Accelerating addiction research via Open Science and Team Science

Jonas Dora*, **Marilyn Piccirillo**, **Katherine T. Foster**, and **Kevin M. King**
University of Washington, Department of Psychology, University of Washington, Seattle, WA, United States
*Corresponding author. e-mail address: jdora@uw.edu

Contents

Abstract

The replication crisis of the past decade has taught us that our literature is severely biased. Many empirical findings underlying our theories cannot be replicated. Although the extent of this problem in addiction science is largely unknown, given several factors we discuss in this chapter (e.g., publication bias), there is little reason to believe that replicability in addiction research is any better. Based on meta-science performed over the past decade, we propose that addiction researchers use Open Science and Team Science practices to improve the replicability of their work and thus reduce the bias in the literature and increase the credibility of our field. We walk the reader through six concrete steps necessary for successful implementation of Open Science and Team Science within the field of addiction research. These steps

Psychology of Learning and Motivation
ISSN 0079-7421, https://doi.org/10.1016/bs.plm.2023.06.004

include, (1) preregister; (2) share data, code, and materials; (3) replicate; (4) train Open Science; (5) change the culture and incentivize Open Science; and (6) collaborate in large teams. We discuss evidence supporting the effectiveness of these practices and highlight some limitations and counterpoints to our proposal.

Addiction researchers aim to understand the development, prevention, and treatment of addictive behaviors and related disorders, such as substance use disorders (e.g., alcohol, tobacco, cannabis, opioids), eating disorders (e.g., binge eating), and other behavioral disorders (e.g., gambling). This is an important field of study; for example, each year, the World Health Organization estimates that 8 million deaths are attributable to tobacco (World Health Organization, 2022) and 3 million deaths are attributable to harmful use of alcohol (World Health Organization, 2022) worldwide. Thus, these two substances alone are estimated to be (partially) responsible for more than 19% of annual deaths worldwide. Among people aged 20–39 years of age, the relative number of deaths are even higher. Beyond the risk of chronic health risk and mortality, addiction causes substantial functional disorder, such as an increased risk of unemployment or other financial problems (Swanton & Gainsbury, 2020), divorce, increased risk for psychiatric comorbidity (Compton, Gfroerer, Conway, & Finger, 2014; Davis, Uezato, Newell, & Frazier, 2008; Schuckit, 2006), as well as more acute risks such as aggression and intimate partner violence (Crane, Godleski, Przybyla, Schlauch, & Testa, 2016; Rothman, Stuart, Temple, & Heeren, 2018). Results from decades of addiction research suggest that these outcomes are preventable and treatable with a variety of approaches (Sher, Grekin, & Williams, 2005). But in order to maintain this progress and effectively disseminate interventions addressing problematic substance use and other addictive behaviors, we need to establish robust findings surrounding mechanisms of addictive behaviors.

In this chapter we will review evidence describing threats to the replicability of addiction research, accumulated by meta-scientists and those affiliated with the Center for Open Science (https://cos.io/) who have and continue to work to develop solutions (Open Science & Team Science) for this multi-faceted problem. We draw on progress made primarily within the field of psychology (as much of the 'replication crisis' has played out in psychology and the four us happen to be psychologists) to discuss how Open Science and Team Science practices can be implemented to improve addiction science more broadly. It is important to note that the replication crisis and problems with transparency in scientific reports is not limited to psychology, but extends to most other

fields of biomedical research (e.g., Motulsky, 2014; Sena, Van Der Worp, Bath, Howells, & Macleod, 2010). We believe that incorporating Open Science and Team Science practices into our research will substantially improve the quality of our science, and consequently, our ability to help people struggling with addiction.

Overwhelming evidence suggests that the scientific literature in psychology is severely biased; biased research erodes public trust in science (National Academy of Sciences, 2018), wastes limited resources (time, effort, money), impedes progress, and leads to the application of interventions that are ineffective or potentially even harmful (IJzerman et al., 2020; Williams, Botanov, Kilshaw, Wong, & Sakaluk, 2021). For example, imagine that scientists develop an intervention that trains people struggling with addiction to improve their self-regulation resources based on the idea that substance use is a form of self-regulation failure (Baumeister, 2003) only for us to learn later that this literature on self-regulation failure was severely biased and self-regulation "success" or "failure" likely does not depend on a limited resource (Hagger et al., 2016; Orquin & Kurzban, 2016; Vohs et al., 2021). It would take us years to design and thoroughly test this intervention, which is based on a biased literature that leads us in the wrong direction. Given how long it takes for effective interventions to become widely adopted (Morris, Wooding, & Grant, 2011), or to stop using ineffective ones (Prasad & Ioannidis, 2014), it is critical that the science on the etiology and treatment of addictions is based upon credible findings. We need to do better.

1. How the replication crisis impacts addiction science

Surveys of major psychological journals suggest that over 95% of hypotheses in the psychological literature are supported (Haeffel, 2022), which either means that all our theories are correct or the literature is biased towards positive results. One way to figure out whether it is more likely to be the former or the latter is to measure the rate at which psychological studies replicate. In simple terms, replicability means that we observe the same outcome again when we repeat a study. Most commonly, researchers test if they can replicate a statistically significant finding or if a replicated estimated effect is in the same direction and similar in size.[a]

[a] See (Nosek et al., 2022) for a much more in-depth discussion of replicability.

Replicability is considered by many to be a core tenet of an empirical finding as it provides evidence that an observation reflects systematic knowledge about the world rather than being a product of chance or study-specific circumstances such as the specific sample in which the original finding was observed (Schmidt, 2009). It is important to note here that replicability is not sufficient to conclude that we have stumbled upon something important, as flawed theories and methods may still produce a replicable effect that misinforms us about the fundamental nature of the world (Devezer, Nardin, Baumgaertner, & Buzbas, 2019). However, if we cannot replicate a finding that is supposed to be generalizable (i.e., holds and can be applied to other samples and situations; Asendorpf et al., 2013), this should decrease our confidence in the original finding to the extent that the replication study is at least comparable in terms of methodological rigor; if multiple rigorous replications fail, this tells us that the original finding likely will not be useful to understand the phenomenon we are studying.

Unfortunately, psychological studies replicate at an exceedingly low rate (prompting reporting on the "replication crisis" or "credibility crisis"). In 2015, a large group of researchers found that of 100 studies published in high-impact psychology journals only roughly one third replicated, and the average effect size in the 100 replications was half the magnitude of the average effect size in the original 100 studies (Open Science Collaboration, 2015). Subsequent large-scale replication projects have found similar estimates (e.g., Camerer et al., 2018; Klein, Hardwicke, et al., 2018). Some of the most prolific psychological phenomena that appeared to be well-established (e.g., priming; Dijksterhuis & van Knippenberg, 1998) turned out to be non-replicable (O'Donnell et al., 2018). Thus, we cannot have high confidence that psychological studies from the past would yield a similar result would we repeat them today, and thus we cannot have high confidence that they teach us something meaningful about the human mind.

The extent to which non-replicability is a problem in addiction science is largely unknown as no large-scale replication project across many research groups from different universities has been conducted in our field. Despite calls for replication efforts in the broader fields of clinical psychology (Tackett, Brandes, Dworak, & Shields, 2020; Tackett et al., 2017) and addiction science (Heirene, 2021; LaPlante, 2019; Pearson et al., 2022; Wohl et al., 2019), almost no replications even of single studies have been

conducted in this field.[b] For example, the Elsevier Journal *Drug and Alcohol Dependence* publishes several hundred research articles every single year, yet a PubMed search for (replication[Title]) AND ("Drug and Alcohol Dependence"[Journal]) indicates that only 9 studies with replication in the title have been published since 2000. Some research has suggested that there are differences across subfields, with personality psychology having relatively high potential for replication (Soto, 2019), while other fields such as cognitive and social psychology have relatively low potential for replication (Camerer et al., 2018; Open Science Collaboration, 2015). In other fields in which addiction is studied (e.g., neuroscience), replication has received even less attention, making it hard to estimate the extent to which replicability is an issue, though the few large-scale projects that exist indicate similar problems (e.g., Nieuwland et al., 2018). Because addiction research is so broad and cross-disciplinary, it is difficult to estimate its potential replicability. There is no reason to believe that replicability in addiction science would be higher than in areas such as cognitive and social psychology where replication has been more common because we conduct research in a similar sociocultural environment of academia and employ similar methods (Tackett et al., 2017).

One way to assess the value of the evidence for a hypothesis in the field of addiction science is to conduct a *p*-curve analysis. A *p*-curve is the distribution of statistically significant *p*-values for a set of studies (Simonsohn, Nelson, & Simmons, 2014), quantifying the degree to which significant effects across studies are likely to represent a true versus a null effect. Consequently, the *p*-curve measures evidential value because we know that studying true effects in an unbiased manner produces a right-skewed *p*-curve; *p*-values smaller than.01 should be much more likely than *p*-values just barely below 0.05 (Cumming, 2008). To illustrate how a distribution of *p*-values can tell us something about the evidential value of a hypothesis, we performed a *p*-curve analysis of the association between negative affect and alcohol use (Fig. 1) in ecological momentary assessment data (EMA); in this case, EMA studies that test the association between negative affect and alcohol use.[c] We chose to focus on this relatively narrow literature for an illustration given our subject matter expertise in this area. According to motivational models of alcohol use (Cooper, Frone,

[b] For an exception, see (Pearson et al., 2017).

[c] We included any statistically significant ($p < 0.05$) test reporting an association between negative affect and alcohol use (main effect, moderation, mediation) as long as the association was studied within-person. We excluded tests from 17 studies that did not report exact *p*-values or did not rely on *p*-values for inference.

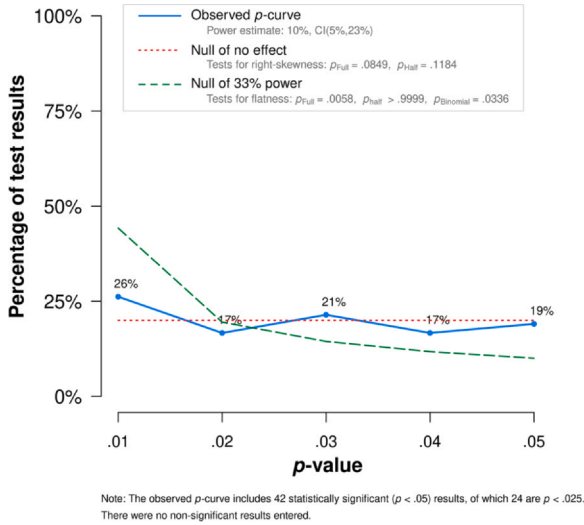

Fig. 1 *P*-curve plot. The blue, solid line indicates the distribution of significant *p*-values across studies in the negative affect-alcohol use EMA literature. The red, dotted line indicates the expected distribution of *p*-values under the null hypothesis. The green, dashed line indicates the expected distribution of *p*-values when the average study has 33% statistical power. The *p*-curve is not right-skewed (there are fewer studies that report $p < 0.025$ than we should expect when studying a true effect), which indicates no evidential value for the hypothesis.

