

# Questionable Research Practices and Open Science in Quantitative Criminology

Jason M. Chin

School of Law, University of Sydney  
Sydney Institute of Criminology, University of Sydney

Justin T. Pickett

State University of New York at Albany

Simine Vazire

School of Psychology, University of Melbourne

Alex O. Holcombe

School of Psychology, University of Sydney

## Abstract

*Objectives.* Questionable research practices (QRPs) lead to incorrect research results and contribute to irreproducibility in science. Researchers and institutions have proposed open science practices (OSPs) to improve the detectability of QRPs and the credibility of science. We examine the prevalence of QRPs and OSPs in criminology, and researchers' opinions of those practices.

*Methods.* We administered an anonymous survey to authors of articles published in criminology journals. Respondents self-reported their own use of 10 QRPs and 5 OSPs. They also estimated the prevalence of use by others, and reported their attitudes toward the practices.

*Results.* QRPs and OSPs are both common in quantitative criminology, about as common as they are in other fields. Criminologists who responded to our survey support using QRPs in some circumstances, but are even more supportive of using OSPs. We did not detect a significant relationship between methodological training and either QRP or OSP use. Support for QRPs is negatively and significantly associated with support for OSPs. Perceived prevalence estimates for some practices resembled a uniform distribution, suggesting criminologists have little knowledge of the proportion of researchers that engage in certain questionable practices.

*Conclusions.* Most quantitative criminologists in our sample have used QRPs, and many have used multiple QRPs. Moreover, there was substantial support for QRPs, raising questions about the validity and reproducibility of published criminological research. We found promising levels of OSP use, albeit at levels lagging what researchers endorse. The findings thus suggest that additional reforms are needed to decrease QRP use and increase the use of OSPs.

**Keywords:** Questionable research practices, meta-research, open science, reproducibility

**Acknowledgements:** We thank Jonathan Brauer for helpful comments on an earlier draft.

## Introduction

It is not hard for scientists, including criminologists, to get whatever research findings they want—evidence that a criminal justice policy or program is effective, support for a favored theory or new hypothesis, statistical significance for a surprising interaction effect (Ritchie, 2020; Sweeten, 2020). Sufficient use of questionable research practices (QRPs) (Simmons, Nelson, & Simonsohn, 2011) will do the trick. QRPs represent the steroids of the scientific publishing game because they artificially boost researchers' performance (i.e., their ability to produce exciting, and therefore publishable, results; Bakker, van Dijk, & Wicherts, 2012). Tortured data, as the saying goes, "will tell you whatever you want to hear" (Mills, 1993, p. 1196). QRPs are torture tools for making data confess false positives; their use makes it easy for scientists to turn "ugly initial results ... into beautiful articles" (O'Boyle, Banks, & Gonzalez-Mulé, 2017, p. 376). They are common in every field where they have been assessed, including psychology (John, Loewenstein, & Prelec, 2012), political science (Franco, Malhotra, & Simonovits, 2015), management (O'Boyle et al., 2017), education (Makel et al., 2019), quantitative communication (Bakker et al., 2020), and ecology and evolutionary biology (Fraser et al., 2018).

Are QRPs common in criminology? Some signs suggest the answer is yes (Burt, 2020). Criminologists are more likely to find desired effects when using weaker research designs that give them more opportunities for undisclosed flexibility (Weisburd & Lum, 2001; Welsh et al., 2011). Indeed, this is a woefully consistent finding in criminological meta-analyses: quasi-experiments produce much larger effects and results that are more often statistically significant than RCTs (e.g., Braga, Papachristos, & Hureau, 2014; Braga, Weisburd, & Turchan, 2018). The same is true in psychology (Kvarven, Strømmland, & Johannesson, 2020), where QRP use has been well-documented (John et al., 2012; Agnoli et al., 2017; Rabelo et al., 2020). Additionally,

there is a sizable inverse relationship between sample size and effect size in criminology (Nelson, Wooditch, & Dario, 2015), which is a telltale sign that something is amiss (Gelman, Skardhamar, & Aaltonen, 2020; Levine, Asada, & Carpenter, 2009). As importantly, published experiments in criminology often differ from the plan described at the proposal stage, and the more they differ the larger the published effect size, suggesting criminologists deviate selectively from research protocols in a way that exaggerates results (Wooditch et al., 2020). Despite this suggestive evidence, however, there have been no surveys on QRP use in criminology, as there have been in other disciplines.

If QRPs are the steroids of scientific publishing (Bakker, van Dijk, & Wicherts, 2012), then Open Science Practices (OSPs) are the drug test (Ritchie, 2020; Simmons et al., 2011). OSPs, such as preregistering analysis plans and publicly posting data and replication code, make it possible for outside researchers not only to replicate findings, but also to evaluate the effects of specific analytical decisions, such as deviations from preregistered protocols. They may also deter QRPs, just as the prospect of a drug test may deter steroid use. A movement is underway in many fields to increase the use of OSPs, and even to institutionalize them at journals and funding organizations (Ritchie, 2020; Vazire, 2018). However, even less survey evidence exists about the prevalence of OSPs than QRPs across disciplines (Bakker et al., 2020; Makel et al., 2019). As with QRPs, there have been no surveys on OSP use in criminology.

To address this void, in 2020 we administered an anonymous survey on QRPs and OSPs to a sample of criminologists. We designed the survey to mirror those fielded in other disciplines (e.g., John et al., 2012; Fraser et al., 2017). The survey measured behavior and relevant attitudes. In what follows, we first outline the various QRPs that criminologists may use and review the evidence from other disciplines about their prevalence and effects. We then discuss the open

science movement and the recommended pro-transparency research practices that have emerged from it. Next, we describe our study and its results, which provide the first large scale estimates of the prevalence of QRPs and OSPs in criminology.

### **Questionable Research Practices (QRPs)**

QRPs include the practice known as *p*-hacking (Bishop, 2019), other inappropriate uses of researcher degrees of freedom (Simmons et al., 2011) or the exploitation of analytic flexibility (Beerdsen, 2020), and publication bias (Fanelli, 2012). All of these terms refer to a set of practices that distort the accuracy of research reports when not reported transparently, typically in a way that exaggerates effect sizes or produces statistically significant results. QRPs often involve hidden research decisions that are based on whether they yield statistically significant results, including decisions regarding when to stop collecting data, which analytic method to use and report, which variables to include in a model, how to code those variables, whether to exclude outliers, and whether to write up a study (John et al., 2012; see Table 1). Such practices produce biases because undisclosed flexibility (e.g., trying out several different covariates, outlier exclusion thresholds, or subgroup analyses) allows researchers to selectively under- or over-fit models and exploit noise in a way that goes uncorrected (e.g., through *p*-value corrections like Bonferroni) and unreported (Gelman & Loken, 2014; Simmons et al., 2011), inflating the false positive rate.

Some authors engage in QRPs without being aware of their pernicious effects. Editors and reviewers sometimes even encourage QRPs, such as testing for non-hypothesized interaction effects and presenting them as planned, or only including them in the manuscript if they are significant, or conducting and selectively reporting post-hoc subgroup analyses. For example, just last year (2020) one reviewer at a top criminology journal asked one of us to run additional

analyses disaggregated by race and to maybe report them, depending on the findings, writing: “And if there ARE significant differences across race in the raw opinions, I think that should be highlighted” (emphasis in original). Other times editors and reviewers may merely incentivize the use of QRPs by devaluing non-significant findings, for example by selectively applying critiques to non-significant results that might just as easily be applied to significant results. As an editor of a top criminology journal wrote in a 2015 decision letter regarding a manuscript that one of us co-authored: “null effects are often viewed as poor evidence because they may result from weak manipulations rather than an incorrect hypothesis.”

A growing body of evidence points to QRPs as a primary reason that many studies are proving difficult to replicate. Some researchers have admitted using QRPs, after their findings failed to replicate (Carney, 2016; Rohrer et al., 2018). For instance, the lead author of psychology’s controversial power pose studies later stated: “The self-report DV was *p*-hacked in that many different power questions were asked and those chosen were the ones that ‘worked’” (Carney, 2016). Similarly, a group of authors studying bilingualism and cognitive advantage admitted that they selectively reported findings that confirmed their hypothesis: “We ourselves are guilty ... the only experiment that we submitted for publication was the one showing an effect of bilingualism” (de Bruin, Treccani, & Sala, 2015, pp. 99-100).

Systematic research supports the association between QRPs and irreproducible research. For instance, Simmons and colleagues (2011) used simulations to estimate the effect of using four QRPs (selective reporting of two DVs, deciding whether or not to add 10 extra observations based on the statistical significance of the result, selectively adding or removing covariates, selectively including or dropping a condition) on the rate of false positive findings. They applied the QRPs to randomly generated data under the null hypothesis (i.e., when all statistically

significant results are false positives), and found that using just those four QRPs inflated the false positive rate from the nominal 5% (alpha) to over 60%. One study in the field of management found evidence that the use of QRPs after dissertation defenses, but before resultant articles were published, led to a 21-percentage-point increase in statistically significant results, which corresponded “to more than a doubling of the ratio of supported to unsupported hypotheses” (O’Boyle et al., 2017, p. 388). Cairo et al. (2020) found similar results in psychology.

How common is QRP use? There are several methods that can be used to investigate this question (Bakker et al., 2020). One is to compare the shape of the distribution of  $p$ -values among the significant results in the literature to what would be expected if the results were all true positives. Simonsohn and colleagues (2014) compared studies with no obvious *indicia* of QRPs to psychology studies reporting results only with covariates (indicating potential selective use of covariates, a QRP) and found the latter contained excess  $p$ -values close to .05 and too few  $p$ -values close to zero. This is consistent with the pattern we would expect if covariates are reported selectively when they help produce significant results (i.e., a QRP), though other explanations, such as fraud, are also possible. Brodeur, Cook, and Heyes (2020) reported similar findings in their analysis of articles published in 25 leading economic journals. They also found that studies using methods that give researchers more methodological discretion, such as instrumental variable analysis, showed more extensive evidence of p-hacking.

Another method for estimating the prevalence of QRPs is to use anonymous surveys to measure self-reported QRP use. Results from such surveys have been published for other disciplines (e.g., psychology, education, ecology) and for multiple regions (Italy, Brazil, US). John et al. (2012), for example, found that over 90% of US psychologists admitted to using at least one QRP. Multiple studies have asked about identical or very similar QRPs to each other

(see a comparison of these studies, <https://osf.io/wm7aq/>), making it possible to draw broad comparisons across fields. Table 1 lists the comparable results from seven prior studies (and we will compare our results to theirs in the forthcoming sections). A clear takeaway is that QRPs are common in many fields, with most QRPs being used, according to self-reports, by more than 20% of scientists, and some being used by the majority of scientists. To illustrate, 45% of US psychologists and 62% of education researchers self-report that they file drawer studies with null results (John et al., 2012; Makel et al., 2019). Similarly, most past surveys found that over 30% of scientists self-report HARKing (hypothesizing after the results are known) and over 20% self-report selectively rounding  $p$ -values. Depending on the discipline, 22-58% of scientists self-report data peeking with optional stopping (deciding whether to stop collecting data after looking at  $p$ -values), and 20-43% self-report that they exclude data selectively after looking at how the exclusion affects the results.

[Insert Table 1 about here]

### **The Credibility Revolution and Open Science Practices (OSPs)**

The threat posed by QRPs has been discussed most extensively in the field of psychology, arguably the eye of the storm of the “replication crisis.” In the wake of the “False Positive Psychology” paper (Simmons et al., 2011), Daryl Bem’s paper claiming to find evidence of Extra Sensory Perception (ESP; Bem, 2011), and several cases of fraud, the field of psychology entered a period of intense self-examination. The outcome has been a large and growing movement pushing for more attention to the quality and rigor of research, and faster progress on raising standards (Fidler & Wilcox, 2018). This loosely-defined movement has been called a “credibility revolution” (Vazire, 2018), “revolution 2.0” (Spellman, 2015), and “psychology’s renaissance” (Nelson, Simmons, & Simonsohn, 2018).

