

Collider bias in administrative workers' compensation claims data: A challenge for cross-jurisdictional research

Tyler J Lane (✉ tyler.lane@monash.edu)

School of Public Health and Preventive Medicine, Monash University <https://orcid.org/0000-0001-6089-1827>

Research Article

Keywords: Collider bias, left-censoring, workers' compensation, cross-jurisdictional comparisons

Posted Date: May 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-383505/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Purpose Workers' compensation claims consist of occupational injuries severe enough to meet a compensability threshold. Theoretically, systems with higher thresholds should have fewer claims but greater average severity. For research that relies on claims data, particularly cross-jurisdictional comparisons of compensation systems, this results in collider bias that can lead to spurious associations and confound analyses. In this study, I use real and simulated claims data to demonstrate collider bias and problems with methods used to account for it.

Methods Using Australian claims data, I used a linear regression to test the association between claim rate and mean disability durations across Statistical Areas. Analyses were repeated with nesting by state/territory to account for variations in compensability thresholds across compensation systems. Both analyses are repeated on left-censored data. Simulated claims data are analysed with Cox survival analyses to illustrate how left-censoring can reverse effects.

Results The claim rate within a Statistical Area was inversely associated with disability duration. However, this reversed when Statistical Areas were nested by state/territory. Left-censoring resulted in an attenuation of the unnested association to non-significance, while the nested association remained significantly positive. Cox regressions on simulated data showed left-censoring can also reverse effects.

Conclusions Collider bias can seriously confound work disability research, particularly cross-jurisdictional comparisons. Work disability researchers must grapple with this challenge by using appropriate study designs and analytical approaches, and considering how collider bias affects interpretation of results.

Introduction

Why do the best-looking Hollywood stars often seem to have the least talent? Do taller NBA players have a scoring advantage? And what do this have to do with cross-jurisdictional work disability comparative research, the subject of this special issue of the *Journal of Occupational Rehabilitation* (1)? In this paper, I will link these three examples around a unintuitive yet vexing problem known as collider bias (2–4).

Motivating examples

In the general population, beauty and talent are unrelated. But it seems the opposite is true among stars in Hollywood: greater beauty predicts less talent (5). The reason is that stardom is conditional on some combination of beauty and talent. The more beautiful need less talent, and vice-versa. This is illustrated in Figure 1.

In this example, Hollywood stardom is a collider between beauty and talent. In the language of Directed Acyclic Graphs (DAG), this would be expressed as *beauty* → *Hollywood stardom* ← *talent* (for an introduction to DAGs, see Rohrer 2018 (2) or Pearl and Mackenzie 2018 (3)). Testing an association between beauty and talent among Hollywood stars is an example of *conditioning on the collider*, which

can lead to a spurious association between the two factors contributing to the collider. What is useful about this example is that by using simulated data, we can observe excluded cases, which makes the bias in Figure 1's right-hand obvious. The next example uses real data and while it has floated around the epidemiology community for a while (6), it is worth reproducing here because it can only be explained using collider bias.

Height is a generally considered an advantage in basketball. Taller players should therefore score more points. But as Figure 2 shows, this is not the case among NBA players (7). Once we understand that we are looking at a dataset that has been conditioned on extreme ability (being in the NBA), it becomes clear that whatever advantage height provides, other skills can compensate. If skill was quantifiable and the data were available, we would almost certainly see an inverse relationship with height and no association with points per game. In this case, the DAG language would describe the collider as: *height -> NBA player <- skill*.

Workers' compensation claims data and collider bias

In work disability research, injured workers are often identified through claims data. However, only a fraction of workplace injuries become a claim. Of an estimated 563,000 Australians who experienced a work injury in 2017/18, only 174,000 (31%) applied for workers' compensation and 154,000 received it (27%) (8). Compensated injuries skew towards the more severe; the most common reason injured workers gave for not lodging a claim was because the injury was too minor (43%) (8).

Each compensation system has its own compensability threshold, a term I use to refer to the formal and informal system settings that determine whether a claim is accepted. Waiting periods are a type of formal compensability threshold that is common in American compensation systems and associated with significantly longer disability durations (9). In Australia, most compensation systems have employer excess periods, which are similar to waiting periods except employers are obligated pay lost wages until compensation benefits start (10). There are also informal practices such as purposely delaying liability decisions to cause injured workers to abandon their claim (11). Others plausibly exist, though may be hard to identify because they are unethical or even illegal.

In the DAG language, the collider that shapes workers' compensation claims data can be described as: *compensability threshold -> compensation <- injury severity*. Unfortunately, compensability thresholds do not lend themselves to simple quantification, and while injuries are often rated to determine what benefits an injured worker is entitled to, this is highly subjective and biased (12). This study therefore relies on proxies. The claim rate stands in for compensability thresholds on the assumption that higher thresholds correspond to lower claim rates and vice-versa. Disability duration stands in for injury severity, which it is strongly related to (13,14). The DAG to be tested in this study is the following: *claim rate -> compensation <- disability duration*. Systems with higher compensability thresholds will have fewer claims but higher average severity and longer disability durations.

