Title: Questionable Research Practices in Ecology and Evolution

Authors: Hannah Fraser[1], Tim Parker[2,3], Shinichi Nakagawa4, Ashley Barnett[1], Fiona Fidler[1,5]

Affiliations:

[1]School of BioSciences, University of Melbourne, Parkville, VIC, Australia

[2]Biology Department, Whitman College, Walla Walla, WA, USA

[3]Department of Biological Sciences, Macquarie University, North Ryde, NSW 2109, Australia

[4]School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW, Australia

[4]School of Historical and Philosophical Studies, University of Melbourne, Parkville, VIC, Australia

Abstract

We surveyed 807 researchers (494 ecologists and 313 evolutionary biologists) about their use of Questionable Research Practices (QRPs), including cherry picking statistically significant results, *p* hacking, and hypothesising after the results are known (HARKing). We also asked them to estimate the proportion of their colleagues that use each of these QRPs. Several of the QRPs were prevalent within the ecology and evolution research community. Across the two groups, we found 64% of surveyed researchers reported they had at least once failed to report results because they were not statistically significant (cherry picking); 42% had collected more data after inspecting whether results were statistically significant (a form of *p* hacking) and 51% had reported an unexpected finding as though it had been hypothesised from the start (HARKing). Such practices have been directly implicated in the low rates of reproducible results uncovered by recent large scale replication studies in psychology and other disciplines. The rates of QRPs found in this study are comparable with the rates seen in psychology, indicating that the reproducibility problems discovered in psychology are also likely to be present in ecology and evolution.

Key Words

P-hacking, HARKing, cherry-picking, reproducibility, replicability, transparency, open science, ecology, evolutionary biology

Introduction

All forms of science communication, including traditional journal articles, involve transforming complicated, often messy data into a coherent narrative form. O'Boyle et al [1] likened the process to a Chrysalis effect, turning "ugly initial results into beautiful articles". Repeated failures to reproduce a large proportion of results in the published literature of other disciplines (e.g. [2,3]) has triggered reflection and meta-research about the ways in which this transformation process is susceptible to confusion and corruption.

Problems with publication bias and inflated type I errors have also been discussed [4–7]. Forstmeier et al [8] explain how, under the conditions of publication bias, practices like *p* hacking and underpowered research can inflate the number of false positive results in the literature. They offer a table of solutions for a range of problematic practices, all specifically relevant to research in ecology and evolution. The dearth of studies replicating previous ecology and evolution research, and associated difficulties, have also been highlighted [9,10]. The majority of advice concerns changes that individual researchers can make to improve the quality of their own research. However, some initiatives look to change things at a broader scale by calling for improvements in the reporting standards in ecology and evolution journals [11,12].

The widespread prevalence of Questionable Research Practices (QRPs) is now well documented in psychology [13–15]. However, this is the first attempt (to the best of our knowledge) to document the prevalence of such practices in ecology and evolution.

What are Questionable Research Practices (QRPs)?

QRPs refer to activities such as *p* hacking, cherry picking, and Hypothesizing After Results are Known (HARKing), all of which have been well documented in other fields including psychology and medicine. *Cherry picking* includes failing to report dependent or response variables or relationships that did not reach statistical significance or other threshold and/or failing to report conditions or treatments that did not reach statistical significance or other threshold. *P hacking* refers to a set of activities: checking the statistical significance of results before deciding whether to collect more data; stopping data collection early because results reached statistical significance; deciding whether to exclude data points (e.g., outliers) only after checking the impact on statistical significance and not reporting the impact of the data exclusion; adjusting statistical models, for instance by including or excluding covariates based on the resulting strength of the main effect of interest; and rounding of a *p* value to meet a statistical significance threshold (e.g., presenting 0.053 as $p < .05$). *HARKing* includes presenting ad hoc and/or unexpected findings as though they had been predicted all along [16]; and presenting exploratory work as though it was confirmatory hypothesis testing [17].

John et al [14] surveyed over 2000 psychological researchers in the US, asking them about the prevalence of several Questionable Research Practices (QRPs), this included asking researchers whether they had used any of these practices, six of which are listed in Table 2. Agnoli et al [13] repeated John et al's survey with a sample of Italian psychologists, and found strikingly similar results (also shown in Table 2). Failure to report outcome measures and stopping rules has also been documented by LeBel et al [18]. O'Boyle et al [1] found that in the process of translating PhD theses' results to published articles the proportion of results supporting statistical hypotheses doubled; a change accounted for by the cherry picking of significant results.

Publication bias and publish-or-perish research culture

Publication bias in this context refers to a bias towards publishing statistically significant, 'positive' results and not publishing statistically non-significant ('negative' or null results). The bias exists in many sciences [19], has been documented for decades in some disciplines (e.g., in psychology, see Sterling, 1959 [20]) and appears to be getting stronger across science, with a detectable increase in the proportion of statistically significant results over the last 25 years [21].

The intersection of increasing publication bias and a growing publish-or-perish culture in science may well impact the frequency with which researchers employ QRPs [13,22]. In a publish-or-perish research culture, studies that were once relegated to a file drawer upon failing to reach statistical significance may now be more likely to be cherry picked, *p* hacked and HARKed back into the literature. In a simulation study, Smaldino & McElreath [23] demonstrate how selection for higher output can speed up the dissemination of poor methods within a research community.