Russell, & Mudar, 1995; Cox & Klinger, 1988), experiencing negative affect should lead to an increase in drinking. As is illustrated in Fig. 1, a *p*-curve based on 42 significant effects reported in this literature indicates that direct replications of the included studies cannot be expected to succeed and the literature does not contain evidence that negative affect is associated with alcohol use in EMA studies. This conclusion mirrors our recent meta-analysis of this literature (Dora et al., 2023).

As we stated previously, the implications of bias in addiction science are serious. Our field would be substantially undermined if findings that appear well-supported and influence current directions turned out to be non-replicable. For example, credible replication failures of the importance of self-efficacy in smoking cessation (Gwaltney, Metrik, Kahler, & Shiffman, 2009), the effectiveness of naltrexone in the treatment of alcohol use disorder (Jonas et al., 2014), or the accuracy of transdermal alcohol sensors in assessing blood alcohol concentration (Yu et al., 2022) would dramatically undermine broad research investments in those areas. We need to become aware of factors that contribute to a biased literature.

2. Why do we have a replication crisis?

"*The first principle is that you must not fool yourself and you are the easiest person to fool.*" Richard Feynman, 1974.

First, it is important to recognize that research is always performed by fallible humans who have internalized biases. Especially in psychology and related fields where data are often noisy and effect sizes are small, these biases can lead us to fool ourselves and others. Several well-documented cognitive biases that might contribute to a biased literature are *confirmation bias* (the tendency to be more likely to see, interpret, and recall information that support one's existing beliefs), *hindsight bias* (the tendency to think that things have been more likely to occur after the fact than one would have predicted a priori), *apophenia* (the tendency to see patterns in randomness), and *motivated reasoning* (the tendency to adopt a lower level of scrutiny for things one wants to believe and a higher level of scrutiny for things one does not want to believe). Each of these biases (and especially in combination) increase the risk of fooling ourselves and others.

"*What we need is a system for evaluating research based only on the procedures employed. If the procedures are judged appropriate, sensible, and sufficiently rigorous to permit conclusions from the results, the research cannot then be judged inconclusive on the basis of the results and rejected by the referees or editors. Whether the procedures were adequate would be judged independently of the outcome.*" Robert Rosenthal, 1966.

Second and much more importantly, it is important to recognize that research is always performed by these fallible humans in an environment that actively does not promote rigor and accuracy (and sometimes directly discourages it). Although most scientists will be intrinsically motivated to produce rigorous and accurate research, this intrinsic motivation is often undermined by extrinsic motivations that do not reward scientists for precision, and instead reward scientists for reporting significant findings. The primary reward in science is getting one's research published (ideally in a journal that maximizes the number of people who will read the study). However, rigor is not the most important criterion for getting one's research published, especially not in prestigious journals (Dubben & Beck-Bornholdt, 2005), incentivizing scientists to produce work that yields a simple yet highly novel narrative. These publications then are used to determine who is hired, promoted, tenured, and receives funding. This whole ecosystem makes it much more likely that novel, exciting but ultimately inaccurate results are published (Nosek, Spies, & Motyl, 2012; Smaldino & McElreath, 2016).

When human biases meet the current state of academia in which quantity is rewarded over quality and novel results are rewarded over rigorous methods, numerous things reduce the replicability of the reported research and increase the bias in the literature. Several major factors (publication bias, questionable research practices, low statistical power, lack of data sharing, lack of large-scale collaboration) have been identified that contribute to low replicability and ultimately a biased literature (Bishop, 2019; Hardwicke & Wagenmakers, 2023; Munafo et al., 2017). These factors arise from a combination of human fallibility and the current landscape of academia mentioned above.

2.1 Publication bias

Publication bias has been identified as a major contributing factor in several failed replications (e.g., Hagger et al., 2016; Shanks et al., 2015). Studies that report statistically significant results are more likely to get published than those reporting non-significant results (Fanelli, 2010; Franco, Malhotra, & Simonovits, 2014), creating a strong bias in the literature where an effect appears more robust than it actually is and effect sizes get inflated (Nissen, Magidson, Gross, & Bergstrom, 2016). This bias appears to be primarily perpetuated by journal editors and reviewers who curate the studies that appear in their journal, often prioritizing novelty and statistically significant findings (Giner-Sorolla, 2012). However, an older study found that researchers also directly contribute to this bias as they report to be more likely to submit studies in which statistically significant effects were obtained (Cooper, DeNeve, & Charlton, 1997), presumably because these studies have a higher chance of being published. While it is easier to publish null results nowadays than it was 25 years ago, such decisions likely contribute to this bias today (Scheel, Schijen, & Lakens, 2021).

2.2 Questionable research practices (QRPs)

QRPs refer to a set of common practices researchers engage in to obtain a statistically significant result (John, Loewenstein, & Prelec, 2012) and in fact many psychological researchers have been taught to engage in these practices during their training (Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). Although much of the discussion surrounding QRPs has been focused on researchers engaging in them with the conscious intention to mislead others, QRPs do not require ill intent to be damaging (Gelman & Loken, 2013) and likely are mostly engaged in unknowingly (i.e., they are a product of the cognitive biases and environment of academia

described above). John et al. (2012) list a total of 10 QRPs, with the most egregious one being the falsification of data. However, while falsification of data clearly is a terrible and intentional offense, it is estimated to be extremely rare and thus likely does not have the biggest impact on the bias in the literature. It is worth mentioning though that addiction science has been the center of one of the most prolific data falsification cases in recent times (Golden, Mazzotta, & Zittel-Barr, 2023).

One of the most common QRPs is *p*-hacking, which refers to any measure a researcher takes to turn a non-significant result into a significant one. Because our theories and hypotheses tend to be vague (van Rooij & Blokpoel, 2020), researchers have to make many (often idiosyncratic and sometimes arbitrary) decisions when analyzing their data (also called researcher degrees of freedom; Gelman & Loken, 2013; Simmons, Nelson, & Simonsohn, 2016), such as how to operationalize the independent and dependent variable, whether to control for covariates, and whether to exclude outliers, all of which can substantially increase the false positive rate when researchers allow themselves analytical flexibility (Stefan & Schönbrodt, 2023). The problem is that statistical significance can be found for virtually any hypothesis when the data are flexibly analyzed and researcher degrees of freedom are high due to natural randomness around the null hypothesis (Gelman & Loken, 2013; Simmons et al., 2016). Thus, *p*-hacking greatly increases the chance of making a Type-I error or in other words erroneously rejecting the null hypothesis. A second common QRP is hypothesizing after the results are known (HARKing; Kerr, 1998). Because confirmatory research tends to be valued higher than exploratory research (Fife & Rodgers, 2022; Scheel, Tiokhin, Isager, & Lakens, 2021), researchers are often tempted to present *post*dictions as *pre*dictions (i.e., they generate and test a hypothesis after looking at the data). This confuses the order of the scientific method, makes the hypothesis unfalsifiable in this dataset, and capitalizes on chance, thus further promoting false positives and bias in the literature (Nosek, Ebersole, DeHaven, & Mellor, 2018).

2.3 Low statistical power

Statistical power is the probability of observing an effect if the effect indeed exists. The sample size that is needed to ensure adequate statistical power can be approximated prior to conducting a study and requires to make a few assumptions (typically surrounding the error rates one deems acceptable, the effect size one wants to detect, and the variance of the effect). In psychology and related fields, effect sizes tend to be small

(Richard, Bond, & Stokes-Zoota, 2003; Schäfer & Schwarz, 2019). All else being equal, smaller effects require larger sample sizes to be detected. When studies are underpowered (i.e., they study small effects in small samples), they are more likely to come to the wrong conclusion (i.e., making a Type-I or Type-II error). Statistically significant effects in underpowered studies *necessarily* overestimate replicability and the true effect size because of the dichotomous nature of significance testing (Vasishth, Mertzen, Jager, & Gelman, 2018). For example, only pursuing effects that were shown to be significant in pilot studies can dramatically overestimate the true effect (Kraemer, Mintz, Noda, Tinklenberg, & Yesavage, 2006). Thus, low statistical power is another factor that makes capitalizing on chance more likely and especially in combination with publication bias and QRPs leads to a large number of false positives in the literature. Low statistical power is likely to be a common issue in addiction science. For example, a meta-analysis estimated that people with higher self-efficacy were more likely to quit smoking successfully (meta-analytic Cohen's $d = 0.21$; Gwaltney et al., 2009). A simple power analysis tells us that 358 participants are needed for any one individual study to detect this effect size with 80% power. Of the studies included in this meta-analysis, only 5/54 (9.26%) studies sampled at least this many individuals. A review found that average statistical power in clinical psychology journals is insufficient to detect small-to-medium-sized effects (Cohen's $d = 0.20$), and that statistical power did not increase from 2000 to 2015 (Reardon, Smack, Herzhoff, & Tackett, 2019).

2.4 Lack of data sharing

Scientists are discouraged to share their data proactively and rarely do so (Wicherts, Borsboom, Kats, & Molenaar, 2006). Sometimes, authors will write that data are available upon request. However, one analysis found that in 2012 not even 40% of authors who published a paper in the APA journals *Emotion*, *Experimental and Clinical Psychopharmacology*, *Journal of Abnormal Psychology* (now called the *Journal of Psychopathology and Clinical Science*), and *Psychology and Aging* (journals which regularly feature research on addictions) shared their data when requested, even though this decision violates the APA ethics code (Vanpaemel, Vermorgen, Deriemaecker, & Storms, 2015). The fact that data are rarely shared is problematic for at least two reasons. First, it makes it impossible to catch human errors (innocent or otherwise). At one point, one out of three genetics papers published in *Nature* contained errors because Microsoft Excel automatically recoded certain genes into dates (Ziemann, Eren, & El-Osta, 2016). Analyzing

hundreds of thousands of psychology papers published between 1985 and 2013, Nuijten and colleagues (2016) found that roughly 50% of these papers contained at least one unambiguous statistical error (2016). Sharing data would make it easier to find and correct these human errors. Second, lack of data sharing means that these data are not available for follow-up research, secondary data analysis and re-analysis, and meta-analyses, and thus never live up to their maximum potential usefulness.[d] While data are not commonly shared alongside research articles, in a recent meta-analytic project that we will describe in detail later many addiction researchers agreed to share their data with us, which potentially indicates a positive shift in our field.

