Meta-Psychology, 2022, vol 6, MP.2020.2601
https://doi.org/10.15626/MP.2020.2601
Article type: Original Article

Open data: Yes
Open materials: Yes
Open and reproducible analysis:Yes
Open reviews and editorial process: Yes
Preregistration: Yes

Edited by: Alex O. Holcombe
Reviewed by: Jason M. Chin, Hannah Fraser
Analysis reproduced by: Lucija Batinović
All supplementary files can be accessed at OSF:
https://doi.org/10.17605/OSF.IO/HA6KD

Better Understanding the Population Size and Stigmatization of Psychologists Using Questionable Research Practices

Nicholas W. Fox
Rutgers University, New Brunswick, New Jersey, USA

Nathan Honeycutt
Rutgers University, New Brunswick, New Jersey, USA

Lee Jussim
Rutgers University, New Brunswick, New Jersey, USA

Abstract

There has been low confidence in the replicability and reproducibility of published psychological findings. Previous work has demonstrated that a population of psychologists exists that have used questionable research practices (QRPs), or behaviors during data collection, analysis, and publication that can increase the number of false-positive findings in the scientific literature. Across two survey studies, we sought to estimate the current size of the QRP-using population of American psychologists and to identify if this sub-population of scientists is stigmatized. Using a self-report direct estimator, we estimate approximately 18% of American psychologists have used at least one QRP in the past 12 months. We then demonstrate the use of two additional estimators: the unmatched count estimate (an indirect self-report estimator) and the generalized network scale up method (an indirect social network estimator). Additionally, attitudes of psychologists towards QRP users, and ego network data collected from self-reported QRP users, suggest that QRP users are a stigmatized sub-population of psychologists. Together, these findings provide insight into how many psychologists are using questionable practices and how they exist in the social environment.

Keywords: Questionable Research Practices, QRPs, Replication Crisis, Social Networks, Stigma, Person Perception

Introduction

It is the psychology researcher's job to generate theories, test hypotheses, collect and interpret data, interpret results, and to publish their findings. This is all done to learn more about the world and how it works. In pursuing these tasks, the researcher has many decisions to make: How many observations will I collect? How will I operationalize my variables? What is my population of interest for this given study? Should I exclude any observations from the final analysis?

Each decision point is a "researcher degree of freedom" (Simmons et al., 2011) with the potential to introduce error and bias. Since there is a high level of ambiguity in academic research, these degrees of freedom can be resolved in a variety of ways. In reviewing how researchers handle outlying observations, Simmons et al. (2011) found different research groups made different decisions on what was most correct. When researchers cleaned their data and removed participants who made responses that were "too fast", some defined this as two standard deviations below the mean response speed, some defined this as any observation smaller than 200 milliseconds, and others removed the fastest 2.5% of observations. None of these definitions are an inherently incorrect interpretation of "too fast", which creates a problem – without clear standards of reporting in place, this type of flexible decision making can blur the lines between what decision is right, what decision produces a desired result, and what decision is most likely to help a finding get published.

There are many "researcher degrees of freedom" that exploit the grey areas of acceptable practice and may bias research findings (John et al., 2012; Wicherts et al., 2016). Some examples include trying different ways to score the chosen primary dependent variable and deciding how to deal with outlying observations in an ad hoc manner. Ten of these types of behaviors have been collectively called "questionable research practices" (QRPs) and have been defined as behaviors during data collection, analysis, and reporting that have the potential to increase false-positive findings in the published scientific literature. For this study, nine of the ten QRPs were considered (Table 1). We did not include "fabricated data" (QRP item 10) as the authors consider this a fraudulent, not questionable, behavior. Not only can QRP use increase the number of false-positive findings (e.g., taking a "non-significant" result and pushing it over a threshold into being "significant"), but using multiple QRPs can also influence the reported effect size of a given finding due to sampling bias and low power (Button et al., 2013). Thus, QRP use can lead to field-wide interpretations of findings that are not warranted by the data.

Prevalence of Questionable Research Practice Users

Consider one of the most basic questions to ask about the current replication crisis in psychology: How many people are contributing to it? John et al. (2012) found 63% of psychologists admitted to publishing work without all the dependent measures included (at some point in their academic career). As articulated by Simmons et al. (2011), this is highly problematic, as increasing the number of dependent variables is correlated with an increase in the probability of finding a significant result. Without reporting all dependent measures, readers are left with a false impression of the research activity underlying the reported findings. This estimate from John et al. (2012) was contested by Fiedler and Schwarz (2016). In their conceptual replication that used differently worded questions, a different conceptualization of "prevalence", and tested a German (as opposed to an American) cohort of psychologists, they found less than 10% prevalence of the same questionable practice (omitting dependent variables). Agnoli et al. (2017) recently replicated the original John et al. (2012) study in an Italian cohort of psychologists and found somewhat higher levels of QRP use (47.9% of respondents had omitted dependent variables, see Table 1). Consequently, there is currently no consensus on the prevalence of QRP use in psychology.

Given these inconsistencies in assessing the prevalence of questionable research practices, the present work seeks to expand on this existing literature in several ways. First, we investigate current QRP users, operationalized as an "individual" who has used at least one of nine QRPs "in the past 12 months". This is different from the previous literature as it shifts the attention to individuals who perform questionable practices and away from the behavior as a concept.

Second, it addresses the recent use of QRPs by defining behaviors performed within a specified time period of one year. Previous work estimating QRP prevalence has done so by either estimating lifetime prevalence or via estimating frequency of QRP use, both providing limited insight on recent use of questionable research practices. Put another way, knowing whether a researcher has used a QRP at some point during their career does not tell us much about how many researchers currently use QRPs, nor does it provide an accurate estimate of the size of the current QRP-using population.

A third unique contribution of the present research is that it addresses prevalence of QRP users with three different estimating methodologies. One is a direct estimate heavily based on prior research (Agnoli et al., 2017; Fiedler and Schwarz, 2016; John et al., 2012). We directly ask researchers if they had used any of the 9 behaviors assessed in Table 1 at least once in the past 12 months.

Previous work to estimate the prevalence of QRP use in psychology has also relied on direct self-report of behavior from participants. It is well known that asking participants to self-report on their socially undesirable behaviors can lead to an underestimation, as participants can lie about their behaviors to researchers to avoid potential negative consequences of their actions (Fisher, 1993; Holbrook and Krosnick, 2010; Salganik and Heckathorn, 2004). Even when survey responses

Table 1
The 10 behaviors commonly described as "questionable research practices", including previous estimates of the prevalence of these behaviors across participants' careers from John et al. (2012) & Agnoli et al. (2017). Items 1-9 are used in the present work, as item 10, falsifying data, is fraudulent behavior rather than questionable.

	Questionable Research Practice	**John et al. (2012) prevalence estimate (US sample, control group)**	**Agnoli et al. (2017) prevalence estimate (Italian sample)**
1	Failing to report all of a study's dependent measures	63.4%	47.9%
2	Collecting more data after looking to see if the results were significant	55.9%	53.2%
3	Failing to report all of a study's conditions	27.7%	16.4%
4	Stopping data collection earlier than planned because one found the result one was looking for	15.6%	10.4%
5	Rounding off p-values to achieve significance	22.0%	22.2%
6	Selectively reporting studies that "worked"	45.8%	40.1%
7	Deciding whether to exclude observations after seeing the effect of doing so on the results	38.2%	39.7%
8	Reporting unexpected findings as being predicted from the start	27.0%	37.4%
9	Reporting results are unaffected by demographics when actually unsure or not tested	3.0%	3.1%
10	Falsifying data	0.6%	2.3%

are completely anonymous, many participants may feel pressure to respond in the socially desirable way (Makimoto et al., 2001). For this reason, we felt it was important to attempt to address this known bias by using two different indirect methods of estimation, in addition to the self-report estimate.

The first indirect method, called the unmatched count technique, is an estimating technique aimed at reducing social desirability response bias in self-reports (Arentoft et al., 2016) (see Method for details). The second method generates an indirect estimate of the population size of QRP users by using social network information from the general population of psychologists (Jing et al., 2014; McCormick et al., 2010; Salganik et al., 2011; Zheng et al., 2006), circumventing the need for a participant to report on their own behavior entirely. Neither the unmatched count technique nor this social network method require participants to identify as belonging to a potentially stigmatized group (i.e., QRP users), thereby reducing the risk of socially desirable response bias compared to more traditional direct estimates. While network methods are expected to provide insights into QRP use prevalence, they have yet to be used in psychology. Thus, this work produced three estimates of QRP use prevalence.

Stigmatization of Questionable Research Practice Users

The term "stigma" was formally described by Erving Goffman as "an attribute that makes [a person] different from others in a category of persons available for [them] to be, and of a less desirable kind" (Goffman, 1963). Goffman describes two states of a stigmatized identity: "discredited", where the stigmatizing attribute is outwardly identifiable to strangers (i.e., race, gender, physical handicap – sometimes referred to as "spoiled identities"), and "discreditable", where the stigmatizing attribute can be concealed from others (i.e., sexual orientation, medical condition, certain mental disorders, behaviors). Since discredited people suffer from a reduced social status, it is potentially beneficial for discreditable people to conceal their stigmatized attribute and to continue being considered "normal" (Goffman, 1963). This is controlled through the process of impression management, where the actor (a person with a concealable stigma) communicates with an audience (others in a social group unaware of the actor's "true" identity) in a manner to convince the audience of the appropriateness of their assumed role in society (Goffman, 1959).