In response to that movement, several large-scale replication projects have been conducted in the social sciences (Camerer et al., 2016; Camerer et al., 2018; OSC, 2015; Ebersole et al., 2016; Klein et al., 2014; Klein et al., 2018). Overall, the rate of “successful” replication (defined as any statistically significant effect in the same direction as the original, which is a fairly liberal criterion in most of these projects as they often had high statistical power to detect effects even much smaller than the original effect) is around 45%. Given that over 90% of published studies in the social sciences claim to find a positive (i.e., statistically significant) key result, this suggests that there are a lot of false positives in the published literature (Scheel, Schijen, & Lakens, 2020). Another key finding from large-scale replications is that effect sizes in published articles tend to be substantially inflated (Kvarven, Strømmland, & Johannesson, 2020). Camerer et al. (2018, p. 637), for example, replicated 21 social science experiments published in *Nature* and *Science* and found that “the effect size of the replications is on average about 50% of the original effect size.” The most likely culprit is QRPs.

Registered Reports, articles that are reviewed and accepted (or rejected) by journals before the data have been collected, are designed to reduce or eliminate avenues for QRPs. Journals cannot decide whether or not to accept a manuscript based on how exciting the results are, and authors cannot change their plan for data collection or analysis after the plan is approved (nor do they have much incentive to do so, if the manuscript has already been accepted for publication) (Ritchie, 2020). Two different analyses comparing the results of Registered Reports to those of traditional journal articles (Allen & Mehler, 2018; Scheel, Schijen, & Lakens, 2020) both find that only about 45% of Registered Reports present a positive (i.e., statistically significant) key finding, compared to over 90% of traditional articles. This strongly suggests that QRPs account for much of the false positives in the traditional literature (though there are



alternative explanations, e.g., that research hypotheses tested in Registered Reports have lower prior probabilities).

As a response to QRPs and related concerns about the credibility of research findings, researchers have proposed greater use of OSPs (Simmons et al., 2011, 1362-63), among other reforms (Vazire, Schiavone, & Bottesini, 2020). These practices include the sharing of data and code, as well as preregistration, replication, and efforts to make articles themselves publicly available. Open practices make it easier to scrutinize a finding, for example by attempting to replicate the study (i.e., collect new data following the same procedures) or attempting to reproduce the results (i.e., re-analyze the original data). Of course, openness does not guarantee that findings will be robust; openness, rather, makes it easier to assess robustness. Critical appraisal is then necessary to identify robust versus weak results (Vazire & Holcombe, 2020). Together, transparency and critical appraisal can curb the use of QRPs by incentivizing rigorous practices and deterring practices that inflate, exaggerate, or increase the risk of false positives.

Open practices also increase the probability that honest errors in research are uncovered. Research consistently finds that honest statistical errors are quite common (Ritchie, 2020). In psychology, for instance, about 50% of published articles were flagged by the statcheck program as containing a statistical error (i.e., an inconsistency among the statistics reported within a single test), and about 12% were flagged as containing an error that changed the statistical significance of the result (Nuijten et al., 2016; see also Bakker & Wicherts, 2011). In order for a field to be credible, it must make its errors detectable (i.e., make research transparent), and incentivize the actual detection and correction of those errors. Open science practices indicate a commitment to self-correction, and are one of the hallmarks of a credible science. A science can

only claim to be credible if it makes transparency a priority, and implements mechanisms to detect and correct errors.

The good news is that in the wake of the credibility revolution, researchers appear to be adopting OSPs. For instance, a recent study asked researchers in psychology, economics, political science, and sociology about the first time they had used one of three open practices: open data, open research materials, and preregistration (Christensen et al., 2019). Overall, they found considerable upticks in these practices over recent years. Similarly, journal policies encouraging or requiring open data seem to be having the desired effect, with articles published in those journals being more likely to have made their data available in a public repository (Hardwicke et al., 2018; Kidwell et al., 2016; c.f., Rowhani-Farid, Aldcroft, & Barnett, 2020). Unfortunately, the evidence on the prevalence of OSPs in different fields is still scant. In Table 1, we list estimates from two prior studies (from education and quantitative criminology). Both found that posting public copies of articles was the most common open practice and that about half of respondents had shared data and/or analysis code.

### **Application to Criminological Research**

Criminology should be just as concerned as other fields about avoiding QRPs and ensuring research is credible, especially given the societal implications of its findings. As Gelman et al. (2020, p. 296) explain, “criminological research findings have considerable potential to influence (for better or worse) citizens’ lives, given the immense reach of the criminal justice system.” Indeed, there is a growing movement toward a “public criminology,” or, in other words, research that is useful to the individuals, communities, and social and governmental institutions (Uggen & Inderbitzin, 2010). Beyond informing policy and other social interventions, criminological research is also relied on in courts. For example, QRPs were

used by a criminologist who studied gangs in the wrongful conviction of a young man in Ontario (Chin, 2019; Chin et al., 2019). That criminologist's use of analytical flexibility (e.g., a shifting definition of "gang member" across studies) resulted in an overstatement of the evidence suggesting that a certain tattoo indicated the bearer was involved in a gang killing.

Despite progress and meta-research in other fields, levels of use and endorsement of QRPs and OSPs among criminologists remain unclear. Usefully, however, one recent study of 75 terrorism researchers found evidence that they supported the use of OSPs, although, unfortunately, they rarely used them (Schumann et al., 2019). And a recent scandal in the field suggests that even among co-authors, data and code are not always shared (Pickett, 2020). This may be exacerbated by disciplinary norms – for example, no criminology journal (to our knowledge) requires the sharing of data or code.

Lack of OSP use is unfortunate because they produce demonstrable benefits (Allen & Mehler, 2019). For instance, in one study of the effects of police violence, open data allowed a reader to find a coding error that changed the study's main finding (American Association for the Advancement of Science, 2019). The author retracted the study before it could have downstream effects. Moreover, OSPs make it easier to conduct replications, which are a cornerstone of scientific knowledge, and are also exceedingly rare in criminology, constituting .5% to 2% of published articles depending on the definition of replication (Pridemore, Makel, & Plucker, 2018; McNeeley & Warner, 2015). Finally, OSPs make it harder to use QRPs, because many QRPs are, by definition, about withholding relevant information from readers.

### **Study Overview**

To provide initial evidence about how criminologists view QRPs and OSPs, and about whether they use them, we conducted a preregistered study of researchers who publish in

criminology journals. Our study, the first survey research on QRPs and OSPs in criminology, can be used to shed light on whether there are particular strengths and weaknesses in criminology's current practices. The findings can also be used as benchmarks to be revisited as the field changes. As we will describe, we asked participants about 10 QRPs (Table 2) that have been widely studied elsewhere, which include two that border on research fraud (filling in missing data without reporting it, and hiding known problems with the data) and 5 OSPs recently studied in surveys of education (Makel et al., 2019) and communication (Bakker et al., 2020) researchers.

As stated in our preregistration (<https://osf.io/fbhkq>), our study's primary aim was descriptive. Specifically, we aimed to provide estimates of criminologists' self-reported use of the 10 QRPs and 5 OSPs examined ("use"), their perceptions of other criminologists' use of these practices ("prevalence"), and their levels of endorsement of these practices ("support"). We also specified two hypotheses in advance of data collection. Our first hypothesis was that use of and support for QRPs would be negatively correlated with use of and support for OSPs. This hypothesis flows from a deterrence theory of open practices; they arose, in part, to make transparent, and therefore discourage, QRP use (Simmons et al., 2012, 1362-63). Our second hypothesis was that methodological training would be associated with use of and support for both QRPs and OSPs, controlling for career stage. Training might make researchers more aware of the negative effects of QRPs (and benefits of OSPs). Alternatively, QRP use could be enabled by greater methodological knowledge and skill. Given the effect of training could plausibly go in either direction, we refrained from making a directional hypothesis.

## Methodology

### Sample

Our research design follows those used to study QRPs and OSPs in other fields (Table 1). Our materials and de-identified data are publicly available in the Open Science Framework (OSF) repository (<https://osf.io/qvcdg/>). Our study received human ethics approval from the University of Sydney (<https://osf.io/n5svq/>). We used a computerized, self-administered survey because research suggests that it is the best mode for obtaining honest answers (Tourangeau, Conrad, & Couper, 2013).

As in Fraser et al. (2018), Makel et al. (2019), and Bakker et al., (2020), we surveyed researchers who had published in journals in the field of interest (criminology). We selected criminology journals using the Web of Science's "Criminology and Penology" category (Web of Science, 2018) and two academic studies of criminology journals (DeJong & St. George, 2018; Sorenson, 2009). From these lists, we excluded 23 journals we determined were not sufficiently related to criminology (e.g., Journal of Forensic Psychiatry & Psychology), and 14 journals for other reasons (e.g., language other than English). As a result, we sampled from 67 criminology journals. This process and exclusion justifications were detailed in our preregistration (<https://osf.io/fbhkq>). They are further explained in our supplementary materials (<https://osf.io/myhx9/>).

From the 67 journals, we extracted 16,157 unique author email addresses. For journals indexed by the Web of Science, we obtained emails through its database of article information. For others, we adapted code written by Makel et al. (<https://osf.io/83mwk/>) that scrapes journal websites for e-mail addresses (<https://osf.io/qvcdg/>). In some cases, we also obtained email addresses by hand-coding author information (<https://osf.io/myhx9/>). Survey invitations and follow-up reminders were sent on August 10, 20, and 28, 2020. We closed data collection on

September 12, 2020. Of the 16,157 obtained email addresses, 17 failed, and 2,370 bounced back, resulting in a total of 13,770 successful email account contacts. However, some of those accounts may not have been actively monitored by their owners during the time period of our survey (August, 2020).

In total, we received 1,612 responses. This response rate (12%) is small, but similar to other recent studies sampling authors or editors (Makel et al., 2019; Hopp & Hoover, 2017; Horbach & Halfman, 2020), and exceeds those often obtained by professional polling organizations (Keeter et al., 2017). A large body of research shows that “nonresponse bias is rarely notably related to [the] nonresponse rate” (Krosnick et al., 2015, p. 6). However, in our survey, given its topic (research behavior), nonresponse may have resulted in bias. Any nonresponse bias, however, is likely to result in underestimates of QRP use and support, and overestimates of OSP use and support, because, if anything, support for the credibility revolution would likely have increased individuals’ likelihood of responding to our survey.

As in Makel et al. (2019), we asked respondents at the start of the survey: “Have you conducted quantitative research that involves null-hypothesis significance testing?” Unlike Makel et al., (2019), we excluded from our main report those who reported they did not do quantitative research involving null-hypothesis significance testing ( $n = 479$ ), because they were not asked all of the questions (they were asked about: HARKing, underreporting results, hiding data problems, hiding imputation and all the OSPs). This exclusion is not listed in our preregistration because we did not anticipate the difficulties created by only asking a subset of the questions to the subsample of non-quantitative respondents. After collecting the data, but before looking at the results, we decided it would increase comparability to limit the analysis to respondents who received the same questionnaire. However, the data for all respondents,

quantitative and non-quantitative, is provided online in the supplementary materials (<https://osf.io/8me9w/>) and, where possible, the analyses below have been reproduced on the whole dataset and on the non-quantitative sample.<sup>1</sup>

Another 50 respondents are excluded because they indicated they did not want their data used. Additionally, one senior criminologist at a top-five institution (per U.S. News rankings) sent an email opining that the survey was “useless” and “a greater ethical lapse than fiddling with p values.” The criminologist informed us that s/he had “selected arbitrary answers out to two decimal places for all answers.” We identified that respondent’s responses (based on the decimals) and dropped them (see markdown file in github directory). Finally, there was item non-response, which further reduced the full analytic sample to between 579 and 711, depending on the analysis.<sup>2</sup>

## Measures

We asked participants about 10 QRPs (Table 2) that were also included in prior surveys in other fields (Table 1). We also asked about five OSPs (Table 3) that Makel and colleagues (2019) included in their survey. The order of the presented practices was randomized between participants. Tables 2 and 3 provide the exact question wording for the specific QRPs and OSPs, along with the abbreviations (variable names) that we use in the figures. For each practice, as in prior research, we measured self-reported use, perceived prevalence, and support.

Use was measured with two questions. The first asked: “Have you ever engaged in this practice?” (1 = yes, 0 = no). The second was a contingency question asked to those who answered affirmatively to the first question: “What PERCENT of studies you have conducted—

<sup>1</sup> <https://github.com/alexholcombe/ChinHolcombePickettVazireCrimSurvey/>.