2.5 Lack of large-scale collaboration

Finally, scientists in psychology rarely work in large-scale teams despite the success of this model in other disciplines (Coles, Hamlin, Sullivan, Parker, & Altschul, 2022). For example, over the past decades behavioral geneticists conducted hundreds of small studies on the link between the 5-HTTPLR gene and depression, until they eventually realized that each one of their studies was severely underpowered (Border et al., 2019). Nowadays, behavioral geneticists predominantly work in large teams pooling their resources to produce robust findings (Corvin, Craddock, & Sullivan, 2010). Thus, a series of failed high-profile replications gave rise to an improved scientific model of collaboration. In psychology, a few large-scale Team Science initiatives have been launched, and all of them have been extremely successful in advancing our knowledge (Dora et al., 2023; Klein et al., 2018; Moshontz et al., 2018; Open Science Collaboration, 2015). Unfortunately, addiction science large-scale collaboration is rare. A notable exception that we are aware of is Project Harmony, which is a executed by a collaborative team of alcohol use researchers (mostly working in psychiatry departments) and which aims to integrate and analyze data from over 30 treatment trials for AUD (Saavedra et al., 2021).

In summary, there is no reason to believe that the field of addiction research is safe from the replication crisis. This blind spot may partially explain why uptake of tools that limit bias in the literature has been comparatively slow in addiction science. As we reviewed, it is not easy to

[d] See Dora, van Hooff, Geurts, Kompier, and Bijleveld (2021) for a study from Jonas's PhD project where the open data were subsequently used for another analysis by an independent group (Lekkas, Price, & Jacobson, 2022).

produce unbiased research in the larger sociocultural context of academia paired with human biases. Nonetheless, we consider it a valuable, even necessary aspiration for our work, as we have a responsibility to the people struggling with addiction-related disorders that rely on scientific progress towards harm reduction and the prevention of deaths. We will now outline a roadmap for addiction science consisting of six global steps that we can take to reduce the bias in our field. Successful implementation of these steps will require buy-in from researchers, journal editors and reviewers, funding organizations, and hiring committees.

3. Improving addiction science via Open Science and Team Science

Open Science can be understood as an umbrella term that encompasses steps researchers can take to reduce individual and collective bias from their research improving their work's replicability (same or other researchers are able to obtain the same result with other random samples drawn from the population), reproducibility (other researchers are able to obtain the same reported result from the same data using the same analysis), and availability (other researchers or people interested in the science can access it). The success of Open Science depends on the efforts of multiple stakeholders; the individual researcher's responsibility to engage in Open Science practices, the journal editor's responsibility to highlight work using Open Science practices, and the funding agencies and hiring committee's responsibility to fund and hire those who engage in Open Science. We discuss six steps (Table 1) that are critical to the reduction of bias in the field of addiction science.

3.1 Preregister studies and write registered reports

As we reviewed above, even when we as researchers have good intentions, cognitive biases will often threaten the credibility of our research (Gelman & Loken, 2013; Munafo et al., 2017; Simmons et al., 2016). In common situations where a psychological theory is vague, there are many 'correct' ways to design, analyze, and interpret a study meant to test the theory. Thus, researchers have many degrees of freedom in these matters. For example, if we imagine that there are five decisions to be made, and for each decision we have to choose from five justifiable options, our study might have ended up with 3125 unique results (Hardwicke & Wagenmakers, 2023). If you

Table 1 Open Science and Team Science practices to improve addiction science.

Practice	Stakeholder	Benefits
Preregister studies and write Registered Reports	Individual researcher	Reduces publication bias, *p*-hacking, & HARKing Improves average statistical power Clearly separates confirmatory from exploratory research
	Journal editors & reviewers	
Share data, materials, and code	Individual researcher	Improves reproducibility and replications Maximizes data utility Makes it easier to find and correct errors early
Replicate studies of your own and others	Individual researcher	Reduces publication bias Improves confidence in findings
	Journal editors & reviewers	
	Funding organizations	
Train the next generation in Open Science	Individual researcher	Open Science becomes the default Improves knowledge and skills
Change the culture and incentivize Open Science	Individual researcher	Makes it easier to engage in Open Science Rewards scientific rigor in publishing, grant applications, hiring processes, and tenure and promotion decisions
	Journal editors & reviewers	
	Funding organizations	
	Hiring committees	
Collaborate in large teams	Individual researcher	Greatly improves statistical power Greatly improves generalizability
	Funding organizations	

combine this with a relatively high amount of noise relative to the effect we are studying (e.g., because sample size is low or measurement error and variance are high), it is easy to mistake random noise around the null hypothesis for an effect that supports one's hypothesis.

Meta-scientists recommend that researchers preregister confirmatory studies [i.e., studies that are designed to test an a priori hypothesis][e] to eliminate these biases from our research (Nosek et al., 2018). A good preregistration ideally describes the study's hypotheses, study design, sampling plan, variables, and analysis plan in thorough detail to sufficiently reduce the researcher degrees of freedom (ideally to 1). A preregistration is then uploaded with a time stamp to a registry, such as the Open Science Framework (OSF; https://osf.io). We recommend the OSF as a place to plan a first preregistration as it contains several useful preregistration templates to choose from. After submitting the preregistration, the researchers proceed with their study as outlined in the preregistration, keeping a record of all deviations to their stated protocol. In the drafting of the manuscript, they report the primary analyses as described in the preregistration, report any unplanned or exploratory analyses, and transparently report all deviations to the preregistration that were made in the conduct of their study. Preregistering studies can be difficult (Nosek et al., 2019) and it may take a few attempts before one preregisters a study without having to deviate from the preregistered plan. However, even in this situation the preregistration is extremely worthwhile, as it provides us with clarity regarding decisions and realizations before and after seeing the data (and thus separates confirmatory from exploratory analyses). Thus, while deviations from the preregistration are not ideal as they increase the researcher degrees of freedom after looking at the data, preregistration permits transparency regarding deviation from the preregistered plan and helps authors and readers understand which decisions were made when and why.

Preregistrations can be written and uploaded at multiple stages in the research process. In an ideal scenario, a preregistration is prepared before data are even collected. When this is done and the preregistration includes sufficient detail, this practice protects us against biases leading to unintentional *p*-hacking and HARKing, thus improving the rigor and replicability of our research. However, in addiction science data collection may take years to complete and multiple papers are sometimes written on a

[e] Of course, exploratory research can be preregistered as well, but here we see it as optional rather than mandatory.

single dataset. In this situation, it may not be feasible to complete all preregistrations prior to data collection. In this case, we can disclose in the preregistration document (e.g., the standard OSF preregistration template offers this option) the stage at which the study is being registered (e.g., while data collection is ongoing, after data collection has finished but before data are accessed, after data are accessed). Disclosure about the degree to which we have knowledge of the data provides transparency and context for the extent that knowledge of the data collected influenced the decisions we made for this project. Preregistrations for secondary data analysis are just as useful as for primary analyses as they also reduce biases as intended, so long as the preregistration is finalized prior to data analysis (Petersen, Apfelbaum, & McMurray, 2022).

Strong evidence demonstrates that preregistration is an extremely effective tool to reduce biases and combat inflated effect sizes as well as that preregistered research is highly replicable and credible. For example, the National Heart, Lung, and Blood Institute (NHLBI) made it mandatory to preregister clinical trials in 2000. Prior to this date (when clinical trials were not preregistered), effects of clinical trials funded by the NHLBI varied widely and most found a statistically significant effect in support of the experimental treatment. However, following the change in reporting policy, results of clinical trials became much more consistent and increasingly showed that most experimental treatments had no associated benefit (Kaplan & Irvin, 2015), suggesting that registration of study results increases the veracity of study findings and illuminates the proportion of null results. Second, a comparison of effect sizes across preregistered and non-preregistered studies found that effect sizes in preregistered research were roughly half as large as in non-preregistered research (Schäfer & Schwarz, 2019), mirroring the results from the first large-scale replication project (Open Science Collaboration, 2015). Third, a replication project of 16 significant findings where the original studies were preregistered found that 86% of these findings replicated, greatly improving on the ~50% replicability in previous projects (Protzko et al., 2020). These examples demonstrate that high replicability in psychological and related research is in fact attainable with best practices.

What does a good preregistration look like? Although this will differ from one project to another, our guiding principle is that a preregistration should contain as much detail as necessary to ensure that all decisions relevant to the inferences were made a priori—but not more. It should not contain additional unnecessary details as that will make it harder for

reviewers and readers to find the relevant information in the preregistration. For example, a preregistration does not typically need to contain a justification for the decisions that were made (an exception is a sample size justification); the justification can be provided in the paper as long as the decision is clearly described in the preregistration. In Table 2 we give examples of a strong preregistration for two common research designs, a lab experiment and an ecological momentary assessment (EMA) study, both testing whether people are more likely to drink alcohol when they are in a negative emotional state.[f] For illustrative purposes, we assumed that we did not collect the data for the lab study yet but that the preregistration of the EMA study is a secondary data analysis.

An even stronger practice is to submit one's work as a Registered Report (Chambers & Tzavella, 2021; Chambers, 2013). A Registered Report is written in two stages. In the first stage, authors submit a Stage 1 report to a journal that then gets peer reviewed. The Stage 1 report usually includes an introduction section and a method section and optionally pilot results (in other words, the Stage 1 is a preregistration-of-sorts at the journal). If reviewers assess this Stage 1 report positively, the study is then accepted-in-principle. This means that, as long as the authors execute the research as preregistered, the study is guaranteed publication. Thus, a Registered Report improves on the traditional publishing system in two major ways. First, a study is being peer reviewed at the most critical time, namely while it is planned, and in that way peer reviews can strengthen the research design. Second, Registered Reports do not only limit bias through *p*-hacking and HARKing (similar to a regular preregistration) but additionally combat publication bias as the study is guaranteed publication regardless of the results. A Stage 2 report then describes the full study including results and discussion, which then gets peer reviewed once more (but cannot get rejected anymore). This second round of reviews is meant to ensure that the preregistered research is executed, and that the discussion of the results is meaningful. If the authors have to deviate from the preregistered plan, this needs to be highlighted, explained, and discussed (for a nice example from addiction science, see Grubbs, Floyd, Griffin, Jennings, & Kraus, 2022). Fig. 2 visualizes the workflow of a Registered Report. Increasingly, journals now offer Registered Reports, some of which

[f] For "real" preregistered studies in addiction science, see for example Copeland, Stafford, Acuff, Murphy, and Field (2022); Dora et al. (2023); Dora, Copeland, Field, and King (2022); Dora, Schultz, Shoda, Lee, and King (2022); Pennington et al. (2020); Pennington et al. (2023).