Reactions towards stigmatized members of society can differ depending on the perceived controllability the stigmatized individual has over their stigma. For example, people with lung cancer tend to be blamed more for their condition compared to other cancer patients due to the link between cigarette smoking (a controllable behavior) and lung cancer (Chapple et al., 2004). This effect persists even if the individual with lung cancer never smoked. Corrigan (2000) describes differing affective responses by population members towards stigmatized individuals depending on whether or not that person is responsible for their stigma. Those seen as responsible are met with anger and potential punishment, while those seen as not responsible are met with pity and potential helping behaviors. QRP use could be framed as either externally or internally attributed. One could argue that QRP use is an inevitable outcome of working in a stressful academic career where success is measured in scientific output (here, QRP use is externally attributed to stress). It could also be argued that QRPs are only used by those unfit to be academics and result to using QRPs to make up for their own inadequacies (here, QRP use is internally attributed to low ability).

There are ways that stigmatized individuals may attempt to manage their identity while minimizing negative effects. One way is through social withdrawal. By interacting with fewer people, there are fewer moments when a concealed identity can be revealed (Ilic et al., 2014). Another way is through selective disclosure of their stigmatized identity. Selective disclosure to trusted others (often those who share this concealed identity) is an adaptive identity management strategy – it allows the stigmatized individual to control their social interactions in a beneficial way and reduces stigmatizing experiences. Social withdrawal, on the other hand, depends more from the individual by asking them to continuously monitor their social network and anticipate their potential social interactions. This additional burden results in worse mental health outcomes and no reduction in stigmatizing experiences (Ilic et al., 2014).

Considering the potential stigmatization of QRP users is important: determining if QRP users are stigmatized will enable the development of interventions that either decrease or increase stigmatization. It is generally accepted that increased stigmatization of tobacco smokers has decreased the number of people who smoke (Bayer, 2008), though it is unclear whether the group or the individual should bear more of the stigma burden (Courtwright, 2013). For these reasons, it is important to first understand how QRP users exist within their social environment prior to implementing interventions aimed to reduce QRP use.

To assess whether QRP use is stigmatizing, we attempt to measure stigma in two ways. First, we assess the attitudes held by the general population of American psychologists towards QRP users, focusing on four theoretical domains: attribution theory and stigma, social norms and stigma, fear and stigma, and power and stigma (Stuber et al., 2008). These domains are important for understanding if QRP use is stigmatized by psychologists. For instance, population members may fear that QRP users will damage the reputation of psychology as a scientific field and thus look down on those who they perceive to be negative contributors. Additionally, Link and Phalen (2001) argue that individuals who are stigmatized must have less power than those doing the stigmatizing, which is investigated in this study.

In addition to measuring the attitudes of the general population of psychologists towards QRP users, this study also measures social withdrawal and selective disclosure behaviors of self-identified QRP users. By using this two-pronged approach, this study attempts to answer the following research questions:

1. Are QRP users stigmatized by the general population of psychologists?

2. Do QRP users behave as a stigmatized group?

Better understanding the size of the QRP-using population of psychologists, and how psychologists view their peers using QRPs, will set a foundation for future interventions aimed at reducing QRP use.

Study 1: Sizing the QRP-Using Population of Psychologists

Methods

Preregistration statement. This study, which describes three estimates of QRP prevalence in the US psychologist population, was preregistered on May 15, 2017. The preregistration is available here: https://osf.io/xu25n.

Population of interest and target group. The population of interest for this work was all tenured or tenure-track researchers associated with a PhD-granting psychology department in the United States. QRP users (the target group) are therefore a subgroup of this population, with a size greater than zero and maximally the size of the population of interest.

A complete list of names and contact information for the population of interest was obtained via private correspondence with Dr. Leslie John (John et al., 2012). The list provided was current as of 2010, so name and email contact data was updated in May, 2017 as this research program was beginning. This was done by reviewing the faculty at each PhD-granting psychology department in the United States and then adding or removing individuals as appropriate.

Survey distribution. Members of the population of interest were invited via email to participant in a brief survey on personal social network size and attitudes towards researchers. All invitations were sent and all surveys were administered using the Qualtrics web tool (Qualtrics, 2005). All members of the population of interest (N = 7,101) were solicited via email to participate. Emails were sent in 10 waves, with each wave consisting of 200-400 invitations. All initial emails were sent to potential participants on a Thursday, and a single follow-up "reminder" email was sent on the following Monday. Participants who had finished the survey were sent a "thank you" email on the Thursday following the initial solicitation. All invitations were sent between September 2017 and December 2017.

Three surveys were distributed simultaneously. This was to facilitate the different types of direct and indirect estimates that will be described in the following sections. Surveys 1 and 2 were each distributed to 1,775 members of the population of interest. Survey 3 was distributed to 3,551 members of the population of interest. Survey 3 included the self-report direct estimator. To maximize the number of self-reported QRP users observed that would then receive additional questions about their social networks, we distributed Survey 3 to half of the total 7,101 population members and split the remaining half between Surveys 1 and 2. All surveys included relevant instructions and definitions (i.e., defining behaviors identified as QRPs). See https://osf.io/2zwqf/ for the survey materials distributed as well as supplemental materials which describes the deviations from the preregistration.

In these surveys, "QRP use" was defined as having used at least one of the nine QRPs in Table 1 in the past 12 months. Similarly, a "QRP user" was defined as a person who has used at least one of the nine items in Table 1 (excluding item 10 for reasons described previously) in the past 12 months. Participants were presented the definition of QRP use at the start of the survey and the definition was always available by hovering over text in the survey by using their computer mouse.

Survey responses. Of the 7,101 email solicitations sent, 214 emails bounced (3%). Six hundred thirteen full responses were collected (9% full response rate), and 296 partial responses were collected. There was no compensation offered for participation. Only full responses were used in the generation of population size estimates. Additionally, 26 participant responses were removed for either being marked complete erroneously by the Qualtrics webtool, or due to breaking estimate-specific criteria. For example, if a respondent claimed to know 290 individuals who have used a QRP in the past 12 months, yet the estimate of the size of their total social network was only 150 individuals, that respondent would be excluded from analysis. Two hundred ninety nine (49%) participants identified as female, 279 (46%) identified as male, and 19 (3%) chose not to identify their gender. One hundred thirty one (21%) participants identified as Assistant Professor, 141 (23%) identified as Associate Professor, and 208 (34%) identified as Full Professor. One hundred thirteen participants chose not to disclose their tenure level.

Estimating Methods

Estimate 1: direct estimate. The self-report direct estimate involved asking members of the target population whether they have used at least one QRP in the past 12 months, and was calculated as the number of respondents who self-identified as using at least one QRP divided by the total number of respondents.

Estimate 2: unmatched count technique estimate. The unmatched count technique (UCT) is an indirect way of measuring base rates of concealable and potentially stigmatized identities (Gervais and Najle, 2017). In this estimate, two groups of participants are given a list of innocuous items that could apply to them (e.g., I own a dishwasher; I exercise regularly). The list of items for both groups is the same except for one additional item that one group receives and the other does not. This extra item asks about the concealable identity (e.g., I own a dishwasher; I exercise regularly; I smoke crack cocaine – examples from Gervais and Najle, 2017). See Table 2 for the full list of items used. Participants are asked to count and report the number of items in the list that apply to them. At no point does a participant identify which items they are counting. The proportion of participants that identify with the extra item is calculated as the mean difference between the innocuous and concealable identity lists.

Estimate 3: generalized network scale up estimate. Network methods estimate population sizes using information about the personal networks of respondents, based on the assumption that personal networks are, on average, representative of the population (Salganik et al., 2011). Each participant's social network provides a sample of the general population, and by collecting network data on many participants, those accumulated social networks provide access to the larger population. Participants were asked about how many psychologists they "know" in the population of interest. In this study, "know" was defined as: the person knows you by face or by name, you know them by face or by name, you could contact the person if you wanted to, and you've been in contact with them in the past two years (Bernard et al., 2010). Participants were then asked a series of questions to estimate the total size

Table 2
Items used in the Unmatched Count Technique (UCT). Items 1-9 were included on both lists, while only item 10 was used in the "sensitive item" list (list 2).

