Title

Overstating trophic cascade strength following large carnivore restoration in Yellowstone: A comment on Painter et al. (2025)

Authors

Daniel R. MacNulty[1]* | Elaine M. Brice[2] | Nicholas J. Bergeron[1] | Eric J. Larsen[3]

[1]Department of Wildland Resources and Ecology Center, Utah State University, Logan, Utah, USA

[2]Department of Biological Sciences, Wellesley College, Wellesley, Massachusetts, USA

[3]Department of Geography and Geology, University of Wisconsin – Stevens Point, Stevens Point, Wisconsin, USA

*Corresponding Author: dan.macnulty@usu.edu

Abstract

Painter et al. (2025) claim that large-carnivore recovery in Yellowstone National Park has produced a strong trophic cascade compared to other systems, citing a 152-fold increase in aspen sapling density and widespread recruitment of new trees. We show that these conclusions substantially overstate the cascade's strength because of key methodological and interpretive flaws. First, Painter et al. miscalculated the baseline density in their dataset, inflating the reported log response ratio from a true 17.5-fold increase to 152-fold. Second, they analyzed repeated measurements of the same stands as if they were independent samples, overstating standardized effect sizes by 30–40%. Third, because sapling densities are highly zero-inflated and right-skewed, mean-based metrics (log response ratios and standardized differences) are disproportionately influenced by a small minority of plots, while most plots showed little or no change. Fourth, Painter et al. conflated stand-level occurrence with magnitude, treating the presence of one or a few tall stems ("43% of stands contained small trees") as evidence of widespread recruitment. Finally, their assumptions that stems ≥ 2 m have escaped browsing and that reduced browsing alone drives height growth are contradicted by long-term data showing substantial browsing through 2–2.5 m and strong height–browsing feedbacks. Taken together, these shortcomings exaggerate the magnitude and pace of aspen recovery. The evidence supports the occurrence of a trophic cascade in Yellowstone, but not the magnitude of strength claimed. Accurate assessment of trophic cascade strength in Yellowstone is vital to ensure that this iconic system reliably informs ecological understanding and restoration practice.

Keywords

Aspen, carnivore, restoration, trophic cascade, ungulate, wolf, Yellowstone National Park

Main text

Understanding how ecosystems respond to the restoration of large carnivores following their long absence is a central challenge in ecology and natural resource management. A prevailing hypothesis is that restored carnivores promote the regeneration of degraded plant communities by reducing ungulate abundance and/or altering ungulate foraging behavior—an indirect effect of predators termed a trophic cascade (Ford and Goheen 2015). A primary test case is northern Yellowstone National Park, where reintroduced wolves (*Canis lupus*) and naturally recovering cougars (*Puma concolor*) and grizzly bears (*Ursus arctos*) interact with elk (*Cervus canadensis*) and woody deciduous plants such as aspen (*Populus tremuloides*) (Peterson et al. 2014, 2020).

Progress in understanding this case hinges on distinguishing between two related but separable dimensions of trophic cascades: occurrence and strength (Brice et al. 2022; MacNulty et al. 2024, 2025). Occurrence refers to whether a trophic cascade has taken place—whether elk browsing pressure has relaxed and plants regenerate. Strength refers to the magnitude and pace of these changes across the plant population—the extent that browsing declines and plant growth accelerates, and the rate at which these shifts occur. Researchers largely agree that a trophic cascade has occurred in northern Yellowstone; the debate now centers on how strong it is and how to apportion causation among different predators, including humans (*Homo sapiens*) whose hunting outside the park affects elk density inside the park (MacNulty et al. 2020; White et al. 2024). Knowledge about trophic cascade strength is vital to understanding the potential for large carnivore restoration—either on its own or in conjunction with human hunting—to restore plant communities to historical baselines.

Painter et al. (2025; hereafter Painter et al.) recently contributed to this debate with a study of aspen in northern Yellowstone. Rather than directly measuring a trophic cascade by quantifying predator effects on elk browsing and linking them to aspen regeneration (e.g., Brice et al. 2025), they inferred one by documenting changes in aspen regeneration over a 25-year span and attributing those changes to predator effects. Painter et al. concluded that recruitment of new aspen trees has been "widespread" across "many stands," representing "an example of widespread ecological change resulting from large carnivore restoration" and a "strong effect" relative to trophic cascades in other systems.

Painter et al.'s focus on cascade strength is an important advance beyond their earlier work, which mainly documented occurrence (Brice et al. 2022; Painter et al. 2023; MacNulty et al.

2024). However, their analysis and interpretation contain errors, ambiguities, and assumptions that overstate cascade strength, particularly with respect to the recruitment of new aspen trees. We address these issues below to clarify the actual magnitude of the Yellowstone carnivore-elk-aspen trophic cascade. In doing so, we aim to improve understanding of how large carnivore restoration can support recovery of degraded plant communities.