<sup>2</sup> Randomization of question ordering (see below) meant that breakoffs equally (on expectation) affected all practices, but also meant that item nonresponse was not concentrated at the end of the survey. As a result, there are many respondents who answered questions about only one randomly presented QRP or OSP.

that is, how many out of 100—would you say that you used this practice?” Perceived prevalence was measured with the question: “What percent of criminologists—that is how many out of 100—would you say have engaged in this practice on at least one occasion?” Finally, support for the practice was measured by asking: “How frequently SHOULD criminologists use this practice?” There were four response options: Almost always (coded 4), often (3), rarely (2), and never (1).<sup>3</sup>

To maintain respondents’ anonymity, we asked only two background questions. The first assessed their career stage: “Which of the following best describes your current position?” There were four response options: Senior research academic/researcher (coded 4), mid-career academic/researcher (3), earlier career academic/researcher (including post-doctoral fellows) (2), and graduate student (1). The second question measured methodological training: “How many university courses (undergraduate or graduate) on methodology or statistics have you taken?” There were eleven numerical response options, ranging from “0” to “10 or more.”

### **Analytic Strategy**

As described in our preregistration (<https://osf.io/fbhkq>), most of our analyses are descriptive, examining the distribution of responses to the individual questions about each practice. Obtaining these estimates was the primary aim of this study. For the analyses examining associations among variables, we constructed indices for QRPs and OSPs by averaging responses for those respondents who answered at least three of the component items in each index. In our preregistration, we did not specify the item-missingness criterion we would use to construct the mean indices. However, prior to looking at the results, we decided to create

<sup>3</sup> In using the 0-100% response scale, our assumption was that criminologists would be as able to use it as laypeople, who regularly respond on this scale in major surveys (Manski, 2004).



the indices only for respondents who answered at least three of the component items. These indices were used in correlations and in ordinary least squares regressions, where we also use robust standard errors to account for heteroskedasticity. This also represents a deviation from our preregistration, which planned to use negative binomial and ordered logistic regression, based on assumptions about the distribution of the outcomes, which turned out to be incorrect. Because of these departures from our analysis plan, we urge caution in interpreting the associated findings.

## **Results**

### **Questionable Research Practices (QRPs)**

#### *Use of QRPs*

How many criminologists report using QRPs? Table 2 presents the self-reported use of QRPs among quantitative criminologists during their career (see the supplementary materials for a comparison of criminologists to other fields). Use of specific QRPs ranged from 7% (hide imputation) to 53% (underreport results) (see Table 2 for 95% confidence intervals around all point estimates reported here). It is concerning that 7% of respondents, by their own admission, do not always disclose when they impute (fill in) missing values, given that it is arguably a form of data fraud—specifically, falsification of data (see Fraser et al., 2018, p. 5). About 10% of respondents admitted to not disclosing known problems with the method, data, or analysis that potentially impact conclusions (hide data problems). QRPs with higher rates of self-reported use were: failing to report null results (omit non-significant studies or variables, 43%), changing the analysis after an earlier one failed to yield significant findings (switch analysis selectively, 39%), using *p*-values to select covariates (drop covariates selectively, 32%), hypothesizing after the results are known (HARK, 29%), and excluding data after checking how it impacts results (exclude data selectively, 24%).

The QRP responses were combined into a summary index for each participant. Respondents were free to leave any question blank, however, and many respondents did not answer every QRP question. Among those that answered at least three QRP questions, the majority (87%) admitted using at least one QRP. Among respondents who answered every QRP question, the average number of QRPs used was three. These metrics were not preregistered, although they are commonly reported in research on QRPs (John et al., 2012; O’Boyle et al., 2017). Overall, the findings indicate that most respondents have used QRPs and that the average respondent has used more than one. Respondents who reported using QRPs also tended to report using them repeatedly. Specifically, QRP-using respondents reported using the different QRPs in 29% to 47% of their studies. Even for the two most serious QRPs (hiding known data problems and filling in missing values without reporting it), users reported regular use (on average, in 31% and 34% of studies, respectively).

It is instructive to compare QRP use in criminology (Table 2) to that in other fields (Table 1; Supplementary materials, <https://osf.io/kj3bf/>), although this comparison should be considered only suggestive because these studies were conducted in different times and countries, and used different sampling methods and question wording (in some cases). Our findings for criminologists are generally in line with those from studies in other disciplines. For example, John et al. (2012) found that 91% of psychologists admitted using at least one QRP, whereas 87% of criminologists admitted doing so in our study. Turning to specific QRPs, 1-10% of scientists in other fields said they changed data without reporting it (hide imputation), compared to 7% of criminologists in our sample. Similarly, 40-62% of scientists in other disciplines said they failed to publish studies with null findings; the figure is 43% for criminologists in our sample. In other disciplines, 20-43% of scientists decided whether to

exclude data after looking to see how it affected results, a range that includes our prevalence estimate for criminology (24%). The most notable difference between our results and those of other studies is for selective sampling (using  $p$ -values to decide when to stop data collection). Comparatively few criminologists in our sample (15% vs. 22-58% in other disciplines) use this QRP. This finding may reflect a greater reliance on secondary data among criminologists, which would reduce their opportunities for selective sampling.

[Insert Table 2 about here]

### *Perceived prevalence of QRPs*

Do criminologists believe that QRPs are common? Perceived prevalence was measured by asking respondents what percent (0 to 100) of criminologists they would say have engaged in the practice at least once. Figure 1 shows respondents' perceptions of prevalence for each QRP. Respondents perceive a relatively high prevalence of QRPs among other criminologists. Respondents perceive that 21-59% of other criminologists have used each QRP at least once. Thus, not only have most criminologists in our sample used QRPs, but most also believe that many other criminologists use them.

[Insert Figure 1 about here]

In general, the pattern of perceived prevalence across QRPs is similar to the pattern of self-reported use, with hiding data problems, hiding imputed data, and selective sampling perceived as relatively rare, whereas omitting non-significant studies or variables, underreporting results, and selectively switching analyses perceived as most common.

However, this similarity between the pattern of mean self-reported use and the mean perceived prevalence may be misleading. Figure 1 shows that the distribution of participants' responses to some of the prevalence questions resembles a uniform distribution. That is, for some

of the practices, the reason the mean prevalence response is close to 50% is not because 50% was a typical answer, but rather because participants gave answers nearly uniformly distributed throughout almost the entire range. This suggests that most participants know very little about the prevalence of these practices in their field. One possibility is that they have a good estimate of the prevalence among some of their peers, but not the field as a whole, and these peer communities are very heterogeneous. As we will elaborate on in the discussion, we obtained data from six previous QRP studies and found a similar pattern of results, suggesting researchers may wish to rethink the diagnosticity of asking researchers to estimate their colleagues' practices, and that descriptive norms about research behavior may only be weakly tied to reality.

#### *Support for QRPs*

Do criminologists believe QRPs are defensible? Figure 2 shows the distribution of respondents' answers, ordered from those with the least support (highest proportion of "never" answers) to the most support. Most respondents support using some QRPs in some circumstances.<sup>4</sup> For example, 67% of respondents support (in at least some circumstances) selectively choosing not to publish null findings (omit non-significant studies or variables), 65% support looking at *p*-values before deciding whether to collect more data (sample selectively), and 45% support framing unexpected findings as if they were hypothesized a priori (HARK) (see Figure 2). Perhaps most concerning is that 25% of respondents believe it can (even if rarely) be okay to hide known data problems, and 18% of respondents say it can be okay to fill in missing values without disclosing it to readers. Unlike the use questions, which may capture behavior from years ago, the support questions measure criminologists' current support.

[Insert Figure 2 about here]

<sup>4</sup> We define support as any answer other than "never."

## Open Science Practices (OSPs)

### *Use of OSPs*

How widespread is OSP use in criminology? One might assume that if QRPs are common in criminology, then OSPs would not be, but prior research suggests this may not be the case. Makel et al. (2019), for example, found that both QRPs and OSPs were common in education research, with most scientists using both. Is the same true of criminologists? Table 3 displays the five OSPs we asked about and the percentage of researchers saying they had used them at least once, and among those, the percentage of studies they had used them in.

Our findings mirror those of Makel et al. (2019) and Bakker et al. (2020) (Table 1; Table 3). They found that the most common OSP was posting articles publicly, so that they are not behind a paywall, with 78% of education researchers and 85% of communication researchers using this practice. The same is true in our survey of criminology, where 68% of respondents have posted articles publicly. Previous studies also found high levels of preregistration (54% in education, 47% in communication), data sharing (59% in education, 64% in communication), and attempting a replication (43% in education, 58% in communication). A further 59% in education reported sharing code at least once (the communication study did not ask about code). The numbers in our survey are similar, although in every case they are lower (we did not perform any inferential statistics): 45% of respondents said they have preregistered studies, 43% have shared data, 40% have attempted a replication, and 43% have shared code. It bears noting that if 40% of criminologists have attempted replications, then the finding in prior research that only .5-2% of published criminology articles are replications (Pridemore, Makel, & Plucker, 2018; McNeeley & Warner, 2015) suggests there may be a large unpublished replication literature in

the discipline (there may also be differences in how our respondents defined replication, see the discussion).

Overall, 89% of the respondents who answered at least three OSP questions said they had used at least one OSP. Among those who answered all the OSP questions, the average respondent reported using two OSPs. Additionally, respondents who have used OSPs have used them frequently. Those who have posted articles publicly have done so for most (59%) of their studies. Those who have preregistered studies have done so for half (50%) of their studies. Those who have shared data and code have done so for about a third of their studies.

It is notable that the vast majority of quantitative criminologists in our sample have used OSPs (89%) and used them frequently, but a similarly large majority have also used QRPs (87%). Although this may seem contradictory, because OSPs signal transparency whereas many QRPs involve hiding crucial information from readers, there are several possible explanations, which we will discuss in the conclusion. Two main possibilities are changing research practices (i.e., shifts over time from QRPs to OSPs) and selective transparency, where OSPs and QRPs are used by the same researchers in the same time period for different articles.

[Insert Table 3 about here]

### *Perceived prevalence of OSPs*

Do criminologists believe that OSPs are common in the discipline? Figure 3 shows respondents' perceptions of the percentage of other criminologists who have used each OSP at least once. The perceived prevalence of OSPs is around 25-30% for most OSPs, but somewhat higher (48%) for posting articles publicly. Each distribution has a fairly prominent peak, suggesting more agreement (or knowledge) about the prevalence of OSPs than QRPs. A possible exception is the "Post article publicly" OSP, which has a very wide distribution. On average,

respondents seem to perceive OSPs as slightly less prevalent than QRPs. As with QRPs, the pattern of perceived prevalence across OSPs matches the pattern of self-reported use, with posting articles publicly perceived as more prevalent than the other OSPs.

[Insert Figure 3 about here]

### *Support for OSPs*

Do criminologists support the use of OSPs? Recall that the response options for the support question (“How frequently should criminologists use this practice?”) were “never”, “rarely”, “often”, and “almost always”. Figure 4 shows the distribution of respondents’ answers. The findings are striking. Respondents are *much* more supportive of OSPs than of QRPs. For each OSP, more than 75% of respondents reported that OSPs should be used “often” or “almost always,” and over 95% supported their use at least “rarely”. There is evidently a strong consensus in criminology that OSPs are important and should be used. However, there are also some sobering patterns in the results. For example, for sharing code or posting articles publicly, which seem to us to be universally good practices (i.e., we cannot think of cases where these would be harmful), only about 25-35% of respondents selected “almost always”. Another striking pattern is that 99% of respondents said that criminologists should at least sometimes (even if “rarely”) attempt replications, but this OSP had the lowest rate of self-reported use (40%, Table 3).

[Insert Figure 4 about here]

### **Correlations between QRP and OSP responses**

Table 4 presents the bivariate correlations among the variables. Interestingly, and contradicting our first hypothesis, we find a significant and positive correlation between QRP and OSP use ( $r[667] = .21, p < .001, 95\% \text{ CI} = .14 \text{ to } .29$ ), though note that we preregistered a

one-tailed test in the opposite direction, so this finding should be considered exploratory, albeit quite strong. This correlation is descriptively larger than the small positive correlation ( $r = .06$ ) found by Makel et al. (2019) among education researchers. Similar to Makel et al. (2019), however, and consistent with the part of our first hypothesis having to do with support, we find a significant and negative correlation ( $r[661] = -.15, p < .001, 95\% \text{ CI} = -.22 \text{ to } -.07$ ) between support for QRPs and support for OSPs. The comparable correlation in Makel et al. (2019) was  $r = -.20$ . In sum, we find mixed evidence for our first hypothesis. While there is some consistency in support (more support for OSPs is associated with less support for QRPs), behavior appears to be inconsistent—respondents who have used more QRPs have also used more OSPs. This is somewhat counterintuitive as OSPs increase transparency whereas most QRPs involve obfuscation or hiding information.