Simmons et al [24] used simulated experimental data to demonstrate how QRPs such as reporting only the subset of dependent/response/outcome variables or experimental conditions that reached statistical significance can inflate the false positive error rate of the research literature. They warned of 'researcher degrees of freedom' in experimental reports, including failing to report the sampling stopping rule. This has been further demonstrated in an ecology and evolution context by Forstmeier et al [8]. For this reason, QRPs have been implicated as a contributing factor to the well-publicised reproducibility crisis in psychology and other disciplines [2,22,25].

Aims

Publication bias in a publish-or-perish research culture incentivises researchers to engage in QRPs, which inflate the false positive rate leading to a less reproducible research literature. In this sense, QRP rates might be indicators of future reproducibility problems. Arguments about the difficulties in directly evaluating the reproducibility of the ecology and evolution literature have been made elsewhere (e.g., Schnitzer & Carson [26] but see Nakagawa & Parker [9]). However, the link between QRPs and irreproducibility is rooted in fundamental statistical theory [27] and so even in the absence of direct replication measures, a high prevalence of QRPs should alone raise sufficient concern to trigger editorial and institutional action.

The specific aims of our research were to:

1. Survey ecology and evolution researchers' own self-reported rate of QRP frequency
2. Survey ecology and evolution researchers' estimated rate of QRP use in their field
3. Compare these rates to those found in other disciplines, particularly psychology, where serious reproducibility problems have been established
4. Explore, through researchers' open-ended comments on each QRP in the survey, attitudes, (mis)understandings, pressures and contexts contributing to QRP use in the discipline

Methods

Survey participants

We collected the email addresses of corresponding authors from 11 'ecology' and 9 'evolutionary biology' journals (see Table 1). Journals were chosen from the highest ranking (assessed by 5-year impact factor) journals within the categories defined by the ISI 2013 Journal Citation Reports [28]. From the highest impact journals, we selected those that publish a broad range of work and excluded those limited to narrower sub-fields. We manually collected 3000 email addresses from ecology journal issues between January 2014 and May 2016 and 3000 from evolutionary biology journal issues between January 2015 and March 2017. After deleting duplicate email addresses so that individual researchers did not receive our survey more than once, we finally emailed a total 5386 researchers with a link to our online survey which returned 807 responses (response rate = 15%).

Table 1: Journals used to identify researchers working in ecology and evolution

Ecology Journals	**Evolution Journals**
Trends in Ecology and Evolution	Evolutionary Application
Ecology Letters	Evolution
Annual Review of Ecology and Evolution	BMC Evolutionary Biology
Frontiers in Ecology and the Environment	Evodevo
Global Change Biology	Am Naturalist
Ecological Monographs	Journal of Evolutionary Biology
Methods in Ecology and Evolution	Evolutionary Biology
Journal of Ecology	Evolutionary Ecology
Global Ecology and Biogeography	Behavioural Ecology
ISME	
Journal of Applied Ecology	

Of the 807 responses, 71% (n=573) were identified through our ‘ecology’ journal sample and 37% (n=299) from our ‘evolution’ journal sample. This imbalance is a product of the number of journals in each sample and the order in which email addresses were collected and duplicated. We first targeted ecology journals, and then after beginning that process, decided to add a second group of evolution journals. Therefore, classifying responses as being from ecology or evolution researchers purely based on the journal classification in Table 1 is problematic. The distinction between what constitutes ecology vs evolutionary research is fuzzy, but we were able to delineate between the two disciplines according to researchers' self-identified discipline. 411 researchers specified their discipline and we re-classified their responses into ecology or evolutionary research categories as follows. First, we classified responses associated with sub-disciplines including the following terms as being made by evolution researchers: 'evolut*', 'behav*', 'reproductive', or 'sexual'. From the remaining set of descriptions (i.e. those that did not mention any of the above terms), we classified all responses associated including the following terms as being made by ecology researchers: 'plant', '*population', 'marine biology', 'biodiversity', 'community', 'environment*', 'conservation', 'ecology', 'botany', 'mycology', or 'zoology'. Researchers who did not use any of these terms and those who did not complete the self-identified sub-discipline question (n=396) were left in their original journal discipline category as outlined in Table 1. At the end of this reclassification process, the sample (n=807) consisted of 61% (n=494) ecology researchers and 39% (n=313) evolution researchers.

Only 69% (558-560/807) of our sample completed the demographic questions at the end of our survey. Of the 560 who completed the gender question, 69% identified as male, 29% as female, 0.2% identified as non-binary and 1% preferred not to say. Of the 558 who completed the career status question, 6% identified as graduate students, 33% as post-doctoral researchers, 24% as midcareer researchers/academics and 37% as senior researchers/academics. The 559 who completed the age question were divided between age categories as follows: under 30 (11.5%), 30-39 (46.7%), 40-49 (25.9%), 50-59 (9.8%), 60-69 (4.8%), and over 70 (1.3%).

Survey instrument

Our research practices survey was administered via Qualtrics (Provo, UT, USA) and sent between Nov 2016 and July 2017. The survey (Supplementary Material S1) included questions about

Questionable Research Practices (QRPs), some of which were modified from those used in John et al [14] and Agnoli et al [13] to make them more relevant to ecology and evolutionary research.

