleep/28.1.33
3. Arendt J. Managing jet lag: some of the problems and possible new solutions. *Sleep Med. Rev.* 2009;13(4):249-256. https://doi.org/10.1016/j.smrv.2008.07.011
4. Reilly T, Waterhouse J, Edwards B. Jet lag and air travel: implications for performance. *Clin. Sports Med.* 2005;24(2):367-380. https://doi.org/10.1016/j.csm.2004.12.004
5. Reilly T, Atkinson G, Waterhouse J. Travel fatigue and jet-lag. *J. Sports Sci.* 1997;15(3):365-369. https://doi.org/10.1080/026404197367371
6. Forbes-Robertson S, Dudley E, Vadgama P, et al. Circadian disruption and remedial interventions. *Sports Med.* 2012;42(3):185-208. https://doi.org/10.2165/11596850-000000000-00000
7. Manfredini R, Manfredini F, Fersini C, et al. Circadian rhythms, athletic performance, and jet lag. *Br. J. Sports Med.* 1998;32(2):101-106. https://doi.org/10.1136/bjsm.32.2.101
8. Lee A, Galvez JC. Jet lag in athletes. *Sports. Health.* 2012;4(3):211-216. https://doi.org/10.1177/1941738112442340
9. Janse van Rensburg DC, Jansen van Rensburg A, Fowler PM, et al. Managing travel fatigue and jet lag in athletes: a review and consensus statement. *Sports Med.* 2021;51(10):2029-2050. https://doi.org/10.1007/s40279-021-01502-0
10. Reilly T, Atkinson G, Edwards B, et al. Coping with jet-lag: a position statement for the European College of Sport Science. *Eur. J. Sport Sci.* 2007;7(1):1-7. https://doi.org/10.1080/17461390701216823
11. Fowler P, Duffield R, Vaile J. Effects of simulated domestic and international air travel on sleep, performance, and recovery for team sports. *Scand. J. Med. Sci. Sports.* 2015;25(3):441-451. https://doi.org/10.1111/sms.12227
12. Fowler P, Duffield R, Vaile J. Effects of domestic air travel on technical and tactical performance and recovery in soccer. *Int. J. Sports Physiol. Perform.* 2014;9(3):378-386. https://doi.org/10.1123/ijspp.2013-0484
13. Richmond LK, Dawson B, Stewart G, et al. The effect of interstate travel on the sleep patterns and performance of elite Australian rules footballers. *J. Sci. Med. Sport.* 2007;10(4):252-258. https://doi.org/10.1016/j.jsams.2007.03.002
14. Song A, Severini T, Allada R. How jet lag impairs Major League Baseball performance. *Proc. Natl. Acad. Sci.* 2017;114(6):1407-1412. https://doi.org/10.1073/pnas.1608847114
15. Leota J, Hoffman D, Czeisler MÉ, et al. Eastward jet lag is associated with impaired performance and game outcome in the National Basketball Association. *Front. Physiol.* 2022;13. https://doi.org/10.3389/fphys.2022.892681
16. Glinski J, Chandy D. Impact of jet lag on free throw shooting in the National Basketball Association. *Chronobiol. Int.* 2022;39(7):1001-1005. https://doi.org/10.1080/07420528.2022.2057321
17. Roy J, Forest G. Greater circadian disadvantage during evening games for the National Basketball Association (NBA), National Hockey League (NHL) and National Football League (NFL) teams travelling westward. *J. Sleep Res.* 2018;27(1):86-89. https://doi.org/10.1111/jsr.12565
18. Smith RS, Efron B, Mah CD, et al. The impact of circadian misalignment on athletic performance in professional football players. *Sleep.* 2013;36(12):1999-2001. https://doi.org/10.5665/sleep.3248
19. Recht LD, Lew RA, Schwartz WJ. Baseball teams beaten by jet lag. *Nature.* 1995;377(6550):583-583. https://doi.org/10.1038/377583a0
20. Smith RS, Guilleminault C, Efron B. Orcadian rhythms and enhanced athletic performance in the National Football League. *Sleep.* 1997;20(5):362-365.
21. Pearl J. An introduction to causal inference. *Int J Biostat.* 2010;6:7. https://doi.org/10.2202/1557-4679.1203
22. Teucher, A., Rudis, B. lutz: Look Up Time Zones of Point Coordinates. 2019.
23. *List of NCAA Division I institutions*. Wikipedia; 2023.