### **Methodological Training, Career Stage, and Research Practices**

Turning to our second hypothesis, Table 5 presents the relevant regression results. We find no evidence that methodological training is significantly related to either research behavior or attitudes, which runs contrary to our hypotheses (Table 5). Controlling for career stage, the relationship of methodological training to QRP use is small and non-significant ( $b = -.123, p = .453, 95\% \text{ CI} = -.445 \text{ to } .199$ ), as is its relationship to OSP use ( $b = .228, p = .328, 95\% \text{ CI} = -.229 \text{ to } .684$ ).<sup>5</sup> Similarly, the relationship between methodological training and support for QRPs and OSPs is small and non-significant (QRPs:  $b = -.005, p = .412, 95\% \text{ CI} = -.017 \text{ to } .007$ ; OSPs:  $b = .011, p = .089, 95\% \text{ CI} = -.001 \text{ to } .024$ ). The confidence intervals in all results exclude unstandardized effects with absolute values greater than .7, suggesting that we have

<sup>5</sup> CIs calculated as  $\pm 1.96 * (\text{sd} / (\text{sqrt}(n)))$ .



enough precision to confidently rule out any meaningfully-sized association between methodological training and responses to our QRP and OSP items.

We also examined the association between career stage and responses to QRP and OSP questions, with later career stages coded with higher scores. Like methodological training, career stage is not significantly related to QRP and OSP use (QRPs:  $b = .496$ ,  $p = .270$ , 95% CI =  $-.387$  to  $1.380$ ; OSPs:  $b = 1.164$ ,  $p = .088$ , 95% CI =  $-.174$  to  $2.502$ ). However, we do find some significant associations between career stage and QRP and OSP support. Compared to respondents at early career stages, those at later stages are: 1) significantly more supportive of QRPs ( $b = .055$ ,  $p = .003$ , 95% CI =  $.019$  to  $.091$ ), 2) significantly less supportive of OSPs ( $b = -.040$ ,  $p = .021$ , 95% CI =  $-.074$  to  $-.006$ ), and 3) significantly less likely to perceive QRPs as common in the discipline ( $b = -3.309$ ,  $p < .001$ , 95% CI =  $-4.880$  to  $-1.738$ ). The findings for career stage should be interpreted as exploratory because, while we did preregister the use of career stage as a variable in our models (<https://osf.io/fbhkq>), we did not predict a relationship between career stage and QRP or OSP support.

[Insert Table 5 about here]

## Discussion

We found widespread self-reported use of QRPs among criminologists at levels similar to what has been found for other fields. Criminologists also reported supporting the use of most QRPs in at least some circumstances (Figure 1), even though some of these practices explicitly entail hiding or misrepresenting information. Moreover, respondents estimated that others were more likely to use QRPs than they were themselves. As we discuss below, this pattern of results, and other design features of our study, suggest we may be underestimating the prevalence of QRP use (and overestimating the prevalence of OSP use).

The results of our survey may provoke both concern and relief. Beginning with concern, the high rate of QRP usage is disappointing because QRPs contribute to false and misleading findings. Our evidence is consistent with the conclusion that many findings in criminology are likely false positives (Sweeten, 2020; Gelman, Skardhamar, & Aaltonen, 2020; Wooditch et al., 2020). Existing efforts to make criminologists aware of the pitfalls of QRPs (see Burt, 2020) should be strengthened. On the other hand, our results (which again, may underestimate QRP prevalence) suggest that the circumstances in criminology are not much different than in fields that may have already been partially reformed by various initiatives and cultural shifts.

Based on our experience, we expected a lower level of self-reported adoption of OSPs among criminologists. Instead, the levels are in line with prior research in other disciplines (Makel et al., 2019). Our survey may have overestimated OSP use, in part because it is likely seen as socially desirable behavior. In any event, it is promising that so many respondents report using OSPs and encouraging that nearly all support these practices. We were especially surprised by the high level of self-reported use of preregistration. Prior studies have found that preregistration lags other OSPs (Christensen et al., 2019), whereas we found it was used at similar rates as other OSPs. One explanation for this difference may be that some participants in our study interpreted preregistration more liberally than we intended (recall we asked about “preregistering hypotheses and analysis plans prior to data collection”). They may have taken this to mean recording research plans anywhere, such as in grant applications or in communications with collaborators. We suggest future studies should use a more specific and detailed definition of preregistration (e.g., specifying that the plan should be recorded in a time-stamped repository). Subsequent research should also ask whether criminologists adhere to their

preregistrations, which we did not ask about; some research suggests they often do not (Wooditch et al., 2020).

Almost 70% of respondents reported publicly posting at least one article in their career, and those indicating they had done so reported doing it for about 60% of their studies. Despite this seemingly high rate, open access to criminology articles remains low. Ashby (2020) recently found that less than 25% of criminology articles published from 2017-2019 were available in open access format (despite all criminology journals he studied allowing preprints). The explanation for the discrepancy between his results and ours is unclear. It may be that our sample is overestimating their own pattern of publicly posting articles. It may be that articles posted publicly (e.g., on ResearchGate.net) were subsequently taken down, either by the website or the researcher. ResearchGate.net, for example, has removed many public full texts because of journal policies. Another possible explanation may be that our sample is overrepresenting criminologists who have used OSPs. Regardless, it is encouraging that 295 criminologists recently signed an open letter to the American Society of Criminology requesting that criminology journals allow authors to post full-text versions of their articles publicly.<sup>6</sup>

Contrary to our preregistered hypothesis, we did not find a negative relationship between QRP and OSP use, and in fact, found an unexpected positive relationship. This relationship should be interpreted with some caution given that we preregistered a one-tailed test, though the result is quite strong even for an unexpected finding. This positive relationship runs contrary to a deterrence perspective of open practices. One possible explanation for why we did not observe the expected negative relationship is that there may be an unmeasured common cause, such as research productivity, that is positively related to both practices, distorting their true association.

<sup>6</sup> <https://www.criminologyopen.com/pub/open-letter-to-asc-concerning-green-access-to-its-journals/release/21>.

Criminologists who publish more articles have more opportunities to use QRPs and OSPs, and thus may be more likely to use both practices.

There are at least three other possible explanations for the unexpected positive correlation between the use of QRPs and OSPs. One is that it reflects selective transparency, whereby criminologists use both practices, but do so in different articles. It may be that criminologists who are more focused on their careers and prestige are more likely to use QRPs to get articles published in top journals. O'Boyle et al. (2017) found that QRPs were more common in articles published in top journals. At the same time, with knowledge of the importance of OSPs comes awareness that using them has reputational benefits. Thus, the same career-oriented criminologists may also want to cash in on the reputational benefits of using OSPs when they can (e.g., when findings are initially significant). The evidence that criminologists deviate selectively from their preregistrations to increase effect sizes is consistent with this selective transparency explanation (Wooditch et al., 2020).

Another possibility is that the positive relationship between OSPs and QRPs may be due to changing research practices, whereby criminologists who once used QRPs have stopped and now use OSPs. Similar to recovering alcoholics who regularly attend alcoholics anonymous meetings, these criminologists may regularly use OSPs to avoid falling back into bad research habits and/or to reinforce their new identity as careful researchers. Our survey did not ask when criminologists used the practices, only whether they had, so we cannot test this possibility. Of course, it is also possible that the positive correlation between QRPs and OSPs is simply due to stylistic responding, whereby some respondents tended to answer affirmatively regardless of question content (Pickett & Baker, 2014).

Although QRP use and OSP use are positively related, for whatever reason, the relationship between QRP support and OSP support is negative, as hypothesized. This negative correlation for attitudes was also found for education researchers (Makel et al., 2019). The finding indicates that those who endorse one set of practices are less likely to endorse the other. Unlike behavior, which was measured retrospectively (“Have you ever engaged in this practice?”), attitudes were measured contemporaneously, at the time of the survey. Thus, the negative correlation between current support for QRPs and OSPs, combined with the positive correlation between past use of QRPs and OSPs, is seemingly consistent with the idea that some criminologists have “made good” or “gone straight,” abandoning QRPs for OSPs. Perhaps their regrets about using QRPs in the past, along with their awareness of the adverse effects of what they did (false positives) and the incentives (e.g., publication) that pushed them to do it, is what convinced them that OSPs are important for the discipline.

Contrary to our preregistered hypothesis, we did not find a relationship between methodological training and QRP or OSP use. It is somewhat disappointing that methodological training does not predict better research behavior. It could be that our sample is rather uniformly highly-trained—half of respondents had taken six or more methods courses—dampening any associations. Of course, that would also mean that highly-trained criminologists still use QRPs at high rates, which would not be good news. One reason that methodological training may not be related to QRP use is that QRP use may not reflect ignorance—QRP users may be fully aware of what they are doing. Alternatively, this lack of association could reflect historic norms in methodological training. Traditionally, methodological training in criminology likely did not emphasize the dangers of QRPs, but more recent training may do so, thus in the future a negative correlation between methodological training and QRP use may emerge.

Turning to career stage, which we did not preregister as a variable of interest, we found that career state was positively correlated with QRP support and negatively correlated with OSP support. More senior researchers were more supportive of using QRPs, and less supportive of using OSPs. Although we did not preregister any hypotheses about these relationships, we will speculate on one possible explanation: they might be a result of more junior researchers learning research practices in an era of greater awareness of the replication crisis and of the dangers of QRP use. Note, however, that previous studies have not found a relationship between career stage and QRP support in other disciplines (Agnoli et al., 2017, p. 9; Makel et al., 2017; Rabelo et al., 2020, p. 680). Set against that background and the fact that we did not preregister relevant hypotheses, we would suggest that conclusions about how senior and junior criminologists differ in their support for QRPs and OSPs await future replications.

Finally, recall that we found that participants' estimates of others' QRP use were quite uniformly distributed from 0 to 100% for multiple practices (Figure 1). We obtained and re-analyzed the data for published studies that used questions almost identical to those of our study and found a similar pattern of results (Figure 5). We also obtained data from studies using questions more similar to those in the original John et al. (2012) study (Figure 6). Both show a visibly uniform distribution for the perceived prevalence of some practices, like we found in our study. For HARKing, for example, four of the six datasets show distributions that are close to uniform, and the two exceptions (Agnoli et al. 2017; Rabelo et al., 2019) are still extremely dispersed, such that the mean is not representative.

These response patterns suggest that in psychology and education research as well as criminology, survey respondents do not have much knowledge about the prevalence of QRPs used by their colleagues. One implication is that there may be at best weak descriptive norms –

norms based on what behavior is actually common – governing research ethics in these fields. Criminologists have shown that people misperceive their peers' deviant behavior, and that personal offending and self-control shapes such perceptions (Young et al., 2011). Along these lines, future research is needed that explores the nature and correlates of scientists' perceptions of their disciplinary colleagues' research behavior. The finding that scholars are unaware of others' research behavior also supports the benefit of studies like ours, which make public and explicit the research practices actually being used in their field.

[Insert Figures 5 and 6 about here]

### **Limitations**

While our study benefited from a large sample size (especially in relation to previous survey studies, <https://osf.io/kj3bf/>), confidence in our results should be qualified by the low response rate. The low response rate increases the possibility that our sample is biased in some way and thus non-representative. However, note that the nonresponse rate is not always related to nonresponse bias (Krosnick et al., 2015, p. 6), and in our study any bias would likely be towards underestimating the use of QRPs and overestimating the use of OSPs. This is because it was clear from survey items (although not the recruitment material) that they pertained to QRPs (which are increasingly proscribed) and OSPs (which are increasingly endorsed). We expect that researchers would be both more motivated to participate if they were the type to be concerned about QRPs (Dahlgaard et al., 2019), and that respondents would be inclined to portray their practices along those lines as well. Future research is thus needed that replicates our study with a high-response rate survey.

