Transparency and reproducibility in invasion science

Running head: Invasion Science: Transparency and Reproducibility

Fabio Mologni[1, 2]

*Jason Pither[1]

* Co-corresponding authors

*1. Department of Biology, and Institute for Biodiversity, Resilience, and Ecosystem Services

I.K. Barber Faculty of Science, University of British Columbia Okanagan

3187 University Way, Kelowna, BC V1V 1V7, Canada

Email: jason.pither@ubc.ca

ORCiD: https://orcid.org/0000-0002-7490-6839

2. Department of Botany and Biodiversity Research

University of Vienna, Rennweg 14, 1030, Vienna, Austria

Phone: +43 670 1812022

ORCiD: https://orcid.org/0000-0003-4750-9974

Abstract

Policymakers and practitioners overseeing invasive species management depend on reliable research for guidance. Transparency and reproducibility are core features of reliable research, and prerequisites for outcomes to be independently replicated within the same or different systems. These features are evidently lacking in many science disciplines, including Ecology. In this Discussion paper, we first report the findings of an assessment of 49 primary research studies that were part of a systematic mapping effort, showing that invasion science research exhibits the same shortfalls as ecology research more broadly. For instance, only one study explicitly considered statistical power in the methods describing study design, and only 2 studies provided access to both data and code, which is the minimum requirement for computational reproducibility. We then discuss the implications that low statistical power has for published invasion science research, for designing studies, and for policymakers and practitioners relying on primary research to inform their decisions. We then make specific recommendations, targeting the same stakeholders as well as publishers, on how to maximize the reliability of invasion science research moving forward. This includes explicitly considering and ideally estimating statistical power, undertaking a study pre-registration, making all relevant code and non-sensitive raw data accessible and useable, and devising and upholding clear and consistent policies on transparent reporting and open materials.

Keywords: Invasion science, Reproducibility, Open Science, Non-native species, Invasive species, Impacts, Replicability, British Columbia, Reed Canarygrass, Russian Olive

Introduction

Globally, biological invasions are estimated to cost over 423 billion USD each year, a number that has quadrupled every decade since 1970 (Roy et al., 2024). Ideally, policymakers and managers tasked with addressing the threats and impacts of invasive species base their decisions and actions on reliable research, thus maximizing potential success, and minimizing waste in time, effort, and money. However, evidence shows that much of the scientific literature is compromised by biases, methodological flaws, and statistical misuse (Ioannidis 2005), calling into question the reliability of the very research policymakers depend on.

"Reliable" means that the research adheres to foundational principles of scientific inquiry, i.e. it is transparent, honest, and thorough in its reporting of methods and results, such that it can be reproduced (i.e. repeating the study using the same materials and data) and its findings independently replicated (repeating the study using new materials and generating new data, and arriving at the same general conclusions) (National Academies of Sciences, Engineering, and Medicine (2019) and Nosek et al. (2022). In reality most research informing policy and management is unlikely to meet these standards. Since Ioannidis' provocative article "Why most published research findings are false" was published two decades ago (Ioannidis, 2005), meta-research studies have firmly established that shortfalls and questionable research practices in the study design, implementation, and reporting stages pervade many disciplines (Begley & Ioannidis, 2015; Ioannidis, 2024), including ecology and evolution (e.g. Parker et al. 2016; Fraser et al. 2018; Purgar et al. 2022; Yang et al. 2022, 2023, 2024; Kimmel et al. 2023). For example, failing to fully and transparently report all procedures (e.g. details of sampling and/or experimental design and statistical analyses, Parker et al., 2016; Davis & Kay, 202), and failing to make code and non-sensitive data available and useable (Roche et al., 2015; Culina et al., 2020; Roche et al., 2022: Popovic et al., 2024) each undermine reproducibility.

Addressing such shortfalls does not, however, guarantee that findings will be independently replicated. For example, a study may be reproducible thanks to thorough and transparent reporting, but if it suffers from low statistical power its findings are less likely to be replicated (Button et al., 2013; Yang et al., 2023). It is recommended that studies achieve a power of at least 80% (Button et al., 2013), but a recent estimate of the average power in primary ecology and evolutionary biology studies was 15% (Yang et al., 2023), slightly lower than other disciplines (e.g. 20% in medical sciences, Lamberink et al., 2018).

Low statistical power in invasion science can reduce the likelihood of detecting impacts of invasive species. In their assessment of experiments investigating the impacts of introduced algal and crustacean species, Davidson and Hewitt (2014) found that only 1 of 31 analyses reporting non-significant impacts achieved an estimated power of 80%, and 25 studies (81%) had an estimated power of less than 40% (Davidson & Hewitt, 2014). Additional consequences of low statistical power that appear less-well appreciated by invasion scientists (but see Catford et al. 2022) are increases in the rate of false positives (Type-I errors), errors in effect magnitudes (Type M) and errors in effect signs (Type S, Button et al., 2013; Parker et al., 2018; Parker & Yang, 2023). When combined with the potential influences of P-hacking and HARKing (i.e. hypothesizing after the results are known, Parker et al. 2016), and demonstrably high rates of selective reporting and publication bias (e.g. Kimmel et al. 2023), it is conceivable that a substantial proportion of published studies in invasion science present findings that exaggerate impacts or even misrepresent the true direction of effects.

In this Discussion, paper we first present the results of a small literature analysis suggesting that – like the general discipline of ecology– invasion science has much room to improve with respect to practices that enhance reproducibility and reliability. We then consider the implications that low power has for published research and for study design, with specific attention to the challenge of natural variability (or heterogeneity) in field study systems. We also provide guidance for researchers and those relying on primary research (practitioners, policymakers) on how to navigate problems associated with low power. Next we emphasize the importance of distinguishing confirmatory from exploratory research, and describe how pre-registering study protocols and making code and non-sensitive data accessible and useable can maximize research reliability. We conclude with recommendations for how publishers and journal editors in invasion science can encourage progress towards a more transparent, reproducible, and reliable field.

Assessing practices relevant to reproducibility in invasion science

We were motivated to undertake an assessment of research practices in invasion science while conducting a systematic map of the evidence on the impacts of a selection of plant species invasive to riparian ecosystems in British Columbia. The protocol for that systematic map is now published (Mologni et al., 2023), and the systematic map itself is in progress. During this work we found substantial variation in reporting practices, including among studies on the same species. We therefore decided to evaluate a subset of the publications for their adherence to practices related to transparency, reproducibility and replicability. We used a structured, checklist-based approach, consistent with content analysis methods used in meta-research (e.g. Hardwicke et al. 2020).

Our assessment focused on studies concerning two plant species invasive to riparian ecosystems in British Columbia, Canada: *Elaeagnus angustifolia* (Russian Olive) and *Phalaris arundinacea* (Reed Canarygrass, thereafter RO and RCG respectively). We systematically identified all studies investigating their impacts on riparian ecosystems within North America (excluding Mexico), following the ROSES reporting standards for systematic maps and reviews (Haddaway et al., 2018). A detailed description can be found in the published protocol (Mologni et al., 2023). For the present study, we applied an additional filter: we removed studies published before the year 2000. This reflects a balance between excluding years in which there was limited awareness around reproducibility in ecology, with the need for a sufficient time period over which to reveal any obvious shifts in reporting practices.

At the outset we acknowledge that our study cannot reliably generalize to all invasion science, as it did not randomly sample from a representative literature base. Rather, our primary goal was to complement a discussion about transparency and reproducibility in invasion science by providing a descriptive overview of practices observed in studies contributing to our systematic map. Nevertheless, we can think of no obvious reason why studies concerning the impacts of RO and RCG would systematically differ in their research practices from studies examining impacts of other species.

Our checklist is a streamlined version of a checklist developed previously in our lab (Kast et al., 2023 see Appendix 1), which combined features of the "transparency and openness promotion" guidelines developed by the Center for Open Science (https://www.cos.io/initiatives/top-guidelines) and the checklist for peer reviewers developed by (Parker et al., 2018). It was also informed by the "transparency checklist" of Aczel et al. (2020). The checklist includes 14 items, and one of us (FM) assigned a score of "0" to each item not addressed, "0.5" to items partially addressed, "1" to items fully addressed, or "n/a" if the item was not applicable. We calculated the final study score by summing the item scores and dividing by the number of applicable questions, then converting to a percentage. We also calculated the overall, by-year and by-item average scores.

All items are equally weighted, though we acknowledge that some items could be considered more critical than others (see below). The full list and description of items can be found in the supplementary material (Appendix 1). The items fall into four categories and include the following questions:

- Pre-registration (Glossary): was the study pre-registered and were deviations from planned methods described?

- Materials availability: were availability statements included? Were data and code stored in openly accessible repositories and links and licences provided? Was the description of the experimental protocol sufficiently comprehensive to enable a replication attempt?
- Study design: was sample size determined (e.g. with a priori power analysis) and randomization described? If applicable, was the study authorized (i.e. a relevant permit was obtained if required)?
- Data analysis: were statistical methods justified and associated assumptions checked?

While it would have been ideal to have pre-registered this study, this was not done because the decision to undertake a checklist-based assessment was made after some of the literature search and evaluation activities had already been completed (as part of the systematic mapping). We also acknowledge that the scoring was completed by a single person (FM), so we lack an assessment of repeatability.