	Item	List
1	I am a vegetarian.	1 & 2
2	I own a dog.	1 & 2
3	I work on a computer nearly every day.	1 & 2
4	I have a dishwasher in my kitchen.	1 & 2
5	I can drive a motorcycle.	1 & 2
6	My job allows me to work from home at least once a week.	1 & 2
7	I jog at least four times a week.	1 & 2
8	I enjoy modern art.	1 & 2
9	I have attended a professional soccer match.	1 & 2
10	I have used at least one QRP in the past 12 months.	2 only

of their social network, and the number of people they know who have used at least one QRP in the past 12 months. Together, the network scale-up can be used to estimate the proportion of QRP users, and was calculated as follows:

$$\rho = \frac{\sum y_i}{\sum d_i} \tag{1}$$

where ρ is the proportion estimate of people who have used at least one QRP in the past 12 months, y_i is the number of people known in the target group y by participant i, and d_i is the estimated total number of people known d by participant i within the population of interest (see Killworth et al., 1998 for more on estimating d). This equation makes two assumptions: that members of the population of interest know all identity information about all members of their ego networks, and that QRP users have the same size social networks as the general population of interest.

Since QRP use is concealable and potentially stigmatizing, the assumptions made for the previous estimate may not be appropriate. For that reason, data was collected from self-identifying QRP users to estimate how QRP-use identity information transmits through ego networks. This estimate is called the transmission rate, or tau (τ), and estimates the social transmissibility of a person's identity information. This data was collected using the game of contacts method (Salganik et al., 2012), described below.

To estimate the QRP use identity transmission rate tau, we performed the game of contacts with participants who self-identified as using at least one QRP in the past 12 months. Briefly, this method has participants answer a set of questions about what they know about the QRP use of several others (called "alters") in their social network, and what those alters know about the participant's QRP use. The questions are semi-graphical and responses are recorded on a digital 2x2 grid, representing the four possible ways information can flow through a given ego-alter relationship (both the participant and the alter know of each others' QRP use, the alter knows of the participant's QRP use only, the participant knows of the alter's QRP use only, or neither the participant nor the alter have insight on the QRP behaviors of the other). The transmission rate τ is then calculated as:

$$\tau = \frac{\sum w^i}{\sum x^i} \tag{2}$$

where w^i is the number of alters that know the ego is a member of the target group, and x^i is the total number of alters generated by the ego. This produces a value between zero and one, where one represents complete transparency of information (all alters are aware of the participant's QRP use) and zero represents the identity being completely hidden from all alters. For a full description of the game of contacts, see Salganik et al. (2012).

The current study utilized a digital distribution of the game of contacts. This method is typically performed in a face-to-face interview setting with the participant (Salganik et al., 2012). Due to the distributed nature of our frame population, this was not feasible. Instead, participants were presented with the game of contacts via Qualtrics (Qualtrics, 2005). These questions were pretested with several academics not within the population of interest for question clarity. A comparison between an in-person and digital game of contacts has been pre-registered by the authors (https://osf.io/yf4xc/) for future study.

Additionally, to relax the assumption of equal social network sizes between the general population of psychologists and QRP users, a popularity ratio (delta, δ) was calculated as:

$$\delta = \frac{dE}{dT} \tag{3}$$

where dE is the average network size for the target group (QRP users), and dT is the average network size for the population of interest (tenure or tenure-track faculty associated with PhD granting psychology departments in the United States).

Together, tau and delta adjust the network scale-up estimate into the generalized network scale-up as follows:

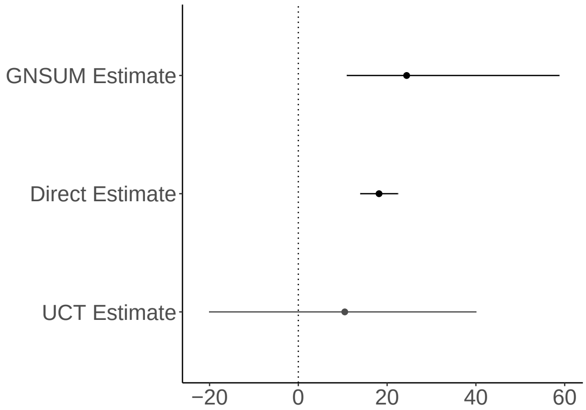

Figure 1. QRP user prevalence estimates (%) using three estimating techniques: the Generalized Network Scale Up Estimate (GNSUM), the Direct Estimate, and the Unmatched Count Technique (UCT). Bars represent 95% percentile bootstrapped confidence interval.

$$\rho = \frac{\sum y_i}{\sum d_i} * \frac{1}{\tau} * \frac{1}{\delta} \tag{4}$$

where ρ is the proportion estimate of people who have used at least one QRP in the past 12 months, $\frac{\sum y_i}{\sum d_i}$ is the network scale-up estimate, τ is the transmission rate, and δ is the popularity ratio. All network scale-up results are calculated using this equation, incorporating τ and δ.

Results

The three estimates of recent QRP use in the frame population of American tenured or tenure-track faculty are summarized in Figure 1 and described in detail below.

Direct estimate. To ensure the highest number of participants in our game of contacts, half of the total population were asked to participate in Survey 3, which contained our direct estimate question. Thus, 3,551 psychologists were solicited, and we received 308 responses able to be analyzed. Of the 308 participants 56 indicated they had used at least one QRP in the past 12 months. We calculated QRP prevalence to be 18.18% (percentile bootstrapped 95% confidence interval [13.96%, 22.40%]).

It is possible this estimate underestimates the true number of psychologists using QRPs. For one, social desirability may lead some scientists who have used QRPs to be unwilling to admit it. This estimate is only generated by those participants willing to reveal their identity as a QRP user. Given the somewhat critical social environment for QRP users (Fiske, 2016; Teixeira da Silva, 2018), it is reasonable to believe some participants withheld their identity when we asked directly. The following indirect estimation methods sought to mitigate this social desirability bias.

Unmatched count technique estimate. The remaining 3,550 psychologists contacted were asked to participate in our unmatched count estimate with 1,775 randomized into the innocuous list condition, and 1,775 randomized into the sensitive list condition. From this, we received 279 responses for analysis.

The average number of list items corresponding to participants in the innocuous list condition was 4.28. The average number of list items corresponding to participants in the sensitive list condition was 4.39. We calculated QRP user prevalence to be 10.46% (percentile bootstrapped 95% confidence interval [-20.19%, 22.40%]).

It was unexpected that the calculated UCT estimate would be lower than our direct estimate. Typically, due to reducing response bias, UCT estimates are larger than direct estimates when the behavior or identity in question is concealable and potentially stigmatized (Gervais and Najle, 2017; Starosta and Earleywine, 2014; Wolter and Laier, 2014). Given the bootstrapped 95% confidence interval crosses zero, it is likely the relatively low number of participants in our UCT (n = 279) led this calculation to be overly sensitive to individual responses.

Upon reviewer suggestion, we calculated the 95% confidence interval using three additional bootstrapping methods: basic, normal, and BCa using the R package 'boot' (Ripley, 2021). These three additional methods produced similar CI ranges (basic = [-19.3%, 41.1%], normal = [-19.4%, 40.1%], BCa = [-19.2%, 41.1%]). Since the UCT estimate is calculated as the mean difference between the two item list means, and because both our sample size and the observed mean difference (0.11) were small, bootstrapping the two item lists and then calculating the UCT estimate can produce replicates where the mean for the innocuous item list is larger than the mean for the concealable identity list, producing a negative population size estimate. The fact that this estimate's confidence interval crosses zero should be indicative that, although the mean difference can be used to generate a point estimate population size, the variability of responses within each list group is sufficient enough to make this estimate uninterpretable.

Generalized network scale up estimate. All participants who were randomized into the UCT estimate were also asked to answer questions about their social networks, and to estimate how many researchers they know who have used at least one QRP in the past 12 months. Participants who were randomized into the direct estimate and who self-identified as a QRP user in that estimate were also asked to answer questions about their social network and to participate in the game of contacts method. Participants in the direct estimate who did not self-identify as a QRP user were asked questions about their social network as well, but were not asked how many researchers they know who have used at least one QRP in the past 12 months. This was because these participants would later be asked about their views on those who use QRPs (see Study 2) and we did not want to prime these participants to think about QRP users in their own social network in an effort to reduce response bias. Therefore, we collected social network responses from 531 participants from the general frame population (to be used in estimating δ, 56 responses from participants who self-identified as QRP users who also completed the game of contacts (to be used in estimating τ and δ), and 279 responses from participants who estimated the number of researchers they know who have used at least one QRP in the past 12 months.

These 279 individuals identified a sum total of 664 QRP users, and know a sum total of 46,828 researchers. Given the total frame population is 7,101 we are fairly confident all or nearly all members were identified at least once by our participants. Using the network scale-up estimate (which does not include tau or delta), this generates an estimate of 1.42% (percentile bootstrapped 95% confidence interval [0.85%, 2.14%]). This estimate assumes QRP use is completely transparent and that all participant's would know the QRP use of the members of their social network. Clearly, this is a poor assumption for this population, but this estimate serves as the base starting point for our key network estimate, the Generalized Network Scale-Up Estimator (GNSUM), detailed below.