Table 2 Preregistration examples for two common research designs.

	Lab experiment	EMA study
Hypothesis	People are more likely to drink alcohol following a negative (vs neutral) mood induction.	People are more likely to drink on days they experience higher negative affect.
Study type	Experiment—A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.	Observational study—Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, "natural experiments", and regression discontinuity designs.
Study design	We use a within-subjects design. Each participant will come to the lab twice. In each session, participants will first be subjected to a mood induction (negative vs neutral) so that each mood is induced once in each participant (order randomized). Then, participants will be presented with a choice to either consume an alcoholic drink or a sweet pastry. Each session will take roughly 20 min to complete.	Following screening and training on the EMA protocol, participants first completed a baseline assessment. For the next 30 days, five times per day, participants received texts with a link to a brief EMA survey. EMAs were sent within 5 three-hour blocks in between 9 am and 11 pm, with at least one hour between surveys. Each survey will take roughly 2 min to complete.
Randomization	The order in which participants will be subjected to the two mood inductions will be randomized with the help of a web application (https://random.org)	N/A
Existing data + explanation	Registration prior to creation of data	Registration prior to accessing the data: The 1st author of this project will only receive access to the data after uploading this preregistration. The second author has used the data from this study for previous projects, none of which used the same variables. Neither author is aware of any pattern or summary statistics related to the relationships we are looking at here.

(*continued*)

Table 2 Preregistration examples for two common research designs. (*cont'd*)

	Lab experiment	EMA study
Data collection procedures	Participants will be recruited via the undergraduate subject pool of the University of Washington. They have to be at least 21 years of age, report drinking at least once a week over the past three months, and pass a screening test to prevent adverse events of alcohol administration.	Participants were recruited via online ads. Participants were required to be between 18 and 65 years of age, be fluent in English, own a smartphone, and report drinking at least once a week over the past three months.
Sample size + rationale	200 individuals will participate in this research: A power analysis revealed that $N = 180$ is needed to detect $d = 0.3$ with 80% power. To account for attrition, we will sample an extra 20 participants.	The sample size for this EMA study ($N = 500$) was based on a power analysis for a different set of analyses. For that reason, we conducted a set of simulations to determine the power we had to detect our smallest effect size of interest (OR = 1.1). […] These simulations showed that with this sample size the 95% CI excludes 0 in 90% of simulations. Thus, power = 0.9.
Stopping rule	We will stop data collection either when we have collected data from 200 participants as planned or by the time the spring quarter 2023 ends at the University of Washington.	N/A
Manipulated variables	Mood induction (negative vs neutral). Negative: Participants will first be instructed to get into a sad mood, then watch a depressing four-minute-long video clip and finally listen to the first four minutes of a depressing song. Neutral: Participants will first be instructed to get into a neutral mood, then watch a four-minute-long documentary clip about magnets and finally listen to the first four minutes of a neutral song.	N/A

Measured variables	Choice: Participants will be presented with a choice to either consume one of three alcoholic beverages (inducing a BAC of 0.04) or consume one of three sweet pastries after both mood inductions.	Negative affect: During EMA assessments, participants reported the extent to which they felt five negative emotions (irritable, unhappy, angry, anxious, bored) since the last assessment on a 100-point visual analog scale (0 = not at all, 100 = very much). Alcohol use: During EMA morning assessments, participants reported on a visual analog scale how many drinks they had the night before (0–30). If participants missed the morning assessment, they reported alcohol use at the second assessment of the day.
Indices	N/A	Negative affect: On days participants drank, the 5 items will be averaged into a pre-drinking score. On days participants did not drink, the temporal cutoff for this score will be the median onset of alcohol use on drinking days. Alcohol use: This will be dichotomized since we are only interested in the likelihood to drink, not the quantity consumed on drinking days.
Statistical models	The hypothesis will be tested with a generalized linear mixed-effects model with the lme4 package in R. The syntax for the model: glmer(choice ~ 1+ mood.induction + (1|subject), family = binomial)	The hypothesis will be tested with a generalized linear mixed-effects model with the lme4 package in R. The syntax of the model: glmer(alc.use ~ 1+ negative.affect + (1+ negative.affect|subject), family = binomial)
Transformations	Categorical predictors will be sum-to-zero coded.	Continuous predictors will be group-mean standardized so that they have a mean of 0 and a standard deviation of 1.

(*continued*)

Table 2 Preregistration examples for two common research designs. (*cont'd*)

	Lab experiment	EMA study
Inference criteria	We will compute a bootstrapped *p*-value (two-tailed, $\alpha = 0.05$) with the mixed() command from the afex package.	We will compute a bootstrapped p-value (two-tailed, $\alpha = 0.05$) with the mixed() command from the afex package.
Data exclusion	We will exclude participants who do not show up to the second lab session.	We will not exclude any data or participants.
Missing data	N/A	We will address missing data via multiple imputation with the mice package in R. As we have limited experience with multiple imputation, we do not preregister our imputation model but will disclose all imputation models we attempt post-hoc.
Exploratory analysis	We will explore whether results are robust when we control for the order in which participants were subjected to the two mood inductions.	N/A

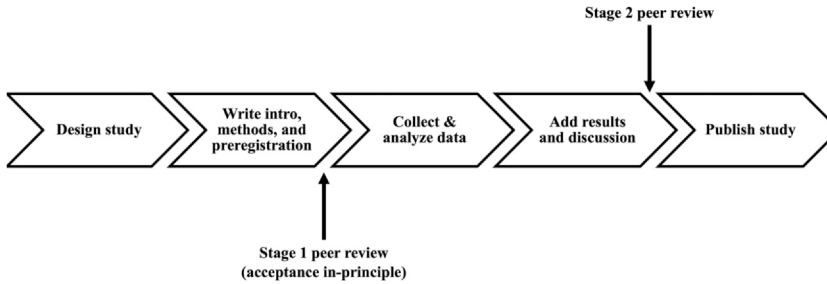

Fig. 2 Workflow of a Registered Report. The study is peer-reviewed twice, once before the data is collected and analyzed (or in case of a secondary data analysis before the data are analyzed) and once after the data are collected and analyzed.

routinely publish addiction-related work [e.g., *Clinical Psychological Science, Psychology of Addictive Behaviors*, *Addiction Research and Theory*, *Collabra: Psychology*, *Journal of Psychopathology and Clinical Science, Nature Human Behavior*, *Royal Society Open Science*][g]. We call on editors-in-chief from journals that frequently publish addiction science to follow suit and offer this highly successful (Hummer, Singleton Thorn, Nosek, & Errington, 2017) model in the near future.

3.2 Share data, materials, and code

Sharing one's research materials greatly improves the contribution one makes to science. Ideally, we should strive to share (e.g., on the OSF and then link to it in the paper) anonymized data, codebooks, analysis code, items, stimuli, and protocols whenever it is feasible and ethical. First, sharing study materials enables researchers to reproduce analyses reported in the paper and to explore alternative analyses of the same hypothesis, which greatly improves the understanding of the reported research and enables readers to properly critique the study during peer review. If data, code, and other materials are not available, large parts of the presented research cannot be thoroughly peer reviewed (Munafo et al., 2017). Second, sharing study materials makes it easier for other research groups to conduct a replication or extension of the study. We will argue in the following why we should replicate findings more often in addiction science. Third, it makes it much easier to perform meta-analyses and research syntheses when

[g] *Drug and Alcohol Dependence* also offers Registered Reports, but we do not recommend to publish here as it is an Elsevier Journal.

the data of studies that fit the inclusion criteria are openly available and linked to in the final manuscript.

Sometimes we collect sensitive data that cannot be shared directly, for example because they imply the risk of being able to identify a participant. A few solutions exist for this problem so that in most cases we can still share at least parts of the dataset that allow others to reproduce the analyses. First, identifiable information could be removed from the dataset or binned together to eliminate this concern. Second, sometimes it is possible to create a synthetic dataset in which the statistical properties and relationships between variables are maintained but no real record of people is shared (e.g., via the *synthpop* package in *R*; Quintana, 2020). If neither of these options is possible, it might still be better to share metadata from the study rather than not sharing anything (e.g., O'Neil et al., 2020). In general, we recommend including mention of data sharing in the informed consent document, see below for an example:

By taking part in this study, you are aware and agree that the research team at Institute X will use the anonymous data from this study to write a research report. Upon publication of that report, the anonymous data from this study will be made available to other researchers. These data will not contain any information that could identify you.

Beyond confidentiality, there are several additional ethical issues to consider surrounding the sharing of participant data, especially data collected among vulnerable and disadvantaged populations. These include the potential for the data to be used repeatedly without informed consent being obtained for each of these uses. In the worst case, data might be used to promote stigma and discrimination towards vulnerable groups. Individuals who struggle with addiction often are vulnerable due to the stigma that is attached to addictive behaviors (Volkow, 2020) and to addiction-related outcomes (e.g., homelessness, HIV and other chronic medical conditions). For these reasons, data sharing is a complicated issue in addiction science. We must balance the ethical risks with data sharing while also continuing to take steps towards improving the rigor in our field. Thus, we believe addiction researchers should maintain an active awareness of the risks of sharing their data, participate in educational opportunities to increase their responsible conduct of research, and maintain active discussions with colleagues in the field around potential solutions to decrease the risk and unintended consequences associated with data sharing. A promising solution to address ethical concerns around data sharing is to develop a shared governance model rooted in community-based participatory research to

engage participants more directly as stakeholders in data use and sharing (for an example, see The MindKind Consortium, 2022). Results from this shared governance model may help to bring about a more balanced approach to data sharing—identifying *which* elements of data may be shared (and how).