Our systematic literature search yielded a total of 877 studies, 402 for RO and 475 for RCG (Figure 1). The majority were extracted from Web of Science (n = 428), followed CABI Invasive Species Compendium (n = 173). The reviews yielded 209 studies, while an additional 67 were identified from other sources. After removing studies published before the year 2000, we obtained 50 studies. One appeared in both searches, so the final list contained 49 studies (citations for which are provided in Appendix 2), with an overall relevance rate (i.e. the percentage of studies included after screening and removing duplicates, Ridley et al., 2022) of 5.6%.

The average score across all 49 studies was 26% and ranged from 11 to 50%. The number of publications and scores increased slightly over time (Figure 2). Ten of the fourteen items averaged a score below 20% (Figure 3), two items were between 20% and 80% (randomization 38%; and statistical assumptions 75%) and the remaining two items scored high: describing the experimental protocol (95%) and rationalizing statistical methods (89%) (Figure 3). Ten studies (20%) made data available, eight via the supplemental material, and two through a dedicated repository (Price et al., 2018; Rowe et al., 2019). Price et al. (2018) was the single study that provided a data licence. Three studies (6%) mentioned or provided code: one provided a code availability statement but did not provide access to their code; one provided code in their supplemental material but did not refer to it in the main paper; and one referred to the code that was made available in their supplementary material. Only one of the 49 studies addressed the sample size item; Cordell et al. (2023) mentioned power in their methods section. One additional study considered low power in their discussion as a potential reason for some non-significant findings (Valente et al. 2019). Lastly, two studies provided access to both data and code (Rowe et al. 2019; Valente et al. 2019), and thus represent the only studies that could potentially be independently computationally reproduced.

A slightly more positive outlook emerges if we (i) focus on seven checklist items (data availability, code availability, sample size determination, randomization, detailing experimental protocol, detailing statistical methods, and checking statistical assumptions) that could arguably be considered "good enough" practices (sensu Wilson et al. 2017), and (ii) apply a more permissive scoring rubric, providing a "1" to the item being addressed in some fashion and zero otherwise. With this approach the average score across the 49 studies increases to 53% (minimum 29%, maximum 86%), with three studies (6%) scoring 6 out of 7 (sample size determination not discussed), and some indication of an improvement in practices in recent years (Appendix 3, Figure S2). We note that a study from our own lab (Collette and Pither, 2016) scored only 4/7 in this more generous assessment, reflecting our own lack of awareness at the time of the importance of some of the practices relevant to transparency and reproducibility.

These findings are consistent with what is observed in ecology and evolution research more generally (Parker et al., 2016; Roche et al., 2015), and indicate considerable room for improvement with respect to implementing practices that enhance transparency and reproducibility.

It was beyond the scope of our literature assessment to critically evaluate each study for its internal validity (e.g. the soundness of study design and statistical analyses), or its external validity, i.e. the degree to which its conclusions are likely to generalize to other locations or systems (Stanhope and Weinstein, 2023). We also emphasize that the biological findings presented in the individual research papers could be valuable despite any shortfalls in methodologies or reporting. It is also possible that some of the studies we evaluated could yield successful, independent conceptual replications (sensu Filazzola and Cahill, 2021). However, in their analysis of more than 88000 effects from almost 12000 primary research studies in ecology and evolution, Yang et al. (2024) found that, in the absence of selective reporting, studies reporting weak (P = 0.05) to moderate (P = 0.01) statistical evidence against a null hypothesis had an approximate successful replication probability of 0.38 (95% confidence limits: 0.34 – 0.41) and 0.56 (95% confidence limits: 0.51 – 0.58). These estimates are optimistic because selective reporting has long been identified as a systemic problem across disciplines (e.g. Bartoš et al. 2024), and was recently evidenced by a thorough meta-research study in ecology (Kimmel et al. 2023). A primary cause of low replication probabilities in ecology and evolution is low statistical power (Lemoine et al. 2016; Yang et al. 2023), which we address below.

Statistical power and published invasion science

Power represents the likelihood of detecting a true effect when it is present. Ideally, studies achieve a power of 80% (for a given alpha level, e.g. 0.05). In reality, logistical constraints, small to modest effect sizes, high inherent variability in natural systems (Kumschick et al. 2015; Catford et al. 2022), and poor study design severely limit power in ecology research – to as low as 15% or 13% according to two recent and extensive meta-research studies (Kimmel et al. 2023; Yang et al. 2023). Thus, severely underpowered studies are likely more common than not in invasion science (e.g. Davidson and Hewitt, 2014). This has profound practical implications for the discipline.

First, as described above, low statistical power reduces the likelihood of detecting real impacts when they are present (i.e. increase the likelihood of “false negatives”). This problem is made worse when researchers implicitly or explicitly interpret non-significant results as “no impact” or “no effect” without due consideration of power (Fidler et al. 2006; Davidson & Hewitt, 2014). Similary, the lower the power of the study (all else being equal), the more likely it is that claims of statistically significant impacts reflect false positives (Button et al. 2013). Second, if an under-powered study detects a true effect, it is likely that its estimate of magnitude is exaggerated (a type “M” error; Button et al. 2013). Exaggerated estimates of effect size (both positive and negative in direction) are expected under low power due to greater variability around the true effect size (Nakagawa and Cuthill, 2007). Combined with the bias towards publishing statistically significant findings, this results in an ecology literature rife with “exaggeration bias”, as evidenced by recent, large-scale meta-research studies in ecology (Kimmel et al. 2023; Yang et al. 2023). For instance, Kimmel et al. (2023) concluded that 63% of the estimates reported in under-powered studies have magnitudes that exaggerate the true effect size by a factor of two or more.

On the other hand, it is possible that field research studies in invasion science are less underpowered, on average, than the typical study in ecology and evolution. In their meta-research study of 3847 field-based experiments and observational studies investigating anthropogenic stressors, Yang et al. (2022) analyzed 316 effect sizes from meta-analyses on studies evaluating the impacts of invasive species. They found that the power achieved by single field studies could, by some estimates, be greater than 50%, which is noticeably higher than the power achieved by studies addressing other global change stressors (see Figure 3 in Yang et al. 2022). Correspondingly, rates of type “M” errors were lower than those observed in studies on other stressors (their Figure 4). The authors did not discuss these specific findings, but one possibility is that the true magnitudes of impacts (i.e. effect sizes) of invasive species within the primary research study systems (such as the degree to which alpha diversity is reduced with the presence of an invasive species) may have been greater, on average, than the magnitudes of impacts of other types of stressors. Indeed, Vilà et al. (2011) point out in their meta-analysis (which contributed to the study by Yang et al. 2022) that "in the vast majority of studies, invaded sites had high alien abundance", implying that the "treatment" (i.e. abundance of invader) was commonly pronounced, which yields greater power, all else being equal. Nevertheless, and despite an apparent focus on study systems where invader abundance (and thus the potential for real impacts) is high, the average power of invasion studies remains far below the ideal.

Ways forward

Recommendations for individual researchers, research institutions, funders, and journals for making progress towards more transparent, reproducible, and reliable research in ecology are provided elsewhere (Hampton et al., 2015; Parker et al. 2016; Parker et al. 2019; Powers and Hampton, 2019; O'Dea et al. 2021; Nakagawa et al. 2024; Purgar et al. 2024), and include helpful tools such as checklists. We also encourage researchers to consult recent guidance on describing and implementing statistical analyses, and reporting their outcomes (e.g. Forstmeier et al. 2017; Fidler et al. 2018; Davis and Kay, 2023; Popovic et al. 2024). Below, we focus on three specific issues within the context of invasion science: statistical power and study design in light of high natural variability in the field, confirmatory versus exploratory research and the value of pre-registration, and ways for publishers / journals to facilitate progress.

Statistical power and study design

Consideration of power is most relevant within the context of null hypothesis significance testing, but more generally wherever binary thresholds of significance (e.g. $P < 0.05$) are applied (including when deciding to eliminate or retain predictors in a statistical model; Fidler et al. 2018). There have been repeated calls for ecologists to shift away from this framework towards an emphasis on effect sizes and uncertainty (e.g. Johnson 1999; Fidler et al. 2006; Nakagawa and Cuthill, 2007), and while progress is being made (e.g. Fenesi et al. 2023), a perusal of recent issues of *Biological Invasions*, for example, show that binary thresholds of significance remain common. When such a framework is applied, we encourage researchers to conduct a prospective power analysis in the study design phase, and to be transparent about power shortfalls, as this facilitates objective interpretations of outcomes and helps raise awareness among scientists and practitioners about the importance of statistical power. Even more valuable would be to estimate rates of type S and type M errors (Gelman and Carlin, 2014), which focus more on estimates rather than significance (Nakagawa et al. 2024). Similarly, we urge policymakers and mangers to routinely consider power when interpreting primary research that employs this framework, and more specifically, to account for possible false negative, false positive, errors in magnitude or direction.

Conventional study designs aim to maximize power and precision by (among other ways) minimizing variation in potentially confounding factors (e.g. soil nutrients, water availability) (e.g. Kumschick et al. 2015). However, there is growing recognition that intentionally and strategically incorporating heterogeneity into field study designs may ultimately yield more robust and generalizable outcomes (e.g. Milcu et al. 2018; Nakagawa et al. 2024). This is especially important when considering the context-dependence of invasive species impacts (Kumschick et al. 2015; Catford et al. 2022). For instance, studies included in our literature assessment found that the Russian Olive both reduces (Mahoney et al., 2019) and has no effect on the richness of birds (Fischer et al. 2012) in Western U.S. riparian ecosystems. The Russian Olive and the Reed Canarygrass both negatively impact aquatic macroinvertebrate species; however, only in mountainous regions of the American West (Ringold et al., 2008). The Reed Canarygrass either reduces (Weilhoefer et al., 2017) or maintains (De Jager et al., 2015) soil organic content in the U.S. Ideally, such inconsistencies in species-specific impacts reflect "mechanistic context dependence" (Catford et al. 2022). In practice, they may also reflect methodological differences among studies, confounding factors, or shortfalls in study design (Catford et al. 2022). This presents a serious challenge to those aiming to synthesize and generalize research for the purpose of informing policy and management. Ways to evaluate and quantify, in a repeatable fashion, the reliability of primary studies contributing to evidence synthesis is an active area of research (e.g. Spake et al. 2022; Stanhope and Weinstein, 2023).