The GNSUM relaxes the assumptions of equal network size (delta) and total information transmission (tau) by incorporating these estimates into the equation. Using the 531 responses from the general population and the 56 responses from the participants who indicated using a QRP in the past 12 months and Equation 3, we estimate δ, which is the ratio of average social network sizes between self-identified QRP users and the general population of psychologists, to be 0.97. This means that, on average, the social network size of a self-identified QRP-using psychologist is 97% the size of a psychologist that has not identified as a QRP-user. Using the game of contacts and Equation 2, we estimate τ, which is the transmissibility of QRP use identity information, as 0.06 (percentile bootstrapped 95% confidence interval [0.03, 0.10]). Using Equation 4 to calculate the generalized network scale up estimate, we estimate QRP user prevalence to be 24.40% (percentile bootstrapped 95% confidence interval [10.93%, 58.74%]).

Additional analyses assessed the validity of the network scale-up method in this population by using it to generate estimates of other populations of known size. NSUM estimates (which do not include tau or delta, see Equation 1) were then compared to those actual population sizes. If NSUM estimates correspond well with the actual size of these populations, it would suggest that the GNSUM network scale up method most likely provides a good estimate of population size in this group of participants.

To this end, we generated additional estimates of 24 populations of known size; the number of psychologists with particular first names (the number of psychologists named David, named Janet, etc). The 24 names were gender balanced and represented common, uncommon, and rare names that exist within the census of the population of interest. The size estimates of these populations of known size can be seen in Figure 2 compared to their actual size. The estimates made by our participants of the size of these 24 populations are similar to the actual prevalence of these groups, see Figure 2. The correlation between our participants' estimates of those group sizes and the actual group sizes is r = 0.91.

The NSUM estimate we calculated for the proportion of QRP-using psychologists was 1.42%. Based on the validity estimate, it is possible this NSUM value is an underestimate of the true proportion of psychologists that have used a QRP in the past 12 months. This would result in our GNSUM estimate of 24.4% also being an underestimation. Even the most common first name for our population of interest (David) only had a true prevalence of 2.5%, so understanding the relationship between NSUM-estimated and actual prevalence beyond this value cannot be determined with our data.

We cannot know for certain whether the NSUM and GNSUM estimates accurately identified the true proportion of QRP users in psychology, given we are estimating several variables that can effect the population size estimate. Nonetheless, that using the NSUM with the same participants generated estimates similar to their known values across multiple populations is consistent with the conclusion that our GNSUM estimate may have also generated an estimate similar to the true proportion of QRP users in psychology.

Discussion

Because of inconsistencies in previous research, this study generated three estimates of current QRP use, using three estimating procedures. While the point estimates generated by our three estimators range from

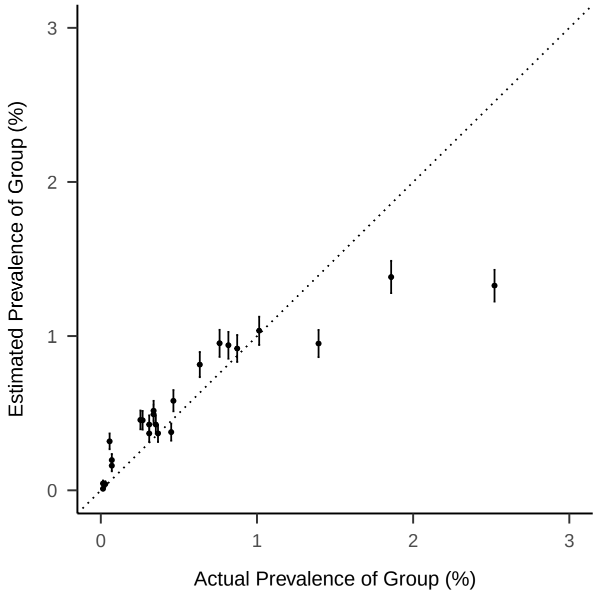

Figure 2. Comparison of estimates made using the GNSUM estimate to the actual prevalence of populations (researchers with specific first names). Dotted line represents when the estimate equals the actual prevalence. Larger groups have a tendency to be underestimated, a phenomenon observed in other published GNSUM estimations (such as Salganik et al., 2011). Correlation between estimated prevalence and actual prevalence r = 0.91.

10.4% to 24.4%, the large confidence intervals generated for both the GNSUM and the UCT estimates make it difficult to make a precise assessment based on these two estimating methods. These large confidence intervals are most likely due to two reasons: first, compared to the direct estimate, both the GNSUM and the UCT estimating equations have more values being estimated (two in the UCT and six in the GNSUM). Second, we observed a high amount of variance, which may be due to the small size of the population of interest (7,101 individuals total) and the low response rate we recorded within this population (8.63%). This in turn effected the precision of our estimates. However, we have more confidence that our direct estimate of 18.18% [13.96%, 22.40%] is an accurate estimation of the proportion of psychologists who have used a QRP in the past 12 months, knowing that it may be an underestimate due to the weakness of self-report measures to response bias.

To the best of our knowledge, this is the first report of the prevalence of QRP users in a proximal timespan. As such, it is difficult to draw conclusions about the magnitude of our estimates compared to previous estimates in the literature. Compared to John et al. (2012), Fraser, Parker, Nakagawa, Barnett, and Fidler (2018), Makel et al., (2019), and Agnoli et al. (2017), we estimate lower rates of questionable research practices. Compared to Fiedler and Schwarz (2016), however, we estimate higher rates of these practices. One often discussed reason for inconsistent QRP use estimates is how QRP behaviors are defined. In this work, we defined questionable research practices using the same language as John et al. (2012) and Agnoli et al. (2017), though restricted use to a timespan of only 15 months (question wording of “in the past 12 months” with data collection lasting 3 months). It should have been expected that our estimates would be lower than some of those reported previously that used an unrestricted timespan of QRP use. Additionally, our estimate may be lower than other reported estimates due to lower usage of QRPs – increased attention to replication failures in psychology may have led to a decrease in these behaviors.

This is also the first report to use the generalized network scale up estimator to investigate the prevalence of QRP users in psychology. Previous use of this estimator in the domains of public health (those most at risk of HIV/AIDS) and oncology (cancer prevalence in Iran) have both shown the usefulness of using social networks to measure hard-to-reach populations (Salganik et al., 2011; Vardanjani et al., 2015). A major strength of this estimating technique is that it can incorporate estimates of information transmissibility, or how available information is to an observer. Direct estimates, on the other hand, rely on an individual’s willingness to participate and their willingness to honestly share their identity to the researcher. Pressure to appear a certain way (social desirability bias) can distort a direct estimate downward.

Social network methods, on the other hand, enable researchers to better understand the social processes at work that produce an environment where members vary in their identity and the identity information they share with others (Zheng et al., 2006). In the process of producing the reported population size estimate for current QRP users, we also report the first estimate of the social transmissibility of QRP-use identity of 0.06 [0.03, 0.10], or 6.02%. This means that only 6% of the population of QRP users is “visible” through the social networks of the general population of psychologists. This estimate suggests that, for each one QRP user a psychologist knows, there are approximately 16 other psychologists in their social network who also are QRP users.

These population size estimates can serve as a baseline to measure the effectiveness of current initiatives, as well as a foundation for new ones. While much work is being done to grow support for interventions such as pre-registration (Wagenmakers and Dutilh, 2016) and Registered Reports (Chambers et al., 2014), it is unknown what quantitative effect these are having on curbing behaviors associated with inflated Type I error such as QRPs. By performing follow-up estimates at future time points, the field can use the baseline estimates presented here to measure the effectiveness of these programs at reducing QRP use.

As noted previously, QRPs exist in a grey area of accepted scientific practice. Therefore, it is difficult to interpret the severity of QRP use. This difficulty, along with the high variability among previous estimates of QRP prevalence, has led to a number of different conclusions. Some have concluded that the problems are overstated (Fanelli, 2018), while others argue QRP use presents a real threat to the viability of several scientific fields, such as education and political science (Bosco et al., 2016). Although our estimates move the field forward in understanding the prevalence of those that use these behaviors, it provides less guidance on the severity of the consequences of QRP use on the whole.

Study 2: Assessing the Stigmatization of QRP-Using Psychologists

Methods

Preregistration statement. This study was not preregistered and should be considered an exploratory assessment of the stigmatization of QRP users within the US psychologist population. Future preregistered studies should be conducted to confirm the relationships described in this study.

Population of interest and target group. The population of interest for this study was all tenured or tenure-track researchers associated with a PhD-granting psychology department in the United States. As this was the same population of interest for Study 1, data was collected for both studies simultaneously.

Survey distribution. Survey material was distributed as described previously (Study 1). In total, 1,775 population members were solicited to participate. Stigma-related survey items were restricted to Survey 1, which did not ask individuals about their own QRP use. One hundred thirty responses were collected from this survey, of which 98 were full responses without missing data. These 98 responses were used for analysis.

Dependent measure. Because there was no existing measures of QRP-related stigma, questionnaire items measuring stigma related to being a QRP user were developed from a scale designed to assess perceived devaluation and discrimination related to smoking cigarettes (Link and Phelan, 2001; Stuber et al., 2008). The measure assesses respondent perceptions of what most other researchers believe. These items were modified to frame them in terms of QRP use. For example, the item “Most people think less of a person who smokes” was modified to “Most people think less of those who use QRPs”. Cronbach’s alpha was calculated to assess the reliability of the items as a scale and alpha = 0.78, suggesting acceptable internal consistency (Tavakol and Dennick, 2011). Responses to each question were on a four-point Likert scale that ranged from “strongly disagree” to strongly agree”. The dependent measure was constructed as the sum of these four item responses, where larger values indicated higher QRP stigma.