1. Not reporting studies or variables that failed to reach statistical significance (e.g. p ≤0.05) or some other desired statistical threshold.
2. Not reporting covariates that failed to reach statistical significance (e.g. p ≤0.05) or some other desired statistical threshold.
3. Reporting an unexpected finding or a result from exploratory analysis as having been predicted from the start.
4. Reporting a set of statistical models as the complete tested set when other candidate models were also tested.
5. Rounding-off a p value or other quantity to meet a pre-specified threshold (e.g., reporting p = 0.054 as p = 0.05 or p = 0.013 as p = 0.01).
6. Deciding to exclude data points after first checking the impact on statistical significance (e.g. $p \leq 0.05$) or some other desired statistical threshold.
7. Collecting more data for a study after first inspecting whether the results are statistically significant (e.g. $p \leq 0.05$).
8. Changing to another type of statistical analysis after the analysis initially chosen failed to reach statistical significance (e.g. $p \leq 0.05$) or some other desired statistical threshold.
9. Not disclosing known problems in the method and analysis, or problems with the data quality, that potentially impact conclusions.
10. Filling in missing data points without identifying those data as simulated.

For each of these 10 practices, researchers were asked to:

(i) estimate the percentage of ecology (evolution) researchers who they believe have engaged in this practice on at least one occasion (0-100%)

(ii) specify how often they had themselves engaged in the practice (never, once, occasionally, frequently, almost always)

(iii) specify how often they believe the practice *should* be used (never, rarely, often, almost always)

At the end of each QRP, researchers had the opportunity to make additional comments under the open-ended question: 'why do you think this practice should or shouldn't be used?'.

At the end of the set of 10 QRP questions, researchers who asked "have you ever had doubts about the scientific integrity of researchers in ecology (evolution)?", and asked to specify the frequency of such doubts, if any, for different sub-groups (see Table 3). Finally, the survey included demographic questions about participants' career stage, gender, age and sub-discipline, discussed above.

Data analysis

Analyses were preregistered [29] and performed in R version 3.3.3 [30]. For each of the 10 QRPs we plotted the proportion (with 95% Confidence Intervals, CIs) of researchers in each discipline who stated that they had used the practice 'never', 'once', 'occasionally', 'frequently', and 'almost always' in response to question (ii) above using *ggplot2* [31] (Figure 2). For the QRPs also covered in the John et al [14] and Agnoli et al [13] surveys, we directly compared proportions of researchers who had engaged in each QRP at least once (Table 2), as this is the primary frequency measure reported in those surveys. In Figure 1, we plotted the proportion of researchers reporting that they had used the practice at least once for each QRP against the researchers' estimates of prevalence in the field, i.e., researchers' responses to question (i) above. We examined correlations between how frequently each participant had engaged in a practice and how acceptable they found the practice, and their age and career stage using Kendall's Tau correlation. All 95% CIs are Wilson Score Intervals except for those on Kendall's Tau, which are bootstrapped based on 1000 bootstrapped samples using *NSM3* [32].

Results

Overall, researchers in ecology and evolution reported high levels of Questionable Research Practices (Table 2, Figure 1). However, the frequency with which researchers reported using these regularly was much lower (Figure 2) and qualitative analyses reveals use of these practices in ways that may be less questionable (Supplementary Material S2).

Comparing Ecology, Evolution and Psychology Researchers

The responses for ecology and evolution researchers were broadly similar to those from the samples of psychologists studied by John et al. [14] and Agnoli et al [13] (Table 2). One exception to this is that ecologists were less likely than psychologists or evolution researchers to report 'collecting more data after inspecting whether the results are statistically significant' (see also Figure 1). Both ecology and evolution researchers were also less likely to report excluding data points after checking significance than psychologists. On the other hand, both ecology and evolution researchers were more likely to acknowledge reporting an unexpected finding as expected than both samples of psychologists.

Table 2: Percentage (with 95% CIs) of researchers in psychology, ecology and evolution who reported having used each Questionable Research Practice at least once. n=555-626.

Questionable Research Practice	**Psychology Italy** Agnoli et al. [13]	**Psychology USA** John et al. [14]	**Ecology**	**Evolution**
Not reporting response (outcome) variables that failed to reach statistical significance#	47.9 *(41.3-54.6)*	63.4 *(59.1-67.7)*	64.1 *(59.1-68.9)*	63.7 *(57.2-69.7)*
Collecting more data after inspecting whether the results are statistically significant#	53.2 *(46.6-59.7)*	55.9 *(51.5-60.3)*	36.9 *(32.4-42.0)*	50.7 *(43.9-57.6)*
Rounding-off a *p* value or other quantity to meet a pre-specified threshold#	22.2 *(16.7-27.7)*	22.0 *(18.4-25.7)*	27.3 *(23.1-32.0)*	17.5 *(13.1-23.0)*

Deciding to exclude data points after first checking the impact on statistical significance	39.7 *(33.3-46.2)*	38.2 *(33.9-42.6)*	24.0 *(19.9-28.6)*	23.9 *(18.5-30.2)*
Reporting an unexpected finding as having been predicted from the start#	37.4 *(31.0-43.9)*	27.0 *(23.1-30.9)*	48.5 *(43.6-53.6)*	54.2 *(47.7-60.6)*
Filling in missing data points without identifying those data as simulated*	2.3 *(0.3-4.2)*	0.6 *(0.0-1.3)*	4.5 *(2.8-7.1)*	2.0 *(0.8-5.1)*

#note that these statements began with “in a paper,” in John et al. [14] and Agnoli et al [13].

*note that this was referred to as “falsifying data” in John et al. [14] and Agnoli et al [13] which may have influenced the difference in response rates.

Self-reported QRP use compared to expected QRP use amongst colleagues

Broadly, researchers’ self-reported QRP use were closely related to their estimates of prevalence of QRPs in the scientific community (Figure 1). However, in the case of QRPs 2, 5, 6, 9 and 10, expected prevalence was substantially higher than individual self-reported use, suggesting that these may be considered the least socially acceptable QRPs in the set.