The most effective way to address the challenge of heterogeneity (natural variability) while simultaneously maximizing power is through collaborative efforts such as coordinated distributed experiments (Fraser et al. 2013; Borer et al. 2014, Yahdjian et al. 2021), wherein standardized protocols are used by all teams to maximize repeatability among and precision within each location. Such collaborations are gaining momentum across disciplines (Knollová et al., 2024; Guerrero-Ramírez et al., 2025), including within invasion science (see numerous examples in Packer et al. 2017), and should be prioritized by funding agencies (Nakagawa et al. 2024).

For individual research groups with limited resources, accommodating heterogeneity is more challenging. However, it could ultimately be more beneficial and informative to conduct multiple, less-well replicated (but more affordable/feasible) experiments that incorporate heterogeneity than to focus all resources into one large, well-powered study (IntHout et al. 2016; Nakagawa et al. 2024). Collaboration amongst researchers in a single region can facilitate this. Local collaborations can also help address the needs of practitioners focused on local, context-specific challenges, where within-study replication (i.e. the same team repeating an experiment across years or locations within a single study) can improve confidence in research outcomes (Filazzola and Cahill, 2021).

It is also important to emphasize here that interaction effects require substantially greater replication to achieve the same power for a given effect size (Nakagawa et al. 2024), so hypotheses about interactions are less amenable to reliable testing by individual research groups. Collaborations can also help with this by facilitating within-study replication.

Given practical challenges of field-based invasion science research, rather than striving for an unrealistic goal (e.g. 80% power), researchers could consider adopting the recently proposed “AHARP” principle (as high as reasonably practical) (Nakagawa et al. 2024), and focus more on ensuring that the study’s design is as strong as possible given the constraints at hand.

Regardless of the study design, it is crucial that researchers clearly communicate their target population (e.g. to what other regions, species, or systems is the study designed to generalize? Spake et al. 2022), as this influences interpretations of power and replicability (Ives, 2018), and facilitates independent evaluations of external validity or generalizability during evidence synthesis (O’Dea et al. 2021; Spake et al. 2022).

Confirmatory versus exploratory research and the need for (pre)registration

Confirmatory studies are experimental or observational studies specifically designed to test a-priori hypotheses. A-priori hypotheses are often central to risk or impact assessments, and are also commonplace because the field is rich with research hypotheses (e.g. invasion meltdown hypothesis; Enders et al, 2018; Jeschke and Heger, 2018). When designed properly (e.g. appropriate controls, sufficient replication, blinding, randomization), manipulative experiments are the most effective way to test hypotheses because – unlike non-manipulative, observational studies – they minimize the influence of confounding variables, and can ascribe cause and effect (Kumschick et al. 2015; Catford et al. 2022). Unfortunately, logistical and ethical constraints often limit opportunities for manipulative field experiments in invasion science. In their meta-analysis of the ecological impacts of exotic invasive plants, covering 199 articles with 1041 cases of invasion across 135 taxa, Vilà et al. (2011) reported 14% of studies involved manipulative experiment.

In contrast with confirmatory studies, *exploratory studies* are crucial for generating hypotheses, and therefore have more flexible workflows (Parker et al. 2016). However, the line between exploratory and confirmatory research is often blurred, with researchers passing off "significant" results found through exploration as support for a pre-specified hypothesis (i.e. HARKing) (Parker et al. 2016; Fraser et al. 2018). We urge invasion science researchers to clearly identify and distinguish analyses that are pre-planned from those that are unplanned and thus exploratory, so that readers can adjust interpretations accordingly (e.g. of *P*-values/significance tests reported as part of exploratory analyses; Forstmeier et al. 2017). This advice is equally relevant within the context collaborative distributed experiments or observation networks: despite their designs being inspired by one or a few specific questions or hypotheses, the data from such efforts are often used to address numerous additional questions and hypotheses (Borer et al. 2014). This is expected and valuable, if study design and data are fit for purpose and the exploratory/confirmatory distinction is clear. This applies, for example, to any future studies emerging from the Global Invader Impact Network (GIIN): the paper describing its study design describes several motivating hypotheses, but also identifies numerous, less well-defined avenues rich for exploration (Barney et al. 2015).

We recommend that invasion science researchers planning confirmatory studies – be they experimental *or* observational – preregister their protocols. Preregistration reduces the risk of bias by increasing transparency and minimizing the potential for outcomes to influence analytical choices or reporting (Hardwicke and Wagenmakers, 2023). In other words, preregistration can effectively reduce research waste (Purgar et al. 2024). Briefly, a preregistration involves completing a detailed description and rationale of the study design in advance of initiating the study, and submitting the protocol to a registry such as OSF registrations (https://osf.io/prereg). Among other things, the registration should clearly describe (i) the rationale for the planned level of replication (e.g. number of independent sampling plots), (ii) the statistical analyses to be used to test each hypothesis (and optionally, example code), and (iii) what would constitute evidence in support of, or contrary to said hypotheses (Hardwicke and Wagenmakers, 2023). Although reservations exist regarding pre-registration (see examples in Nosek et al. 2019; Simmons et al. 2021), its efficacy is increasingly clear (Hardwicke and Wagenmakers, 2023), including in ecology (Purgar et al. 2024).

With respect to planning analyses and reporting evidence, we urge invasion science researchers to heed long-standing calls (e.g. Nakagawa and Cuthill, 2007) to report effect sizes and confidence intervals instead of, or alongside, *P*-values, and to clearly interpret and communicate them in terms of biological significance (e.g. Fenesi et al. 2023). For observational, confirmatory studies, registrations should also clearly describe potential confounding variables and how these might limit the scope of inference.

Lastly, when combined with high-quality study designs and pre-registration, making code and non-sensitive, raw data available according to FAIR guidelines ensure that studies achieve their greatest value (Powers and Hampton, 2019; Culina et al. 2020; Roche et al. 2022), though we acknowledge barriers remain (Roche et al. 2014; Soeharjono and Roche, 2021; Gomes et al. 2022). Even with shortfalls in power, studies implementing this practice can meaningfully contribute to subsequent evidence synthesis efforts, which are increasingly being used by practitioners and policymakers in invasion science. Thus, we urge invasion science researchers to consult recent guidelines on sharing code and non-sensitive data (Abdill et al. 2024).

Publishers should standardize and enforce policies on transparency and open materials

Open science policies of the main journals in invasion science vary considerably (i.e. Biological Invasions, Neobiota, Invasive Plant Science and Management, and Wetlands, the latter being the most common journal in our bibliographic sample). These journals consistently encourage data sharing, sometimes in specialized repositories; however, this is not mandatory. Biological Invasions is the only journal that requires a data availability statement, albeit this does not require data themselves to be made available. Invasive Plant Science and Management encourages best practices in reporting methodologies, while Neobiota strongly encourages depositing methods and “protocols” in a repository (we consider a protocol akin to a pre-registration). Yet again, these policies are not mandatory. Springer journals (Biological Invasions and Wetlands) require sharing voucher specimens and identifiers. Wetlands also require the submission of specific data and materials (i.e. proteins, DNA, and RNA sequences) to appropriate repositories. Thus, policies vary greatly across journals, are typically non-mandatory, and address only a fraction of open science practices.

In recent years, data availability statements have increasingly become mandatory, and some journals now require authors to share their code too (Proceedings of the Royal Society B). American Naturalist now enlists data editors to evaluate data and code sharing of each manuscript submitted to the journal. Finally, Environmental Evidence requires registering a protocol (i.e. preregistration) in an open-access repository. Still, consistency in journal policies concerning open science good practices is lacking. In a pre-print awaiting peer-review, Ivimey-Cook et al. (2025) examined data and code sharing policies for 275 ecology and evolution journals, and found initial compliance by authors around code and data sharing was substantially higher among journals with mandatory versus optional polices. They speculate that inconsistencies in policy and wording of policies contribute to low compliance overall. Similarly, although code-sharing policies (whether mandatory or not) increase reproducibility potential, they are likely insufficient without enforcement (Culina et al. 2020; Sánchez-Tójar et al. 2025). We therefore urge editors of invasion science journals to agree on a consistent, explicit, and accessible policy that (i) requires making code and non-sensitive data accessible and useable, (ii) requires authors to clearly distinguish between confirmatory from exploratory analyses, and (iii) encourages pre-registration of study protocols.

Conclusion

Invasive species managers and policymakers rely on research outputs to inform their decisions and actions. Our findings highlight an opportunity and need to improve the transparency, reproducibility, and reliability of invasion science research. Ultimately, addressing this challenge in invasion science will help ensure resources are allocated optimally, and in ways that maximize success. We emphasize, however, that adapting research protocols towards the ideal takes time and is not adequately supported or incentivised (O'Dea et al. 2021). Researchers should embrace and celebrate every advance made, however small.