Problems in assessing trophic cascade strength

Painter et al. used plant density as their metric of trophic cascade strength, drawing on two datasets: their own and that of Brice et al. (2022). Their dataset comprised surveys of 87 randomly selected aspen stands resampled in 2020–21, with earlier measurements from the same stands in 2012 (Painter et al. 2014, 2015) and 1997–98 (Larsen 2001). Each stand was sampled using a single georeferenced, physically unmarked 2 × 30 m (0.006 ha) belt transect. Brice et al. (2022) assembled a similar dataset based on physically marked 1 × 20 m (0.002 ha) belt transects (N = 113 stands), sampled annually from 2007 to 2017 (minus 2015). By using the Brice et al. data as a comparable base of inference, Painter et al. implicitly reject their group's earlier claim that these data are an unrepresentative sample of aspen stands that should be "viewed cautiously" (Beschta et al. 2023).

Painter et al. defined cascade strength as the change in mean plant density between the first and last years of each time series—1997–98 to 2020–21 for their own data, and 2007 to 2017 for the Brice et al. data. They quantified these changes using effect size statistics: the log response ratio (LRR) and Cohen's *d* for their dataset, and the LRR for the Brice et al. dataset. As we show below, these analyses and their interpretations are problematic in multiple respects, and together they overstate cascade strength.

Ambiguous measurement unit

An overarching problem is that Painter et al. did not clearly specify which types of aspen plants were counted in the density estimates. In the *Methods*, they identified the unit of measurement as the sapling, defined as aspen "≥ 2 m tall and < 5 cm dbh" (diameter at breast height). In the *Results*, however, the measurement broadens to "young aspen (≥ 2 m tall)" in their Figure 5, and to "young aspen > 2 m height (saplings and young trees)" in their Table 3. A "young tree" was defined in the *Methods* as "≥ 5 cm dbh but < 20 cm dbh." Thus, while the area basis of the density estimate is fixed, the numerator—what is counted—apparently shifts from saplings alone to saplings plus young trees. This shift obscures which plants underlie the reported densities and effect sizes, complicating comparisons across datasets. For clarity in the following discussion, we use "sapling" to refer to the unit of measurement, while recognizing that its precise meaning is ambiguous.

Miscalculated baseline

Painter et al. reported a log response ratio (LRR) of 5.02, corresponding to a 152-fold increase in mean sapling density between 1997–98 (9.6 saplings ha^{-1}) and 2020–21 (1460 saplings ha^{-1}; their Table 3). Because no saplings were recorded in 1997–98, they added a pseudo-count of 0.5 sapling per plot to permit calculation of the log ratio. However, they miscalculated the baseline density derived from this adjustment. The correct pseudo-baseline is 83.3 saplings ha^{-1} (0.5 sapling ÷ 0.006 ha plot area), not 9.6 saplings ha^{-1} as they reported. This error is consequential: it inflates the LRR by more than 75% (from 2.86 to 5.02) and the fold increase by 768% (from 17.5 to 152.1). The impact is magnified by the fact that the inflated statistic made headline news in international outlets, including the Washington Post: "Since wolves' return, Yellowstone's aspens are recovering, study finds. The researchers found a 152-fold increase in sapling and young-tree density*"* (Blakemore 2025).

More broadly, this error illustrates the sensitivity of LRR to the choice of pseudo-baseline (Fig. 1). Even small adjustments to the pseudo-count can generate large shifts in effect size, raising doubts about the reliability of this metric for assessing trophic cascade strength when baseline densities are zero. Contrary to Painter et al.'s assertion that their "log ratio likely underestimated the actual effect strength," correcting the pseudo-baseline reduces the effect size substantially. As shown in Figure 1, smaller pseudo-counts inflate the LRR rather than dampen it, meaning that Painter et al.'s calculation overstates, rather than understates, effect strength.

Even if the baseline calculation were correct, interpretation of the LRR would remain problematic. Painter et al. compare their effect size directly to results from other systems, including the global meta-analysis by Borer et al. (2005), without accounting for critical differences in ecological context. That analysis assumed plant communities were at or near equilibrium, such that observed changes in plant biomass or density reflected stable predator–herbivore–plant interactions. In northern Yellowstone, by contrast, aspen regeneration is ongoing, nonlinear, and influenced by multiple biotic and abiotic drivers in addition to carnivore restoration (Peterson et al. 2014, 2020; Brice et al. 2022, 2025; MacNulty et al. 2024). These transient dynamics mean that changes in aspen density cannot be attributed solely to stable trophic interactions. Applying equilibrium-based benchmarks in a non-equilibrium context therefore overstates cascade strength.

Overlooked paired design

Both datasets analyzed by Painter et al. tracked the same stands through time, making the design repeated-measures (paired). However, they treated the first and last years of each time series as independent groups, discarding the within-stand correlation and preventing each plot from serving as its own control. For the Painter et al. dataset comparing 1997–98 to 2020–21, this inflated the standardized mean difference (Cohen's *d*) because the independent-groups formula ignores the covariance between repeated measures (Morris and DeShon 2002; Lakens 2013). The inflation arose from misspecifying the denominator: using the pooled standard deviation (SD = 2068) of two independent groups instead of the standard deviation of within-stand differences (SDΔ = 2925). With a numerator equal to the mean change in sapling density (1460), the independent-groups effect size was 1460/2068 ≈ 0.71 (their Table 3), whereas the paired effect size was 1460/2925 ≈ 0.50—overstated by about 42%. A similar pattern occurred with the Brice et al. dataset comparing 2007 to 2017, where the independent-groups approach yielded an effect size of 0.62 compared to a paired value of 0.48—overstated by about 30%.