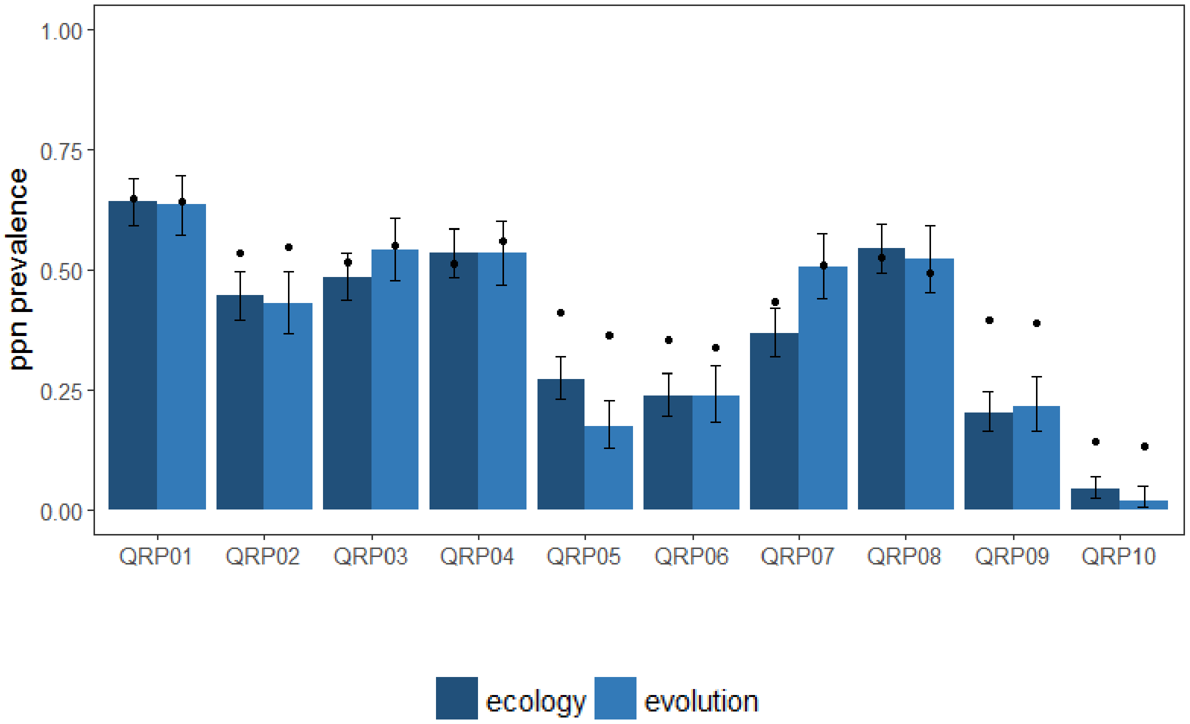

Figure 1: The prevalence of Questionable Research Practices in ecology and evolution. Columns represent the proportion of researchers who reported having used a practice at least once. The dots show researchers' mean estimates of suspected use by colleagues in their field. Dots that are much higher than bars may suggest that the QRP is considered particularly socially unacceptable. Error bars are 95% confidence intervals.

Frequency of individual researchers' QRP use

It was extremely rare for researchers to report high frequencies (frequently, almost always) use of QRPs. Most reported usage was at low frequency (once, occasionally), with many researchers reporting they had never engaged in these practices (Figure 2).

Age and career stage were not strong predictors of how frequently researchers used Questionable Research Practices (Kendall's Tau of 0.05, 95% CI = 0.001-0.069 and 0.04, 95% CI 0.011-0.058 respectively) but there was a considerable correlation between how often participants thought the practice should be used and how often they used it (Kendall's Tau = 0.6, 95% CI = 0.61-0.65). Those who used practices frequently or almost always were much more likely to indicate that they should be used often.