References

Abdill, R. J., Talarico, E., & Grieneisen, L. (2024). A how-to guide for code sharing in biology. *PLOS Biology*, *22*(9), e3002815. https://doi.org/10.1371/journal.pbio.3002815

Aczel, B., Szaszi, B., Sarafoglou, A., Kekecs, Z., Kucharský, Š., Benjamin, D., Chambers, C. D., Fisher, A., Gelman, A., Gernsbacher, M. A., Ioannidis, J. P., Johnson, E., Jonas, K., Kousta, S., Lilienfeld, S. O., Lindsay, D. S., Morey, C. C., Munafò, M., Newell, B. R., … Wagenmakers, E.-J. (2020). A consensus-based transparency checklist. *Nature Human Behaviour*, *4*(1), Article 1. https://doi.org/10.1038/s41562-019-0772-6

Barney, J. N., Tekiela, D. R., Barrios-Garcia, M. N., Dimarco, R. D., Hufbauer, R. A., Leipzig-Scott, P., Nuñez, M. A., Pauchard, A., Pyšek, P., Vítková, M., & Maxwell, B. D. (2015). Global Invader Impact Network (\textlessspan style="font-variant:small-caps;"\textgreaterGIIN\textless/span\textgreater): Toward standardized evaluation of the ecological impacts of invasive plants. Ecology and Evolution, 5(14), 2878–2889. https://doi.org/10.1002/ece3.1551

Bartoš, F., Maier, M., Wagenmakers, E., Nippold, F., Doucouliagos, H., Ioannidis, J. P. A., Otte, W. M., Sladekova, M., Deresssa, T. K., Bruns, S. B., Fanelli, D., & Stanley, T. D. (2024). Footprint of publication selection bias on meta-analyses in medicine, environmental sciences, psychology, and economics. Research Synthesis Methods, 15(3), 500–511. https://doi.org/10.1002/jrsm.1703

Begley, C. G., & Ioannidis, J. P. A. (2015). Reproducibility in Science. *Circulation Research*, *116*(1), 116–126. https://doi.org/10.1161/CIRCRESAHA.114.303819

Borer, E.T., Harpole, W.S., Adler, P.B., Lind, E.M., Orrock, J.L., Seabloom, E.W. and Smith, M.D. (2014), Finding generality in ecology: a model for globally distributed experiments. *Methods in Ecology and Evolution*, 5: 65-73. https://doi.org/10.1111/2041-210X.12125

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Catford, J. A., Wilson, J. R. U., Pyšek, P., Hulme, P. E., & Duncan, R. P. (2022). Addressing context dependence in ecology. Trends in Ecology & Evolution, 37(2), 158–170. https://doi.org/10.1016/j.tree.2021.09.007

Collette, L. K. D., & Pither, J. (2016). Insect assemblages associated with the exotic riparian shrub Russian olive (Elaeagnaceae), and co-occurring native shrubs in British Columbia, Canada. The Canadian Entomologist, 148(3), 316–328. https://doi.org/10.4039/tce.2015.63

Cordell, J. R., Kidd, S. A., Toft, J. D., Borde, A. B., Cullinan, V. I., Sagar, J., Corbett, C. A. (2023). Ecological effects of reed canarygrass in the lower Columbia River. Biological Invasions, 25, 3485–3502. https://doi.org/10.1007/s10530-023-03119-y

Culina, A., Berg, I. van den, Evans, S., & Sánchez-Tójar, A. (2020). Low availability of code in ecology: A call for urgent action. *PLOS Biology*, *18*(7), e3000763. https://doi.org/10.1371/journal.pbio.3000763

Davidson, A. D., & Hewitt, C. L. (2014). How often are invasion-induced ecological impacts missed? Biological Invasions, 16(5), 1165–1173. https://doi.org/10.1007/s10530-013-0570-4

Davis, A. J., & Kay, S. (2023). Writing statistical methods for ecologists. Ecosphere, 14(5), e4539. https://doi.org/10.1002/ecs2.4539

De Jager, N. R., Swanson, W., Strauss, E. A., Thomsen, M., & Yin, Y. (2015). Flood pulse effects on nitrification in a floodplain forest impacted by herbivory, invasion, and restoration. *Wetlands Ecology and Management*, *23*(6), 1067–1081. https://doi.org/10.1007/s11273-015-9445-z

Enders, M., Havemann, F., Ruland, F., Bernard-Verdier, M., Catford, J. A., Gómez-Aparicio, L., Haider, S., Heger, T., Kueffer, C., Kühn, I., Meyerson, L. A., Musseau, C., Novoa, A., Ricciardi, A., Sagouis, A., Schittko, C., Strayer, D. L., Vilà, M., Essl, F., … Jeschke, J. M. (2020). A conceptual map of invasion biology: Integrating hypotheses into a consensus network. Global Ecology and Biogeography, 29(6), 978–991. https://doi.org/10.1111/geb.13082

Fenesi, A., Botta-Dukát, Z., Miholcsa, Z., Szigeti, V., Molnár, C., Sándor, D., Szabó, A., Kuhn, T., & Kovács-Hostyánszki, A. (2023). No consistencies in abundance–impact relationships across herbaceous invasive species and ecological impact metrics. Journal of Ecology, 111(5), 1120–1138. https://doi.org/10.1111/1365-2745.14085

Fidler, F., Burgman, M. A., Cumming, G., Buttrose, R., & Thomason, N. (2006). Impact of Criticism of Null-Hypothesis Significance Testing on Statistical Reporting Practices in Conservation Biology. Conservation Biology, 20(5), 1539–1544. https://doi.org/10.1111/j.1523-1739.2006.00525.x

Fidler, F., Fraser, H., McCarthy, M. A., & Game, E. T. (2018). Improving the transparency of statistical reporting in Conservation Letters. Conservation Letters, 11(2), e12453. https://doi.org/10.1111/conl.12453

Filazzola, A., & Cahill, J. F. (2021). Replication in field ecology: Identifying challenges and proposing solutions. Methods in Ecology and Evolution, 12(10), 1780–1792. https://doi.org/10.1111/2041-210X.13657

Fischer, R. A., Valente, J. J., Guilfoyle, M. P., Kaller, M. D., Jackson, S. S., & Ratti, J. T. (2012). Bird Community Response to Vegetation Cover and Composition in Riparian Habitats Dominated by Russian Olive (*Elaeagnus angustifolia*). *Northwest Science*, *86*(1), 39–52. https://doi.org/10.3955/046.086.0104

Forstmeier, W., Wagenmakers, E.-J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings – a practical guide. Biological Reviews, 92(4), 1941–1968. https://doi.org/10.1111/brv.12315

Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. PLOS ONE, 13(7), e0200303. https://doi.org/10.1371/journal.pone.0200303

Fraser, L. H., Henry, H. A., Carlyle, C. N., White, S. R., Beierkuhnlein, C., Cahill, J. F., Casper, B. B., Cleland, E., Collins, S. L., Dukes, J. S., Knapp, A. K., Lind, E., Long, R., Luo, Y., Reich, P. B., Smith, M. D., Sternberg, M., & Turkington, R. (2013). Coordinated distributed experiments: An emerging tool for testing global hypotheses in ecology and environmental science. *Frontiers in Ecology and the Environment*, *11*(3), 147–155. https://doi.org/10.1890/110279

Gelman, A., & Carlin, J. (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. Perspectives on Psychological Science, 9(6), 641–651. https://doi.org/10.1177/1745691614551642

Gomes, D. G. E., Pottier, P., Crystal-Ornelas, R., Hudgins, E. J., Foroughirad, V., Sánchez-Reyes, L. L., Turba, R., Martinez, P. A., Moreau, D., Bertram, M. G., Smout, C. A., & Gaynor, K. M. (2022). Why don’t we share data and code? Perceived barriers and benefits to public archiving practices. *Proceedings of the Royal Society B: Biological Sciences*, 289(1987), 20221113. https://doi.org/10.1098/rspb.2022.1113

Guerrero-Ramírez, N., Ah-Peng, C., Borges, P., Camperio, G., Nascimento, L. de, Elias, R., Essl, F., Gabriel, R., Lenzner, B., Mologni, F., Morgado, L., Patiño, J., Suter, S., Martin, W. W., Zemp, C., & Kreft, H. (2025). Biodiversity monitoring of island ecosystems (BioMonI). ARPHA Conference Abstracts, 8, e155875. https://doi.org/10.3897/aca.8.e155875

Haddaway, N. R., Macura, B., Whaley, P., & Pullin, A. S. (2018). ROSES RepOrting standards for Systematic Evidence Syntheses: Pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environmental Evidence*, *7*(1), 7. https://doi.org/10.1186/s13750-018-0121-7

Hampton, S. E., Anderson, S. S., Bagby, S. C., Gries, C., Han, X., Hart, E. M., Jones, M. B., Lenhardt, W. C., MacDonald, A., Michener, W. K., Mudge, J., Pourmokhtarian, A., Schildhauer, M. P., Woo, K. H., & Zimmerman, N. (2015). The Tao of open science for ecology. Ecosphere, 6(7), art120. https://doi.org/10.1890/ES14-00402.1

Hardwicke, T.E., Wallach, J.D., Kidwell, M.C., Bendixen, T., Crüwell, S. & Ioannidis, J.P.A. (2020). An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014–2017). *R. Soc. open sci.*, 7, 190806.