3.3 Replicate studies of your own and others

As we mentioned several times so far, the extent to which we should expect studies to replicate in addiction science is largely unknown. But to know which findings to pursue further and build cumulative knowledge on, we need to know which findings are robust and we can place high confidence in. This is central to scientific progress in our discipline.

One type of replication that is relatively common in addiction science are studies which test the same or a similar hypothesis in a similar manner (which is called a *conceptual* replication). However, while conceptual replications advance our knowledge about the phenomena we study (e.g., they test the extent to which effects are generalizable to different populations, methods, statistical analyses, etc.), they do not test replicability directly. That is because when a conceptual replication study does not produce a similar result, this difference may well be due to differences between the studies, leaving us with more uncertainty. Therefore, more *direct* replications are needed where a study is repeated as similarly as possible to the original (i.e., same inclusion criteria/recruitment, same hypothesis, same operationalizations, same analysis, …). Three things are important to consider when planning a replication project. First, given that we do not have the resources to replicate every study, which studies should be replicated? Meta-scientists define the replication value of a study as the combination of [a] the value of being certain about the effect, [b] the current uncertainty about the effect, [c] the ability of a replication to reduce this uncertainty, and [d] the cost of the replication (Isager et al., 2023). Simplified, we should replicate impactful findings for which the evidence is not that strong yet, if we are able to and can afford to perform a high-quality replication. Second, it is generally recommended to contact the original authors to get their input on the preregistration, as they might have valuable feedback that improves the planned study. Third, given that effect sizes tend to be inflated in the literature, it is crucial to think deeply about the sample size that is needed to reduce uncertainty (and thus likely a larger sample size than in the original study will be needed). For a more in-depth discussion of how to replicate well, see Brandt et al. (2014).

3.4 Train the next generation in Open Science

The best way to increase uptake of Open Science is to make it part of the training at every graduate program. Good training is the most straightforward way to high-quality research. We cited a study earlier that showed that many researchers were trained in QRPs (Wagenmakers et al., 2011); we hope that in the future a similar study would show that many researchers were trained in practices that maximize the credibility and replicability of psychological research. Many resources already exist to learn more about Open Science, such as tutorial papers (Klein, Vianello, et al. 2018; Meyer, 2018; Tackett et al., 2020; Van den Akker et al., 2021) and massive open online courses (e.g., https://www.coursera.org/learn/reproducible-research; https://www.coursera.org/learn/statistical-inferences). Such free resources are critical to reduce barriers and to increase the ease of use of Open Science among grad students across the globe. Mentors can point their trainees towards these resources and help them to master the materials. Additionally, programs can more directly incorporate Open Science practices into their research curriculum. For example, courses on research methods can include didactics on how to prepare a preregistration; an excellent way to develop their understanding of methods and statistics. Graduate trainees could also be encouraged to perform a preregistered replication of an interesting finding in their field as an initial research project. Lab directors can also set precedents for each lab member to share responsibility in curating and sharing lab research materials, data, and code with public repositories, so that individual lab members build graduated experience with Open Science practices.

We caution readers who remain hesitant towards Open Science that avoiding engagement with open science issues in training can directly impede the careers of trainees. Trainees may pursue lines of research based on published findings that are not replicable because the majority of findings were null and thus deemed unpublishable. Without extensive knowledge of a field and the research that was not published (which will disproportionately benefit well-connected mentors), trainees may waste time and resources on research projects that neither benefit science nor their careers, posing a cascading problem around equity in career development. For example, early in his career, the senior author (King) collected data from over 1000 college students using "established" task measures of response and choice impulsivity (Hamilton et al., 2015) which were supposedly empirically based and psychometrically strong, with the goal of linking them to self-report measures of impulsivity and substance use.

Despite a relatively voluminous literature reporting associations between task and self-report measures, the senior author's data failed to show any compelling evidence of such associations. After presenting these findings at conferences, this author was told by other senior investigators in the field that "of course they should not be expected to correlate, they measure constructs at different levels" or "Yes, we tried including those measures in our study and they never worked", which starkly contrasted with the optimism presented in the published literature. Indeed, only in the last decade that research emerged showing that behavioral task measures tend to have extremely poor reliability *by design*, and thus should not be expected to be associated with other reliable measures of individual differences (e.g., Dang, King, & Inzlicht, 2020).

3.5 Change the culture and incentivize Open Science

Wide-spread adoption of Open Science requires a strong shift in how we conduct research and evaluate others' work. The responsibility to perform rigorous research cannot lie solely with the individual researcher. Thus, we call on those in power (i.e., journal editors, funding agencies, members of hiring and promotion committees) to incentivize and reward Open Science.

Journal editors can contribute to this cultural shift in several ways. First, they can introduce Registered Reports as a submission option in their journal. Second, they can adopt the Transparency and Openness Promotion (TOP) guidelines (Nosek et al., 2015) at their journal, which broadly describe Open Science standards for scientific studies. Ideally, we call on journals to adopt Level 3 TOP guidelines, which require authors to preregister their *confirmatory* studies and to share their anonymized data, code, and materials unless a strong rationale is given why that is not possible. Third, they can to commit to the "pottery barn rule" (Srivastava, 2012), which states that a journal should generally accept well-conducted replications of original research that they published.

Similarly, we hope to see large funding agencies such as the *National Institutes of Health* (NIH) promote Open Science by offering explicit funding opportunities in the future to replication studies in addiction research and to explicitly require engagement in Open Science practices. NIH recently took a first step in this direction by requiring that all data coming out of NIH-funded projects be shared openly and timely, and has for some time required all clinical trials to be registered. Finally, to incentivize Open Science we would like to see scientific rigor rewarded in hiring and promotion decisions over the counting of publications or other

quantitative metrics. This will be a much more useful metric than for instance the impact factor of published papers, as journal impact factor has been shown repeatedly to be unassociated or even negatively associated with various quality indicators, such as replicability, accuracy of statistical reporting, and retractions (Brembs, Button, & Munafò, 2013; Dougherty & Horne, 2022).

One principle of Open Science we have not discussed in detail yet is that research should be made openly accessible. Open Access guarantees that studies can be read by everyone, not only privileged researchers working at institutions that can afford to subscribe to hundreds of different outlets. Within the Open Access framework, authors of articles also retain copyright, and all articles are openly accessible to the public (for an example journal following this model in psychology, see *Collabra*). Scientific publishers such as *Elsevier* make exorbitant profits through journal subscriptions and high Open Access fees while nowadays providing little-to-no value to science or the taxpayers who fund the science. Editorial boards have the capacity to initiate change from within (see the example of the *Elsevier* journal *Lingua* whose editorial board resigned due to high Open Access fees and re-launched the journal under a new name, *Glossa*, while keeping Open Access fees as low as possible; Bakovic & von Fintel, 2015). We encourage editorial boards of addiction-related journals that are published by *Elsevier* to consider doing the same.

3.6 Collaborate in large teams

Finally, a step we see as having tremendous potential to accelerate addiction science is to engage in Team Science or in other words to start large-scale collaborations across research groups, countries, and cultures. In our field, we aspire to conduct high-powered and generalizable research. Achieving high power is difficult due to the small and variable effects inherent to psychology (Schäfer & Schwarz, 2019). Generalizability is difficult to achieve when we only collect data from undergraduate students at one university or community members in one city (Henrich, Heine, & Norenzayan, 2010; Yarkoni, 2020). At the same time, there is a growing recognition within addiction research and related fields of psychology that there are likely few universal laws and instead etiological mechanisms and treatment effects are likely highly contextual and person-specific (Foster & Beltz, 2021; Hamaker, 2012; Piccirillo & Rodebaugh, 2019; Wright & Woods, 2020), increasing the need for transparent research practices given the added complexity with parsing contextual and person-specific effects.

Combined, this implies that vast amounts of resources and adequately powered research designs are necessary to achieve even a minimal amount of reliability and generalizability. Such desired outcomes can be achieved by centralized, consortium efforts like the ABCD study (Volkow et al., 2018) have the potential to provide conclusive answers to substantive questions while bringing together researchers with different expertise and a shared goal.

Team Science can improve upon each of the problems we face that are outlined in this chapter even beyond the uptake of Open Science practices (Forscher et al., 2023). Instead of single research groups having to make important decisions such as what question(s) to study, how to measure important constructs, how to design a study, and how to analyze the data, a Team Science project allows a large number of people involved in a field to make these decisions collectively, which maximizes the chances that scientists with various expertise (e.g., theory, measurement, analysis) are part of the study planning. By pooling resources, much larger sample sizes are achievable; not only that, but also it is now possible to make use of the networks of the participating research groups to recruit participants across the country or ideally across the globe, vastly improving generalizability. Team Science approaches can also be adopted for secondary data analyses and meta-analyses collaboratively, where many research groups pool data and then perform the meta-analytic research together.

In fact, we recently led such a collaborative meta-analysis of the daily association between affect and alcohol use bringing together more than 60 alcohol use researchers (Dora et al., 2023). The four of us initiated this project because we felt that a meta-analysis of this literature was long overdue. However, for this project the raw data from studies were needed due to repeated measures nested in participants paired with a zero-inflated count outcome (number of alcoholic drinks consumed). Following a systematic literature review, we wrote the corresponding author of every study meeting the inclusion criteria an email and asked them if they would be willing and able to share their data with us for this project. Based on accounts of people who have attempted something similar before us in other fields, we were not optimistic that we would receive a lot of data. To our surprise, most researchers we contacted were enthusiastic about our proposal and ended up sharing their data with us and joined our authorship team, which in our opinion shows how open to collaboration addiction researchers are (there were a handful of people who were unwilling to share, a handful of people who wanted to but ultimately were unable to

share, and a handful of people who never responded). This allowed us to ultimately analyze the individual participant data from more than 12,000 individuals across 69 studies and coming from three different continents. Co-authors not only shared their data with us but also provided two rounds of input on the preregistration, multiple rounds of feedback on the manuscript, and helped us respond to three rounds of reviews before the paper was ultimately accepted for publication. As part of this process, we made several decisions as a group (via an anonymous poll), such as which smallest effect size of interest to preregister and to which journal to submit the paper. Beyond the vast amount of data we were able to meta-analyze, involving diverse perspectives strengthened the resulting manuscript and generated constructive discussion surrounding the interpretations of our findings. Our analyses showed conclusively that, contrary to theory, negative affect is not associated with same-day alcohol use in daily life. At the same time, extensive conversations with our co-authors helped us to realize the limitations of our data and to identify next steps for this field, which we were able to feature in the discussion of our paper.