Collider bias makes it extremely difficult to differentiate cohort-shaping effects (*who* gets into the compensation system) and outcome-shaping effects (*how* the system changes the individuals within them). However, there is some conflation in the work disability research base. For instance, in a study of differences in sick leave across several European countries, the authors conclude that “less strict compensation policies to be eligible for long-term (partial) benefits, contributed to sustained RTW [return to work].” While easier access to compensation benefits could plausibly minimise the iatrogenic effects of being on compensation (15–17), it would expand system access to those who would fare better anyway. Collider bias can make it impossible to tell the difference. For some purposes, this is not much of a problem. When the aim is to reduce system costs, it matters little if savings were achieved by restricting access or improved outcomes. But if the aim is to improve injured worker outcomes, which I will take an editorial opportunity to endorse, it is essential to disentangle cohort and outcome-shaping effects.

In this study, I test whether claim rates and disability duration are in fact inversely correlated. I also demonstrate what happens when variations in compensability thresholds are accounted for by treating compensation systems as fixed effects. Finally, I test the effects of left-censoring disability duration, which I have previously applied to overcome the problem of differences in compensability thresholds (18–20), and adding arbitrary low disability durations to censored claims to include them in survival analysis, which has been suggested as a way to overcome biases due to waiting /employer excess (21).

Methods

Setting

Australia has nine major compensation systems: one in each of the six states and two main territories plus Comcare covers federal government employees and grants self-insurance licenses to certain interstate private employers (22). All are cause-based, meaning compensation is only provided if the injury or illness can be demonstrably linked to work (23). Collectively, these cover 94% of Australia’s workforce (24).

Data

Claim records are from the *National Data Set for Compensation-based Statistics* (NDS), a minimum dataset compiled by Safe Work Australia from each of the major compensation systems (25). Records are limited to those lodged between 2010 and 2015. As of writing, records are updated for six years to July 2017, providing a minimum of 1.5 years of follow-up to calculate disability duration.

I calculated the rate of claims using a labour force denominator (26) and mean disability duration with individual records capped at five years. As the nine state and territory compensation systems provided insufficient data points for analysis, claims were instead aggregated at Statistical Area of residence at Level 4 (27). This also provided an opportunity to examine the association between claim rates and disability duration within a Statistical Area while accounting for compensability threshold fixed effects

(i.e., by state/territory). The Australian Capital Territory was excluded because it is composed of a single Statistical Area at Level 4 and labour force estimates were unavailable.

Statistical analyses

I used linear regression to test the association between claim rates and disability duration. Analyses were repeated in a multi-level linear regression with Statistical Areas nested within states and territories. Claim rates were z-transformed, which mean-centred the distribution at zero and scaled variance in terms of standard deviations, to provide a more meaningful scale of effects. Disability duration was log-transformed to estimate percent rather than absolute changes.

Sensitivity analyses tested the effects of left-censoring at two weeks to account for known compensability thresholds (the longest employer excess periods in Australia, found in Victoria and South Australia). The first sensitivity analysis replicated the analysis above with the left-censored data. The second used a Cox regression model on simulated claims data to demonstrate how left-censoring can cause effects to reverse. A third analysis adopted an approach developed by Sears & Heagerty (21) to counter waiting periods in the Washington compensation system. In the original, the authors assumed many “medical-only” claim had time loss that did not exceed the waiting period. They were arbitrarily assigned low disability durations values (.001) to allow their inclusion in survival analysis. In this study, censored cases are assigned the arbitrary low disability duration value (medical-only and left-censored cases are similar because they reflect an artificial suppression of less severe injuries in time loss outcomes).

Statistical software and packages

Aggregate data and analytical code have been archived on a public repository (28). Data cleaning and analyses were conducted in R (29) using RStudio (30) with the following packages: *broom* (31), *broom.mixed* (32), *janitor* (33), *lubridate* (34), *magick* (35), *magrittr* (36), *nlme* (37), *readxl* (38), *see* (39), *sf* (40), *tidyverse* (41), and *zoo* (42).

Results

Claim rate as a predictor of disability duration

There was an inverse association between claim rates and disability durations within a Statistical Area. A standard deviation increase in the claim rate was associated with a 15.9% decrease in disability duration (95% CI: -22.9% to -8.8%). However, this reversed when Statistical Areas were nested by state and territory. A standard deviation increase in the claim rate was associated with a 6.4% increase in disability duration (95% CI: 2.6–10.1%). These effects are illustrated in Fig. 4, which also shows a distinct clustered ordering of data points by state and territory.

Sensitivity analysis: Effects of left-censoring

When the above analyses were limited to claims with at least two weeks of compensated time loss, the unnested association attenuated by two-thirds and became non-significant (-5.2%; 95% CI: -10.9–0.5%). This suggests a reduction in collider bias. However, the nested association was a 6.4% increase, same point estimate as for all time loss claims, though confidence intervals narrowed, indicating greater precision (95% CI: 4.2–8.6%). The clustered ordering by state and territory remained intact (see Supplementary Fig. 1).

Figure 4 illustrates the results of Cox regressions of all simulated claims data, left-censored, and left-censored with arbitrary low values. Systems A and B were assigned mean disability durations of 0.5 weeks (SD: 0.9) and 1 week (SD: 0.6), both on a log scale to reflect the heavy right skew often found in disability duration (21). Admittedly, I tested multiple iterations to achieve the effects described below, meaning the simulation is somewhat contrived.

When analysing all simulated claims data, those in System B were 23% less likely to exit compensation at each time point (Hazard Ratio: 0.77; 95% CI: 0.74–0.81) relative to System A. When left-censored at two weeks, claims in System became 30% *more likely* to exit compensation (HR: 1.30; 95% CI: 1.22–1.38), a reversal of effects. When left-censored cases were assigned an arbitrary low value to allow their inclusion in analysis, the results were similar to analysis of all claims: those in System B were 18% less likely to exit compensation at each time point. However, the confidence interval did not include the original point estimate (HR: 0.82; 95% CI: 0.79–0.86).