A comparable assessment is not possible for Painter et al.'s LRR estimates. All baseline values in 1997–98 (Painter et al. data) were zero and only two were non-zero in 2007 (Brice et al. data). Adding a pseudo-count to the baseline does not resolve the issue, since more than half of the follow-up values in 2021 and 2017 were also zero. With so many zeros, log ratios become undefined or unstable, rendering the LRRs uninterpretable as measures of cascade strength.

Misleading means

The high frequency of zeros and the strong right-skew among non-zero values in the final year of both datasets (Fig. 2) reveal a more basic problem with Painter et al.'s analysis. Mean-based effect sizes (LRR, Cohen's d) do not reflect a widespread change in sapling density; instead, they are dominated by a minority of plots that disproportionately determine the final-year mean. A decile decomposition of the mean (Cowell 2011) makes this clear: the top 20% of plots contributed 79% of the Painter et al. mean sapling density ($\bar{x}$ = 1,460 stems ha^{-1}) and 82% of the Brice et al. mean ($\bar{x}$ = 1,597 stems ha^{-1}; Fig. 3).

Moreover, in the Brice et al. data, the median (50th percentile) remained zero throughout 2007–2017, meaning that half of all plots had no saplings even as the mean increased (Fig. 4). The 75th percentile also stayed near zero until the final years, while sharp increases occurred only at the 90th and 95th percentiles. The distribution of the final year of the Painter et al. data shows the same pattern: the median is zero and only the upper tail is elevated (Fig. 4). Thus, Painter et al.'s claim that mean sapling densities in their data and the Brice et al. data reflect a "rapid increase in sapling and young tree density over the last two decades (Fig. 5)" is a selective interpretation that applies at most to the upper quartile of plots in either datasets.

Together, these results show that the plot-level increases in mean sapling density that form the basis of Painter et al.'s effect size estimates are not representative of most plots. Instead, they reflect gains in a small minority of plots while the majority changed little, if at all. In this context, mean-based effect sizes overestimate the central tendency and inflate inferences of "widespread ecological change resulting from large carnivore restoration."

Blurring trophic cascade strength and occurrence

Beyond quantifying cascade strength, Painter et al. also reported numerous results describing its occurrence. These were based on plot measurements combined with non-systematic stand-level surveys that recorded whether a stand contained at least one tall young aspen—described variously as a "sapling," "young tree," or "small tree." By design, this non-systematic approach, known as 5T sampling because it targets the five tallest (5T) young aspen in a stand, detects presence but does not yield a representative measure of stand-level conditions or of the broader aspen population (Ripple and Beschta 2007; Kauffman et al. 2013; Brice et al. 2022). Independent evaluation has shown that 5T sampling overestimates overstory regeneration relative to random sampling (by 4–7× in northern Yellowstone) because it captures only the fastest-growing ~10–15% of young aspen rather than the typical individual in a stand (Brice et al. 2022). Thus, the 5T method is useful for detecting occurrence but not for quantifying the strength of a trophic cascade (MacNulty et al. 2024).

Although Painter et al. acknowledged that "[t]he (5T) method confirms the presence of saplings and young trees in a stand, but does not distinguish a stand with many saplings from a stand with few," they nevertheless interpreted their occurrence-based results as evidence of a strong trophic cascade. Chief among these was their finding that "43% of sampled stands contained small trees," which they described as "the first substantial recruitment since the 1940s." They repeatedly highlighted this result as evidence that recruitment was "widespread" and that "[m]any aspen stands in northern Yellowstone now have historically and ecologically significant amounts of new tall saplings…growing into new overstory trees." This interpretation is problematic because it treats the 43% result as a valid measure of the magnitude of aspen recovery, i.e., trophic cascade strength. Closer inspection shows that the result is a biased indicator of cascade strength for two main reasons.

First, reducing stand-level recruitment to a presence–absence outcome obscures magnitude. Each stand contributed one "vote" regardless of how many trees it contained, and a stand was classified as "recovering" if at least one young tree was detected anywhere within it (Painter et al., Table 1). A stand with a single small tree was weighted equally to a stand with many, even though those cases do not represent equivalent progress toward overstory regeneration. Thus, the claim that 43% of stands contained new small trees establishes only that at least one qualifying tree was present in 43% of 87 stands; it does not indicate how strongly, or how broadly across the landscape, recruitment is occurring. Notably, only 25% of stands contained ≥ 5 young trees (Painter et al., Table 1), underscoring that much of the detected "recovery" consisted of isolated stems rather than robust stand-level replacement.