Independent measures. The independent measures were: *Age:* Participants self-reported their age in years. *PhD year:* Participants self-reported the year in which they obtained their PhD. Although collected, this measure was not used in subsequent analyses. *Acceptability:* To access descriptive and injunctive social norms at a peer level, one question was asked to participants: “How do most of your colleagues feel about using QRPs? Do they find it acceptable, unacceptable, or that they don’t care one way or another?” The 17 participants who responded “they don’t care one way or another” were excluded from analyses that included this measure due to ambiguity in whether this response indicated positive or negative attitudes about QRP use. *Attribution:* Two items were used to assess what participants believe were the causes of QRP use: “QRP use is due to weak character”, which was used to assess internal attribution, and “QRP use is due to stress”, which was used to assess external attribution. *Fear:* To access fear related to the academic hazards posed by QRP users in their capacity as mentors, one item was: “QRP users are a threat to their students”. *Power:* Socioeconomic status was assessed by tenure level (assistant professor, associate professor, or full professor), and by individual income level (measured with six bins: less than \$49,999, \$50,000 - \$74,999, \$75,000 - \$99,999, \$100,000 - \$149,999, \$150,000 - \$199,999, \$200,000 or more). Although collected, tenure level was not used in subsequent analyses.

Control variables. Racial/ethnic status was assessed by self-identification of categories planned to be used in the 2020 U.S. Census (White, Black or African American, Latino, Hispanic or Spanish Origin, American

Indian or Alaska Native, Asian, Middle Eastern or North African, Native Hawaiian or Other Pacific Islander, None of the Above, or Prefer Not to Say). Political orientation ("politics") was assessed on a 6-point scale (Very Conservative, Somewhat Conservative, Middle-of-the-road, Somewhat Liberal, Very Liberal, and Not Sure). Gender was assessed as either Female, Male, or Prefer Not to Say.

Behavioral measures. To assess behaviors associated with concealing a stigmatized identity, social withdrawal and selective information transmission were measured. The average social network size of QRP users was measured and used in the calculation of the Generalized Network Scale Up Method in Study 1. If QRP users socially withdrawal as an adaptation to living and working with a stigmatized identity, we would predict that their average social network size would be smaller than the average social network size of the general population of psychologists. Selective transmission was assessed by measuring the number of social network alters in each QRP-user's social network who are aware of the QRP-use identity of the participant and assessing which alters are also QRP users. If a QRP user selectively discloses their identity information to in-group members, we predict that another QRP user is more likely to know the QRP-use status of a QRP-using participant compared to a psychologist whose QRP use identity is unknown to the QRP user. In other words, QRP users disclose their QRP-use identity information to other QRP users rather than disclose to individuals with an unknown QRP-use status.

Statistical analyses. For descriptive analyses, responses answered on a four-point Likert scale were reduced to two bins ("agree" and "disagree"). Linear regression was used to assess the direct relationship between independent measures and the dependent measure using the statistical program R (version 4.0.2 – RMarkdown files with full analyses and R packages used are available on our project OSF page: https://osf.io/2zwqf/). A possible curvilinear relationship between power and QRP stigma was tested by introducing the squared power predictor to an additional model. Data points depicted in linear regression graphs were jittered to provide increased clarity. An odds ratio was calculated to determine the odds of a QRP-using alter knowing the participant's QRP-use identity compared to an alter with unknown QRP-use status knowing the participant's QRP-use identity. An independent samples t test was calculated to determine the mean difference between the average social network size of QRP users compared to the average network size of the general psychologist population.

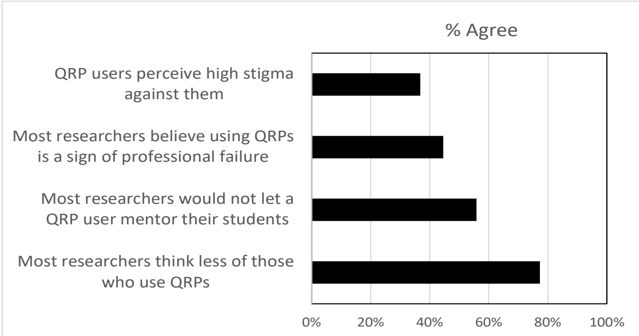

Figure 3. Prevalence of perceived stigma against QRP users among the general population of psychologists. Fewer than half believe QRP users perceive stigma against them, though nearly 80% of respondents believe the researcher community thinks less of QRP users.

Results

Figure 3 shows the prevalence of perceived stigma against QRP users among the general population of psychologists. Participants agreed that "most researchers think less of those that use QRPs" (77.3% of participants agree) and that "most researchers would not let a QRP user mentor their students" (55.8% of participants agree). Furthermore, 44.6% of participants agreed that using QRPs is a sign of professional failure. Interestingly, only 36.7% of participants agreed with the statement that QRP users perceive high stigma against them. It could be argued that the gap between "most researchers think less of those who use QRPs" and "QRP users perceive high stigma against them" speaks to the nature of stigma itself; that it is a negative process established at the environmental level (as opposed to the individual level) by those free of the stigmatizing mark.

Table 3 reports the multiple regression output of all independent variables of interest regressed on the dependent variable. For this analysis, income was used as the operationalization of power, and age (in years) was used as the operationalization of age (as opposed to PhD conferral year) as these were more interpretable variables and have been used in previous literature (Stuber et al., 2008). This model also included the control variables of gender, ethnicity, and political orientation.

In this model, age and fear are both significant predictors of stigmatization of QRP users. Here, younger participants gauged QRP use as significantly more stigmatizing than older participants (p = 0.03), and those who feared QRP users as a threat to their students were significantly more stigmatizing to QRP users (p = 0.0069). As we are interested in whether specific theoretical domains of stigma predict stigma against

Table 3
Multiple regression output of the single model that includes all stigma domains (age, acceptability, external attribution, internal attribution, fear, and power (linear), as well as control variables gender, ethnicity, and political orientation.

Coefficients	**Estimate(β)**	**Estimate(b)**	**SE**	**t-value**	***p*-value**
(Intercept)	—	7.8437	2.4069	3.26	.0016**
Age	-0.23245	-0.0408	0.0185	-2.21	.03*
Acceptability	-0.01056	-0.0494	0.4917	-0.1	.9202
Internal attribution	0.13725	0.4822	0.3803	1.27	.2083
External attribution	-0.00675	-0.022	0.3511	-0.06	.9502
Fear	0.31118	0.9136	0.3299	2.77	.0069**
Power (linear)	0.16157	0.368	0.2328	1.58	.01178
Gender	-0.07151	-0.258	0.4253	-0.61	.5458
Ethnicity	-0.05825	-0.1381	0.2852	-0.48	.6296
Political orientation	0.02853	0.0508	0.1909	0.27	.7908

Table 4
β coefficient outputs of the seven individual regressions run to test domains of stigma. Each model was specified as follows: Stigma DV regressed on domain (age, acceptability, internal attribution, external attribution, fear, power (linear), or power (quadratic)) + gender + ethnicity + political orientation.

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	***p*-value**
Age	-0.22							.031*
Acceptability		0.43						.000***
Internal attribution			0.28					.008***
External attribution				0.15				.160
Fear					0.4			.000***
Power (linear)						0.12		.250
Power (quadratic)							0.27	.660
Gender	0.01	-0.03	-0.09	-0.03	-0.03	-0.05	-0.04	
Ethnicity	-0.15	-0.04	-0.1	-0.12	-0.03	-0.14	-0.15	
Political orientation	-0.07	-0.03	0	-0.06	-0.01	-0.02	-0.02	

QRP users, it was theoretically important to also look at the direct relationships between the predictors in the multiple regression and the QRP stigma outcome (Mela and Kopalle, 2002). Investigating the direct relationships between each theoretical domain of stigma and QRP stigma provides additional insight into whether QRP use satisfies conditions predicted by stigma theory: namely, that QRP use breaks social norms, is internally attributed, is feared, and that QRP users are in a lower position of power compared to the general population of psychologists. Age is an additional predictor that is outside of classic stigma theory, but interesting in this specific context, as QRP use and the resulting scientific reform movement may unequally affect researchers across age (Everett and Earp, 2015).

The results of the seven direct models are reported in Table 4. Age was a significant predictor of stigma, with younger participants holding greater stigmatizing views of QRP users than older participants (β = -0.22, p = 0.031).

Acceptability was dummy coded, where QRP use being acceptable was coded as "0" and QRP use being unacceptable was coded as "1". In the direct model, acceptability of QRP use was a significant predictor of stigma. Those participants who considered QRP use unacceptable held greater stigmatizing views of QRP users than those who considered QRP use acceptable (β = 0.43, p = 0.000).