Discussion

This study demonstrates how collider bias confounds cross-jurisdictional comparative research. As predicted, there was an inverse association between claim rate and disability duration. These served as proxies for compensability thresholds and injury severity, suggesting systematic baseline differences in cohorts. This makes it extremely difficult to differentiate a compensation systems cohort-shaping and outcome-shaping effects.

The association reversed when compensation systems were treated as fixed effects in a textbook example of Simpson's Paradox, or more technically, Simpson's Reversal (43). If the positive association can be considered the "true" relationship between claim rate and disability duration (i.e., occupational injury frequency and severity are positively associated in the real world), the results demonstrate how collider bias can mask real effects. To be clear, neither the inverse or positive association are inherently misleading or wrong on their own, though Simpson's Paradox highlights the importance of matching analysis to the research question (44). In this study, the question was whether differences in compensability thresholds between compensation systems produce a spurious association between claim rates and disability duration. This makes the unnested approach the appropriate one, though the nested approach adds insight.

The clustered ordering of states/territories in Figure 3 provides additional evidence of collider bias due to compensability thresholds, in this case employer excess periods. Victoria and South Australia have the

longest excess period in Australia at ten days/two weeks, which is twice the next longest. Correspondingly, both are situated in the upper-left quadrant of Figure 2, denoting fewer claims and longer durations. However, the remaining order of states/territories appears unrelated to employer excess: after Victoria and South Australia comes Western Australia, which has no employer excess period, followed by Tasmania (no employer excess period), Northern Territory (part of first day), New South Wales (one week), and Queensland (around one week). This suggests other compensability thresholds are at work.

The effects of left-censoring and potential for further bias

Left-censoring was applied to account for employer excess periods, a known type of compensability threshold. While the inverse association between claim rate and disability duration attenuated to non-significance with left-censored data, the direction of effect remained negative. However, when Statistical Areas were nested by state and territory, which treated compensability thresholds as fixed effects, the association between claim rate and disability duration was again positive, as it had been in uncensored analysis. As above, if this is the “true” association, it remained masked in unnested analyses even when data were left-censored to account for employer excess periods.

Survival analyses on simulated claims data revealed another potential for reversal of effects due to left-censoring. In a real world setting, this could happen where one system successfully resolves many of the low-severity, easy-to-resolve cases, while another delays their exit until after the censored period, a “depletion of the susceptibles” in reverse (45).

Adding an arbitrary low value to left-censored cases largely accounted for the effects of left-censoring, which is in line with the theoretical paper that proposed this approach (21). However, the simulated data include all censored cases, an advantage that real world data likely lacks. For instance, in real world data the arbitrary low value would be applied to medical-only claims, or those without any recorded time loss. However, we run into a similar problem as with time loss durations: such injuries are only recorded as claims if they are compensated for treatment. Medical care benefits also have compensability thresholds that vary across compensation systems. For instance, as of 2019 Victoria required employers to cover the first \$707 of medical costs, while Queensland required employers to pay the first \$1,527.80 of combined costs (10). Even for medical-only claims, compensation status is a source of collider bias.

Alternative approaches

There are other ways to test whether and how compensation systems improve or worsen injured worker outcomes. Randomised controlled trials avoid much of the problem of collider bias by randomly allocating exposures, theoretically balancing baseline differences between cohorts. However, these are often ethically or financially impractical and can lack external validity (2,46). Quasi-experiments may be better suited to answering questions about the impact of system settings. These include study designs such as interrupted time series and difference-in-differences, which compare outcomes before and after an event like legislated changes, and regression discontinuity, which use arbitrary cut-offs like wage

replacement caps. They can also overcome logistical hurdles of randomised controlled trials through exogenous allocation of large populations to experimental/exposure and control conditions in ways that mimic randomisation (46,47). In some circumstances, quasi-experiments have better external validity than randomised controlled trials because they rely on population-level, real-world data (46).

However, quasi-experiments have important limitations. Events like legislative change are infrequent and may not modify policies of interest, or entirely change the compensation system, making it difficult to differentiate the effects of policy change from service disruption (48). Legislative change may also introduce a collider if it alters compensability thresholds. For instance, we previously found that when New South Wales restricted eligibility to its compensation system in 2012, the claim rate decreased while disability duration increased (49). Such an outcome would be consistent with both a change in the cohort towards more severe and complex injuries *and* a system that has increased the stressfulness of compensation. At the very least, analyses based on legislative change should examine whether the event affected claim rates (48,50) to pro-actively test for cohort-shaping effects as an indicator of collider bias.

Strengths and limitations

Study strengths include use of population-level claims data from workers' compensation systems with near-universal coverage of the Australian workforce and the use of simulated data with known characteristics to demonstrate how collider bias can distort statistical associations. To my knowledge, this is the first work disability study to directly engage with the problem of collider bias.

Limitations include the inability to test the proposed mechanism of compensability thresholds directly. I used labour force denominators to estimate claim rates, which are not equivalent to covered worker estimates; variations in the proportion of the workforce who are insured against workplace injury could vary across Statistical Areas or compensation systems and bias claim rate estimates.