Hardwicke, T. E., & Wagenmakers, E.-J. (2023). Reducing bias, increasing transparency and calibrating confidence with preregistration. Nature Human Behaviour, 7(1), 15–26. https://doi.org/10.1038/s41562-022-01497-2

IntHout, J., Ioannidis, J. P., & Borm, G. F. (2016). Obtaining evidence by a single well-powered trial or several modestly powered trials. Statistical Methods in Medical Research, 25(2), 538–552. https://doi.org/10.1177/0962280212461098

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. PLOS Medicine, 2(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Ioannidis, J. P. A. (2024). Transparency, bias, and reproducibility across science: A meta-research view. Journal of Clinical Investigation, 134(22), e181923. https://doi.org/10.1172/JCI181923

Ives, A. R. (2018). Informative Irreproducibility and the Use of Experiments in Ecology. BioScience, 68(10), 746–747. https://doi.org/10.1093/biosci/biy090

Ivimey-Cook, E., Sánchez-Tójar, A., Berberi, I., Culina, A., Roche, D., Almeida, R., Amin, B., Bairos-Novak, K., Balti, H., Bertram, M., Bliard, L., Byrne, I., Chan, Y.-C., Cioffi, W., Corbel, Q., Elsy, A., Florko, K., Gould, E., Grainger, M., … Moran, N. (2025). From Policy to Practice: Progress towards Data- and Code-Sharing in Ecology and Evolution. California Digital Library (CDL). https://doi.org/10.32942/x2492q

Jeschke, J. M., & Heger, T. (2018). Invasion biology: Hypotheses and evidence (1st ed.). CABI.

Johnson, D. H. (1999). The Insignificance of Statistical Significance Testing. The Journal of Wildlife Management, 63(3), 763. https://doi.org/10.2307/3802789

Kast, S., Shelton, B., & Pither, J. (2023, February 15). *Mycorrizhal Research Reproducibility Review*. Open Science Framework. https://doi.org/10.17605/OSF.IO/9BV8Q

Kimmel, K., Avolio, M. L., & Ferraro, P. J. (2023). Empirical evidence of widespread exaggeration bias and selective reporting in ecology. Nature Ecology & Evolution, 7(9), 1525–1536. https://doi.org/10.1038/s41559-023-02144-3

Knollová, I., Chytrý, M., Bruelheide, H., Dullinger, S., Jandt, U., Bernhardt-Römermann, M., Biurrun, I., de Bello, F., Glaser, M., Hennekens, S., Jansen, F., Jiménez-Alfaro, B., Kadaš, D., Kaplan, E., Klinkovská, K., Lenzner, B., Pauli, H., Sperandii, M. G., Verheyen, K., … Essl, F. (2024). ReSurveyEurope: A database of resurveyed vegetation plots in Europe. *Journal of Vegetation Science*, *35*(2), e13235. https://doi.org/10.1111/jvs.13235

Kumschick, S., Gaertner, M., Vilà, M., Essl, F., Jeschke, J. M., Pyšek, P., Ricciardi, A., Bacher, S., Blackburn, T. M., Dick, J. T. A., Evans, T., Hulme, P. E., Kühn, I., Mrugała, A., Pergl, J., Rabitsch, W., Richardson, D. M., Sendek, A., & Winter, M. (2015). Ecological Impacts of Alien Species: Quantification, Scope, Caveats, and Recommendations. BioScience, 65(1), 55–63. https://doi.org/10.1093/biosci/biu193

Lamberink, H. J., Otte, W. M., Sinke, M. R. T., Lakens, D., Glasziou, P. P., Tijdink, J. K., & Vinkers, C. H. (2018). Statistical power of clinical trials increased while effect size remained stable: An empirical analysis of 136,212 clinical trials between 1975 and 2014. *Journal of Clinical Epidemiology*, *102*, 123–128. https://doi.org/10.1016/j.jclinepi.2018.06.014

Lemoine, N. P., Hoffman, A., Felton, A. J., Baur, L., Chaves, F., Gray, J., Yu, Q., & Smith, M. D. (2016). Underappreciated problems of low replication in ecological field studies. Ecology, 97(10), 2554–2561. https://doi.org/10.1002/ecy.1506

Mahoney, S. M., Smith, A. N. B., Motyka, P. J., Lundgren, E. J., Winton, R. R., Stevens, B., & Johnson, M. J. (2019). Russian olive habitat along an arid river supports fewer bird species, functional groups and a different species composition relative to mixed vegetation habitats. *Journal of Arid Environments*, *167*, 26–33. https://doi.org/10.1016/j.jaridenv.2019.04.018

Milcu, A., Puga-Freitas, R., Ellison, A. M., Blouin, M., Scheu, S., Freschet, G. T., Rose, L., Barot, S., Cesarz, S., Eisenhauer, N., Girin, T., Assandri, D., Bonkowski, M., Buchmann, N., Butenschoen, O., Devidal, S., Gleixner, G., Gessler, A., Gigon, A., … Roy, J. (2018). Genotypic variability enhances the reproducibility of an ecological study. Nature Ecology & Evolution, 2(2), 279–287. https://doi.org/10.1038/s41559-017-0434-x

Mologni, F., Moffat, C. E., & Pither, J. (2023). Collating existing evidence on cumulative impacts of invasive plant species in riparian ecosystems of British Columbia, Canada: A systematic map protocol. *Environmental Evidence*, *12*(1). https://doi.org/10.1186/s13750-023-00320-3

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. Biological Reviews, 82(4), 591–605. https://doi.org/10.1111/j.1469-185X.2007.00027.x

Nakagawa, S., Lagisz, M., Yang, Y., Drobniak, S. M. (2024). Finding the right power balance: Better study design and collaboration can reduce dependence on statistical power. *PLOS Biology*, 22 (1), e3002423. https://doi.org/10.1371/journal.pbio.3002423

National Academies Of Sciences, Engineering, And Medicine. (2019). *Reproducibility and Replicability in Science*. The National Academies Press. https://doi.org/10.17226/25303

Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., van ’t Veer, A. E., & Vazire, S. (2019). Preregistration Is Hard, And Worthwhile. Trends in Cognitive Sciences, 23(10), 815–818. https://doi.org/10.1016/j.tics.2019.07.009

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. Proceedings of the National Academy of Sciences, 115(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Struhl, M. K., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology*, *73*(Volume 73, 2022), 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

O'Dea, R.E., Lagisz, M., Jennions, M.D., Koricheva, J., Noble, D.W
.A., Parker, T.H., Gurevitch, J., Page, M.J., Stewart, G., Moher, D. and Nakagawa, S. (2021), Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension. Biol Rev, 96: 1695-1722. https://doi.org/10.1111/brv.12721

O'Dea, R. E., Parker, T. H., Chee, Y. E., Culina, A., Drobniak, S. M., Duncan, D. H., Fidler, F., Gould, E., Ihle, M., Kelly, C. D., Lagisz, M., Roche, D. G., Sánchez-Tójar, A., Wilkinson, D. P., Wintle, B. C., & Nakagawa, S. (2021). Towards open, reliable, and transparent ecology and evolutionary biology. BMC Biology, 19(1), 68. https://doi.org/10.1186/s12915-021-01006-3

Packer, J. G., Meyerson, L. A., Richardson, D. M., Brundu, G., Allen, W. J., Bhattarai, G. P., Brix, H., Canavan, S., Castiglione, S., Cicatelli, A., Čuda, J., Cronin, J. T., Eller, F., Guarino, F., Guo, W.-H., Guo, W.-Y., Guo, X., Hierro, J. L., Lambertini, C., … Pyšek, P. (2017). Global networks for invasion science: Benefits, challenges and guidelines. Biological Invasions, 19(4), 1081–1096. https://doi.org/10.1007/s10530-016-1302-3

Parker, T. H., & Yang, Y. (2023). Exaggerated effects in ecology. *Nature Ecology & Evolution*, *7*(9), Article 9. https://doi.org/10.1038/s41559-023-02156-z

Parker, T. H., Forstmeier, W., Koricheva, J., Fidler, F., Hadfield, J. D., Chee, Y. E., Kelly, C. D., Gurevitch, J., & Nakagawa, S. (2016). Transparency in Ecology and Evolution: Real Problems, Real Solutions. *Trends in Ecology & Evolution*, *31*(9), 711–719. https://doi.org/10.1016/j.tree.2016.07.002

Parker, T. H., Griffith, S. C., Bronstein, J. L., Fidler, F., Foster, S., Fraser, H., Forstmeier, W., Gurevitch, J., Koricheva, J., Seppelt, R., Tingley, M. W., & Nakagawa, S. (2018). Empowering peer reviewers with a checklist to improve transparency. *Nature Ecology & Evolution*, *2*(6), 929–935. https://doi.org/10.1038/s41559-018-0545-z

Parker, T., Fraser, H., & Nakagawa, S. (2019). Making conservation science more reliable with preregistration and registered reports. Conservation Biology, 33(4), 747–750. https://doi.org/10.1111/cobi.13342

Popovic, G., Mason, T. J., Drobniak, S. M., Marques, T. A., Potts, J., Joo, R., Altwegg, R., Burns, C. C. I., McCarthy, M. A., Johnston, A., Nakagawa, S., McMillan, L., Devarajan, K., Taggart, P. L., Wunderlich, A., Mair, M. M., Martínez-Lanfranco, J. A., Lagisz, M., & Pottier, P. (2024). Four principles for improved statistical ecology. *Methods in Ecology and Evolution*, *15*(2), 266–281. https://doi.org/10.1111/2041-210X.14270

Powers, S. M., & Hampton, S. E. (2019). Open science, reproducibility, and transparency in ecology. Ecological Applications, 29(1), e01822. https://doi.org/10.1002/eap.1822