Collaborating in large teams has huge potential to make breakthroughs in addiction research in the future, and this potential has barely been leveraged. This is unsurprising given several barriers to large-scale collaborations. First, within our current scientific culture, there are few incentives to work in large teams. The majority of the credit for a publication goes to the first author and team that takes the lead. For example, in our meta-analysis there was little external incentive for our many co-authors to share their data and to invest their time and effort into this project. There was only the internal incentive of knowing that this will be an important project and a strong scientific study (making it all the more impressive that so many addiction researchers decided to collaborate with us!). We believe this highlights the potential of this model in addiction science. Shifts in academic culture, such as measurable credit for co-authored, team-based work in tenure/promotion review (e.g., engaging with consortium efforts, use of the CRediT system) are needed to promote sustainable change. Shifts within funding agencies are also needed. Additionally, at this point in time there are no funding mechanisms at many large agencies (e.g., *National Institutes of Health*) awarded to teams or consortia; grants usually go to individual researchers. While some mechanisms exist that encourage cooperation (e.g., P30, U01), the fact that they are awarded to a Principal Investigator makes it harder to put together an application as a team of equal contributors. This indirectly disincentivizes

large-scale collaboration. Given the early success of large-scale collaborations across psychology, we hope to see such mechanisms to appear in the future.

Lastly, it is important to highlight a few challenges that come with large-scale collaborations, some of which we experienced first-hand in our meta-analytic project referenced earlier. The first challenge is that all large-scale collaborations need leadership; without structure, it is difficult to make progress. However, this implies the risk that leadership could abuse this position of power for their own gain over that of the collective or to make decisions based on their individual preferences or beliefs. To combat this risk, leadership needs to be held accountable and important decisions need to be made democratically as much as possible. For example, at the start of the project shared goals and strategy should be articulated so that leadership can be held accountable in light of these goals and strategy whenever a member of the team is unsatisfied. A second challenge is that doing research in a large team involves a large number of interpersonal relationships among people with a shared goal but ultimately differences in opinions and beliefs. This makes conflict common, which can be constructive when it remains task-focused but also implies the risk of a toxic environment especially if clear communication is lacking. A solution to this challenge is for leadership to regularly ask for formal and informal feedback from members, creating a feedback loop in which members' concerns are acknowledged and openly discussed. In our experience, engaging in Open Science practices facilitates this, as early project milestones such as preregistration of the study requires dialog and agreement. Problematic behavior by any member needs to be held accountable and an environment should be created in which members feel safe to communicate disagreement, issues, and concerns. Such leadership and communication skills are rarely taught or incentivized, but they are key to the success of large-scale collaborations.

4. Objections to and limitations of Open Science

Despite our optimism around Open Science practices, we acknowledge that some readers may still have concerns. We wish to briefly address a few common concerns here and provide some compromise towards the integration of Open Science practices. One common concern is that preregistration may turn into a "prison" where one can no longer analyze their data outside of the ways described in the preregistration. It is

important to point out that preregistration does not restrict the ways in which you are allowed to analyze your data. You can perform and report on exploratory analyses. Preregistration is merely meant to clearly label those analyses that were planned a priori (i.e., confirmatory, in the sense that all decisions were made before looking at the data and thus the researcher degrees of freedom are controlled for) from those that were conducted post-hoc (i.e., exploratory, requiring further replication).

A second concern we sometimes hear is that proponents of Open Science suggest that preregistration guarantees a high-quality study. We do not believe this to be the case; transparency and openness are necessary *but not sufficient* for high-quality research. Studies might be preregistered, but critical deviations are not disclosed. Study designs and analysis plans might be outlined in great detail but fatally flawed. Preregistered research needs to be peer reviewed and evaluated just as carefully as non-preregistered research. In fact, we would argue it is much easier to critique a preregistered study due to the increased transparency. Similarly, preregistration may well be ineffective in certain scenarios as it does nothing to improve the psychological theory that is tested which might be flawed (Szollosi et al., 2020). While this argument has been used to suggest that preregistration is pointless, we maintain that the success of preregistration in reducing bias reviewed above makes preregistration worthwhile (Nosek et al., 2019).

More generally, our position is that some concern is warranted with regards to any Open Science (or in fact, scientific) practice. Of course, there are ethical concerns with sharing participant data if we cannot properly reduce the risk of identifiability. Similarly, some effects we should not expect to replicate; for example, if we replicate a study on cannabis use behavior performed in Washington State that was conducted in 2010, a non-replication might have to do with the fact that Washington State legalized cannabis in 2012. Thus, a replication failure does not necessarily imply that the original study did not make an important observation that should be a part of the scientific record. We do not want to imply that you should engage in every single Open Science practice described here in every single one of your projects. But we do believe that many projects would benefit from these practices, and ultimately it needs to be decided on a project-by-project basis what results in the strongest science.

Engaging in Open Science can be resource and labor intensive; these factors have differential consequences for those at various stages in their careers. For example, studies may require larger sample sizes (e.g., more

resource intensive), careful preregistration can be labor intensive, both of which delay publication, which might be a concern for early career researchers as they apply for jobs and fellowships. At the same time, Open Science can be beneficial for the very same reasons. In our experience, preregistration does not incur more work but merely *shifts* the work from late in the research process to the early planning stages. Stage 1 Registered Reports that are accepted-in-principle can already be highlighted on a CV along preprints of preregistered manuscripts under review. Registered reports guarantee publication of null results, reducing anxiety around findings that may end up more difficult to publish. Overall, with greater uptake, there will be additional ways to recognize the efforts of those who engage in Open Science.

5. Conclusion

In this chapter, we have outlined key changes to strengthen future addiction research, sharing examples from our experience as addiction psychologists. Making these changes may be challenging; however, the consequences of ignoring the replicability crisis are far too great. Open Science and Team Science are promising solutions and can shape a future in which scientific rigor is valued more than shiny results and research quality is valued more than research quantity. At the end of the day, we believe that the practices described in this chapter are necessary to adopt to accelerate our understanding of addiction.

References

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., & Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality, 27*(2), 108–119. https://doi.org/10.1002/per.1919

Bakovic, E., & von Fintel, K. (2015). Lingua is dead. Long live Glossa! *Language Log*. ⟨https://languagelog.ldc.upenn.edu/nll/?p=22162⟩.

Baumeister, R. F. (2003). Ego Depletion and Self-Regulation Failure: A Resource Model of Self-Control. *Alcoholism: Clinical & Experimental Research, 27*, 281–284. https://doi.org/10.1097/01.ALC.0000060879.61384.A4

Bishop, D. (2019). Rein in the four horsemen of irreproducibility. *Nature, 568*, 435. https://doi.org/10.1038/d41586-019-01307-2

Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. *American Journal of Psychiatry, 176*(5), 376–387. https://doi.org/10.1176/appi.ajp.2018.18070881

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., & Van'T Veer, A. (2014). The Replication Recipe: What makes for a convincing

replication. *Journal of Experimental Social Psychology, 50*, 217–224. https://doi.org/10.1016/j.jesp.2013.10.005

Brembs, B., Button, K., & Munafò, M. (2013). Deep impact: Unintended consequences of journal rank. *Frontiers in Human Neuroscience,* 7, 291. https://doi.org/10.3389/fnhum.2013.00291

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., & Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour, 2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

Chambers, C. D. (2013). Registered Reports: A new publishing initiative at Cortex. *Cortex, 49*(3), 609–610. https://doi.org/10.1016/j.cortex.2012.12.016

Chambers, C. D., & Tzavella, L. (2021). The past, present and future of Registered Reports. *Nature Human Behaviour, 6*(1), 29–42. https://doi.org/10.1038/s41562-021-01193-7

Coles, N., Hamlin, J., Sullivan, L., Parker, T., & Altschul, D. (2022). Build up big team science. *Nature, 601*, 505–507.

Compton, W. M., Gfroerer, J., Conway, K. P., & Finger, M. S. (2014). Unemployment and substance outcomes in the United States 2002–2010. *Drug and Alcohol Dependence, 142*, 350–353. https://doi.org/10.1016/j.drugalcdep.2014.06.012

Cooper, H., DeNeve, K., & Charlton, K. (1997). Finding the missing science: The fate of studies submitted for review by a human subjects committee. *Psychological Methods, 2*(4), 447–452. https://doi.org/10.1037/1082-989X.2.4.447

Cooper, M., Frone, M., Russell, M., & Mudar, P. (1995). Drinking to regulate positive and negative emotions: A motivational model of alcohol use. *Journal of Personality and Social Psychology, 69*, 990–1005. https://doi.org/10.1037/0022-3514.69.5.990

Copeland, A., Stafford, T., Acuff, S. F., Murphy, J. G., & Field, M. (2022). Behavioral economic and value-based decision-making constructs that discriminate current heavy drinkers versus people who reduced their drinking without treatment. *Psychology of Addictive Behaviors, 37*, 132–143. https://doi.org/10.1037/adb0000873

Corvin, A., Craddock, N., & Sullivan, P. F. (2010). Genome-wide association studies: A primer. *Psychological Medicine, 40*(7), 1063–1077. https://doi.org/10.1017/S0033291709991723

Cox, M., & Klinger, E. (1988). A motivational model of alcohol use. *Journal of Abnormal Psychology, 97*, 168–180. https://doi.org/10.1037/0021-843X.97.2.168

Crane, C. A., Godleski, S. A., Przybyla, S. M., Schlauch, R. C., & Testa, M. (2016). The proximal effects of acute alcohol consumption on male-to-female aggression: A meta-analytic review of the experimental literature. *Trauma, Violence & Abuse, 17*(5), 520–531. https://doi.org/10.1177/1524838015584374

Cumming, G. (2008). Replication and *p* intervals: *P* values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science, 3*(4), 286–300. https://doi.org/10.1111/j.1745-6924.2008.00079.x

Dang, J., King, K., & Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated. *Trends in Cognitive Sciences, 24*, 267–269. https://doi.org/10.1016/j.tics.2020.01.007

Davis, L., Uezato, A., Newell, J. M., & Frazier, E. (2008). Major depression and comorbid substance use disorders. *Current Opinion in Psychiatry, 21*(1), 14–18. https://doi.org/10.1097/YCO.0b013e3282f32408