Conclusions

Collider bias is an under-recognised problem in work disability research. In this paper, I present evidence that compensation status is a collider between *injury severity* and *compensability thresholds*, which manifests as an inverse relationship between claim rates and disability duration. This makes it difficult to determine whether differences in disability duration between compensation systems are due to who is compensated or how the system treats them. I also show that left-censoring to account for compensability thresholds such as waiting periods and employer excess may not account for collider bias and may be a source of additional bias.

Randomised controlled trials and quasi-experiments can produce more robust causal estimates of how system factors affect injured worker outcomes, though these have their own theoretical and practical limitations. Work disability researchers must pro-actively engage with the problem of collider bias to improve the reliability of research that depends on compensation data and to better understand the implications of their findings.

Declarations

Acknowledgements

I would like to thank my supervisor and editor of this special edition, Professor Alex Collie, for his comments on an earlier version of this manuscript.

Funding

This study was funded by an Australian Research Council Discovery Project Grant (DP190102473), as part of the Compensation and Return to Work Effectiveness (COMPARE) Project, and by Safe Work Australia, a government statutory agency that develops national work health and safety and workers' compensation policy.

Conflicts of interest

The author previously receives salary support from funding provided by the workers' compensation systems investigated in this study.

Ethics approval

This study received ethics approval from the Monash University Human Research Ethics Committee (CF14/2995 – 2014001663).

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Case-level claims data contain potentially re-identifiable data and are therefore not able to be publicly shared. Aggregate-level claims data used, NBA player data, and simulated data have all available on a public repository (28).

Code availability

All cleaning, data simulation, and analytical code have been archived on a public repository (28).

Author's contributions

This paper is the sole work of Tyler J Lane.

References

1. Journal of Occupational Rehabilitation. Call for Paper: Cross-Jurisdiction Comparisons in Work Disability Systems [Internet]. Springer. 2020 [cited 2020 Nov 23]. Available from: <https://www.springer.com/journal/10926/updates/18025764>.
2. Rohrer JM. Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Adv Methods Pract Psychol Sci*. 2018;1(1):27–42.
3. Pearl J, Mackenzie D. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books; 2018. 432 p.
4. Lee H, Aronson JK, Nunan D. Collider bias. In: *Catalogue of Bias* [Internet]. 2019. Available from: <https://catalogofbias.org/biases/collider-bias/>.
5. Elwert F, Winship C. Endogenous selection bias: The problem of conditioning on a collider variable. *Annu Rev Sociol*. 2014;40:31–53.
6. Hahn M. You can be a professional basketball player, no matter how tall you are! No correlation between height and scoring success in the NBA: <https://t.co/C1XpivlhQc> [Internet]. @3rdreviewer. 2017 [cited 2020 Dec 2]. Available from: <https://twitter.com/3rdreviewer/status/840266678076350464>.
7. Media Ventures NBA. LLC. NBA Advanced Stats - Stats Home / Players / Bio [Internet]. 2020 [cited 2020 Sep 9]. Available from: https://stats.nba.com/players/bio/?Season=2018-19&SeasonType=Regular Season&sort = PLAYER_NAME&dir = 1.
8. Australian Bureau of Statistics. 6324.0 Work-Related Injuries, Australia, Jul 2017 to Jun 2018 [Internet]. 2018 [cited 2020 May 7]. Available from: <https://www.abs.gov.au/ausstats/abs@.nsf/mf/6324.0>.
9. Shraim M, Cifuentes M, Willetts JL, Marucci-Wellman HR, Pransky GS. Length of disability and medical costs in low back pain: Do state workers' compensation policies make a difference? *J Occup Environ Med*. 2015;57(12):1275–83.
10. Safe Work Australia. *Comparison of Workers' Compensation Arrangements in Australia and New Zealand* [Internet]. Canberra, AU: Safe Work Australia; 2020. Available from: <https://www.safeworkaustralia.gov.au/doc/comparison-workers-compensation-arrangements-australia-and-new-zealand-2019>.
11. Spieler EA, Burton JF. The lack of correspondence between work-related disability and receipt of workers' compensation benefits. *Am J Ind Med*. 2012;55(6):487–505.
12. Lax MB, Manetti FA, Klein RA. Medical Evaluation of Work-related Illness: Evaluations by a Treating Occupational Medicine Specialist and by Independent Medical Examiners Compared. *Int J Occup Environ Health*. 2004 Jan;10(1):1–12.
13. Krause N, Dasinger LK, Deegan LJ, Rudolph L, Brand RJ. Psychosocial job factors and return-to-work after compensated low back injury: A disability phase-specific analysis. *Am J Ind Med*. 2001 Oct;40(4):374–92.