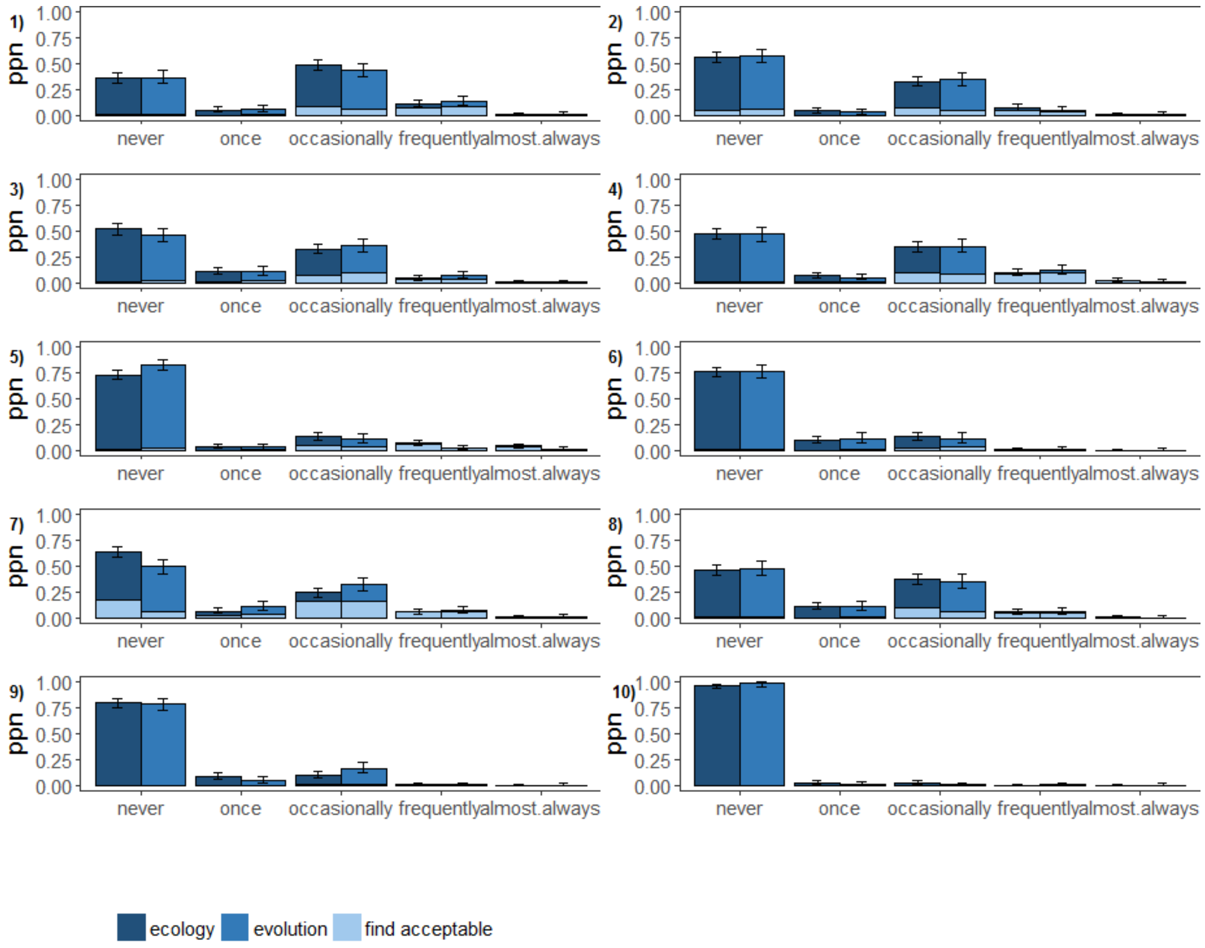

Figure 2: Proportion of researchers in ecology and evolution reporting frequency of use (or not) of 10 Questionable Research Practices. Shading indicates the proportion of each use category that identified the practice as acceptable. Error bars are 95% confidence intervals.

Perceptions of scientific integrity

Researchers in ecology and evolution expressed considerable doubts about their community's scientific integrity (Table 3), mostly in relation to QRPs rather than scientific misconduct. Concern about the integrity of researchers at their own institution was roughly equal to concern about the integrity of other institutions, nor was there any notable difference in concern about graduate students', senior colleagues' or collaborators' behaviour. Our participants expressed least concern about their own integrity, but 44.6% still indicated doubts over their own use of QRPs.

Table 3: Proportion (with 95% CI) of researchers in ecology and evolution (combined) who reported having doubts about scientific integrity.

	Questionable Research Practices			**Scientific Misconduct**		
	Never	Once or Twice	Often	Never	Once or Twice	Often
Researchers from other institutions	8.9 *(6.8-11.6)*	56.6 *(52.3-60.7)*	34.5 *(30.6-38.6)*	39.0 *(34.9-43.4)*	55.5 *(51.0-59.8)*	5.5 *(3.8-7.8)*
Research at your institution	27.9 *(24.2-31.8)*	52.2 *(47.9-56.4)*	20.0 *(16.8-23.6)*	69.2 *(65.0-73.1)*	29.1 *(25.3-33.3)*	1.6 *(0.8-3.1)*
Graduate student research at your institution	31.0 *(27.2-35.1)*	48.6 *(44.3-52.8)*	20.4 *(17.1-24.0)*	72.5 *(68.4-76.3)*	25.6 *(21.9-29.7)*	1.8 *(1.0-3.5)*
Senior colleagues or collaborators	31.5 *(27.6-35.5)*	50.8 *(46.6-55.1)*	17.7 *(14.7-21.2)*	73.3 *(69.2-77.0)*	24.7 *(21.1-28.7)*	2.0 *(1.1-3.7)*
Your own research	52.2 *(48.0-56.4)*	44.6 *(40.5-48.8)*	3.2 *(2.0-5.0)*	97.9 *(96.2-98.8)*	2.0 *(1.1-3.7)*	0.0 *(0.0-0.8)*

*note that not all researchers answered each component of the table above so the total sample size for each of the cells differs slightly, ranging from 488 to 539 samples per cell

Qualitative data analysis

At the end of each QRP question, researchers had the opportunity to make additional comments on the practice. Overall, we were surprised by the proportion of researchers who made comments, for some QRPs half the researchers left comments, and often substantial ones. Here we have summarised the ecology and evolution groups comments together, having not detected any major differences between the groups in a qualitative assessment. We interpret the volume of additional comments positively, as evidence of a research community highly engaged with issues of research practice and scientific integrity.

The most frequently offered justifications for engaging in QRPs were: publication bias; pressure to publish; and the desire to present a neat, coherent narrative (Table 4). A full description of the qualitative analysis is available in Supplementary Material S2.

Table 4: Frequently offered arguments against and justifications for various Questionable Research Practices, summarising qualitative comments provided by ecology and evolution researchers.