Devezer, B., Nardin, L. G., Baumgaertner, B., & Buzbas, E. O. (2019). Scientific discovery in a model-centric framework: Reproducibility, innovation, and epistemic diversity. *PLoS One, 14*(5), e0216125. https://doi.org/10.1371/journal.pone.0216125

Dijksterhuis, A., & van Knippenberg, A. (1998). The relation between perception and behavior, or how to win a game of Trivial Pursuit. *Journal of Personality and Social Psychology, 74*(4), 865–877. https://doi.org/10.1037/0022-3514.74.4.865

Dora, J., Copeland, A., Field, M., & King, K. M. (2022). Modeling the value-based decision to consume alcohol in response to emotional experiences. *Experimental and Clinical Psychopharmacology*. https://doi.org/10.31234/osf.io/puhqj

Dora, J., Piccirillo, M., Foster, K. T., Arbeau, K., Armeli, S., Auriacombe, M., & King, K. M. (2023). The daily association between affect and alcohol use: A meta-analysis of individual participant data. *Psychological Bulletin, 149*(1–2), 1–24. https://doi.org/10.1037/bul0000387

Dora, J., Schultz, M. E., Shoda, Y., Lee, C. M., & King, K. M. (2022). No evidence for trait- and state-level urgency moderating the daily association between negative affect and subsequent alcohol use in two college samples. *Brain and Neuroscience Advances, 6*, 23982128221079556. https://doi.org/10.1177/23982128221079556

Dora, J., van Hooff, M., Geurts, S., Kompier, M., & Bijleveld, E. (2021). Fatigue, boredom and objectively measured smartphone use at work. *Royal Society Open Science, 8*, 201915. https://doi.org/10.1098/rsos.201915

Dougherty, M. R., & Horne, Z. (2022). Citation counts and journal impact factors do not capture some indicators of research quality in the behavioural and brain sciences. *Royal Society Open Science, 9*(8), 220334. https://doi.org/10.1098/rsos.220334

Dubben, H.-H., & Beck-Bornholdt, H.-P. (2005). Systematic review of publication bias in studies on publication bias. *BMJ (Clinical Research ed.), 331*(7514), 433–434. https://doi.org/10.1136/bmj.38478.497164.F7

Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLoS One, 5*(4), e10068. https://doi.org/10.1371/journal.pone.0010068

Fife, D. A., & Rodgers, J. L. (2022). Understanding the exploratory/confirmatory data analysis continuum: Moving beyond the "replication crisis". *American Psychologist,* 77(3), 453–466. https://doi.org/10.1037/amp0000886

Forscher, P. S., Wagenmakers, E.-J., Coles, N. A., Silan, M. A., Dutra, N., Basnight-Brown, D., & IJzerman, H. (2023). The benefits, barriers, and risks of big-team science. *Perspectives on Psychological Science, 18*, 607–623. https://doi.org/10.1177/17456916221082970

Foster, K., & Beltz, A. (2021). Heterogeneity in affective complexity among men and women. *Emotion (Washington, D. C.), 22*, 1815–1827.

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science (New York, N. Y.), 345*(6203), 1502–1505. https://doi.org/10.1126/science.1255484

Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time*. Department of Statistics, Columbia University.

Giner-Sorolla, R. (2012). Science or Art? How Aesthetic Standards Grease the Way Through the Publication Bottleneck but Undermine Science. *Perspectives on Psychological Science,* 7, 562–571. https://doi.org/10.1177/1745691612457576

Golden, J., Mazzotta, C. M., & Zittel-Barr, K. (2023). Systemic obstacles to addressing research misconduct in higher education: A case study. *Journal of Academic Ethics, 21*(1), 71–82. https://doi.org/10.1007/s10805-021-09438-w

Grubbs, J. B., Floyd, C. G., Griffin, K. R., Jennings, T. L., & Kraus, S. W. (2022). Moral incongruence and addiction: A registered report. *Psychology of Addictive Behaviors, 36*(7), 749–761. https://doi.org/10.1037/adb0000876

Gwaltney, C. J., Metrik, J., Kahler, C. W., & Shiffman, S. (2009). Self-efficacy and smoking cessation: A meta-analysis. *Psychology of Addictive Behaviors, 23*(1), 56–66. https://doi.org/10.1037/a0013529

Haeffel, G. J. (2022). Psychology needs to get tired of winning. *Royal Society Open Science, 9*(6), 220099. https://doi.org/10.1098/rsos.220099

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., & Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science, 11*(4), 546–573. https://doi.org/10.1177/1745691616652873

Hamaker, E. (2012). *Why researchers should think "within-person": A paradigmatic rationale. Handbook of research methods for studying daily life.* The Guilford Press. 43–61.

Hamilton, K. R., Littlefield, A. K., Anastasio, N. C., Cunningham, K. A., Fink, L. H. L., Wing, V. C., & Potenza, M. N. (2015). Rapid-response impulsivity: Definitions, measurement issues, and clinical implications. *Personality Disorders: Theory, Research, and Treatment, 6*(2), 168–181. https://doi.org/10.1037/per0000100

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

Heirene, R. M. (2021). A call for replications of addiction research: which studies should we replicate and what constitutes a 'successful' replication? *Addiction Research & Theory, 29*, 89–97. https://doi.org/10.1080/16066359.2020.1751130

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world. *Behavioral and Brain Sciences, 33*, 61–83. https://doi.org/10.1017/S0140525X0999152X

Hummer, L.T., Singleton Thorn, F., Nosek, B.A., & Errington, T.M. (2017). *Evaluating registered reports: A naturalistic comparative study of article impact* [Preprint]. Open Science Framework. ⟨https://doi.org/10.31219/osf.io/5y8w7⟩.

IJzerman, H., Lewis, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., & Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour, 4*(11), 1092–1094. https://doi.org/10.1038/s41562-020-00990-w

Isager, P., van Aert, R., Bahnik, S., Brandt, M., DeSoto, A., Giner-Sorolla, R., & Lakens, D. (2023). Deciding what to replicate: A decision model for replication study selection under resource and knowledge constraints. *Psychological Methods, 28*, 438–451. https://doi.org/10.1037/met0000438

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

Jonas, D. E., Amick, H. R., Feltner, C., Bobashev, G., Thomas, K., Wines, R., ... Garbutt, J. C. (2014). Pharmacotherapy for Adults With Alcohol Use Disorders in Outpatient Settings: A Systematic Review and Meta-analysis. *JAMA, 311*, 1889. https://doi.org/10.1001/jama.2014.3628

Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLoS One, 10*(8), e0132382. https://doi.org/10.1371/journal.pone.0132382

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., & Frank, M. C. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology, 4*(1), 20. https://doi.org/10.1525/collabra.158

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., & 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. https://doi.org/10.1177/2515245918810225

Kraemer, H. C., Mintz, J., Noda, A., Tinklenberg, J., & Yesavage, J. A. (2006). Caution regarding the use of pilot studies to guide power calculations for study proposals. *Archives of General Psychiatry, 63*(5), 484. https://doi.org/10.1001/archpsyc.63.5.484

Lekkas, D., Price, G. D., & Jacobson, N. C. (2022). Using smartphone app use and lagged-ensemble machine learning for the prediction of work fatigue and boredom. *Computers in Human Behavior, 127*, 107029. https://doi.org/10.1016/j.chb.2021.107029

Meyer, M. N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science, 1*(1), 131–144. https://doi.org/10.1177/2515245917747656

Morris, Z. S., Wooding, S., & Grant, J. (2011). The answer is 17 years, what is the question: Understanding time lags in translational research. *Journal of the Royal Society of Medicine, 104*(12), 510–520. https://doi.org/10.1258/jrsm.2011.110180

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., & Chartier, C. R. (2018). The psychological science accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science, 1*(4), 501–515. https://doi.org/10.1177/2515245918797607

Motulsky, H. J. (2014). Common misconceptions about data analysis and statistics. *Journal of Pharmacology and Experimental Therapeutics, 351*(1), 200–205. https://doi.org/10.1124/jpet.114.219170

Munafo, M., Nosek, B., Bishop, D., Button, K., Chambers, C., Percie du Sert, N., & Ioannidis, J. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*, 0021. https://doi.org/10.1038/s41562-016-0021

National Academy of Sciences. (2018). The science of science communication III: Inspiring novel collaborations and building capacity. In *Proceedings of a colloquium*.

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., & Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife, 7*, e33468. https://doi.org/10.7554/eLife.33468

Nissen, S. B., Magidson, T., Gross, K., & Bergstrom, C. T. (2016). Publication bias and the canonization of false facts. *ELife, 5*, e21451. https://doi.org/10.7554/eLife.21451

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., & Yarkoni, T. (2015). Promoting an open research culture. *Science (New York, N. Y.), 348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., & Vazire, S. (2019). Preregistration is hard, and worthwhile. *Trends in Cognitive Sciences, 23*(10), 815–818. https://doi.org/10.1016/j.tics.2019.07.009

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

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology, 73*(1), 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7*(6), 615–631. https://doi.org/10.1177/1745691612459058

Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods, 48*(4), 1205–1226. https://doi.org/10.3758/s13428-015-0664-2

O'Donnell, M., Nelson, L. D., Ackermann, E., Aczel, B., Akhtar, A., Aldrovandi, S., ... Zrubka, M. (2018). Registered replication report: Dijksterhuis and van Knippenberg (1998). *Perspectives on Psychological Science, 13*(2), 268–294. https://doi.org/10.1177/1745691618755704

O'Neil, M. E., Harik, J. M., McDonagh, M. S., Cheney, T. P., Hsu, F. C., Cameron, D. C., & Hamblen, J. L. (2020). Development of the PTSD-repository: A publicly available repository of randomized controlled trials for posttraumatic stress disorder. *Journal of Traumatic Stress, 33*(4), 410–419. https://doi.org/10.1002/jts.22520

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science (New York, N.Y.), 349*, aac4716. ⟨https://doi.org/10.1126/science.aac4716⟩.

Orquin, J. L., & Kurzban, R. (2016). A meta-analysis of blood glucose effects on human decision making. *Psychological Bulletin, 142*(5), 546–567. https://doi.org/10.1037/bul0000035

Pearson, M. R., Prince, M. A., & Bravo, A. J. (2017). Moderators of the effects of alcohol protective behavioral strategies: Three attempts of replication and extension. *Substance Use & Misuse, 52*(7), 939–949. https://doi.org/10.1080/10826084.2016.1267222

Pearson, M. R., Schwebel, F. J., Richards, D. K., & Witkiewitz, K. (2022). Examining replicability in addictions research: How to assess and ways forward. *Psychology of Addictive Behaviors, 36*, 260–270. https://psycnet.apa.org/doi/10.1037/adb0000730.