Our sample was also skewed towards mid-career researchers (n = 199) and senior researchers (n = 241) (<https://osf.io/5xyp3/>). We received responses from few graduate students

( $n = 32$ ). This is likely because graduate students are not often named as the corresponding author in published articles, which is how we sampled criminologists' email addresses. This emphasis on mid-career and senior researchers also makes it possible we were surveying researchers who used QRPs earlier in their career, but do not presently (or that their use has tailed off). Still, present support for QRPs suggests that use is enduring. Further surveys may wish to target earlier career researchers to get a better idea of their engagement with and opinions of QRPs and OSPs.

## **Conclusion**

The message of the Rolling Stones' "You can't always get what you want," applies here. If more criminologists forgo QRPs and adopt OSPs, they will get less of the findings they want, but the discipline (and society) will get more of what it needs: reproducible science. We found that most quantitative criminologists in our sample have used QRPs, most believe other researchers use QRPs, and many support the use of QRPs in at least some circumstances. We also found that many criminologists use OSPs, and even more—over 95% of those surveyed—support using OSPs in at least some circumstances.

Our findings, then, provide both bad and good news. The bad news is that QRPs appear widespread and are often condoned in criminology. QRPs bias research by exploiting undisclosed flexibility in the data gathering and analysis process to get findings that are desired but often wrong (Beerdsen, 2020; Simmons, Nelson, & Simonsohn, 2012). As a field that affects policy (Uggen & Inderbitzin, 2010) as well as court decision-making (Chin, Mellor, & Grown, 2019), QRPs in criminology have real-world consequences. When QRPs are widespread, the evidence used for evidence-based policy is not credible. The good news is that there appears to be an opportunity for improvement, given criminologists' support for OSPs. OSPs make errors



detectible (American Association for the Advancement of Science, 2019), disincentivize misconduct (Ritchie, 2020), and promote public access to science (Ashby, 2020).

Behavior is supported, in some part, by beliefs about what others do, which are the foundation of descriptive norms. For some issues, such as undisclosed imputation of data, we found that most criminologists are in agreement that only a small proportion of their colleagues have engaged in the practice. For multiple practices, however, the distribution of prevalence responses reveals that criminologists show little agreement (or knowledge) about the practice's prevalence – indeed, their responses resemble a uniform distribution.

We hope our results inspire further reform in criminology and serve as a benchmark for future studies assessing the state of the field. Reforms to empirical and methodological training may be useful in reducing endorsement of QRPs ([FORRT, 2019](#)). For open practices, we saw a pattern that is familiar from surveys of other fields, whereby researchers' own practices lag endorsed practices (Anderson et al., 2007). In this case, we hope that efforts that make it increasingly easy to upload data and materials to public repositories (Meyer, 2018), preregister hypotheses and methods (<https://osf.io/zab38/wiki/home/>), and upload preprints (see Ashby, 2020), will help bring researcher behavior closer to the norms they subscribe to.

## References

- Allen C. & Mehler D. M. A. (2019). Open science challenges, benefits and tips in early career and beyond. *PloS Biol*, 17(5), e3000246.
- Agnoli, F., Wicherts, J. M., Veldkamp, C. L. S., Albiero, P., & Cubelli, R. (2017). Questionable research practices among italian research psychologists. *PLoS ONE*, 12(3), e0172792.
- Ashby, M. P. J. (2020). The Open-Access Availability of Criminological Research to Practitioners and Policy Makers. *Journal of Criminal Justice Education*, 1-21.
- Bakker, B. N., Jaidka, K., Dörr, T., Fasching, N., & Lelkes, Y. (2020, November 18). Questionable and open research practices: attitudes and perceptions among quantitative communication researchers. <https://doi.org/10.31234/osf.io/7uyn5>
- Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, 43(3), 666-678.
- Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, 7(6), 543-554.
- Beerdson E. (in press). Litigation Science after the Knowledge Crisis. *Cornell Law Review*.
- Bem, D. J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of personality and social psychology*, 100(3), 407.
- Bishop, D. (2019). Rein in the four horsemen of irreproducibility. *Nature*, 568(7753).
- Braga, A. A., Papachristos, A. V., & Hureau, D. M. (2014). The effects of hot spots policing on crime: An updated systematic review and meta-analysis. *Justice Quarterly*, 31(4), 633-663.
- Braga, A. A., Weisburd, D., & Turchan, B. (2018). Focused deterrence strategies and crime control: An updated systematic review and meta-analysis of the empirical evidence. *Criminology & Public Policy*, 17(1), 205-250.
- Brodeur, A., Cook, N., & Heyes, A. (2020). Methods matter: P-hacking and publication bias in causal analysis in economics. *American Economic Review*, 110(11), 3634-60.
- Burt C. (2020). Doing Better Science: Improving Review & Publication Protocols to Enhance the Quality of Criminological Evidence. *The Criminologist*, 45(4), 1-6.
- Cairo, A. H., Green, J. D., Forsyth, D. R., Behler, A. M. C., & Raldiris, T. L. (2020). Gray (literature) matters: Evidence of selective hypothesis reporting in social psychological research. *Personality and Social Psychology Bulletin*, 46(9), 1344-1362.

- Camerer, C. F., Dreber, A., Forsell, E., Ho, T., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), 1433-1436.
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T. ... Wu, H. (2018). Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015. *Nature Human Behavior*, 2, 637–644.
- Carney, D. My position on “Power Poses”.  
[https://faculty.haas.berkeley.edu/dana\\_carney/pdf\\_my%20position%20on%20power%20poses.pdf](https://faculty.haas.berkeley.edu/dana_carney/pdf_my%20position%20on%20power%20poses.pdf).
- Chin, J. M. (2018). *Abbey Road: The (ongoing) journey to reliable expert evidence*. *The Canadian Bar Review*, 96(3), 422-459.
- Chin, J. M., Grown, B., & Mellor, D. T. (2019). Improving expert evidence: the role of open science and transparency. *Ottawa Law Review*, 50, 365-410.
- Christensen, G., Wang, Z., Paluck, E. L., Swanson, N., Birke, D. J., Miguel, E., & Littman, R. (2019, October 18). Open Science Practices are on the Rise: The State of Social Science (3S) Survey. <https://doi.org/10.31222/osf.io/5rksu>
- Dahlgaard, J. O., Hansen, J. H., Hansen, K. M., & Bhatti, Y. (2019). Bias in Self-reported Voting and How it Distorts Turnout Models: Disentangling Nonresponse Bias and Overreporting Among Danish Voters. *Political Analysis*, 27(4), 590-598.
- de Bruin, A., Treccani, B., & Sala, S. D. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, 26(1), 90-107.
- DeJong, C., & St. George, S. (2018). Measuring journal prestige in criminal justice and criminology. *Journal of Criminal Justice Education*, 29(2), 290-309.
- Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, H. L., Brown, E. R., Budiman, N. I., Cairoj, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., ... Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, 67, 68-82.
- Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3), 891-904.

- Fidler, F., & Wilcox, J. (2018). Reproducibility of Scientific Results. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.
- Franco, A., Malhotra, N., & Simonovits, G. (2015). Underreporting in political science survey experiments: Comparing questionnaires to published results. *Political Analysis*, 23, 306-312.
- Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fiddler, F. (2018). Questionable research practices in ecology and evolution. *PLoS One*, 13(7): e0200303.
- Gelman, A., & Loken, E. (2014). The statistical crisis in science: data-dependent analysis--a "garden of forking paths"--explains why many statistically significant comparisons don't hold up. *American scientist*, 102(6), 460-466.
- Gelman, A., Skardhamar, T., & Aaltonen, M. (2020). Type M error might explain Weisburd's paradox. *Journal of Quantitative Criminology*, 36(2), 395-604.
- Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsson, G., Banks, G. C., Kidwell, M. C., Mohr, A. H., Clayton, E., Yoon, E. J., Tessler, M. H., Lenne, R. L., Altman, S., Long, B., & Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society open science*, 5(8), 180448.
- Hopp, C. & Hoover, G. A. (2017). How prevalent is academic misconduct in management research? *Journal of Business Research*, 80, 73-81.  
<https://doi.org/10.1016/j.jbusres.2017.07.003>
- Horbach, S. P., & Halffman, W. (2020). Journal peer review and editorial evaluation: Cautious innovator or sleepy giant?. *Minerva*, 58(2), 139-161.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524-532.
- Keeter, S., Hatley, N., Kennedy, C., & Lau, A. (2017). *What Low Response Rates Mean for Telephone Surveys*. Pew Research Center. Retrieved from:  
<https://www.pewresearch.org/methods/2017/05/15/what-low-response-rates-mean-for-telephone-surveys/>
- Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS biology*, 14(5), e1002456.

- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams Jr, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, S., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E., D., Bocian, K., Brandt, M. J., Busching, R., ... Nosek, B. A. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, 1(4), 443-490.
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M. ... Nosek, B. A. (2014). Investigating variation in replicability. *Social psychology*, 45(3), 142-152.
- Krosnick, J. A., Presser, S., Fealing, K. H., & Ruggles, S. (2015). *The Future of Survey Research: Challenges and Opportunities*. The National Science Foundation Advisory Committee for the Social, Behavioral and Economic Sciences Subcommittee on Advancing SBE Survey Research. Available online at: [http://www.nsf.gov/sbe/AC\\_Materials/The\\_Future\\_of\\_Survey\\_Research.pdf](http://www.nsf.gov/sbe/AC_Materials/The_Future_of_Survey_Research.pdf)
- Kvarven, A., Strømmland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, 4, 423-434.
- American Association for the Advancement of Science. (2019). Retraction of the Research Article: "Police Violence and the Health of Black Infants".
- Levine, T., Asada, K. J., & Carpenter, C. (2009). Sample sizes and effect sizes are negatively correlated in meta-analyses: Evidence and implications of a public bias against nonsignificant findings. *Communication Monographs*, 76(3), 286-302.
- Makel, M. C., Hodges, J., Cook, B. G., & Plucker, J. (2019, October 31). Questionable and Open Research Practices in Education Research. <https://edaxiv.org/f7srb/>
- Manski, C.. (2004). Measuring expectations. *Econometrica*, 72, 1329-76.
- McNeeley, S., & Warner, J. J. (2015). Replication in criminology: A necessary practice. *European Journal of Criminology*, 12(5), 581-597.
- Meyer, M. N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science*, 1(1), 131-144.
- Mills, J. L. (1993). Data torturing. *New England Journal of Medicine*, 329, 1196-1199.
- Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual review of psychology*, 69, 511-534.

- Nelson, M. S., Wooditch, A., & Dario, L. M. (2015). Sample size, effect size, and statistical power: A replication study of Weisburd's paradox. *Journal of Experimental Criminology*, 11, 141-163.
- Nuijten, M. B., Hartgerink, C. H., van Assen, M. A., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior research methods*, 48(4), 1205-1226.
- O'Boyle, E. H., Jr., Banks, G. C., & Gonzalez-Mulé, E. (2017). The chrysalis effect: How ugly initial results metamorphize into beautiful articles. *Journal of Management*, 43(2), 376-399.
- Open Science Collaboration. (2015). *Estimating the reproducibility of psychological science*. *Science*, 349(6251),
- Parsons, S., Azevedo, F., & FORRT (2019). Introducing a Framework for Open and Reproducible Research Training (FORRT). <https://osf.io/bnh7p/>
- Pickett, J. T. (2020). The Stewart Retractions: A Quantitative and Qualitative Analysis. *Econ Journal Watch*, 7(1), 152.
- Pickett, J. T., & Baker, T. (2014). The pragmatic American: Empirical reality or methodological artifact? *Criminology*, 52, 195-222.
- Pratt, T. C., Cullen, F. T., Sellers, C. S., Winfree, L. T. Jr., Madensen, T. D., Daigle, L. E., Fearn, N. E., & Gau, J. M. (2010) The empirical status of social learning theory: A meta-analysis. *Justice Quarterly*, 27, 765-802.
- Pridemore, W. Al., Makel, M. C., Plucker, J. A. (2018). Replication in criminology and the social sciences. *Annual Review of Criminology*, 1, 19-38.
- Rabelo, A. L. A., Farias, J. E. M., Sarmet, M. M., Joaquim, T. C. R., Hoersting, R. C., Victorino, L., Modesto, J. G. N., & Pilati, R. (2020). Questionable research practices among Brazilian psychological researchers: Results from a replication study and an international comparison. *International Journal of Psychology*, 55(4), 674-683.
- Ritchie, S. (2020). *Science Fictions: How Fraud, Bias, Negligence, and Hype Undermine the Search for Truth*. New York: Metropolitan Books.
- Rohrer, J. M., Tierney, W., Uhlmann, E. L., DeBruine, L. M., Heyman, T., Jones, B. C., ... Yarkoni, T. (2018, December 12). Putting the Self in Self-Correction: Findings from the Loss-of-Confidence Project. <https://doi.org/10.31234/osf.io/exmb2>
- Rowhani-Farid, A., & Barnett, A. G. (2018). Badges for sharing data and code at Biostatistics: an observational study. *F1000Research*, 7.