In the direct model, internal attribution of QRP use was a significant predictor of stigma. Participants who more strongly believed that QRP use was due to a researcher's weak character held greater stigmatizing views of QRP users (β = 0.28, p = 0.008). However, we did not observe a statistically significant effect of external attribution on stigma towards QRP users (p = 0.16).

Fear of QRP users was a significant predictor of stigma. Participants who more strongly believed that QRP users were a threat to their students held a greater stigmatizing view of QRP users (β = 0.40, p = 0.000).[1]

Power was operationalized as individual income, and was modeled both linearly and curvilinearly as it was theoretically plausible that those at the very low and high ends of income in the academic workplace would hold more similar views towards QRP users compared to those at middle incomes. In both the linear and quadratic models, we did not observe a statistically significant difference in stigma predicted by power (p = 0.25 and p = 0.66).

Beyond the bivariate relationships, it is also important to consider the frequency of participant responses. Table 5 reports the prevalence of agreement with the independent measures used in the previous regression models.

Although internal attribution was a significant and positive predictor of stigma (see Table 2), only a small number (24%) of participants agreed that QRP use could be internally attributed. Most participants (66.2%) agreed that QRP use could be externally attributed to stress. Similarly, most participants (75%) agreed that QRP use broke social norms and that QRP use was threatening to students (68.5%).

Stigma-related behaviors. To assess whether self-identified QRP-using psychologists behave in ways predicted by social stigma theory, two behaviors were observed and assessed: social withdrawal and selective information transmission (or selective disclosure).

Social withdrawal. The average professional social network size for the general population of psychologists was 184.93 individuals. The average professional social network size for self-identified QRP-using psychologists was 178.60 individuals. We did not observe a statistically significant difference in social network size, t(72) = -0.2, p = 0.8. See Figure 4 for the kernel density plot of professional social network sizes for all participants.

Selective disclosure. The 56 self-identified QRP users in this work produced a total of 1,230 social network alters from the game of contacts procedure (described in Study 1). One hundred of these alters were considered "in-group" members, meaning these were alters that were identified as QRP users by participants in the study who self-identified as QRP users. In other words, the participants and these alters shared a common "QRP user" identity. The other 1,130 social network alters were out-group members, or psychologists with an unknown QRP-use status by the 56 QRP-using participants described in this work.

Participants, or "egos", were asked for each alter whether or not that person knew of the participant's QRP-user identity status (either "this person knows I have used a QRP in the past 12 months" or "I do not know if this person knows I have used a QRP in the past 12 months"). The counts of these responses are depicted in Figure 5.

As seen in Figure 5, 58 out of 100 in-group alters generated know the ego's QRP-use identity (58%, top left panel). Conversely, when the alter's QRP-use status is unknown to the ego, only 16 out of 1,130 alters generated know of the ego's QRP-use identity (1.44%, top right panel). This results in an odds ratio of 96.14 (95% confidence interval [51.03, 181.14], calculation described in Szumilas 2010), indicating that the odds of an in-group alter knowing the ego's QRP-use status is 96.14 times higher than out-group alters. This provides evidence of selective transmission of QRP-identity status to in-group members over out-group members, a

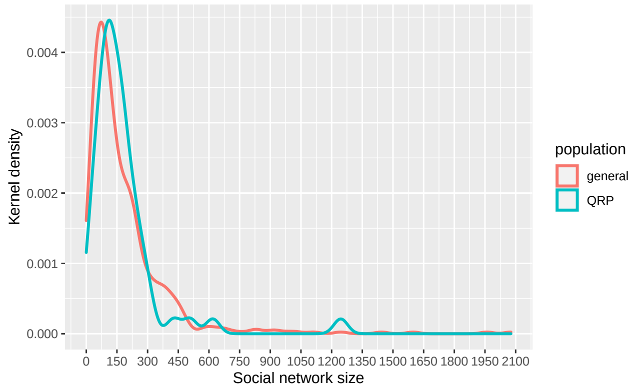

Figure 4. Kernel density plot of social network size for self-identified QRP users and the general population (those who did not self-identify as a QRP user). The social network distributions for these two groups were not significantly different, t(72) = -0.2, p = 0.8. Generated using geom_density() function within the ggplot2 package in R 4.0.2.

[1]Note that one item of the stigma DV, "QRP users are a threat to their students" is similar to the IV item operationalizing the fear component of stigma, "most researchers would not let a QRP user mentor their students". In a post hoc analysis performed during manuscript review, the fear IV item was a significant predictor of 3 of the 4 items in the stigma inventory: "most researchers think less of those who use QRPs", β = 0.44, p < 0.01, "most researchers would not let a QRP user mentor their students", β = 0.37, p < 0.01, and "most researchers believe using QRPs is a sign of professional failure", β = 0.43, p < 0.01. It was not a significant predictor of the fourth item, "QRP users perceive high stigma against them", β = 0.06, p = 0.54. This should provide some additional insight that this relationship is not being driven solely by a similarity between IV and DV item.

Table 5
Percent of participants who agreed or strongly agreed with items in the Stigma dependent measure.

Domain	**Item**	**% Agree**
Acceptability	Most of your colleagues feel using QRPs is unacceptable	75.0%
Fear	QRP users are a threat to their students	68.5%
External Attribution	Most researchers believe using QRPs is due to stress	66.2%
Internal Attribution	Most researchers believe using QRPs is due to weak character	24.0%

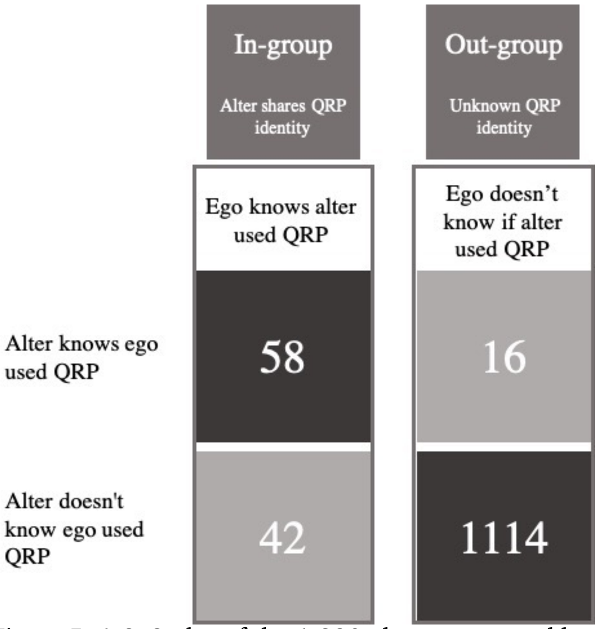

Figure 5. A 2x2 plot of the 1,230 alters generated by the 56 self-identified QRP users in this study. If the participant in our study (the ego) knows the alter is a QRP user, the alter is much more likely to know the QRP-use identity of the ego compared to when the ego does not know the QRP-use behavior of the alter (odds ratio = 96.14 (95% confidence interval [51.03, 181.14])).

behavior also observed in other stigmatized populations (Herman, 1993).

Discussion

This study focused on the relationships between groups of research psychologists and whether QRP-using psychologists were stigmatized by their peers. All analyses except those focused on socioeconomic power (model 6 and 7, Table 4) support the hypothesis that QRP-users are a stigmatized subpopulation of psychologists.

One reviewer was not sympathetic to the prediction made by Link and Phalen's (2001) model of stigma that those with more socioeconomic power would be more stigmatizing in the academic context. We believe that there are some potential reasons why power was not a significant predictor of QRP stigma in this study. It could be that economic power is a poor operationalization of power in the academic social environment. It is possible that the number of published papers, citation count, h-index, or years in a prestigious position could serve as better proxies of power in the academic social setting than income (Bourdieu, 1988). It could also be that there is no difference in power between QRP users and the general population of psychologists. Academia is unlike the typical social environment in some key ways. For instance, success as an academic psychologist has relied more and more on working with others. Collaboration rates in psychology have been rising over the past 90 years (Zafrunnisha and Pullareddy, 2009), and this selective pressure to collaborate may serve as a vehicle for high income and lower income academics to intersect. The academic model is also based on a mentor-mentee relationship, where professors who make an adequate salary often closely work with graduate students, who are either unpaid, paid a modest stipend, or are economically insecure (Ehrenberg and Mavros, 1992). Academia may not support a social environment where those of higher economic power can stigmatize those of lower economic power. It could also be that those in high socioeconomic positions used QRP behaviors to get to that position of power, and thus are in no position to hold stigmatizing attitudes towards other QRP users.

Taken together, the results of these models suggest that QRP-using psychologists are stigmatized by the general population of psychologists. QRP users are seen as breaking social norms and are feared as a threat to their students. When QRP use is internally attributed, stigmatizing attitudes are higher. However, when asked directly, most participants agreed that QRP use was more attributable to external variables (like stress, see Table 5) than internal variables (like weak character).