Pennington, C. R., Monk, R. L., Heim, D., Rose, A. K., Gough, T., Clarke, R., & Jones, A. (2023). The labels and models used to describe problematic substance use impact discrete elemenJonasts of stigma: A registered report. *Psychology of Addictive Behaviors*. https://doi.org/10.1037/adb0000919

Pennington, C. R., Shaw, D. J., Adams, J., Kavanagh, P., Reed, H., Robinson, M., & White, H. (2020). Where's the wine? Heavy social drinkers show attentional bias towards alcohol in a visual conjunction search task. *Addiction (Abingdon, England), 115*(9), 1650–1659. https://doi.org/10.1111/add.14997

Petersen, I. T., Apfelbaum, K. S., & McMurray, B. (2022). Adapting open science and pre-registration to longitudinal research. *Infant and Child Development*. https://doi.org/10.1002/icd.2315

Piccirillo, M. L., & Rodebaugh, T. L. (2019). Foundations of idiographic methods in psychology and applications for psychotherapy. *Clinical Psychology Review, 71*, 90–100. https://doi.org/10.1016/j.cpr.2019.01.002

Prasad, V., & Ioannidis, J. P. (2014). Evidence-based de-implementation for contradicted, unproven, and aspiring healthcare practices. *Implementation Science, 9*(1), 1. https://doi.org/10.1186/1748-5908-9-1

Protzko, J., Krosnick, J., Nelson, L., Nosek, B.A., Axt, J., Berent, M., … Schooler, J. (2020). *High replicability of newly-discovered social-behavioral findings is achievable.* ⟨https://doi.org/10.31234/osf.io/n2a9x⟩.

Quintana, D. S. (2020). A synthetic dataset primer for the biobehavioural sciences to promote reproducibility and hypothesis generation. *ELife, 9*, e53275. https://doi.org/10.7554/eLife.53275

Reardon, K. W., Smack, A. J., Herzhoff, K., & Tackett, J. L. (2019). An N-pact factor for clinical psychological research. *Journal of Abnormal Psychology, 128*(6), 493–499. https://doi.org/10.1037/abn0000435

Richard, F. D., Bond, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology,* 7(4), 331–363. https://doi.org/10.1037/1089-2680.7.4.331

Rothman, E. F., Stuart, G. L., Temple, J. R., & Heeren, T. (2018). Alcohol, marijuana, and dating abuse perpetration by young adults: Results of a daily call study. *Violence Against Women, 24*(10), 1187–1206. https://doi.org/10.1177/1077801218781959

Saavedra, L. M., Morgan-López, A. A., Hien, D. A., López-Castro, T., Ruglass, L. M., Back, S. E., & Hamblen, J. (2021). Evaluating treatments for posttraumatic stress disorder, alcohol and other drug use disorders using meta-analysis of individual patient data: Design and methodology of a virtual clinical trial. *Contemporary Clinical Trials, 107*, 106479. https://doi.org/10.1016/j.cct.2021.106479

Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology, 10*, 813. https://doi.org/10.3389/fpsyg.2019.00813

Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with registered reports. *Advances in Methods and Practices in Psychological Science, 4*(2), 251524592110074. https://doi.org/10.1177/25152459211007467

Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why hypothesis testers should spend less time testing hypotheses. *Perspectives on Psychological Science, 16*(4), 744–755. https://doi.org/10.1177/1745691620966795

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology, 13*(2), 90–100. https://doi.org/10.1037/a0015108

Schuckit, M. A. (2006). Comorbidity between substance use disorders and psychiatric conditions. *Addiction (Abingdon, England), 101*, 76–88. https://doi.org/10.1111/j.1360-0443.2006.01592.x

Sena, E. S., Van Der Worp, H. B., Bath, P. M. W., Howells, D. W., & Macleod, M. R. (2010). Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biology, 8*(3), e1000344. https://doi.org/10.1371/journal.pbio.1000344

Shanks, D. R., Vadillo, M. A., Riedel, B., Clymo, A., Govind, S., Hickin, N., & Puhlmann, L. M. C. (2015). Romance, risk, and replication: Can consumer choices and risk-taking be primed by mating motives. *Journal of Experimental Psychology: General, 144*(6), e142–e158. https://doi.org/10.1037/xge0000116

Sher, K., Grekin, E., & Williams, N. (2005). The development of alcohol use disorders. *Annual Review of Clinical Psychology, 1*, 493–523. https://doi.org/10.1146/annurev.clinpsy.1.102803.144107

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2016). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. In A. E. Kazdin (Ed.). *Methodological issues and strategies in clinical research* (pp. 547–555). (4th ed.). American Psychological Association. ⟨https://doi.org/10.1037/14805-033⟩.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143*(2), 534–547. https://doi.org/10.1037/a0033242

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science, 3*(9), 160384. https://doi.org/10.1098/rsos.160384

Soto, C. J. (2019). How replicable are links between personality traits and consequential life outcomes? The life outcomes of personality replication project. *Psychological Science, 30*(5), 711–727. https://doi.org/10.1177/0956797619831612

Srivastava, S. (2012). A Pottery Barn rule for scientific journals. *The Hardest Science*. ⟨https://thehardestscience.com/2012/09/27/a-pottery-barn-rule-for-scientific-journals/⟩.

Stefan, A. M., & Schönbrodt, F. D. (2023). Big little lies: A compendium and simulation of *p*-hacking strategies. *Royal Society Open Science, 10*(2), 220346. https://doi.org/10.1098/rsos.220346

Swanton, T. B., & Gainsbury, S. M. (2020). Gambling-related consumer credit use and debt problems: A brief review. *Current Opinion in Behavioral Sciences, 31*, 21–31. https://doi.org/10.1016/j.cobeha.2019.09.002

Szollosi, A., Kellen, D., Navarro, D. J., Shiffrin, R., van Rooij, I., van Zandt, T., & Donkin, C. (2020). Is Preregistration Worthwhile? *Trends in Cognitive Sciences, 24*, 94–95. https://doi.org/10.1016/j.tics.2019.11.009

Tackett, J. L., Brandes, C. M., Dworak, E. M., & Shields, A. N. (2020). Bringing the (pre) registration revolution to graduate training. *Canadian Psychology/Psychologie Canadienne, 61*(4), 299–309. https://doi.org/10.1037/cap0000221

Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D., ... Shrout, P. E. (2017). It's Time to Broaden the Replicability Conversation: Thoughts for and From Clinical Psychological Science. *Perspectives on Psychological Science, 12*, 742–756. https://doi.org/10.1177/1745691617690042

The MindKind Consortium. (2022). MindKind: A mixed-methods protocol for the feasibility of global digital mental health studies in young people. *Wellcome Open Research, 6*, 275. https://doi.org/10.12688/wellcomeopenres.17167.2

Van den Akker, O. R., Weston, S., Campbell, L., Chopik, B., Damian, R., Davis-Kean, P., & Bakker, M. (2021). Preregistration of secondary data analysis: A template and tutorial. *Meta-Psychology, 5*. https://doi.org/10.15626/MP.2020.2625

Van Rooij, I., & Blokpoel, M. (2020). Formalizing verbal theories. *Social Psychology, 51*, 285–298. https://doi.org/10.1027/1864-9335/a000428

Vanpaemel, W., Vermorgen, M., Deriemaecker, L., & Storms, G. (2015). Are we wasting a good crisis? The availability of psychological research data after the storm. *Collabra, 1*(1), 3. https://doi.org/10.1525/collabra.13

Vasishth, S., Mertzen, D., Jager, L., & Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language, 103*, 151–175. https://doi.org/10.1016/j.jml.2018.07.004

Vohs, K. D., Schmeichel, B. J., Lohmann, S., Gronau, Q. F., Finley, A. J., Ainsworth, S. E., & Albarracín, D. (2021). A multisite preregistered paradigmatic test of the ego-depletion effect. *Psychological Science, 32*(10), 1566–1581. https://doi.org/10.1177/0956797621989733

Volkow, N. D. (2020). Stigma and the toll of addiction. *New England Journal of Medicine, 382*(14), 1289–1290. https://doi.org/10.1056/NEJMp1917360

Volkow, N. D., Koob, G. F., Croyle, R. T., Bianchi, D. W., Gordon, J. A., Koroshetz, W. J., & Weiss, S. R. B. (2018). The conception of the ABCD study: From substance use to a broad NIH collaboration. *Developmental Cognitive Neuroscience, 32*, 4–7. https://doi.org/10.1016/j.dcn.2017.10.002

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (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. https://doi.org/10.1037/a0022790

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist, 61*(7), 726–728. https://doi.org/10.1037/0003-066X.61.7.726

Williams, A. J., Botanov, Y., Kilshaw, R. E., Wong, R. E., & Sakaluk, J. K. (2021). Potentially harmful therapies: A meta-scientific review of evidential value. *Clinical Psychology: Science and Practice, 28*(1), 5–18. https://doi.org/10.1111/cpsp.12331

Wohl, M. J. A., Tabri, N., & Zelenski, J. M. (2019). The need for open science practices and well-conducted replications in the field of gambling studies. *International Gambling Studies, 19*, 369–376. https://doi.org/10.1080/14459795.2019.1672769

World Health Organization. (2022). *Fact Sheet Tobacco*. https://www.who.int/news-room/fact-sheets/detail/tobacco.

World Health Organization. (2022). *Fact Sheet Alcohol*. https://www.who.int/news-room/fact-sheets/detail/alcohol.

Wright, A., & Woods, W. (2020). Personalized models of psychopathology. *Annual Review of Clinical Psychology, 16*, 49–74. https://doi.org/10.1146/annurev-clinpsy-102419-125032

Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences, 45*, 1–37. https://doi.org/10.1017/S0140525X20001685

Yu, J., Fairbairn, C. E., Gurrieri, L., & Caumiant, E. P. (2022). Validating transdermal alcohol biosensors: a meta-analysis of associations between blood/breath-based measures and transdermal alcohol sensor output. *Addiction, 117*, 2805–2815. https://doi.org/10.1111/add.15953

Ziemann, M., Eren, Y., & El-Osta, A. (2016). Gene name errors are widespread in the scientific literature. *Genome Biology, 17*(1), 177. https://doi.org/10.1186/s13059-016-1044-7