Description	**Complaints about this practice**	**Why this practice is tempting**	**Conditions identified by researchers as justifying this practice**
Cherry-picking			
“Sometimes lots of data are collected and tested. Often non-significant variables are thrown out if they're not integral to the story. I think this is okay.” “Not reporting non-significant results biases the big picture (e.g. meta-analysis), mislead other researchers into thinking that a question is unexplored...This publication bias however, is obviously a result of the publication system.” “If multiple model sets are tested they should all be presented, otherwise we risk presenting misleading results by trying a bunch of stuff until one turns out to be significant”			
QRP 1: Not reporting studies or variables that failed to reach statistical significance *(n=408)* QRP 2: Not reporting covariates that failed to reach statistical significance *(n=350)* QRP 4: Reporting a subset of statistical models as the complete tested *(n=386)*	- increases false positive rate - leads to redundant investigation - impedes interpretation - skews meta-analyses - there is important information in non-significant results - it is unethical	- hard to publish non-significant results - journal word limits - difficult to create a compelling story with non-significant results - complete report makes boring methods and result sections - running extra models improves understanding of the system	- original method was flawed - analyses were exploratory - results from multiple analyses were the same - they were excluded during formal model selection - variables correlated - data did not match model assumptions
HARKing			
“well, this is a difficult one - in the statistical sense, this should not happen, but in current times scientists are forced to market their work as best as possible and this is one way to make it more publishable.” “Encourages, just-so stories, we can always come up with a suitable explanation and prediction. The key point here is to avoid doing so without noticing.” “I believe it should not be used but editors and reviewers often demand that exploratory results are framed as a priori hypotheses”			

QRP 3: Reporting an unexpected finding as having been predicted *(n=371)*	- it is unethical - unexpected results need to be confirmed - increases false positive rate	- makes article sexier - reviewers ask for this - pressure to publish - not always clear exactly what was hypothesised	- new hypotheses arise from better understanding of the system - researchers can explain the result - researchers should have hypothesised something else
P-hacking			
“Attempts to conform to strict cut-off significance thresholds demonstrate an adherence to conventional practice over understanding of probability (e.g. the difference between *p* = 0.013 and 0.010 is and should be viewed as trivial).” “This practice leads to statistical significance overshadowing effect sizes and biological significance.” “Again, one needs to be ethical. Science is about testing hypotheses with experiment, not about publishing p<0.05 in the sexiest journal possible. A priori and post priori hypotheses are both acceptable, but they need to be labelled as such.”			
QRP 5: Rounding- off a *p* value or other quantity to meet a pre-specified threshold *(n=409)* QRP 6: Deciding to exclude data points after first checking the impact on statistical significance *(n=334)* QRP 7: Collecting more data for a study after first inspecting whether the results are statistically significant *(n=364)* QRP 8: Changing to another type of statistical analysis after the analysis initially chosen failed to reach statistical significance *(n=346)*	- it is unethical - increases false positive rate	- the 0.05 threshold is arbitrary anyway - hindsight bias - pressure to publish - reviewers may ask for more data or different analyses	- all results are presented - process is reported - decision not based on significance - additional data collection already planned - original analysis was poorly chosen - data didn’t meet assumptions of original analysis - new analysis better reflects ecological context - tests are conducted to test robustness of result

Discussion

Our results indicate that QRPs are broadly as common in ecology and evolution research as they are in psychology. Of the 807 researchers in our sample, 64% reported cherry picking statistically significant results in at least one publication; 42% reported *p* hacking by collecting more data after first checking the statistical significance of results and 51% acknowledged reporting an unexpected finding as though it had been hypothesised from the start (HARKing). That these are similar to QRP rates in psychology is hardly surprising given that publication bias and the same publish-or-perish culture persists across disciplines. However, it is important to establish the QRP rate in ecology and evolution, as it provides important evidence on which to base initiatives to improve research practices in these disciplines.

Disciplinary differences

Our results are most marked by how similar rates of QRPs were across disciplines, but a couple of differences are worth noting. Ecology researchers were less likely to report 'collecting more data after inspecting whether the results are statistically significant' (QRP7) than evolution researchers or psychologists. We suspect this reflects a difference in the constraints of field versus laboratory research, rather than differences in the integrity of the researchers. It is often not physically possible collect more data after the fact in ecology (field sites may be distant, available sites and budgets may be exhausted). This interpretation seems supported by evidence that many ecologists who stated that they had 'never' engaged in this practice indicated that they found it acceptable.

The first nine of the QRPs we asked about were certainly controversial practices, generating mixed responses. The tenth is qualitatively different; it essentially asks about fraud. The social unacceptability of this practice is well recognised, and we might therefore expect under reporting even in an anonymous survey. The comments volunteered by participants largely reflected this: "Is that the science of 'alternative facts'?" and "It is serious scientific misconduct to report results that were not observed". The proportion of researchers admitting to this was relatively high in ecology

(4.5%) compared to evolution (2.0%), US psychology (2.3%) and Italian psychology (0.6%). However, it's important to note that our wording of this question was quite different to that in the John et al and Agnoli et al surveys. They asked directly about 'falsifying data' whereas we asked a softer, less direct question about 'filling in missing data points without identifying those data as simulated'. Fiedler et al (2015) found that modified question wording changed QRP reporting rates and we suspect our change to the wording has resulted in an elevated reporting rate. We will not speculate further about ecology researchers reporting a higher rate of this than evolution researchers because the numbers of researchers admitting to this action are very small in both groups and the 95%CIs on these proportions overlap considerably.

Novel insights into the usage of QRPs

Our results contribute to the broader understanding of researchers' practices in two important ways. First, our results on reported frequency provide new insight into the regularity with which researchers engage in these practices; previous surveys in psychology did not elicit this information and asked only if the practice had been used 'at least once'. Information about frequency of use allows us to better estimate the disruption these practices may have had on the published literature. We show that while reports of having engaged in QRPs at least once are alarmingly high, virtually no researchers acknowledge using any of the QRPs more than 'occasionally'. Secondly, our qualitative results offer new understanding of the perceived acceptability of these practices, and common justifications of their use.