Second**,** the 43% figure ignores variation in stand size, thereby inflating the extent of recovery. Painter et al.'s analysis treated each of the 87 stands equally regardless of area, even though stands varied greatly in size (their Fig. 6). Without accounting for this variation, the results conflate the proportion of stands showing recovery (prevalence) with the proportion of the total aspen area recovering (extent). If stands counted as "recovering" are disproportionately small, the stand-weighted percentage can imply widespread recovery even when only a small fraction of the total aspen area is involved. For example, if 60 stands were 0.1 ha and 27 were 1.0 ha, and all 37 "recovering" stands were among the smaller stands, the stand-weighted figure would still be 43%. Yet the area-weighted extent of recovery would equal only 11% of the total aspen area (3.7 ha of 33 ha). In this case, a statistic that suggests recovery is widespread across stands would actually correspond to a very limited share of the aspen landscape. Thus, the 43% figure signals occurrence of recruitment but not how much of the aspen landscape has transitioned to new trees—and therefore cannot measure cascade strength.

These general limitations are compounded by two additional issues. First, the 5T method was applied inconsistently across years. In 2012, the authors selected “the five tallest young aspen in each stand (within 60 m of the 2 × 30 m sampling plot)” (Painter et al. 2014), whereas in 2020–21 they selected “five of the tallest young aspen in the entire stand” (Painter et al. 2025). This broader search area in the later period increased the likelihood of encountering tall saplings and could inflate the reported percentage of stands with young trees relative to 2012.

Second**,** ambiguous terminology clouds the meaning of the 43% result. It refers inconsistently to “small trees” (5–10 cm dbh) and to “young trees” (≥5 cm but <20 cm dbh), with shifting definitions across the text, figures, and tables. This lack of clarity makes it uncertain exactly what class of trees the 43% result represents and, therefore, what biological process it reflects. Further confusion stems from the repeated use of the term “new cohort” to describe the presence of one or more trees in these size classes. In demographic usage, a cohort denotes a group (>1) of individuals sharing a common origin. Applying the term to a single tree exaggerates the degree of recruitment implied and risks suggesting that substantial stand-level replacement is underway when the evidence is far more limited.

These issues collectively indicate that the 43% figure reflects isolated occurrences of tall saplings, not a reliable measure of trophic cascade strength.

Inaccurate assumptions about the browsing-height relationship

Painter et al.'s inferences about the trophic cascade rely on two inaccurate assumptions about the relationship between browsing pressure and aspen height, each of which inflates assessments of the cascade's occurrence and strength. The first assumes that plants ≥ 2 m in height have largely escaped browsing and therefore provide a reliable proxy for plants likely to recruit into the overstory. The second assumes that plants < 2 m follow a one-way causal relationship between browsing and height, such that reduced browsing necessarily produces taller plants. Painter et al. do not justify these assumptions with data. Rather, citing Beschta et al. (2023) as supporting evidence, they claim the assumptions are warranted because "numerous studies of aspen have used 2 m…as a general indication that aspen were escaping from elk browsing" and because this height "has functioned well as an indicator of likely recruitment success." Yet neither Beschta et al. (2023) nor any of the studies they cite provide empirical data showing how elk browsing varies with plant height. Thus, the 2 m threshold invoked by Painter et al. rests on convention rather than evidence.

In support of their reliance on convention, Painter et al. reported that "we found little evidence of browsing of aspen above 2 m during our data collection." However, larger and longer-term datasets than theirs document substantial browsing at and above 2 m in northern Yellowstone (Brice et al. 2022, 2025; Hobbs et al. 2024). Modeled browsing probabilities show that, on average, 43% of 2 m aspen (95% CI = 35-52%) and 23% of 2.5 m aspen (95% CI = 17-30%) were browsed, with browsing declining to negligible levels only after aspen exceeded roughly 3 m (Brice et al. 2022, 2025). Likewise, in willow, roughly 30% of shoots were browsed until plants surpassed 2.5 m, with little further decline until they were taller than 3 m (Hobbs et al. 2024).

These results indicate that a non-trivial proportion of stems ≥ 2 m continue to experience browsing at rates (≥ 30%) that Painter et al. themselves acknowledge can impair aspen recruitment.

Moreover, Brice et al. (2022) show that adopting a 2 m browse-escape threshold inflates inferred cascade strength under two metrics: (i) the annual increase in the probability that an individual stem "recruits", i.e., crosses the threshold, and (ii) the percentage of sampled stands whose median stem height exceeds it. In their analyses, the probability that a stem exceeded 2 m rose by 2.1 percentage points per year (95% CI = 1.5–2.8), roughly double the 1.0 percentage point per year increase estimated for a 3 m criterion (95% CI = 0.4–1.6). The share of stands with median height > 2 m increased from 1% to 15%, versus 0% to 6% for > 3 m. Thus, classifying all stems ≥ 2 m as "likely recruits" exaggerates both the apparent rate and spatial extent of recruitment—and, by extension, the strength of the trophic cascade. Simply lowering the browse-escape threshold from 3 m to 2 m roughly doubles the apparent recruitment rate and more than doubles its spatial footprint, magnifying perceived cascade strength without any real change in browsing.