14. Dasinger LK, Krause N, Deegan LJ, Brand RJ, Rudolph L. Physical workplace factors and return to work after compensated low back injury: a disability phase-specific analysis. *J Occup Environ Med.* 2000;42(3):323–33.
15. Kilgour E, Kosny A, McKenzie D, Collie A. Healing or Harming? Healthcare Provider Interactions with Injured Workers and Insurers in Workers' Compensation Systems. *J Occup Rehabil.* 2015 Mar;25(1):220–39.
16. Kilgour E, Kosny A, McKenzie D, Collie A. Interactions between injured workers and insurers in workers' compensation systems: A systematic review of qualitative research literature. *J Occup Rehabil.* 2015 Mar;25(1):160–81.
17. Kilgour E, Kosny A, Akkermans A, Collie A. Procedural Justice and the use of Independent Medical Evaluations in workers' compensation. *Psychol Inj Law.* 2015;8(2):153–68.
18. Collie A, Lane TJ, Hassani-Mahmoei B, Thompson J, McLeod C. Does time off work after injury vary by jurisdiction? A comparative study of eight Australian workers' compensation systems. *BMJ Open.* 2016 May;6(5):e010910.
19. Lane TJ, Sheehan LR, Gray SE, Collie A. Regional differences in time off work after injury: a comparison of Australian states and territories within a single workers compensation system [Internet]. *Occupational and Environmental Health*; 2020 Jul [cited 2020 Nov 13]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2020.07.23.20160416>.
20. Lane TJ, Sheehan LR, Gray SE, Beck D, Collie A. Step-downs reduce workers compensation payments to encourage return to work. Are they effective? *Occup Environ Med.* 2020;77:470–7.
21. Sears JM, Heagerty PJ. Including Injured Workers Without Compensated Time Loss in Cox Regression Models: Analyzing Time Loss Using All Available Data. *J Occup Rehabil.* 2008 Sep;18(3):225–32.
22. Collie A, Lane TJ. Australian workers' compensation systems. In: Willis E, Reynolds L, Judge T, editors. *Understanding the Australian Health Care System.* 4th ed. Chatswood: Elsevier Australia; 2019. pp. 208–22.
23. Lippel K, Lötters F. Public Insurance Systems: A Comparison of Cause-Based and Disability-Based Income Support Systems. In: Loisel P, Anema JR, editors. *Handbook of Work Disability* [Internet]. New York, NY: Springer New York; 2013 [cited 2020 Nov 13]. p. 183–202. Available from: http://link.springer.com/10.1007/978-1-4614-6214-9_12.
24. Lane TJ, Collie A, Hassani-Mahmoei B. *Work-related injury and illness in Australia, 2004 to 2014.* Melbourne, AU; 2016.
25. Safe Work Australia. *National Data Set for Compensation-based Statistics Third Edition.* Canberra, AU; 2004.
26. Australian Government. *Employment Region Data* [Internet]. Labour Market Information Portal. 2020 [cited 2020 Jul 29]. Available from: <https://lmip.gov.au/default.aspx?LMIP/Downloads/EmploymentRegion>.

27. Australian Bureau of Statistics. 1270.0.55.001 - Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas, July 2016 [Internet]. 2016 [cited 2020 Jul 30]. Available from: [https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/bySubject/1270.0.55.001 ~ July 2016 ~ Main Features ~ Statistical Area Level 4 \(SA4\) ~ 10016](https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/bySubject/1270.0.55.001~July2016~MainFeatures~StatisticalAreaLevel4(SA4)~10016).
28. Lane TJ. Cross-jurisdictional comparisons using workers' compensation: illustrating collider bias due to compensability thresholds - data and code [Internet]. Bridges. 2021 [cited 2020 Oct 26]. Available from: <https://doi.org/10.26180/13140374>.
29. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2020.
30. RStudio Team. RStudio: Integrated Development for R. Boston: RStudio, Inc.; 2020.
31. Robinson D, Hayes A. broom: Convert Statistical Analysis Objects into Tidy Tibbles. 2020.
32. Bolker B, Robinson D. broom.mixed: Tidying Methods for Mixed Models. 2020.
33. Firke S. janitor: Simple Tools for Examining and Cleaning Dirty Data. 2020.
34. Grolemund G, Wickham H. Dates and Times Made Easy with lubridate. J Stat Softw. 2011;40(3):1–25.
35. Ooms J: Advanced Graphics and Image-Processing in R [Internet]. 2021. Available from: <https://CRAN.R-project.org/package=magick>.
36. Bache SM, Wickham H. magrittr: A Forward-Pipe Operator for R. 2014.
37. Pinheiro J, Bates D, DebRoy S, Sarkar D. R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. 2020.
38. Wickham H, Bryan J. readxl: Read Excel Files. 2019.
39. Lüdecke D, Makowski D, Waggoner P, Ben-Shachar MS. see: Visualisation Toolbox for “easystats” and Extra Geoms, Themes, and Color Palettes for “ggplot2”. 2019.
40. Pebesma E. Simple Features for R: Standardized Support for Spatial Vector Data. R J. 2018;10(1):439–46.
41. Wickham H. tidyverse: Easily Install and Load the “Tidyverse.” 2017.
42. Zeileis A, Grothendieck G. zoo: S3 Infrastructure for Regular and Irregular Time Series. J Stat Softw. 2005;14(6):1–27.
43. Pearl J. Comment. Understanding Simpson's paradox. Am Stat. 2014;68(1):8–13.
44. Hernán MA, Clayton D, Keiding N. The Simpson's paradox unraveled. Int J Epidemiol. 2011;40(3):780–5.
45. Stovitz SD, Banack HR, Kaufman JS. 'Depletion of the susceptibles' taught through a story, a table and basic arithmetic. BMJ Evid-Based Med. 2018 Oct;23(5):199–9.
46. Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin; 2002.
47. Krause N, Frank JW, Dasinger LK, Sullivan TJ, Sinclair SJ. Determinants of duration of disability and return-to-work after work-related injury and illness: challenges for future research. Am J Ind Med.

2001;40(4):464–84.