Our qualitative analysis highlighted the perception of a detrimental influence of the current publish-or-perish culture and rigid format currently required in many ecology and evolution journals. Researchers' comments revealed the pressure they feel to present a short, cohesive story with statistically significant results that confirm a priori hypotheses, rather than a full (and likely messy) account of the research as it was conceptualised and conducted.

Researchers' qualitative comments also drew attention to grey areas, where the distinction between QRPs and acceptable practice was less clear. For example, in many ecology and evolution articles no hypotheses are overtly stated but the way the background material is described in the introduction can imply that the result was expected; does this constitute HARKing? Similarly, a number of participants answering QRP 6 stated that, although they had technically changed models after investigating statistical significance, their decision to change models was based on finding an error in the original model or discovering that the data did not match the model assumptions. These participants are recorded as using this QRP but whether or not it was 'questionable' in their case is unclear.

Social acceptability of QRPs

Discrepancies between individual researchers' self-identified QRP use and their estimates of others' use suggest that certain practices are less socially acceptable. When average estimates of others' use are much higher than average self-report of the practice, it suggests that the practice is particularly socially undesirable and that self-report measures may underestimate prevalence. In our results, the greatest discrepancies were observed for QRPs 2, 5, 6, 9, and 10 (see Figure 2), suggesting that self-reported prevalence may underestimate the true prevalence of these practices. In contrast, where there is little discrepancy between these two measures we can infer that the practice has gained a degree of social acceptability, for example QRPs 1, 4, 7, 8. These may be harder practices to shift, as researchers may not recognise them as problematic.

Solutions

Our results indicate that there is substantial room to improve research practices in ecology and evolution. However, none of these problems are insurmountable. In fact, the correlation we found between acceptability and prevalence of QRPs and the justifications people provided in text (Supplementary material S2) suggest that the prevalence of these practices could be reduced by educating researchers about their ramifications. These practices are driven by a publish-or-perish research culture that puts emphasis on producing sexy, novel stories over solid science. The open science movement has given rise to a series of solutions that help reduce the temptation for and prevalence of these QRPs [4,8,12,33].

Among the most promising of these solutions is preregistration. A thorough preregistration specifies a researcher's hypotheses, how they will decide on their sample size, data exclusion criteria, and the analyses they will conduct, among other things. This helps researchers think their research through thoroughly, improving its rigor, as well as protecting against HARKing, cherry-picking and p-hacking [34,35]. Of course, preregistration does not change the fact that journals prefer to publish significant results, so this is not a complete solution on its own. Efforts like registered reports, whereby the review process takes place before data collection and analysis, oblige journals to publish good research regardless of the results and may help to avert this issue [35,36].

Another promising solution is increasing the transparency of research; providing full accounts of methods, code for analyses and data so that researchers are accountable for their choices during peer review and once a paper is published. Many researchers are now making their data and code openly available, although there is still some concern about whether this is beneficial to individual researchers [37]. To incentivise this openness, a minority of ecology and evolution journals have put in place rigorous checklists for authors to encourage more transparent reporting (e.g. Conservation Biology, Nature).

Conclusion

The use of Questionable Research Practices in ecology and evolution research is high enough to be of concern. The rates of QRPs found in our sample of 807 ecologists and evolutionary biologists are similar to those that have been found in psychology, where the reproducibility rates of published research have been systematically studied and found to be low (36-47% depending on the measure [2]). Researchers in our survey offered justifications for their practices including: publication bias;

pressure to publish; and the desire to present a neat, coherent narrative. We recommend that all journals in ecology and evolution adopt editing and reviewing checklists to ensure more complete and transparent reporting, encourage preregistration and registered reports article formats to minimise HARKing, and encourage open code and data whenever possible.

Acknowledgements

Fiona Fidler is supported by an Australian Research Council Future Fellowship (FT150100297). T. Parker was supported by a sabbatical provided by Whitman College and was hosted by S. Griffith at Macquarie University. We would like to thank Felix Singleton Thorn and Franca Agnoli for feedback on the manuscript and Aurora Marquette for her assistance in collecting contact author addresses.

References

1. O'Boyle EH, Banks GC, Gonzalez-Mulé E. The chrysalis effect: how ugly initial results metamorphosize into beautiful articles. J Manage. 2017;43: 376–399. doi:10.1177/0149206314527133
2. Open Science Collaboration. Estimating the reproducibility of psychological science. Science (80-). 2015;349. doi:10.1126/science.aac4716
3. Freedman LP, Cockburn IM, Simcoe TS. The economics of reproducibility in preclinical research. PLoS Biol. 2015;13: 1–9. doi:10.1371/journal.pbio.1002165
4. Parker TH, Nakagawa S. Mitigating the epidemic of type I error: ecology and evolution can learn from other disciplines. Front Ecol Evol. 2014;2: 1–3. doi:10.3389/fevo.2014.00076
5. Forstmeier W, Schielzeth H. Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. Behav Ecol Sociobiol. 2011;65: 47–55. doi:10.1007/s00265-010-1038-5
6. Cassey P, Ewen JG, Blackburn TM, Moller AP. A survey of publication bias within evolutionary ecology. Proc R Soc B Biol Sci. 2004;271: S451–S454. doi:10.1098/rsbl.2004.0218

7. Csada RD, Cres S, James CSTPC, Branch W. The “file drawer problem” of non-significant results: does it apply to biological research? OIKOS. 1996;76: 591–593. doi:10.2307/3546355