- Scheel, A. M., Schijen, M., & Lakens, D. (2020, February 5). An excess of positive results: Comparing the standard Psychology literature with Registered Reports. <https://doi.org/10.31234/osf.io/p6e9c>
- Schumann, S., van der Vegt, I., Gill, P., & Schuurman, B. (2019). Towards Open and Reproducible Terrorism Studies: Current Trends and Next Steps. *Perspectives on Terrorism*, 13(15), 61-73.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359-1366.
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of experimental psychology: General*, 143(2), 534.
- Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society open science*, 3(9), 160384.
- Sorensen, J. R. (2009). An assessment of the relative impact of criminal justice and criminology journals. *Journal of Criminal Justice*, 37(5), 505-511.
- Spellman, B. A. (2015). A short (personal) future history of revolution 2.0. *Perspectives on Psychological Science*, 10(6), 886-899.
- Sweeten, G. (2020). Standard errors in quantitative criminology: Taking stock and looking forward. *Journal of Quantitative Criminology*, 36(2), 263-272.
- Tourangeau, R., Conrad, F. G., & Couper, M P. (2013). *The Science of Web Surveys*. New York: Oxford University Press.
- Uggen, C., & Inderbitzin, M. (2010). Public criminologies. *Criminology & Public Policy*, 9(4), 725-749.
- Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*, 13(4), 411-417.
- Vazire, S., & Holcombe, A. O. (2020, August 13). Where Are The Self-Correcting Mechanisms In Science?. <https://doi.org/10.31234/osf.io/kgqzt>
- Vazire, S., Schiavone, S. R., & Bottesini, J. G. (2020). Credibility Beyond Replicability: Improving the Four Validities in Psychological Science. <https://doi.org/10.31234/osf.io/bu4d3>

- Wagenmakers, E. J., Wetzels, R., Borsboom, D., & Van Der Maas, H. L. (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). *Journal of Personality and Social Psychology*, 100(3), 426-432.
- Weisburd, D., Lum, C. M., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *The Annals of the American Academy of Political and Social Science*, 578, 50-70.
- Welsh, B., Peel, M., Farrington, D., Elffers, H., & Braga, A. (2011). Research design influence on study outcomes in crime and justice: A partial replication with public area surveillance. *Journal of Experimental Criminology*, 7, 183-198.
- White, K. M., Smith, J. R., Terry, D. J., Greenslade, J. H., & McKimmie, B. M. (2009). Social influence in the theory of planned behaviour: The role of descriptive, injunctive, and in-group norms. *British Journal of Social Psychology*, 48, 135-158.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wooditch, A., Sloan, L. B., Wu, X., & Key, A. (2020). Outcome reporting bias in randomized experiments on substance abuse disorders. *Journal of Quantitative Criminology*, 36(2), 273-293.
- Young, J. T. N., Barnes, J. C., Meldrum, R. C., & Weerman, F. W. (2011). Assessing and explaining misperceptions of peer delinquency. *Criminology*, 49(2), 599-630.



**Table 1. Prior Research: Prevalence of QRPs and OSPs in Other Disciplines**

Practice <sup>1</sup>	John et al. (2012) <sup>2</sup>	Psychology <sup>3</sup> Agnoli et al. (2017)	Rabelo et al. (2020)	Ecology Fraser et al. (2018)	Evolution Fraser et al. (2018)	Education Makel et al. (2019)	Communication Bakker et al. (2020)
Omit non-significant studies or variables	46 (485)	40 (217)	55 (232)	—	—	62 (783)	60
Underreport outcomes	63 (486)	48 (219)	22 (232)	64	64	—	—
Underreport conditions	28 (484)	16 (219)	35 (232)	—	—	—	—
Underreport results	—	—	—	—	—	67 (871)	64
Sample selectively	56 (490)	53 (221)	22 (232)	37	51	29 (806)	23
Exclude data selectively	38 (484)	40 (219)	20 (232)	24	24	25 (806)	34
Drop covariates selectively	—	—	—	—	—	42 (773)	46
Switch analysis selectively	—	—	—	—	—	50% (811)	45
HARK	27 (489)	37 (219)	9 (232)	49	54	46 (880)	46
Round p-values	22 (499)	22 (221)	18 (232)	27	18	29 (806)	24
Mislead about demographic effects	3 (499)	3 (223)	4 (232)	—	—	—	—
Hide problems	—	—	—	—	—	24 (889)	—
Hide imputation	1(495)	2 (220)	1 (232)	5	2	10 (898)	9
Preregister study						54 (873)	47
Share data						46 (888)	64
Share code						59 (884)	—
Attempt replication						43 (876)	58
Post article publicly						78 (881)	85

NOTES: Percentage of respondents saying they used the practice at least once (n). OSPs are shaded. <sup>1</sup>The specific questions used varied slightly across studies. See the supplementary materials for all differences (<https://osf.io/wm7aq/>). <sup>2</sup>The estimates from John et al. (2012) are those not using their manipulation designed to glean more candid responses. <sup>3</sup>The studies of psychologists have focused on different countries: John et al. (2012) surveyed US psychologists, Agnoli et al. (2017) surveyed Italian psychologists, and Rabelo et al. (2020) surveyed Brazilian psychologists. For a table that includes our results in criminology, see (<https://osf.io/kj3bf/>).

**Table 2. Questionable Research Practices (QRPs): Question Wording and Self-Reported Use Among Quantitative Criminologists**

QRP	Percentage Using		For Users, Percentage of Studies Using		N
	Estimate	95% CI	Estimate	95% CI <sup>3</sup>	
HARK	29%	25–32%	36%	31–40%	686
	“Reporting an unexpected finding or a result from exploratory analysis as having been predicted from the start.”				
Underreport Results	53%	50–57%	47%	44–51%	677
	“Reporting a set of results as the complete set of analyses when other analyses were also conducted.”				
Hide Problems	10%	8–12%	31%	23–39%	696
	“Not disclosing known problems in the method and analysis, or problems with the data quality, that potentially impact conclusions.”				
Hide Imputation	7%	5–9%	34%	24–44%	683
	“Filling in missing data points without identifying those data as simulated.”				
Omit non-significant studies or variables	43%	40–47%	34%	31–38%	681
	“Not reporting studies or variables that failed to reach statistical significance (e.g. $p < 0.05$ ) or some other desired statistical threshold.”				
Drop Covariates Selectively	32%	28–35%	39%	35–44%	670
	“Not reporting covariates that failed to reach statistical significance (e.g. $p < 0.05$ ) or some other desired statistical threshold.”				
Round P-Values	27%	23–30%	46%	40–51%	692
	“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 $p = 0.01$ ).”				
Exclude Data Selectively	24%	20–27%	34%	29–39%	679
	“Deciding to exclude data points after first checking the impact on statistical significance (e.g. $p < 0.05$ ) or some other desired statistical threshold.”				
Sample Selectively	15%	12–18%	29%	23–34%	680
	“Collecting more data for a study after first inspecting whether the results are statistically significant (e.g. $p < 0.05$ ).”				
Switch Analysis Selectively	39%	35–42%	29%	25–32%	681
	“Changing to another type of statistical analysis after the analysis initially chosen failed to reach statistical significance (e.g. $p < 0.05$ ) or some other desired statistical threshold.”				
Used Any QRP <sup>1</sup>	87%	84–89%	—	—	711
Total QRPs Used (mean) <sup>2</sup>	2.7	2.6–2.9	—	—	579

NOTES: <sup>1</sup>Among Respondents who answered at least three QRP questions. <sup>2</sup>Among Respondents who answered all QRP questions. <sup>3</sup>CI's calculated as  $\pm 1.96*(sd/(\sqrt{n}))$ .

**Table 3. Open Science Practices (OSPs): Question Wording and Self-Reported Use Among Quantitative Criminologists**

OSP	Percentage Using		For Users, Percentage of Studies Using		N
	Estimate	95% CI	Estimate	95% CI <sup>2</sup>	
Preregister Study	45%	42–49%	50%	45–54%	680
	“Preregistering hypotheses and analysis plans prior to data collection.”				
Share Data	43%	40–47%	34%	31–37%	689
	“Sharing data you collected to a publicly accessible, online repository.”				
Share Code	43%	40–47%	32%	29–35%	683
	“Sharing code or other research materials to a publicly accessible, online repository.”				
Attempt Replication	40%	37–44%	21%	18–24%	688
	“Sought to replicate the work of other researchers by following their methods as closely as possible with no intentional changes.”				
Post Article Publicly	68%	65–72%	59%	56–62%	680
	“Posted copies of your research so that it is not behind a paywall (e.g., on a publicly accessible, online preprint server).”				
Used Any OSP <sup>1</sup>	89%	87–91%	—	—	682
Total OSFs Used (mean) <sup>2</sup>	2.4	2.3–2.5	—	—	597

NOTES: <sup>1</sup>Among Respondents who answered at least three OSP questions. <sup>2</sup>Among Respondents who answered all OSP questions. <sup>2</sup>CI<sub>s</sub> calculated as  $\pm 1.96 \times (sd / \sqrt{n})$ .

**Table 4. Correlation Matrix**

Variables	1	2	3	4	5	6	7
1. Career Stage	—						
2. Methods Training	-.04	—					
3. Personal QRP Usage <sup>1</sup>	.04	-.03	—				
4. Personal OSP Usage <sup>1</sup>	.07	.04	.21*	—			
5. Support QRP <sup>1</sup>	.12	-.03	.56*	.09	—		
6. Support OSP <sup>1</sup>	-.09	.07	-.01	.40*	-.15*	—	
7. Perceived QRP <sup>1</sup>	-.17*	.07	.44*	.11	.28*	.18*	—
8. Perceived OSP <sup>1</sup>	-.02	-.01	.18*	.42*	.15*	.10	.22*

NOTES: <sup>1</sup>Variable is a mean index calculated for respondents who answered at least three of the items.

\*p < .05 (two-tailed, Bonferroni-corrected).

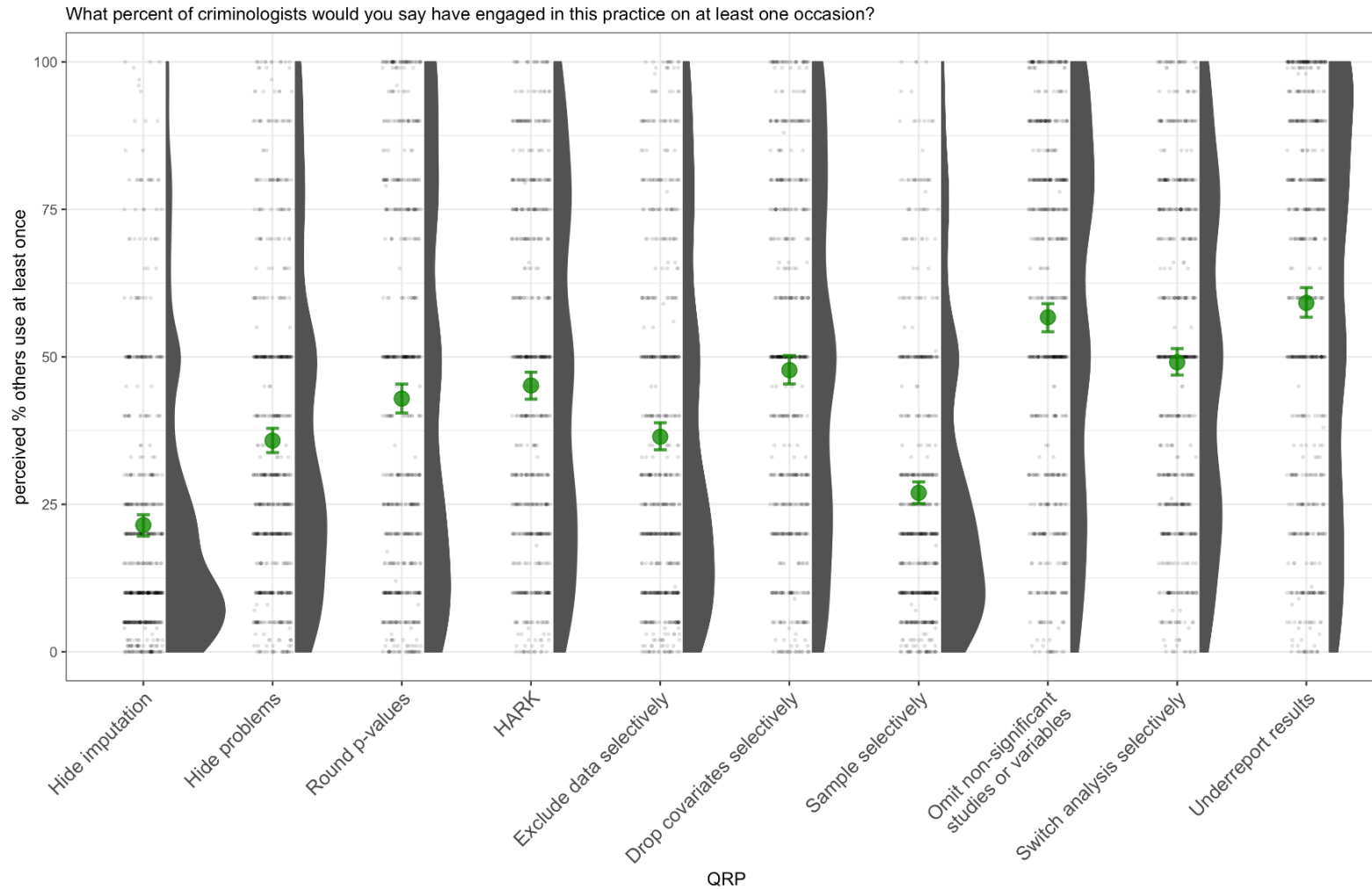
**Table 5. Regressions Predicting Quantitative Criminologists' Perceptions and Use of Questionable Research Practices and Open Science Practices**