Equally problematic is the assumption of a one-way causal link for plants < 2 m—namely that reduced browsing causes taller growth but not vice versa. In reality, height and browsing form a feedback: plant height is both a consequence and a cause of reduced browsing because elk and other cervids browse woody plants within a preferred browsing height (PBH), above which browsing declines (e.g., Rounds 1979; Motta 2003; Renaud 2003; Kuijper et al. 2013; Konôpka et al. 2018, 2022; Maxwell et al. 2019; van Beeck Calkoen 2018; Brice et al. 2022, 2025; Hobbs et al. 2024). In northern Yellowstone, the PBH for elk on aspen is about 120 cm; above this height, the odds of browsing decrease by 9.6% per 10 cm (95% CI = 9.3–9.8%; Brice et al. 2022, 2025). Thus, within the 1.2–2.0 m range, reduced browsing reflects aspen growing beyond elk's PBH rather than, or in addition to, carnivore-mediated reductions in browsing. Painter et al. sought to address height bias by excluding stems > 2 m from their browsing analysis, but the bias begins well below that threshold. As more stems occupy the upper portion of the < 2 m interval, average browsing rates decline even if true browsing is unchanged, giving the illusion of a stronger trophic cascade than actually exists.

Conclusion

Painter et al. aimed to quantify the strength—not merely the occurrence—of the Yellowstone carnivore–elk–aspen trophic cascade. Our assessment shows that their central claims of a "widespread" and "strong" cascade rest on ambiguous measurement, miscalculated baselines, analyses that ignore the repeated structure of the data, and mean-based effect sizes that are unrepresentative of most plots. Each issue independently inflates perceived recovery; together they materially overstate cascade strength.

The question in northern Yellowstone is no longer whether a trophic cascade has occurred but how strong it is, where, and why. Credible answers require clear units, design-consistent estimators, distribution-sensitive metrics, area-weighted summaries, and thresholds grounded in the actual browsing environment. Applying these standards will replace appealing—but fragile—signals with estimates that genuinely reflect the magnitude and pace of plant recovery and, in turn, provide more reliable guidance for both science and management.

References

Blakemore, E., 2025. Since wolves' return, Yellowstone's aspens are recovering, study finds. The Washington Post, January 25, 2025. https://www.washingtonpost.com/climate-environment/2025/01/25/yellowstone-wolves-aspen-recovery/

Beschta, R.L., Painter, L.E., Ripple, W.J., 2023. Revisiting trophic cascades and aspen recovery in northern Yellowstone. Food Webs 36, e00276.

Borer, E.T., Seabloom, E.W., Shurin, J.B., Anderson, K.E., Blanchette, C.A., Broitman, B., Cooper, S.D., 2005. What determines the strength of a trophic cascade? Ecology 86, 528–537.

Brice, E.M., Larsen, E.J., MacNulty, D.R., 2022. Sampling bias exaggerates a textbook example of a trophic cascade. Ecology Letters 25, 177–188.

Brice, E.M., Larsen, E.J., Stahler, D.R., MacNulty, D.R., 2025. The primacy of density- mediated indirect effects in a community of wolves, elk, and aspen. Ecological Monographs 95, e1627.

Cowell, F.A., 2011. Measuring Inequality, 3rd ed. Oxford University Press, Oxford, UK.

Ford, A.T., Goheen, J.R., 2015. Trophic cascades by large carnivores: a case for strong inference and mechanism. Trends in Ecology & Evolution 30, 725–735.

Hobbs, N.T., Johnston, D.B., Marshall, K.N., Wolf, E.C., Cooper, D.J., 2024. Does restoring apex predators to food webs restore ecosystems? Large carnivores in Yellowstone as a model system. Ecological Monographs 94, e1598.

Kauffman, M.J., Brodie, J.F., Jules, E.S., 2013. Are wolves saving Yellowstone's aspen? A landscape-level test of a behaviorally mediated trophic cascade: reply. Ecology 94, 1425–1431.

Konôpka, B., Pajtík, J., Shipley, L.A., 2018. Intensity of red deer browsing on young rowans differs between freshly felled and standing individuals. Forest Ecology and Management 429**,** 511–519.

Konôpka, B., Šebeň, V., Pajtík, J., Shipley, L.A., 2022. Influence of tree species and size on bark browsing by large wild herbivores. Plants 11, 2925.

Kuijper, D.P. J., de Kleine, C., Churski, M., van Hooft, P., Bubnicki, J., Jędrzejewska, B., 2013. Landscape of fear in Europe: wolves affect spatial patterns of ungulate browsing in Białowieża Primeval Forest, Poland. Ecography 36, 1263–1275.

Lakens, D., 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Frontiers in Psychology 4, 863.

Larsen, E.J., 2001. Aspen age structure and stand conditions on elk winter range in the northern Yellowstone ecosystem. PhD dissertation. Oregon State University, Corvallis.