8. Forstmeier W, Wagenmakers EJ, Parker TH. Detecting and avoiding likely false-positive findings – a practical guide. Biol Rev. 2017;92: 1941–1968. doi:10.1111/brv.12315

9. Nakagawa S, Parker TH. Replicating research in ecology and evolution: Feasibility, incentives, and the cost-benefit conundrum. BMC Biol. BMC Biology; 2015;13: 1–6. doi:10.1186/s12915-015-0196-3

10. Kelly CD. Replicating empirical research in behavioural ecology: how and why it should be done but rarely is. Q Rev Biol. 2006;80: 221–236. doi:10.1086/516403

11. Parker TH, Griffith SC, Bronstein JL, Fidler F, Foster S, Fraser H, et al. Empowering peer reviewers with a checklist to improve transparency. Nat Ecol Evol.

12. Ihle M, Winney IS, Krystalli A, Croucher M. Striving for transparent and credible research: Practical guidelines for behavioral ecologists. Behav Ecol. 2017;28: 348–354. doi:10.1093/beheco/arx003

13. Agnoli F, Wicherts JM, Veldkamp CLS, Albiero P, Cubelli R. Questionable research practices among Italian research psychologists. PLoS One. 2017;12: 1–17. doi:10.1371/journal.pone.0172792

14. John LK, Loewenstein G, Prelec D. Measuring the prevalence of Questionable Research Practices with incentives for truth telling. Psychol Sci. 2012;23: 524–532. doi:10.1177/0956797611430953

15. Fiedler K, Schwarz N. Questionable Research Practices revisited. Soc Psychol Personal Sci. 2016;7: 45–52. doi:10.1177/1948550615612150

16. Kerr N. HARKing: hypnothesizing after the results are known. Personal Soc Psychol Rev. 1998;2: 196–217. doi:10.1207/s15327957pspr0203_4

17. Wagenmakers EJ, Wetzels R, Borsboom D, van der Maas HLJ, Kievit RA. An agenda for purely confirmatory research. Perspect Psychol Sci. 2012;7: 632–638.

doi:10.1177/1745691612463078

18. LeBel EP, Borsboom D, Giner-Sorolla R, Hasselman F, Peters KR, Ratliff KA, et al. PsychDisclosure.org: Grassroots support for reforming reporting standards in psychology. Perspect Psychol Sci. 2013;8: 424–432. doi:10.1177/1745691613491437

19. Fanelli D. “Positive” results increase down the hierarchy of the sciences. PLoS One. 2010;5. doi:10.1371/journal.pone.0010068

20. Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of significance - or visa versa. J Am Stat Assoc. 1959;54: 30–34. doi:10.1080/01621459.1959.10501497

21. Fanelli D. Negative results are disappearing from most disciplines and countries. Scientometrics. 2012;90: 891–904. doi:10.1007/s11192-011-0494-7

22. Fidler F, En Chee Y, Wintle BC, Burgman MA, McCarthy MA, Gordon A. Metaresearch for evaluating reproducibility in ecology and evolution. Bioscience. 2017;67: 282–289. doi:10.1093/biosci/biw159

23. Smaldino PE, McElreath R. The natural selection of bad science. 2016; doi:10.1098/rsos.160384

24. Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci. 2011;22: 1359–1366. doi:10.1177/0956797611417632

25. Pashler H, Wagenmakers EJ. Editors’ introduction to the special section on replicability in psychological science: a crisis of confidence? Perspect Psychol Sci. 2012;7: 528–530. doi:10.1177/1745691612465253

26. Schnitzer SA, Carson WP. Would ecology fail the repeatability test? Bioscience. 2016;66: 98–99. doi:10.1093/biosci/biv176

27. Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005;2: 0696–0701. doi:10.1371/journal.pmed.0020124

28. Reuters T. Thomson Reuters Research Analytics Unveils 2013 Edition of Its Journal Citation Reports. PR Newswire. Ipswich, MA; 2013.

29. Fraser H, Parker TH, Nakagawa S, Barnett A, Fidler F. Preregistration: Questionable Research Practices in ecology and evolution [Internet]. Melbourne: Open Science Framework; 2017. doi:10.17605/OSF.IO/RG2NQ

30. R Core Development Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.

31. Wickham H. ggplot2: Elegant graphics for data analysis. New York: Springer-Verlag; 2009.

32. Schneider G, Chicken E, Becvarik R. NSM3: Functions and datasets to accompany Hollander, Wolfe, and Chicken [Internet]. Nonparametric Statistical Methods; 2017. Available: https://cran.r-project.org/web/packages/NSM3/index.html

33. Parker TH, Forstmeier W, Koricheva J, Fidler F, Hadfield JD, Chee YE, et al. Transparency in ecology and evolution: real problems, real solutions. Trends Ecol Evol. 2016;31: 711–719. doi:10.1016/j.tree.2016.07.002

34. Mellor D. Preregistration and increased transparency will benefit science. 2017;preprint. doi:10.17605/OSF.IO/XSFAM

35. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. Proc Natl Acad Sci. 2018; 1–7. doi:10.1073/pnas.1708274114

36. Nosek BA, Lakens D. Registered reports: A method to increase the credibility of published results. Soc Psychol (Gott). 2014;45: 137–141. doi:10.1027/1864-9335/a000192

37. Shaw RG, Moore AJ, Noor M, Ritchie MG. Transparency and reproducibility in evolutionary research. Evolution. 2016;70: 1433–1434. doi:10.1111/evo.12977