Variables	Model 1: Support QRPs		Model 2: Support OSPs	
	<i>b</i>	SE	<i>b</i>	SE
Career Stage	.055**	.018	-.040*	.017
Methods Training	-.005	.006	.011	.007
N	622		614	
Variables	Model 3: Perceived QRPs		Model 4: Perceived OSPs	
	<i>b</i>	SE	<i>b</i>	SE
Career Stage	-3.309***	.800	-.373	.673
Methods Training	.425	.267	-.057	.237
N	585		582	
Variables	Model 5: QRP Usage		Model 6: OSP Usage	
	<i>b</i>	SE	<i>b</i>	SE
Career Stage	.496	.450	1.164	.681
Methods Training	-.123	.164	.228	.232
N	624		615	

NOTES: Models are estimated using ordinary least squares regression with robust standard errors.

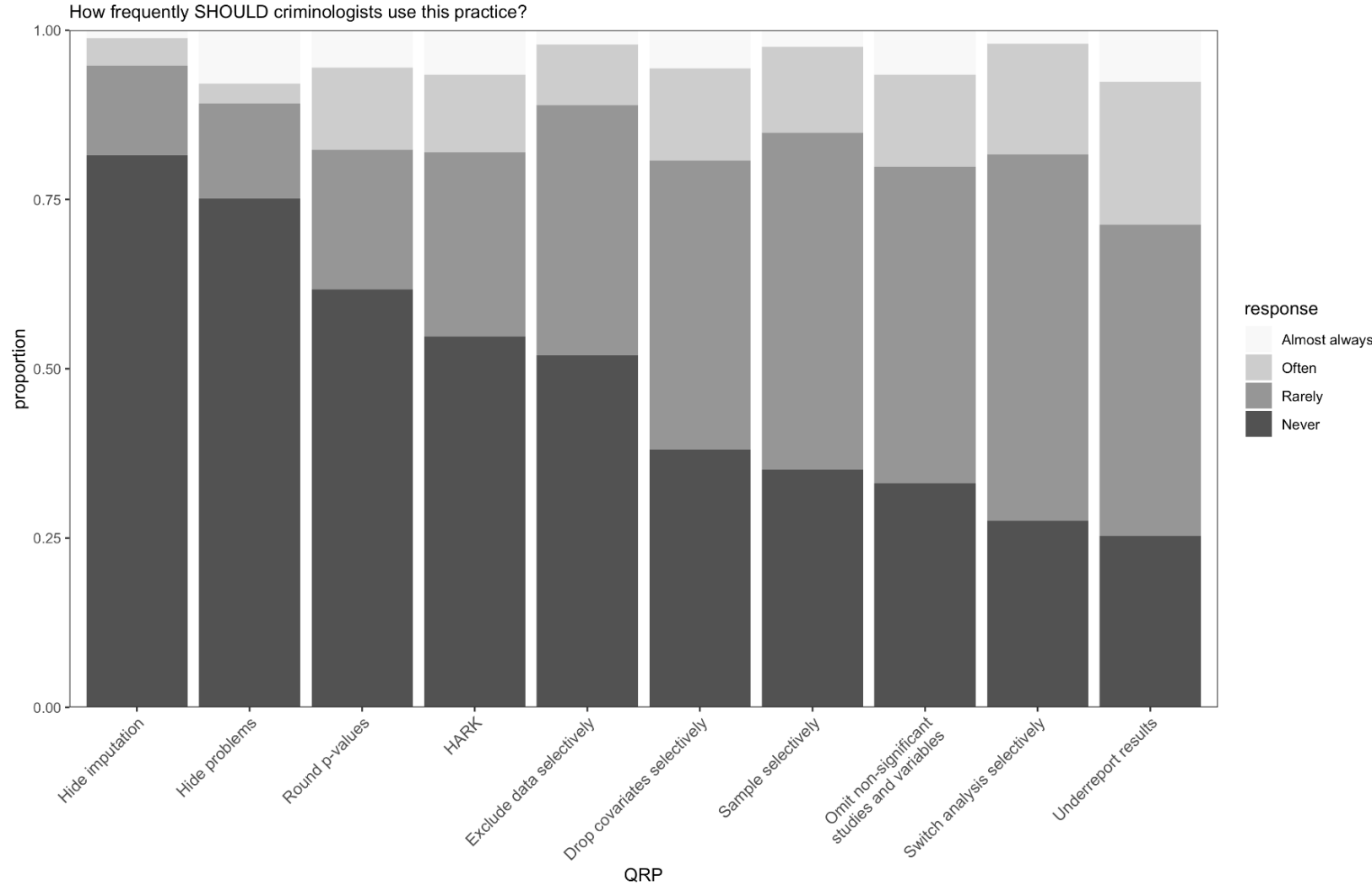
\* $p < .05$ ; \*\* $p < .01$ , \*\*\* $p < .001$  (two-tailed).

**Figure 1. Perceived Prevalence of QRPs**



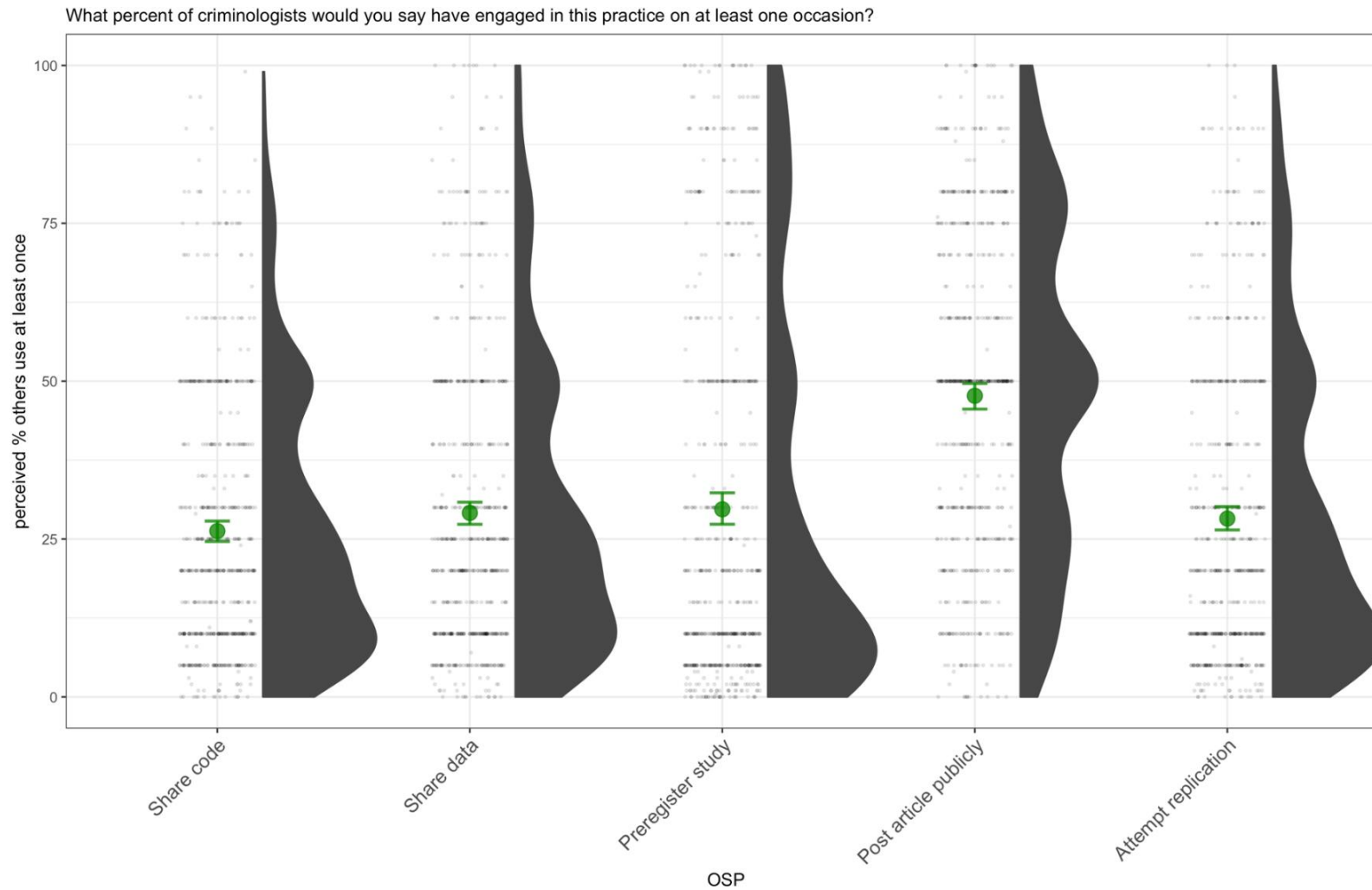
NOTES: Distribution of the perceived % of other researchers using the QRP at least once. Individual responses are plotted in grey, and the density is plotted in dark grey. The mean is plotted in green, with 95% confidence intervals calculated using the percentile bootstrapping method (Efron & Tibshirani, 1994).

Figure 2. Support for QRPs



NOTES: Participants reported whether they thought the stated practice should be used never, rarely, often, or almost always.