Beyond just investigating the attitudes of the general population of psychologists on QRP users, this study also directly observed stigma-related behaviors of QRP users themselves. This is a step forward in determining if QRP users are stigmatized because we can ask the question "Do QRP users act like other stigmatized groups?". There were two-stigma-related behaviors observed in this study: social withdrawal and selective information transmission (selective disclosure).

Figure 4 shows the comparison in social network size between QRP users and the general population of psychologists. Although QRP users have a slightly smaller average social network size (178.6 versus the general population of psychologists' average social network size of 184.93), this difference was not statistically significant. Here too, it is possible that the nature of academic psychology inhibits QRP users from socially withdrawing. As mentioned previously, success as a psychologist has relied more and more on collaboration, therefore restricting one's academic social network directly inhibits success. This outcome may also be due to selection bias, where those QRP-using psychologists who had socially withdrawn to protect their stigmatized identity no longer found success in academia and moved onto other careers. Having a sufficiently large social network may be a key factor to success in academic psychology, and shrinking one's social network to protect a concealed identity may reduce academic success, and the possibility of being solicited for this study.

The other stigma-related behavior studied was selective transmission of QRP-use identity. Figure 5 shows the number of people in QRP users' social networks that either do or do not know about that person's QRP-use identity, given that the social network member either is or isn't of known QRP-use status themselves. It suggests that the social transmission of QRP-use identity is dependent on a shared in-group social status. When both members of a social dyad (ego and alter) are QRP users, they are more likely to know that information about each other. When the QRP-use identity of an alter is unknown (they may or may not be a QRP user), the alter is much less likely to know the QRP-using identity of the ego. This is evidence that QRP users selectively disclose their QRP use to other known QRP users.

Revealing is one significant way individuals can manage an invisible social identity (Goffman, 1963). Being stigmatized is harmful, as it can lead to stereotyping, loss of status, and discrimination (Clair et al., 2005; Link and Phelan, 2001). By selectively revealing an invisible stigmatized identity to in-group members (in this case, other QRP users), one can avoid the harmful effects of stigmatization while minimizing the negative consequences of keeping one's identity a secret from others (Garcia and Crocker, 2008; Ilic et al., 2014).

General Discussion

Contributions

The present research makes a number of important contributions. First, it identifies that approximately 20% of American psychologists are recent users of QRPs. This is a large proportion, especially given the fact that the "replication crisis" is already several years old. The current research suggests that even at this time, a non-negligible number of psychologists are using practices in data collection, but especially in preparing scientific reports that can increase the number of false-positives in the published literature. It shows that more work must be done to change researcher behaviors that are beyond the influence of statistical initiatives like lowering the conventional alpha threshold in null hypothesis significance testing (Benjamin et al., 2017). Six of the ten defined QRPs in Table 1 take place during manuscript writing and preparation, meaning an intervention that goes beyond data analysis is needed to impact and reduce these behaviors.

Second, it contributes to the literature on stigma. We use data from both the general population and from the potentially stigmatized population to determine the stigma status of that group. Being able to observe a group collectively manage their stigma, while simultaneously collecting data on the negative attitudes held by the general population towards that group provides us with additional confidence in the conclusion that QRP-using psychologists are indeed a stigmatized population of scientists. That said, as an observational study, causal relationships between stigmatizing attitudes and potential behavioral responses in QRP users cannot be determined here.

Strengths, limitations, and future research

There were numerous strengths to these studies. Rather than relying solely on self-reports, the population sizing was conducted using three different estimators. For this reason, we learned not only about the size of the population, but how these estimates and their confidence interval ranges can vary according to the estimator selected. This is important, especially within the context of attempting to measure a sub-population (QRP-users) of a small population of interest (American psychologists in PhD-granting departments, total N = 7,101). The social network estimator allowed us to estimate the size of the population, but also provided insight into how QRP users share their identity information with others, a critical insight elaborated on in Study 2. While both studies have elements of self-report (in the self-report estimate in Study 1, and investigating the attitudes of the general population of psychologists in Study 2), each study used multiple approaches to minimize potential social desirability biases.

A major limitation of this work was the low response rate we observed (8.63% full response rate). There are a few possible reasons why the response rate was low. First, we did not offer an incentive of any type for participating in this survey. This was due to the fact that the work was unfunded. Another potential reason for this low response rate was that the window to participate was only open for one week following our email solicitation. We also only included participants who completed the entire survey, further reducing our response rate.

The behaviors of researchers have the potential to shift quickly as norms change with the increased adoption of interventions like preregistration and the Registered Reports format of publishing. Future research should continue to estimate the total number of QRP users to help determine if these interventions are having an effect, or if new, different mechanisms are needed. Future work should also start to use the stigma literature to its advantage when considering how to best reduce the use of questionable research practices. By knowing that QRP users are stigmatized, future research could focus on the causal relationships that may exist between social attitudes and QRP users feeling stigmatized. Future interventions could investigate whether decreasing stigma produces an environment that promotes QRP users revealing their identity or reforming their research behaviors, or if increasing stigma on QRP users limits the number of researchers who believe QRPs are acceptable research practices and thus limits the number of new QRP users (Bayer, 2008).

Conclusions

Much work has shown that there are psychologists who use questionable research practices in the course of analyzing their data and preparing their manuscripts that contribute to the inflated false-positive rates in the published literature. The current studies provide an estimate of the size of this population among tenure or tenure-track American research psychologists (18.18% using a direct estimate, though this study showed that both the point estimate and variance surrounding this estimate can depend on the estimator used). These researchers are a stigmatized subgroup of psychologists; members of the general population of psychologists hold negative attitudes towards them in domains consistent with the stigma literature, and they selectively disclose their QRP-using identity to in-group others, or social network members they have identified as like QRP-users themselves. These results suggest that even after several years of a "replication crisis" and a movement towards reform, the field of psychology has much work to do in curbing the use of questionable research practices and shifting its constituent researchers towards reducing the influence of the researcher on the results of the research.

Author Contact

Corresponding author:
Nicholas Fox
institutional email: nwf7@scarletmail.rutgers.edu
permanent email: nfox423@gmail.com
ORCID: 0000-0002-3772-8666

Conflict of Interest and Funding

At the time of manuscript submission, Nicholas Fox worked at the Center for Open Science as a Research Scientist. The Center for Open Science has an interest in seeing research become more transparent and shareable. All work, including data collection and the writing of initial drafts, was performed while Nicholas was a PhD candidate at Rutgers University, working in Lee Jussim's laboratory. This project was not explicitly funded, though access to Qualtrics was provided by Rutgers University to all faculty and students.

Author Contributions

CRediT taxonomy
Study conceptualization: NF
Resources: LJ
Data curation: NF
Software: NF, NH, LJ
Formal analysis: NF
Supervision: LJ
Funding acquisition: N/A
Validation: NF
Investigation: NF
Visualization: NF
Methodology: NF
Writing, original draft: NF
Writing, editing: NF, NH
Project administration: NF, LJ

Nicholas Fox conceptualized the work, carried it out, and developed the first and final drafts of the manuscript, and thus is first author. Lee Jussim supervised the project from beginning to end, provided feedback and guidance as a PhD candidate advisor throughout the work, and provided physical space for conducting this work, and is thus last author. Nathan Honeycutt provided critical feedback during the editing process of this manuscript, including recommending the addition of Study 2, and is thus second author. He also submitted the manuscript to the journal, and for that Nicholas Fox is extremely thankful.

Open Science Practices

This article earned the Preregistration+, Open Data and the Open Materials badge for preregistering the hypothesis and analysis before data collection, and for making the data and materials openly available. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process, including the open reviews, are published in the online supplement.

References

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), 1–17. https://doi.org/10.1371/journal.pone.0172792

Arentoft, A., Van Dyk, K., Thames, A. D., Sayegh, P., Thaler, N., Schonfeld, D., Labrie, J., & Hinkin, C. H. (2016). Comparing the unmatched count technique and direct self-report for sensitive health-risk behaviors in HIV+ adults. *AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV*, *28*(3), 370–375. https://doi.org/10.1080/09540121.2015.1090538

Bayer, R. (2008). Stigma and the ethics of public health: Not can we but should we. *Social Science and Medicine*, *67*(3), 463–472. https://doi.org/10.1016/j.socscimed.2008.03.017

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., & Johnson, V. E. (2017). Redefine Statistical Significance. *PsyArxiv*, (July 22), 1–18. https://doi.org/10.17605/OSF.IO/MKY9J

Bernard, H. R., Hallett, T., Iovita, A., Johnsen, E. C., Lyerla, R., McCarty, C., Mahy, M., Salganik, M. J., Saliuk, T., Scutelniciuc, O., Shelley, G. A., Sirinirund, P., Weir, S., & Stroup, D. F. (2010). Counting hard-to-count populations: The network scale-up method for public health. *Sexually transmitted infections*, *86 Suppl 2*, ii11–5. https://doi.org/10.1136/sti.2010.044446

Bosco, F. A., Aguinis, H., Field, J. G., Pierce, C. A., & Dalton, D. R. (2016). HARKing's Threat to Organizational Research: Evidence From Primary and Meta-Analytic Sources. *Personnel Psychology*, *69*(3), 709–750. https://doi.org/10.1111/peps.12111

Bourdieu, P. (1988). *Homo academicus*. Stanford University Press.