48. Lane TJ, Gray SE, Hassani-Mahmooei B, Collie A. Effectiveness of employer financial incentives in reducing time to report worker injury: an interrupted time series study of two Australian workers' compensation jurisdictions. *BMC Public Health*. 2018;18(100):1–10.
49. Collie A, Beck D, Gray SE, Lane TJ. Impact of legislative reform on benefit access and disability duration in workers' compensation: An interrupted time series study. *Occup Environ Med*. 2020;77(1):32–9.
50. Pollak C. The impact of a sick pay waiting period on sick leave patterns. *Eur J Health Econ*. 2017 Jan;18(1):13–31.

Figures

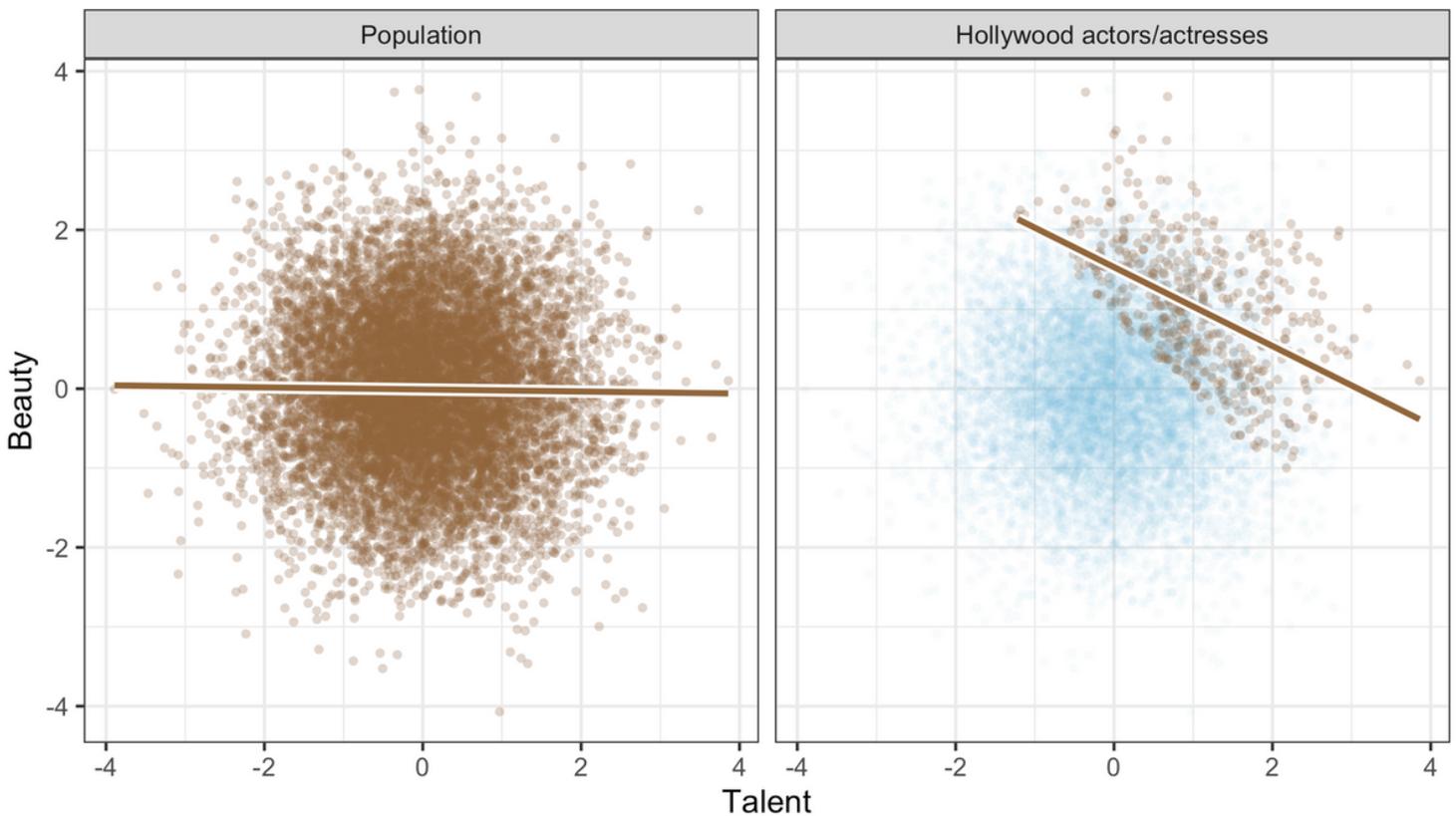


Figure 1

Illustration of collider bias; the plot on the left shows no relationship between beauty and talent in the general population ($p = .203$), while the plot on the right shows a strong inverse relationship when limited to a random weighted sample of those in the top quarter of the sum of beauty and talent ($p < .001$), a proxy for Hollywood success