MacNulty, D.R., Cooper, D., Procko, M.P., Clark-Wolf, T.J., 2025. Flawed analysis invalidates claim of a strong Yellowstone trophic cascade after wolf reintroduction: A comment on Ripple et al. (2025). Global Ecology and Conservation 63, e03899.

MacNulty, D.R., Brice, E.M., Larsen, E.J., 2024. Non-random sampling measures the occurrence but not the strength of a textbook trophic cascade. Ecology Letters 27, e14344.

MacNulty, D.R., Stahler, D.R., Wyman, T., Ruprecht, J., Smith, L.M., Kohl, M.T., Smith, D.W., 2020. Population dynamics of northern Yellowstone elk after wolf reintroduction. In: Smith, D.W., Stahler, D.R., MacNulty, D.R. (Eds.), Yellowstone Wolves: Science and Discovery in the World's First National Park. University of Chicago Press, Chicago, pp. 184–199.

Maxwell, J.D., Rhodes, A.C., St. Clair, S.B., 2019. Human-altered disturbance patterns and forest succession: impacts of competition and ungulate herbivory. Oecologia 189, 1061–1070.

Morris, S.B., DeShon, R.P., 2002. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychological Methods 7, 105-125.

Motta, R., 2003. Ungulate impact on rowan (*Sorbus aucuparia* L.) and Norway spruce (*Picea abies* (L.) Karst.) height structure in mountain forests in the eastern Italian Alps. Forest Ecology and Management 181, 139–150.

Painter, L.E., Beschta, R.L., Larsen, E.J., Ripple, W.J., 2014. After long-term decline, are aspen recovering in northern Yellowstone? Forest Ecology and Management 329, 108–117.

Painter, L.E., Beschta, R.L., Larsen, E.J., Ripple, W.J., 2015. Recovering aspen follow changing elk dynamics in Yellowstone: evidence of a trophic cascade? Ecology 96, 252–263.

Painter, L.E., Beschta, R.L., Ripple, W.J., 2023. Aspen recovery in northern Yellowstone: a comment on Brice et al. (2021). Ecology Letters 27, e14353.

Painter, L.E., Beschta, R.L., Ripple, W.J., 2025. Changing aspen stand structure following large carnivore restoration in Yellowstone. Forest Ecology and Management 594, 122941

Peterson, R.O., Vucetich, J.A., Bump, J.M., Smith, D.W., 2014. Trophic cascades in a multicausal world: Isle Royale and Yellowstone. Annual Review of Ecology, Evolution, and Systematics 45, 325–345.

Peterson, R.O., Beschta, R.L., Cooper, D.J., Hobbs, N.T., Johnston, D.B., Larsen, E.J., Marshall, K.N., Painter, L.E., Ripple, W.J., Rose, J.R., Smith, D.W., Wolf, E.C., 2020. Indirect effects of carnivore restoration on vegetation. In: Smith, D.W., Stahler, D.R., MacNulty, D.R.

(Eds.), Yellowstone Wolves: Science and Discovery in the World’s First National Park. University of Chicago Press, Chicago, pp. 205–222.

Renaud, P.C., Verheyden-Tixier, H., Dumont, B., 2003. Damage to saplings by red deer (*Cervus elaphus*): effect of foliage height and structure. Forest Ecology and Management 181, 31–37.

Ripple, W.J., Beschta, R.L., 2007. Restoring Yellowstone's aspen with wolves. Biological Conservation 138, 514–519.

Rounds, R.C., 1979. Height and species as factors determining browsing of shrubs by wapiti. Journal of Applied Ecology 16, 227–241.

White, P.J., Stahler, D.R., Smith, D.W., MacNulty, D.R., Garrott, R.A., 2024. Northern Yellowstone elk: resilience and adaptation to changes in management policies and the ecosystem. Yellowstone National Park, Mammoth, Wyoming.

van Beeck Calkoen, S.T., Kuijper, D.P., Sand, H., Singh, N.J., van Wieren, S.E., Cromsigt, J.P., 2018. Does wolf presence reduce moose browsing intensity in young forest plantations? Ecography 41, 1776-1787.