24. Abatzoglou JT. Development of gridded surface meteorological data for ecological applications and modelling. *Int. J. Climatol.* 2013;33(1):121-131. https://doi.org/10.1002/joc.3413
25. Gilani, S., Easwaran A, Lee J. et al. *cfbfastR: Access College Football Play by Play Data*. 2022.
26. Attia J, Holliday E, Oldmeadow C. A proposal for capturing interaction and effect modification using DAGs. *Int. J. Epidemiol.* 2022;51(4):1047-1053. https://doi.org/10.1093/ije/dyac126
27. Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. *Am. J. Epidemiol.* 2006;163(12):1149-1156. https://doi.org/10.1093/aje/kwj149
28. Huling JD, Greifer N, Chen G. Independence weights for causal inference with continuous treatments. *J. Am. Stat. Assoc.* 2024;119(546):1657-1670. https://doi.org/10.1080/01621459.2023.2213485
29. Huling JD, Mak S. Energy balancing of covariate distributions. *J. Causal Inference.* 12(1):20220029. https://doi.org/10.1515/jci-2022-0029
30. Fong C, Hazlett C, Imai K. Covariate balancing propensity score for a continuous treatment: application to the efficacy of political advertisements. *Ann. Appl Stat.* 2018;12(1):156-177. https://doi.org/10.1214/17-AOAS1101
31. Austin PC. Assessing covariate balance when using the generalized propensity score with quantitative or continuous exposures. *Stat. Methods Med. Res.* 2019;28(5):1365-1377. https://doi.org/10.1177/0962280218756159
32. Rassen JA, Brookhart MA, Glynn RJ, et al. Instrumental variables II: instrumental variable application—in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J. Clin. Epidemiol.* 2009;62(12):1233-1241. https://doi.org/10.1016/j.jclinepi.2008.12.006
33. Zhu Y, Coffman DL, Ghosh D. A boosting algorithm for estimating generalized propensity scores with continuous treatments. *J Causal Inference.* 2015;3(1):25-40. https://doi.org/10.1515/jci-2014-0022
34. Hox J, Moerbeek M, van de Schoot R. *Multilevel Analysis: Techniques and Applications*. 3rd ed. New York: Routledge; 2018.
35. Schuler MS, Chu W, Coffman D. Propensity score weighting for a continuous exposure with multilevel data. *Health Serv. Outcomes Res. Methodol.* 2016;16(4):271-292. https://doi.org/10.1007/s10742-016-0157-5
36. Arpino B, Mealli F. The specification of the propensity score in multilevel observational studies. *Comput Stat Data Anal.* 2011;55(4):1770-1780. https://doi.org/10.1016/j.csda.2010.11.008
37. Royston P, Ambler G. GAM: Stata module for generalised additive models. 2012.

38. Wood SN. *Generalized Additive Models: An Introduction with R*. 2nd ed. Boca Raton: (Chapman and Hall/CRC; 2017.
39. Wood SN. On confidence intervals for generalized additive models based on penalized regression splines. *Aust. N. Z. J. Stat.* 2006;48(4):445-464. https://doi.org/10.1111/j.1467-842X.2006.00450.x
40. Wang Y, Wahba G. Bootstrap confidence intervals for smoothing splines and their comparison to Bayesian confidence intervals. *J Stat Comput Simul.* 1995;51(2-4):263-279. https://doi.org/10.1080/00949659508811637
41. Gerber AS, Green DP. *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton & Company; 2012.
42. Westreich D, Cole SR. Invited commentary: positivity in practice. *Am. J. Epidemiol.* 2010;171(6):674-677. https://doi.org/10.1093/aje/kwp436
43. Herxheimer A, Petrie KJ, Cochrane Common Mental Disorders Group. Melatonin for the prevention and treatment of jet lag. *Cochrane Database Syst. Rev*. 2010;2010(1):CD001520. https://doi.org/10.1002/14651858.CD001520
44. Burke TM, Markwald RR, Chinoy ED, et al. Combination of light and melatonin time cues for phase advancing the human circadian clock. *Sleep.* 2013;36(11):1617-1624. https://doi.org/10.5665/sleep.3110
45. Crowley SJ, Eastman CI. Phase advancing human circadian rhythms with morning bright light, afternoon melatonin, and gradually shifted sleep: can we reduce morning bright light duration? *Sleep Med.* 2015;16(2):288-297. https://doi.org/10.1016/j.sleep.2014.12.004
46. LeGates TA, Fernandez DC, Hattar S. Light as a central modulator of circadian rhythms, sleep and affect. *Nat. Rev. Neurosci.* 2014;15(7):443-454. https://doi.org/10.1038/nrn3743
47. Boivin DB, James FO, Wu A, et al. Circadian clock genes oscillate in human peripheral blood mononuclear cells. *Blood.* 2003;102(12):4143-4145. https://doi.org/10.1182/blood-2003-03-0779
48. Khalsa SBS, Jewett ME, Cajochen C, et al. A phase response curve to single bright light pulses in human subjects. *J. Physiol.* 2003;549(3):945-952. https://doi.org/10.1113/jphysiol.2003.040477
49. Franken P, Dijk D-J. Circadian clock genes and sleep homeostasis. *Eur. J. Neurosci.* 2009;29(9):1820-1829. https://doi.org/10.1111/j.1460-9568.2009.06723.x

American Journal of Epidemiology, 2025, **00**, 1–8
https://doi.org/10.1093/aje/kwaf120
Breakthroughs in Epidemiology