Price, E. P. F., Spyreas, G., & Matthews, J. W. (2018). Biotic homogenization of regional wetland plant communities within short time-scales in the presence of an aggressive invader. *Journal of Ecology*, *106*(3), 1180–1190. https://doi.org/10.1111/1365-2745.12883

Purgar, M., Glasziou, P., Klanjscek, T., Nakagawa, S., & Culina, A. (2024). Supporting study registration to reduce research waste. Nature Ecology & Evolution, 8(8), 1391–1399. https://doi.org/10.1038/s41559-024-02433-5

Purgar, M., Klanjscek, T., & Culina, A. (2022). Quantifying research waste in ecology. Nature Ecology & Evolution, 6(9), 1390–1397. https://doi.org/10.1038/s41559-022-01820-0

Ridley, F. A., Hickinbotham, E. J., Suggitt, A. J., McGowan, P. J. K., & Mair, L. (2022). The scope and extent of literature that maps threats to species globally: A systematic map. *Environmental Evidence*, *11*(1), Article 1. https://doi.org/10.1186/s13750-022-00279-7

Ringold, P. L., Magee, T. K., & Peck, D. V. (2008). Twelve invasive plant taxa in US western riparian ecosystems. *Journal of the North American Benthological Society*, *27*(4), 949–966. https://doi.org/10.1899/07-154.1

Roche, D. G., Berberi, I., Dhane, F., Lauzon, F., Soeharjono, S., Dakin, R., & Binning, S. A. (2022). Slow improvement to the archiving quality of open datasets shared by researchers in ecology and evolution. Proceedings of the Royal Society B: Biological Sciences, 289(1975), 20212780. https://doi.org/10.1098/rspb.2021.2780

Roche, D. G., Kruuk, L. E. B., Lanfear, R., & Binning, S. A. (2015). Public Data Archiving in Ecology and Evolution: How Well Are We Doing? PLOS Biology, 13(11), e1002295. https://doi.org/10.1371/journal.pbio.1002295

Roche, D. G., Lanfear, R., Binning, S. A., Haff, T. M., Schwanz, L. E., Cain, K. E., Kokko, H., Jennions, M. D., & Kruuk, L. E. B. (2014). Troubleshooting Public Data Archiving: Suggestions to Increase Participation. PLoS Biology, 12(1), e1001779. https://doi.org/10.1371/journal.pbio.1001779

Rowe, J. C., Duarte, A., Pearl, C. A., McCreary, B., Galvan, S. K., Peterson, J. T., & Adams, M. J. (2019). Disentangling effects of invasive species and habitat while accounting for observer error in a long-term amphibian study. *Ecosphere*, *10*(4), e02674. https://doi.org/10.1002/ecs2.2674

Roy, H. E., Pauchard, A., Stoett, P., Renard Truong, T., Bacher, S., Galil, B. S., Hulme, P. E., Ikeda, T., Sankaran, K., McGeoch, M. A., Meyerson, L. A., Nuñez, M. A., Ordonez, A., Rahlao, S. J., Schwindt, E., Seebens, H., Sheppard, A. W., & Vandvik, V. (2024). *IPBES Invasive Alien Species Assessment: Summary for Policymakers* (Version 3). Zenodo. https://doi.org/10.5281/ZENODO.7430692

Sánchez-Tójar, A., Bezine, A., Purgar, M., & Culina, A. (2025). Code-sharing policies are associated with increased reproducibility potential of ecological findings. Peer Community Journal, 5, e37. https://doi.org/10.24072/pcjournal.541

Shackelford, G. E., Martin, P. A., Hood, A. S. C., Christie, A. P., Kulinskaya, E., & Sutherland, W. J. (2021). Dynamic meta-analysis: A method of using global evidence for local decision making. BMC Biology, 19(1), 33. https://doi.org/10.1186/s12915-021-00974-w

Simmons, J., Nelson, L., & Simonsohn, U. (2021). Pre-registration: Why and How. Journal of Consumer Psychology, 31(1), 151–162. https://doi.org/10.1002/jcpy.1208

Soeharjono, S., & Roche, D. G. (2021). Reported Individual Costs and Benefits of Sharing Open Data among Canadian Academic Faculty in Ecology and Evolution. BioScience, 71(7), 750–756. https://doi.org/10.1093/biosci/biab024

Spake, R., O'Dea, R. E., Nakagawa, S., Doncaster, C. P., Ryo, M., Callaghan, C. T., & Bullock, J. M. (2022). Improving quantitative synthesis to achieve generality in ecology. Nature Ecology & Evolution, 6(12), 1818–1828. https://doi.org/10.1038/s41559-022-01891-z

Stanhope, J., & Weinstein, P. (2023). Critical appraisal in ecology: What tools are available, and what is being used in systematic reviews? Research Synthesis Methods, 14(3), 342–356. https://doi.org/10.1002/jrsm.1609

Valente, J. J., McCune, K. B., Tamulonis, R. A., Neipert, E. S., & Fischer, R. A. (2019). Removal pattern mitigates negative, short-term effects of stepwise Russian olive eradication on breeding birds. Ecosphere, 10(5), e02756. https://doi.org/10.1002/ecs2.2756

Vilà, M., Espinar, J. L., Hejda, M., Hulme, P. E., Jarošík, V., Maron, J. L., Pergl, J., Schaffner, U., Sun, Y., & Pyšek, P. (2011). Ecological impacts of invasive alien plants: A meta-analysis of their effects on species, communities and ecosystems. Ecology Letters, 14(7), 702–708. https://doi.org/10.1111/j.1461-0248.2011.01628.x

Weilhoefer, C. L., Williams, D., Nguyen, I., Jakstis, K., & Fischer, C. (2017). The effects of reed canary grass (Phalaris arundinacea L.) on wetland habitat and arthropod community composition in an urban freshwater wetland. *Wetlands Ecology and Management*, *25*(2), 159–175. https://doi.org/10.1007/s11273-016-9507-x

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. PLOS Computational Biology, 13(6), e1005510. https://doi.org/10.1371/journal.pcbi.1005510

Yahdjian, L., Sala, O. E., PiÑEiro-Guerra, J. M., Knapp, A. K., Collins, S. L., Phillips, R. P., & Smith, M. D. (2021). Why Coordinated Distributed Experiments Should Go Global. *BioScience*, *71*(9), 918–927. https://doi.org/10.1093/biosci/biab033

Yang, Y., Hillebrand, H., Lagisz, M., Cleasby, I., & Nakagawa, S. (2022). Low statistical power and overestimated anthropogenic impacts, exacerbated by publication bias, dominate field studies in global change biology. Global Change Biology, 28(3), 969–989. https://doi.org/10.1111/gcb.15972

Yang, Y., Sánchez-Tójar, A., O'Dea, R. E., Noble, D. W. A., Koricheva, J., Jennions, M. D., Parker, T. H., Lagisz, M., & Nakagawa, S. (2023). Publication bias impacts on effect size, statistical power, and magnitude (Type M) and sign (Type S) errors in ecology and evolutionary biology. BMC Biology, 21(1), 71. https://doi.org/10.1186/s12915-022-01485-y

Yang, Y., Van Zwet, E., Ignatiadis, N., & Nakagawa, S. (2024). A large-scale in silico replication of ecological and evolutionary studies. Nature Ecology & Evolution, 8(12), 2179–2183. https://doi.org/10.1038/s41559-024-02530-5

Figure Captions

Figure 1 Screening process for the systematic extraction of the evidence on the impacts of Russian olive and Reed canarygrass in riparian ecosystems in British Columbia, Canada. The screening process follows the ROSES reporting standards (Haddaway et al., 2018). 'Records extracted from bibliographic searches' refers to studies extracted from WOS (Web of Science), CABI and the reviews identified from those sources.

Figure 2 The number of studies on the impacts of Russian olive and Reed canarygrass in riparian ecosystems in British Columbia, Canada and their average reproducibility score by year since the year 2000. On the y-axis are the number of studies (left) and the average reproducibility score (right), while on the x-axis is the year of publication. The number of studies is represented by grey bars, the average reproducibility score by year is represented by the black line.

Figure 3 Plots displaying average reproducibility scores by item category for studies investigating the impacts of Russian olive and Reed canarygrass in riparian ecosystems in British Columbia, Canada.