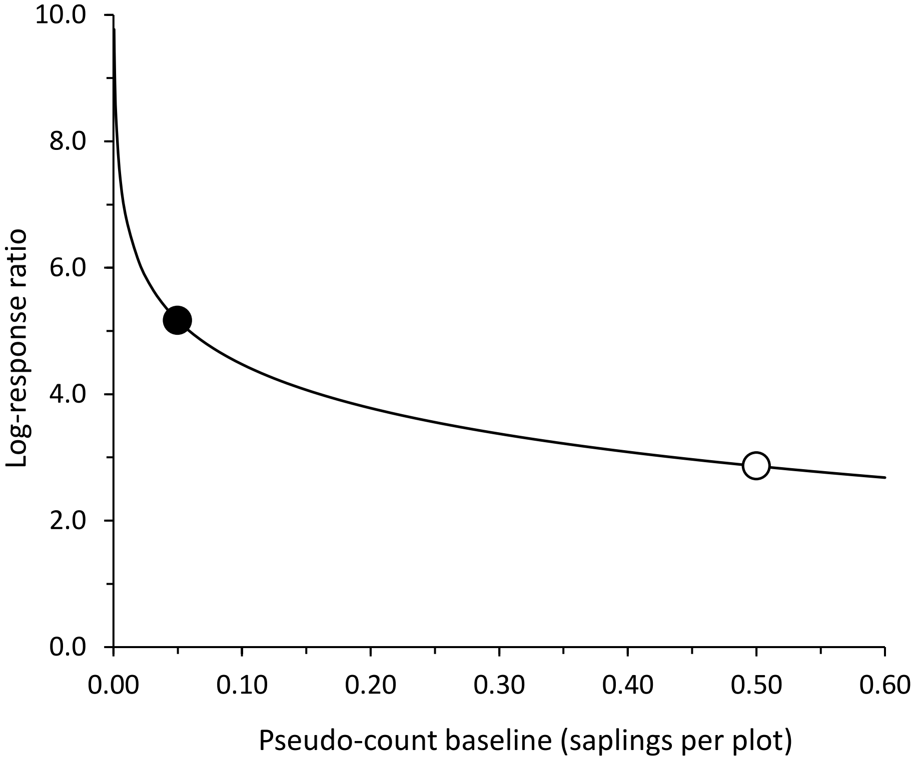

Figure 1. Sensitivity of the log response ratio (LRR) to the choice of pseudo-count baseline when initial sapling density is zero. Painter et al. (2025) reported an LRR of 5.02 by adding a pseudo-count of 0.5 saplings per plot but miscalculating the corresponding density as 9.6 stems ha^{-1}. The correct conversion is 83.3 stems ha^{-1}, which yields an LRR of 2.86 (white dot). By contrast, an LRR of 5.02 is recovered only if the pseudo-count is expressed as 0.0576 saplings per plot (black dot). The nearly twofold difference in LRR arises solely from how the pseudo-count is scaled, illustrating the instability of LRR in zero-inflated contexts where small adjustments to the pseudo-baseline can greatly alter effect size and inflate estimates of trophic cascade strength.

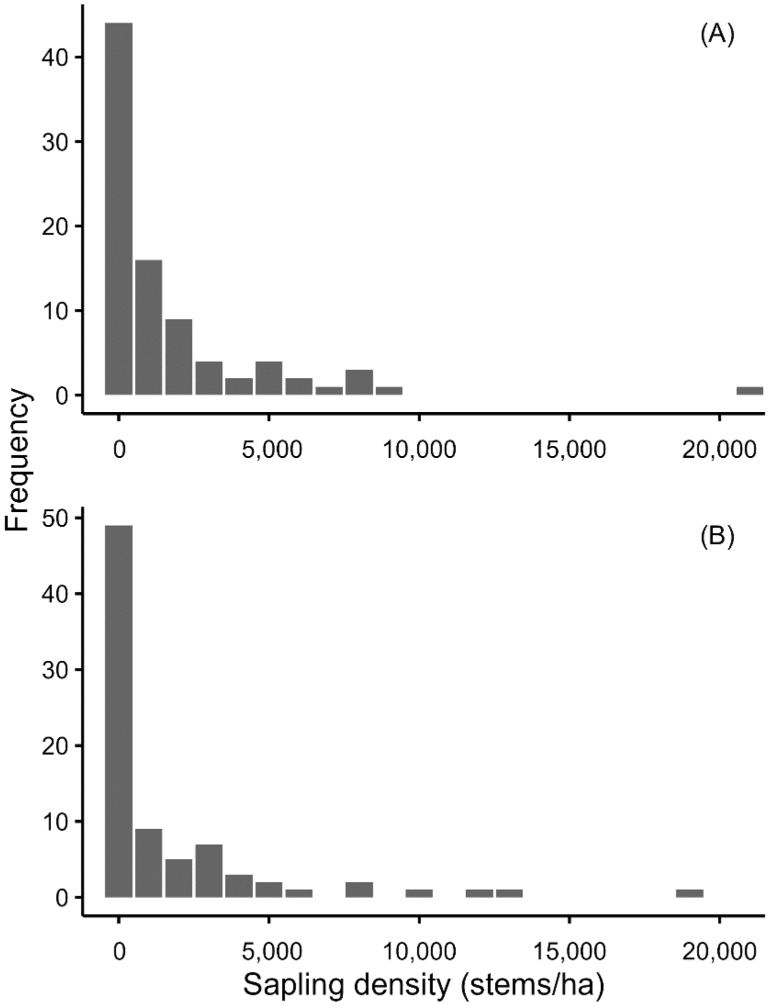

Figure 2. Frequency distributions of sapling density measured in sampling plots within stands. Panel (A) shows the Painter et al. dataset for 2020–21 ($\bar{x}$ = 1,460 stems ha^{-1}; SD = 2,925; range = 0–20,333); panel (B) shows the Brice et al. dataset for 2017 ($\bar{x}$ = 1,597 stems ha^{-1}; SD = 3,289; range = 0–19,000). Bars represent 1,000-stems/ha bins, beginning at 0; the first bin includes only zero-density plots. Frequency indicates the number of sampling plots per bin.