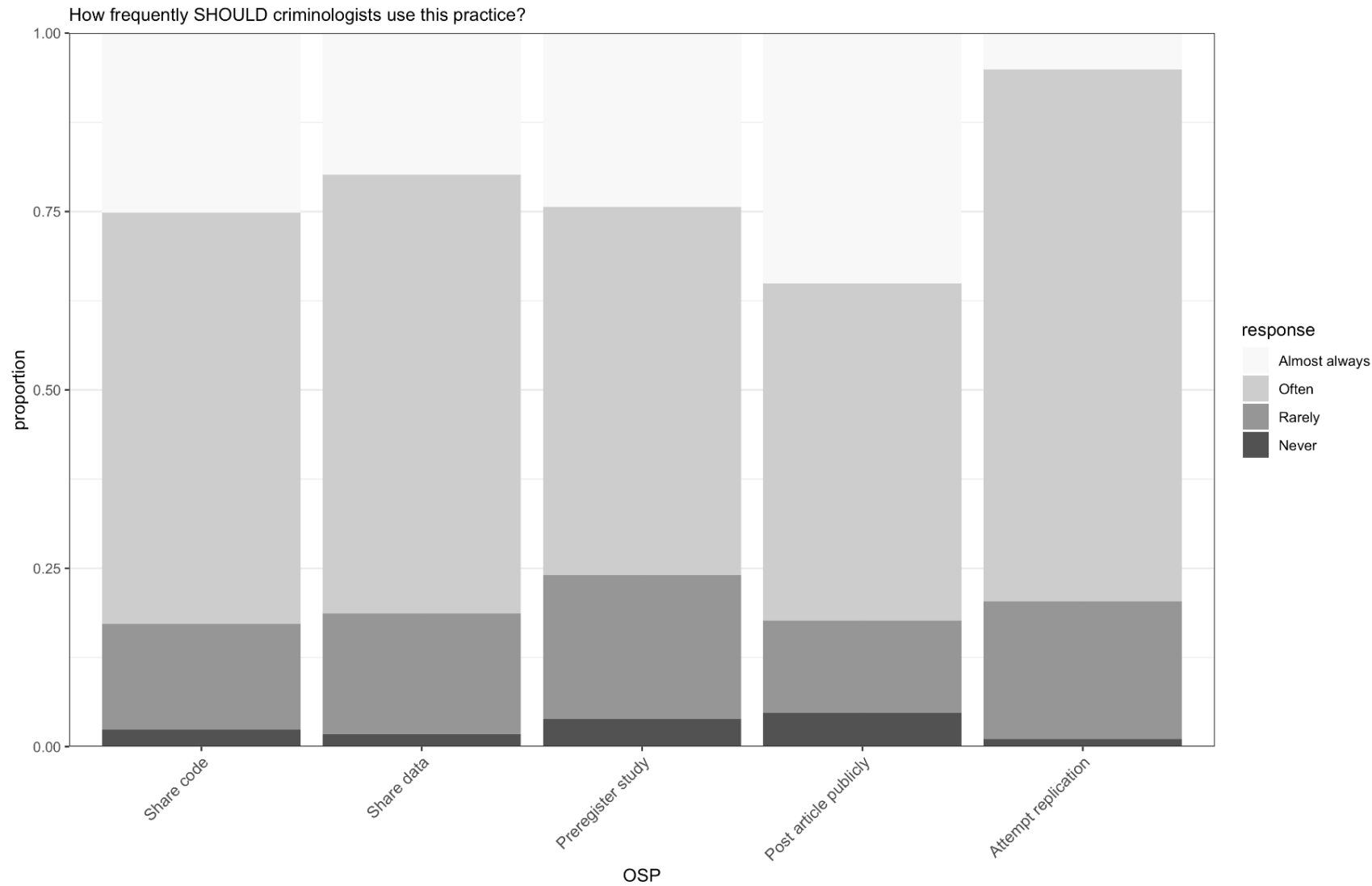
**Figure 3. Perceived Prevalence of OSPs**



NOTES: Distribution of the perceived % of other researchers using the QRP at least once. Individual responses are plotted in grey, and the density is plotted in dark grey. The mean is plotted in green, with 95% confidence intervals calculated using the percentile bootstrapping method (Efron & Tibshirani, 1994).

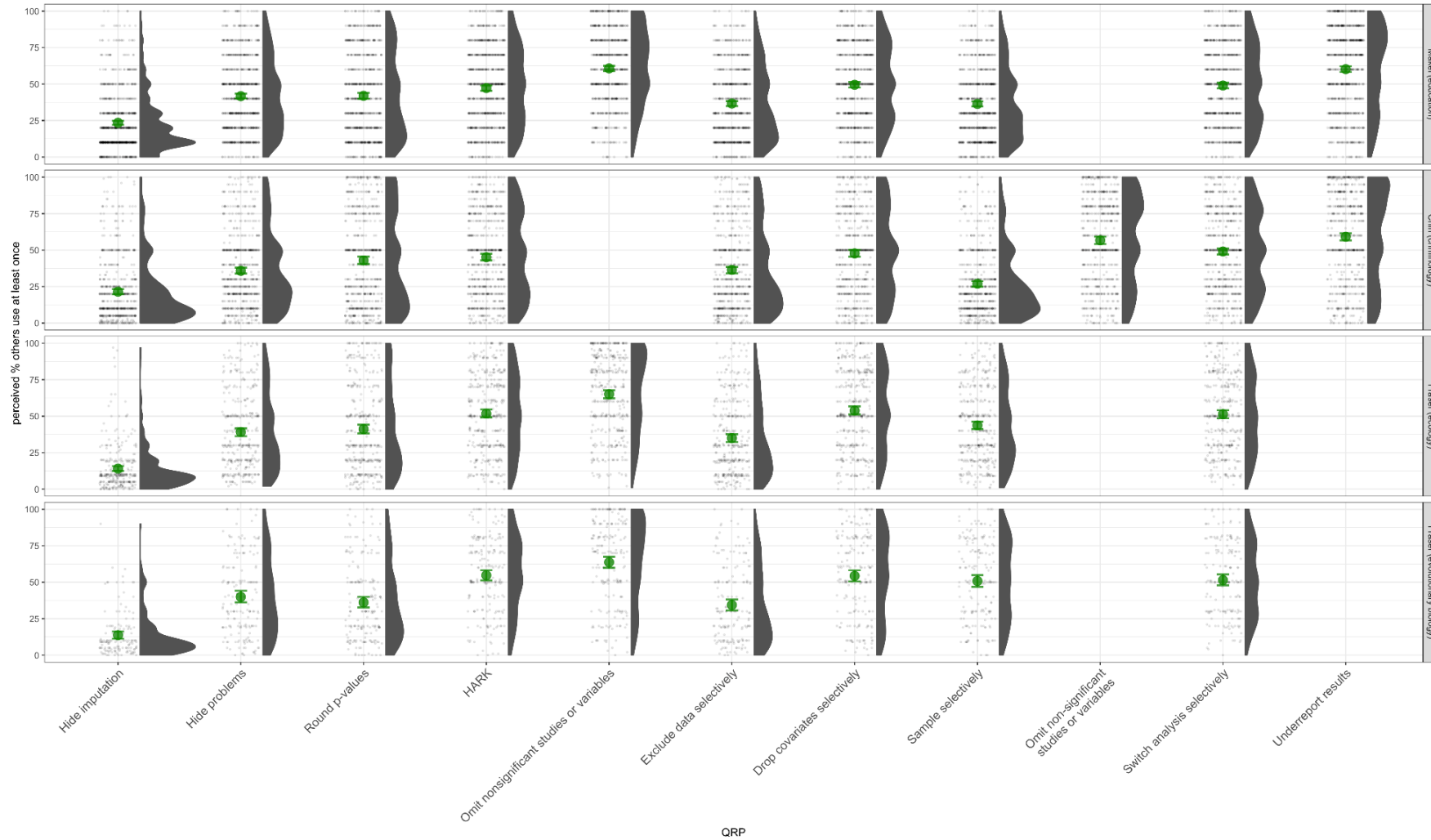


Figure 4. Support for OSPs



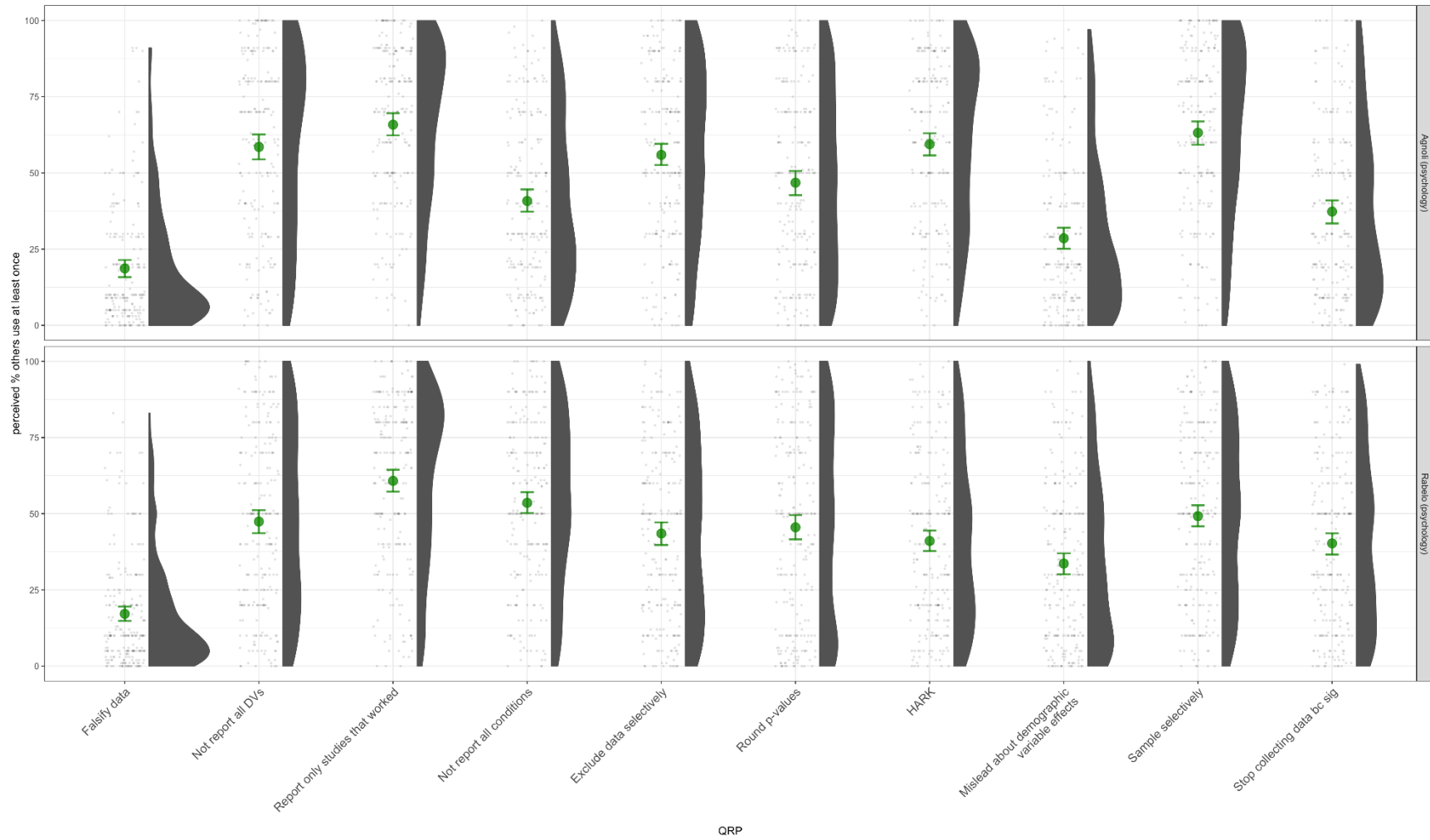
NOTES: Participants reported whether they thought the stated practice should be used never, rarely, often, or almost always.

**Figure 5. Perceived Prevalence of QRPs in the Three Published Studies that used Questions most Similar to the Present Study**



NOTES: Distribution of the perceived % of other researchers in (rows) education (Makel et al. 2020), criminology (current study), ecology, and evolution (Fraser et al., 2018) using the QRP at least once. 95% confidence intervals calculated using the percentile bootstrapping method (Efron & Tibshirani, 1994).

**Figure 6. Perceived Prevalence of QRPs in the Two Published Studies that used Questions most Similar to John et al. (2012)**



NOTES: Distribution of perceived % of other Italian (Agnoli et al., 2017) and Brazilian (Rabelo et al., 2020) psychologists using the QRP at least once. 95% confidence intervals calculated using the percentile bootstrapping method (Efron & Tibshirani, 1994).

## **CRedit Statement**

**Conceptualization:** Jason M. Chin and Justin T. Pickett.

**Data Curation:** Jason M. Chin and Alex O. Holcombe.

**Formal Analysis:** Jason M. Chin and Justin T. Pickett.

**Investigation:** Jason M. Chin.

**Methodology:** Jason M. Chin, Justin T. Pickett, Alex O. Holcombe and Simine Vazire.

**Project Administration:** Jason M. Chin.

**Software:** Jason M. Chin, Justin T. Pickett and Alex O. Holcombe.

**Visualization:** Jason M. Chin, Justin T. Pickett and Alex O. Holcombe.

**Writing - Original Draft Preparation:** Jason M. Chin, Justin T. Pickett, Alex O. Holcombe and Simine Vazire.

**Writing - Review & Editing:** Jason M. Chin, Justin T. Pickett, Alex O. Holcombe and Simine Vazire.