Figure 1

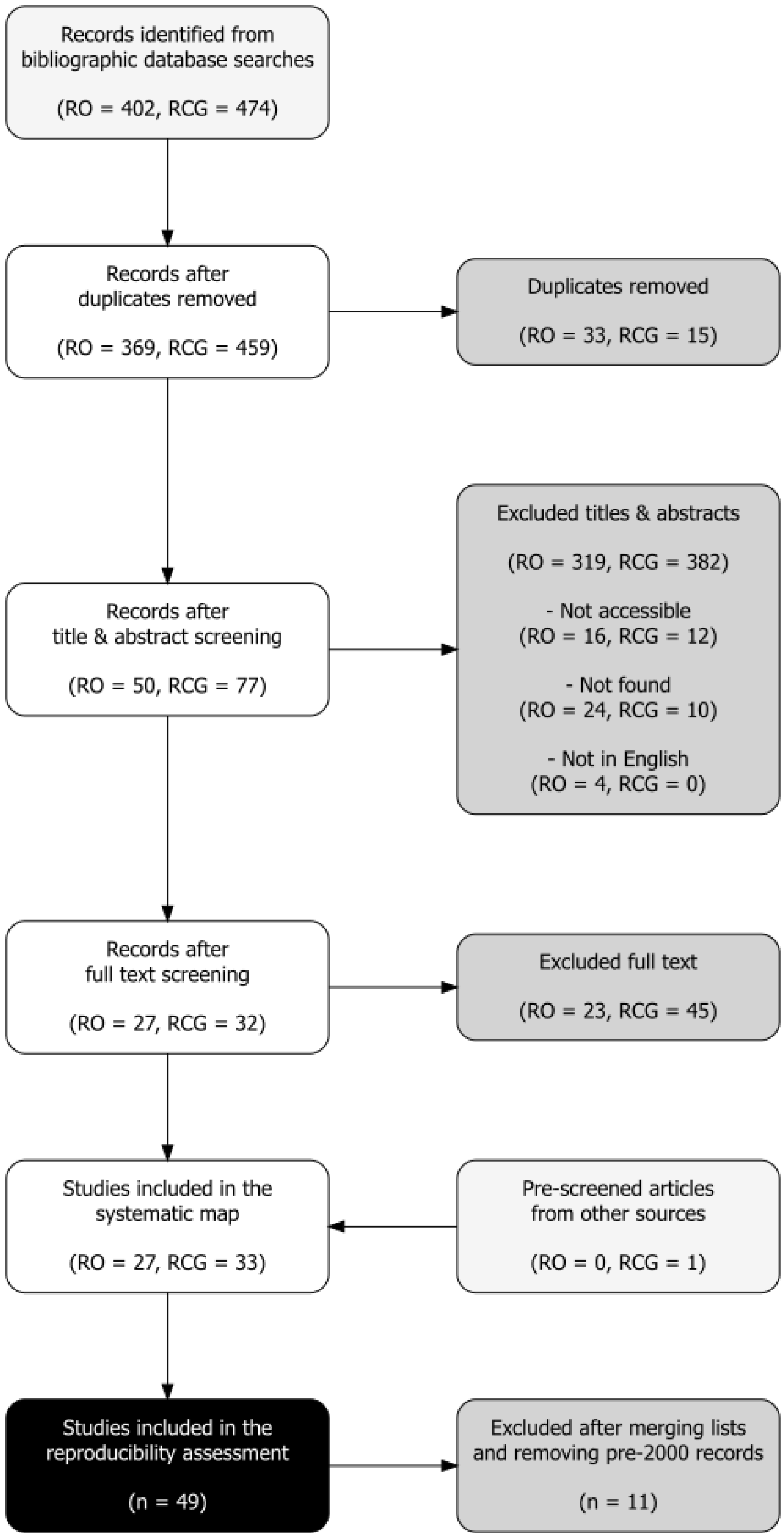
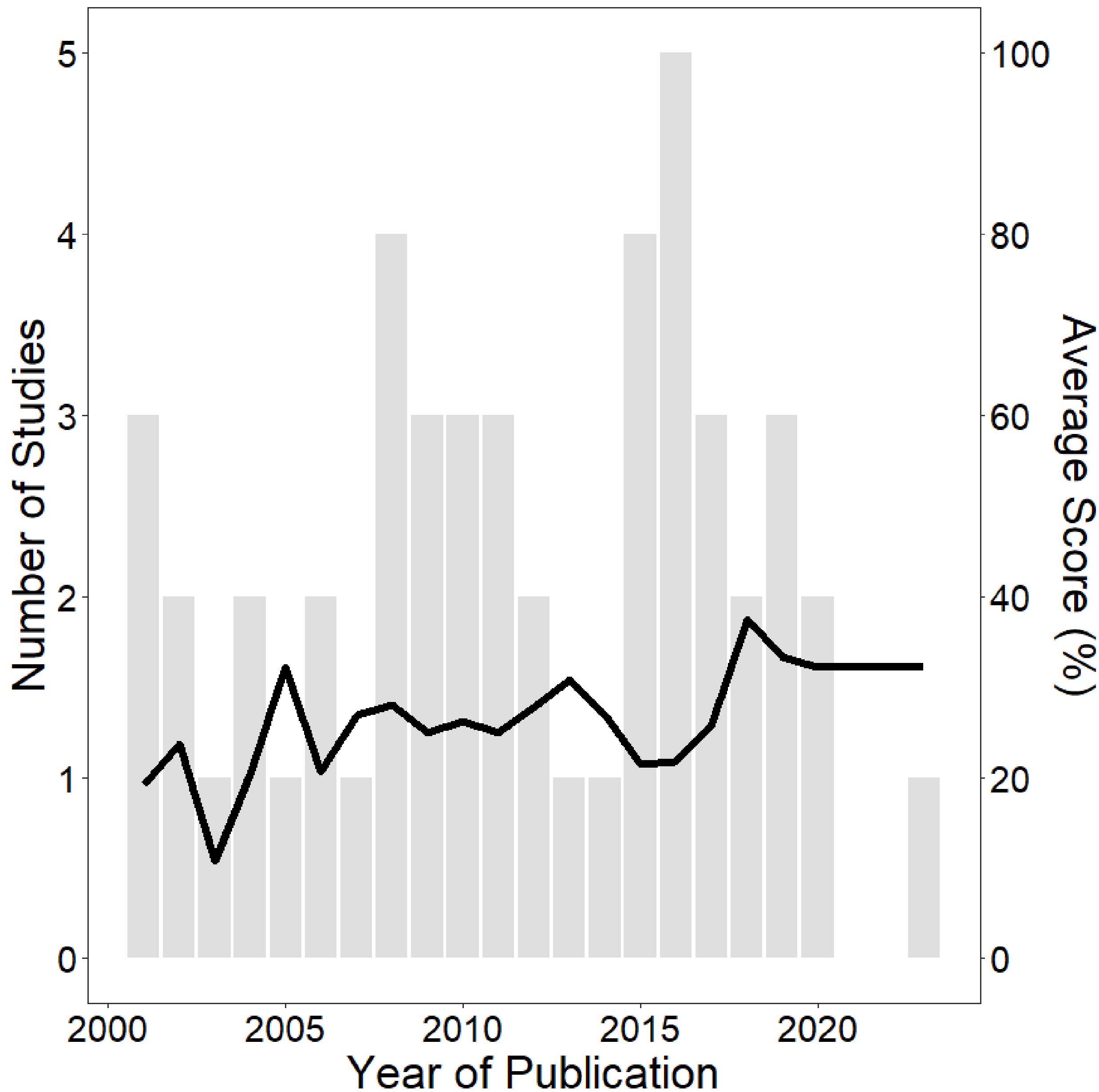
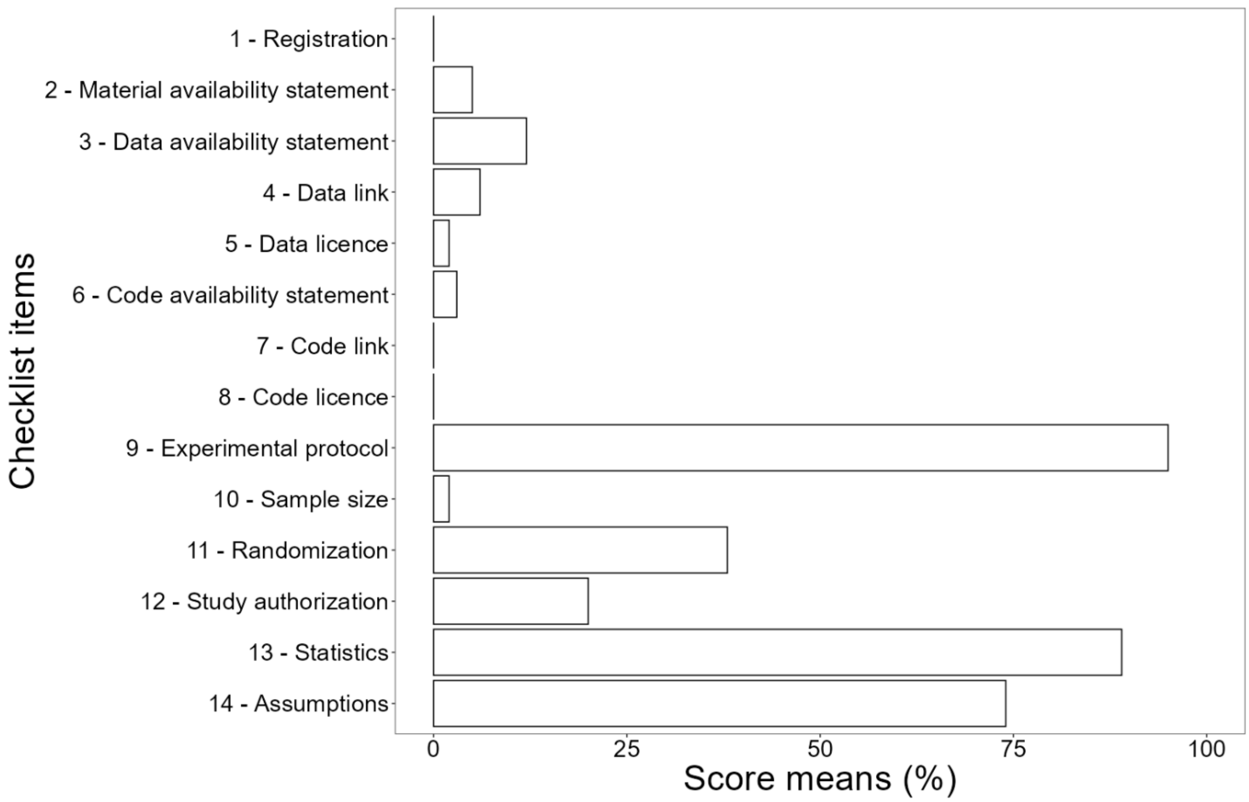