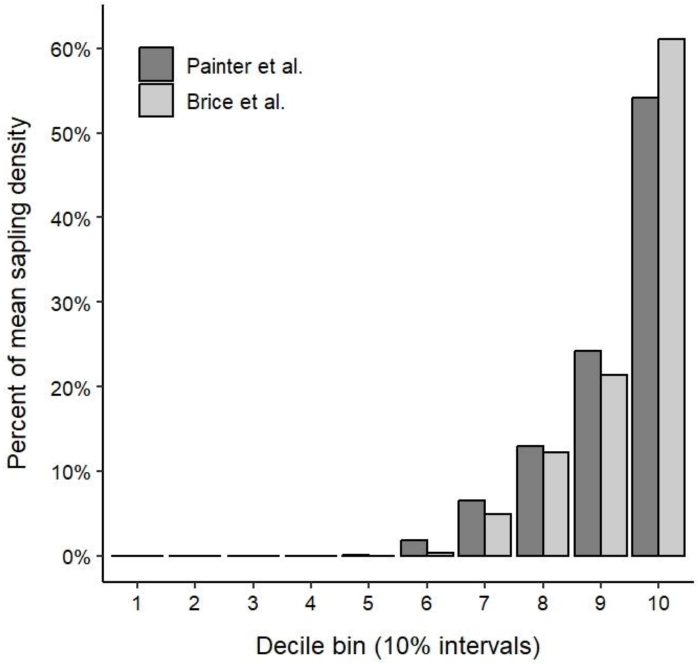

Figure 3. Decile decomposition of mean sapling density in sampling plots within stands. Bars show the percentage of the overall mean sapling density contributed by each decile of plots, ranked by density, for the final year of each dataset—Painter et al. (2020–21) and Brice et al. (2017). Contributions were calculated using a distributional decomposition method (Cowell 2011): plots were ranked by sapling density, assigned to deciles, and densities summed within each decile and expressed as a fraction of the total across plots. In the Painter et al. dataset, the top 10% of plots (≥ ~4,700 stems ha^{-1}) accounted for 54% of the mean sapling density (1,460 stems ha^{-1}), and in the Brice et al. dataset, the top 10% of plots (≥ ~5,000 stems ha^{-1}) accounted for 61% of the mean (1,597 stems ha^{-1}).

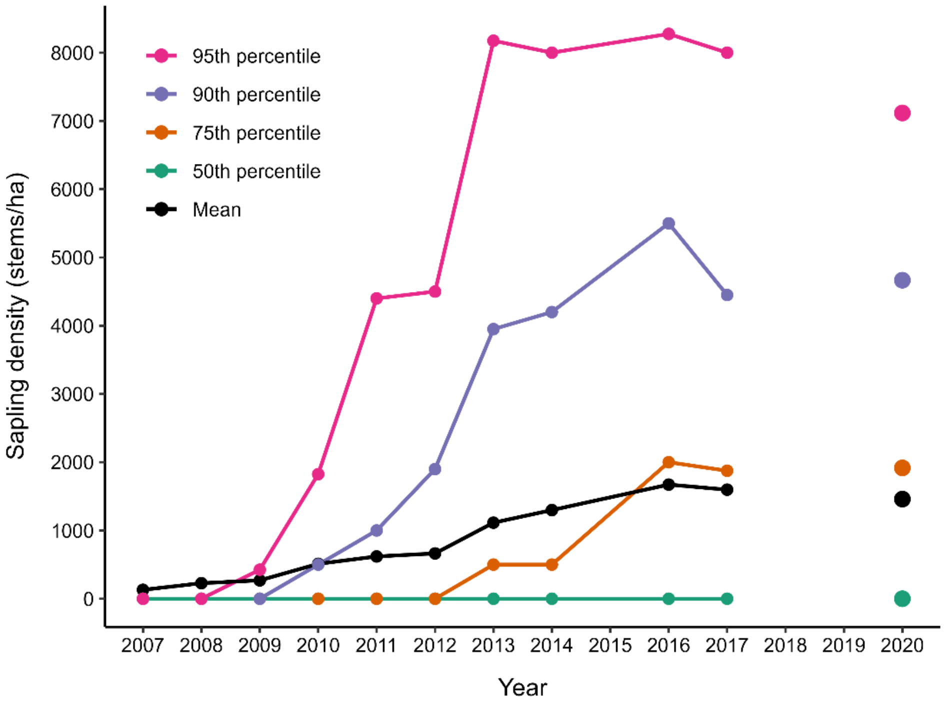

Figure 4. Distributional patterns of sapling density in sampling plots within stands. Brice et al. data (2007–2017) are shown as lines with points (means and the 50th, 75th, 90th, and 95th percentiles), and Painter et al. data (2020–2021) as points only, plotted at 2020 to represent values combined across 2020 and 2021.