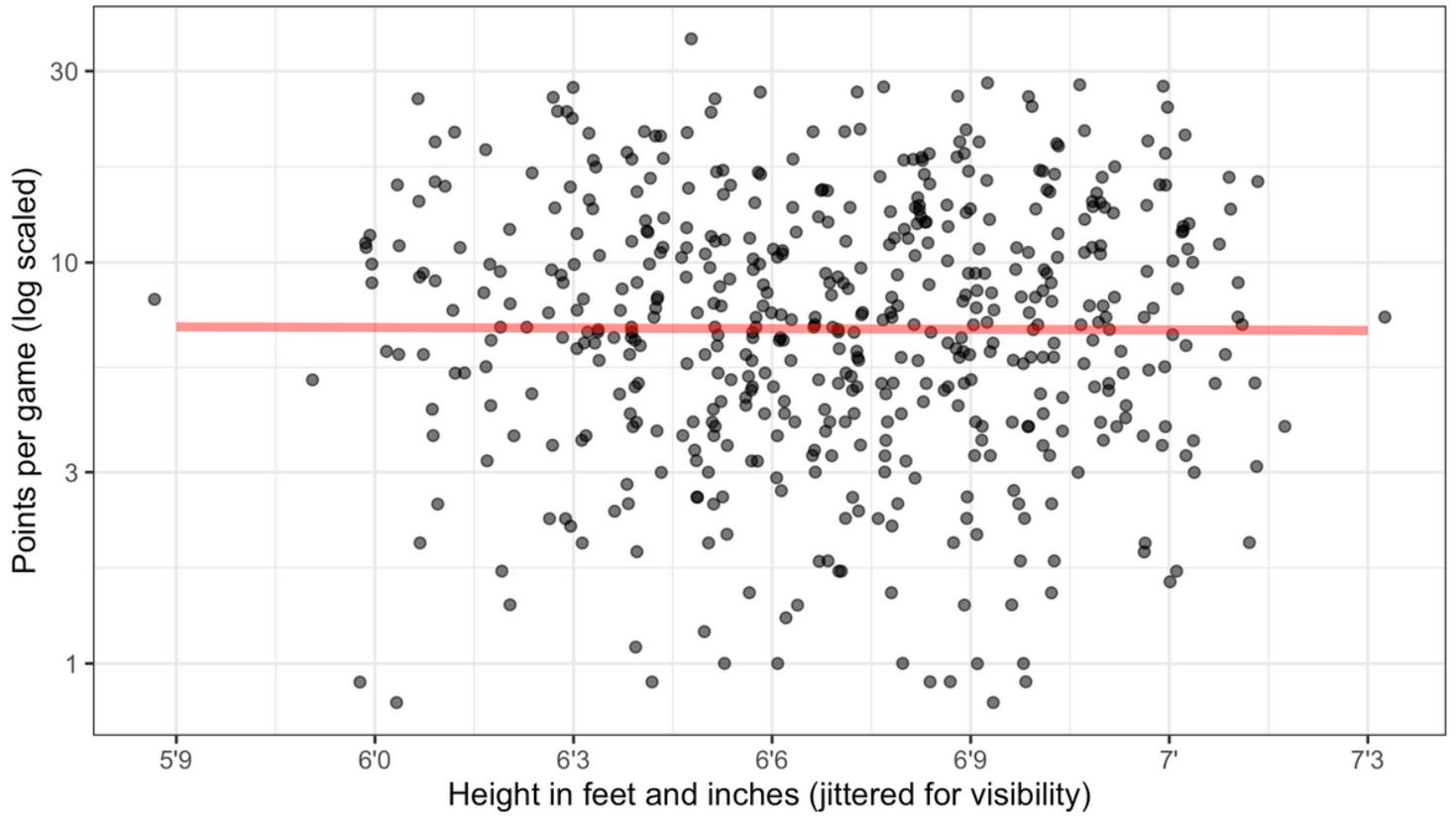


Figure 2

Association between height and points per game among NBA players in the 2018-19 regular season (7)