Records identified from bibliographic database searches
(RO = 402, RCG = 474)
Records after duplicates removed
(RO = 369, RCG = 459)
Duplicates removed
(RO = 33, RCG = 15)
Excluded titles & abstracts
(RO = 319, RCG = 382)
- Not accessible
(RO = 16, RCG = 12)
- Not found
(RO = 24, RCG = 10)
- Not in English
(RO = 4, RCG = 0)
Records after title & abstract screening
(RO = 50, RCG = 77)
Records after full text screening
(RO = 27, RCG = 32)
Excluded full text
(RO = 23, RCG = 45)
Studies included in the systematic map
(RO = 27, RCG = 33)
Pre-screened articles from other sources
(RO = 0, RCG = 1)
Studies included in the reproducibility assessment
(n = 49)
Excluded after merging lists and removing pre-2000 records
(n = 11)

Figure 2



5
4
3
2
1
0
100
80
60
40
20
0
Number of Studies
Average Score (%)
2000
2005
2010
2015
2020
Year of Publication

Figure 3


1 - Registration
2 - Material availability statement
3 - Data availability statement
4 - Data link
5 - Data licence
6 - Code availability statement
7 - Code link
8 - Code licence
9 - Experimental protocol
10 - Sample size
11 - Randomization
12 - Study authorization
13 - Statistics
14 - Assumptions
Checklist items
0
25
50
75
100
Score means (%)

Data availability statement

Data and code are available at https://doi.org/10.5281/zenodo.14288882, under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Data include a digital copy of all the articles assessed in this study.

Acknowledgments

We thank Mathew Vis-Dunbar (UBC librarian) for providing valuable comments.

Funding

FM is funded by the Ministry of Forests, British Columbia, Canada, and by the Irving K. Barber Faculty of Science at the Okanagan campus of UBC. JP acknowledges financial support from the Natural Science and Engineering Council of Canada (Discovery Grant 2020-43906543). Open access funding provided by the University of Vienna.

Competing interests

The authors acknowledge that JP is a coauthor of one of the studies evaluated in this work. As such, he was not involved in any part of the assessment of that study.

Supplementary material

Appendix 1 - Full scoring rubric.

Appendix 2 - List of studies included in this article and associated scores by rubric item.

Appendix 3 - Supplementary information

Appendix 1 – Checklist of Open Science Best Practices

Scoring

0 = Not completed
1 = Partially completed (does not complete all aspects of a given criterion)
2 = Fully completed
NA = This criterion is not relevant

Criteria

Pre-Registration

1. Pre-Registration: Was the study pre-registered? Are protocol deviations/changes from the pre-registration fully described (e.g. experimental procedures)? (1 point each question)

Data, Methods, and Materials Availability

Note: A proper availability statement is a dedicated section that tries to improve the reproducibility of a manuscript by stating what data/code/materials were used during the study and where said data/code/materials can be found/obtained. Availability statements almost always have a dedicated header.
Note: The "code availability statement" refers to newly generated custom computer code (or the software or mathematical algorithm or LCA models).

2. Material_Availability[1]: Is there a materials availability statement (materials to include in a material availability statement are: plants, microbe strains, cell lines, primary cell cultures, and primers, as well as all unique chemical and biological materials, including commercially and non-commercially obtained materials, newly created materials, and unique specimens)?
    a. Sequence_Availability: Is there an accession number in repository or a supplier name, catalog number, clone number, or RRID for DNA/RNA sequences? Is metadata (e.g. hosts, source, collection location, etc.) available for DNA/RNA sequences?
    b. Plant_Availability: If plants were used, are the species, strain, ecotype, cultivar and source (including location for collected wild specimens) provided where relevant? Is there a unique accession number (if available)?

3. Data_Availability: Is there a data availability statement?

[1] Points are given even if material availability is discussed in the text, rather than as a separate section.

4. Data_Link: Is there an accession number or DOI or URL (data)?

5. Data_License: If there are newly created datasets, are licensing details for said datasets provided?

6. Code_Availability: Is there a code availability statement (Often included in data availability statement)?

7. Code_Link: Is there an accession number or DOI or URL (code)?

8. Code_License: If there is newly created code, are licensing details for said code provided?

9. Experimental_Protocol: Were detailed step-by-step experimental protocols made available to allow for replication? Are experimental protocols sufficiently described to allow for replication?

Experimental study design

10. Sample_Size: Is there a description of sample size/replicate determination (e.g., a priori power analysis)?

11. Randomization: Do the authors state if randomization occurred and justify why or why not? (1 point each question)

12. Study_Authorization: If the study involves field sampling, do the authors state if relevant permits were obtained and provide details of authority approving study? If none were required, is there an explanation?

Data Analysis

13. Statistics: Do the authors describe their use of statistical tests and justify their choice of tests?

14. Assumptions: Are appropriate assumptions of the applied statistical tests discussed

Appendix 3

Figure S1 Plots displaying scores by item category. First is the average of each score. Colours indicate *overlooked* (white, < 20%), *in progress* (grey, 20-80%), and *addressed* (black, >80%) items. Then, scores over time are illustrated.

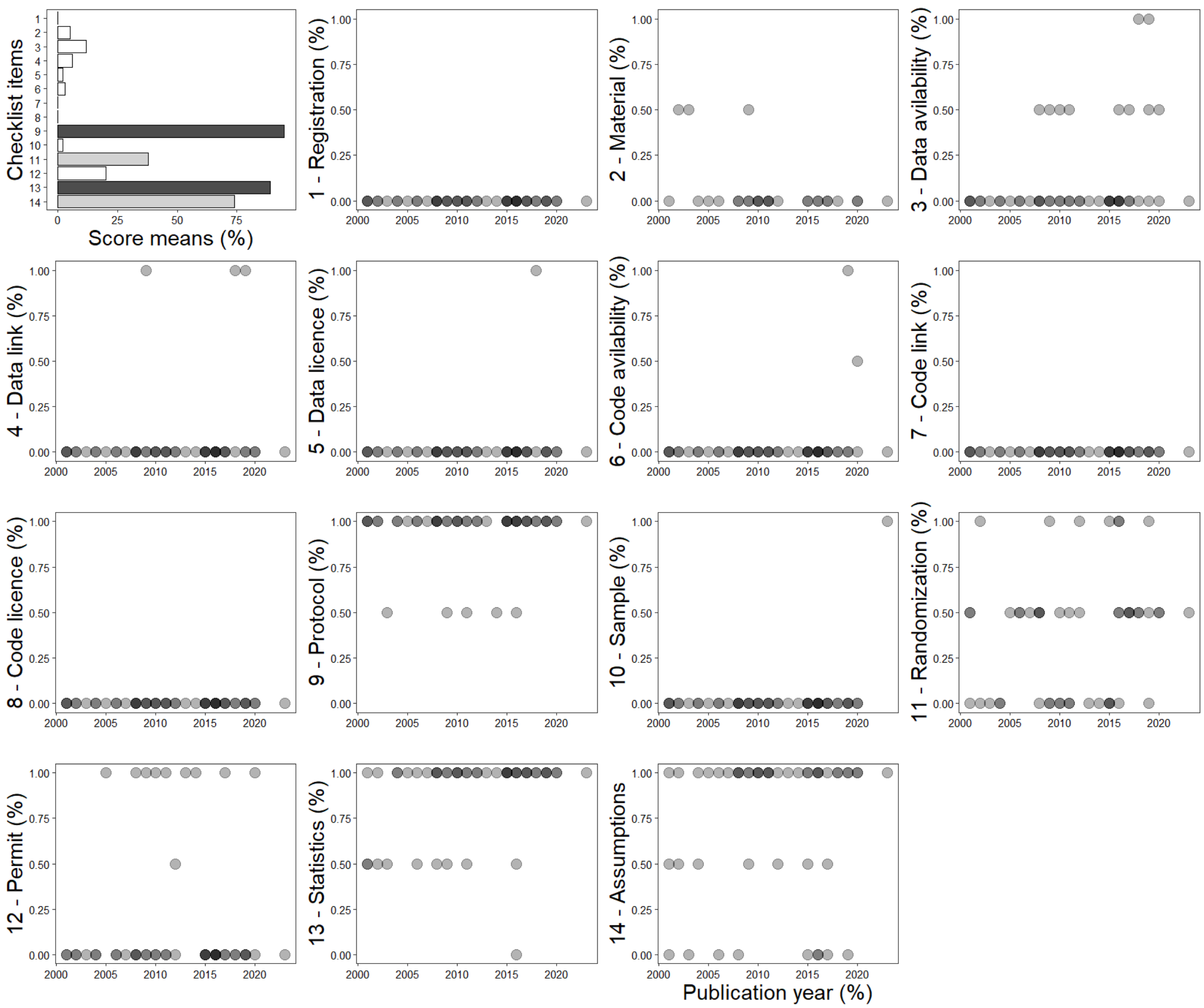


Figure S2 The number of studies on the impacts of Russian olive and Reed canarygrass in riparian ecosystems in British Columbia, Canada and their average reproducibility score by year since the year 2000 focusing only on 7 checklist items (data availability, code availability, sample size determination, randomization, detailing experimental protocol, detailing statistical methods, and checking statistical assumptions). On the y-axis are the number of studies (left) and the average reproducibility score (right), while on the x-axis is the year of publication. The number of studies is represented by grey bars, the average reproducibility score by year is represented by the black line.