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

Chambers, C., Feredoes, E., D. Muthukumaraswamy, S., & J. Etchells, P. (2014). Instead of "playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neuroscience*, *1*(1), 4–17. https://doi.org/10.3934/Neuroscience.2014.1.4

Chapple, A., Ziebland, S., & Mcpherson, A. (2004). Stigma, shame and blame experienced by patients with lung cancer: Qualitative study. *BMJ*, *online fir*(June), 1–5. https://doi.org/10.1136/bmj.38111.639734.7C

Clair, J. A., Beatty, J. E., & Maclean, T. L. (2005). Out of sight but not out of mind: Managing invisible social identities in the workplace. *Academy of Management Review*, *30*(1), 78–95. https://doi.org/10.5465/AMR.2005.15281431

Corrigan, P. (2000). Mental Health Stigma as Social Attribution : Implications for Research Methods and Attitude Change Mental Health Stigma as Social Attribution : Implications for Research Methods and Attitude Change. *Clinical Psychology Science and Practice*, *7*(1), 48–67. https://doi.org/10.1093/clipsy.7.1.48

Courtwright, A. (2013). Stigmatization and public health ethics. *Bioethics*, *27*(2), 74–80. https://doi.org/10.1111/j.1467-8519.2011.01904.x

Ehrenberg, R., & Mavros, P. (1992). Do Doctoral Students' Financial Support Patterns Affect Their Times-To-Degree and Completion Probabilities.

Everett, J. A. C., & Earp, B. D. (2015). A tragedy of the (academic) commons : Interpreting the replication crisis in psychology as a social dilemma for early-career researchers. *6*(August), 1–4. https://doi.org/10.3389/fpsyg.2015.01152

Fanelli, D. (2018). Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences of the United States of America2*, *in press*, 1–4. https://doi.org/10.1073/pnas.1708272114

Fiedler, K., & Schwarz, N. (2016). Questionable Research Practices Revisited. *Social Psychological and Personality Science*, *7*(1), 45–52. https://doi.org/10.1177/1948550615612150

Fisher, R. J. (1993). Social desirability bias and the vailidity of indirect questioning. *Journal of Consumer Research*, *20*(September 1993), 303–315. https://doi.org/10.1086/209351

Fiske, S. T. (2016). Mob Rule or Wisdom of Crowds. *APS Observer*.
Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLOS ONE*, *13*(7). https://doi.org/https://doi.org/10.1371/journal.pone.0200303
Garcia, J. A., & Crocker, J. (2008). Reasons for disclosing depression matter: The consequences of having egosystem and ecosystem goals. *Social Science and Medicine*, *67*(3), 453–462. https://doi.org/10.1016/j.socscimed.2008.03.016
Gervais, W. M., & Najle, M. B. (2017). How many atheists are there? *Social Psychological and Personality Science*, 1948550617707015–1948550617707015.
Goffman, E. (1959). The Presentation of Self in Everyday Life. *Garden City NY*, *A 174 : So*, 259–259. https://doi.org/10.2307/258197
Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity*. Simon; Schuster.
Herman, N. J. (1993). Return to sender: Reintegrative stigma-management strategies of ex-psychiatric patients [Place: US Publisher: Sage Publications]. *Journal of Contemporary Ethnography*, *22*(3), 295–330. https://doi.org/10.1177/089124193022003002
Holbrook, A. L., & Krosnick, J. A. (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly*, *74*(1), 37–67. https://doi.org/10.1093/poq/nfp065
Ilic, M., Reinecke, J., Bohner, G., R??ttgers, H. O., Beblo, T., Driessen, M., Frommberger, U., & Corrigan, P. W. (2014). Managing a stigmatized identity-evidence from a longitudinal analysis about people with mental illness. *Journal of Applied Social Psychology*, *44*(7), 464–480. https://doi.org/10.1111/jasp.12239
Jing, L., Qu, C., Yu, H., Wang, T., & Cui, Y. (2014). Estimating the sizes of populations at high risk for HIV: A comparison study. *PLoS ONE*, *9*(4), 1–6. https://doi.org/10.1371/journal.pone.0095601
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. https://doi.org/10.1177/0956797611430953
Killworth, P. D., McCarty, C., Bernard, H. R., Shelley, G. a., & Johnsen, E. C. (1998). Estimation of seroprevalence, rape, and homelessness in the United States using a social network approach. *Evaluation review*, *22*(2), 289–308. https://doi.org/10.1177/0193841X9802200205
Link, B. G., & Phelan, J. C. (2001). Conceptualizing Stigma. *Annual Review of Sociology*, *27*, 363–385. https://doi.org/10.1146/annurev.soc.27.1.363
Makel, M., Hodges, J., Cook, B., & Plucker, J. (2019). Questionable and Open Research Practices in Education Research. https://doi.org/10.35542/osf.io/f7srb
Makimoto, K., Iida, Y., Hayashi, M., & Takasaki, F. (2001). Response Bias by Neuroblastoma Screening Participation Status and Social Desirability Bias in an Anonymous Postal Survey, Ishikawa, Japan. *Journal of Epidemiology*, *11*(2), 70–73.
McCormick, T. H., Salganik, M. J., & Zheng, T. (2010). How many people do you know?: Efficiently estimating personal network size. *Journal of the American Statistical Association*, *105*(489), 59–70. https://doi.org/10.1198/jasa.2009.ap08518
Mela, C., & Kopalle, P. (2002). The impact of collinearity on regression analysis: The asymmetric effect of negative and positive correlations [Publisher: Taylor & Francis Journals]. *Applied Economics*, *34*(6), 667–677. Retrieved March 17, 2020, from https://econpapers.repec.org/article/tafapplec/v_3a34_3ay_3a2002_3ai_3a6_3ap_3a667-677.htm
Qualtrics. (2005). Qualtrics Survey Software. www.qualtrics.com
Ripley, B. (2021). Package 'boot', 117.
Salganik, M. J., Fazito, D., Bertoni, N., Abdo, A. H., Mello, M. B., & Bastos, F. I. (2011). Assessing network scale-up estimates for groups most at risk of HIV/AIDS: Evidence from a multiple-method study of heavy drug users in Curitiba, Brazil. *American Journal of Epidemiology*, *174*(10), 1190–1196. https://doi.org/10.1093/aje/kwr246
Salganik, M. J., & Heckathorn, D. D. (2004). Sampling and estimation in hidden populations using respondent-driven sampling. *Sociological methodology*, *34*(1), 193–240. https://doi.org/10.1111/j.0081-1750.2004.00152.x
Salganik, M. J., Mello, M. B., Abdo, A. H., Bertoni, N., Fazito, D., & Bastos, F. I. (2012). The Game of Contacts: Estimating the Social Visibility of Groups. *Networks*, *33*(1), 70–78. https://doi.org/10.1016/j.socnet.2010.10.006.The
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology. *Psychological Science*,

22(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Starosta, A. J., & Earleywine, M. (2014). Assessing base rates of sexual behavior using the unmatched count technique. *Health Psychology and Behavioral Medicine*, *2*(1), 198–210. https://doi.org/10.1080/21642850.2014.886957

Stuber, J., Galea, S., & Link, B. G. (2008). Smoking and the emergence of a stigmatized social status. *Social Science and Medicine*, *67*(3), 420–430. https://doi.org/10.1016/j.socscimed.2008.03.010

Szumilas, M. (2010). Explaining Odds Ratios. (August), 227–229.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Teixeira da Silva, J. A. (2018). Freedom of Speech and Public Shaming by the Science Watchdogs. *Journal of Advocacy, Research, and Education*, *5*(1).

Vardanjani, H. M., Baneshi, M. R., & Haghdoost, A. (2015). Cancer Visibility among Iranian Familial Networks: To What Extent Can We Rely on Family History Reports? [Publisher: Public Library of Science]. *PLOS ONE*, *10*(8), e0136038. https://doi.org/10.1371/journal.pone.0136038

Wagenmakers, E.-J., & Dutilh, G. (2016). Seven Selfish Reasons for Preregistration. https://www.psychologicalscience.org/observer/seven-selfish-reasons-for-preregistration/comment-page-1

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., van Aert, R. C., & van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid P-hacking. *Frontiers in Psychology*, *7*(NOV), 1–12. https://doi.org/10.3389/fpsyg.2016.01832

Wolter, F., & Laier, B. (2014). The Effectiveness of the Item Count Technique in Eliciting Valid Answers to Sensitive Questions. An Evaluation in the Context of Self-Reported Delinquency. *Survey Research Methods*, *8*(3), 153–168.

Zafrunnisha, N., & Pullareddy, V. (2009). Authorship pattern and degree of collaboration in psychology. *Annals of Library and Information Studies*, *56*(December), 255–261.

Zheng, T., Salganik, M. J., & Gelman, A. (2006). How Many People Do You Know in Prison? *Journal of the American Statistical Association*, *101*(474), 409–423. https://doi.org/10.1198/016214505000001168