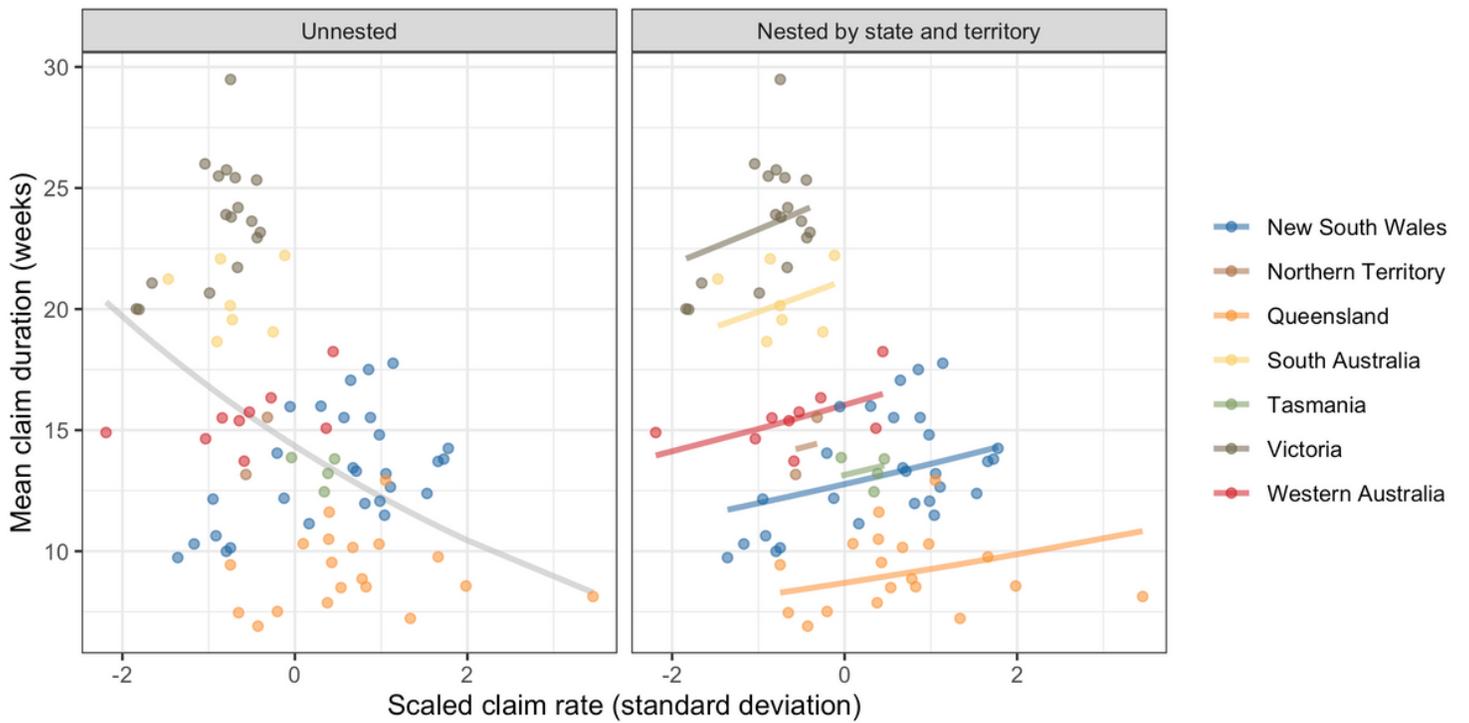


Figure 3

Association between claim rate and disability duration, unnested and nested by state and territory

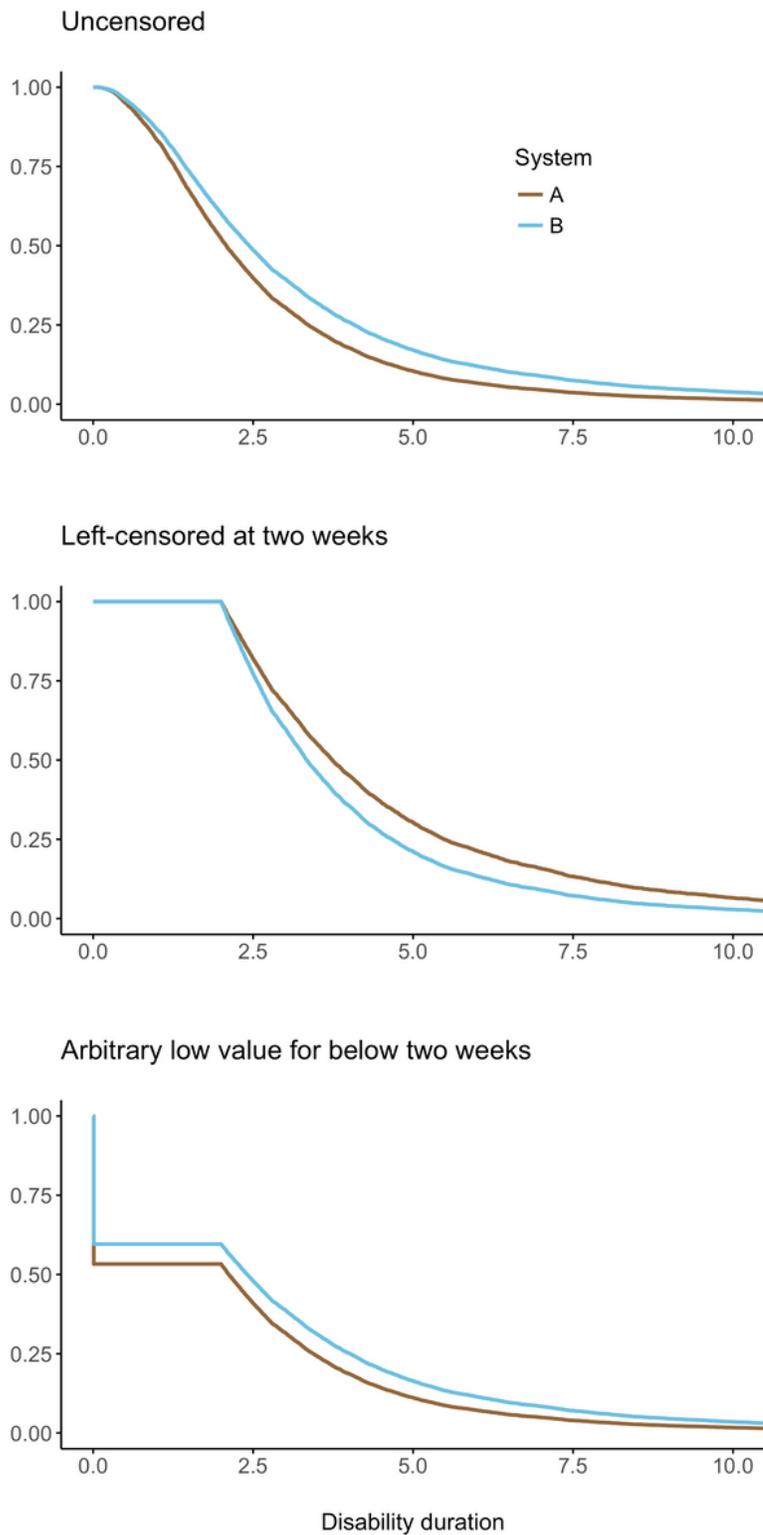


Figure 4

Survival curves of simulated compensation systems, both uncensored and censored at two weeks; System A has a mean of 0.5 weeks (standard deviation: 0.9 weeks) and System B a mean of 1 week (standard deviation: 0.6 weeks) on a log scale

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [s1durclaims.png](#)