The past, present, and future of Registered Reports

Christopher D. Chambers
Loukia Tzavella

Cardiff University Brain Research Imaging Centre (CUBRIC)
School of Psychology
Cardiff University
CF24 4HQ
United Kingdom

Email: chambersc1@cardiff.ac.uk
Tel: +44 (0) 2920 870331

Abstract

Registered Reports are a form of empirical publication in which study proposals are peer reviewed and pre-accepted before research is undertaken. By deciding which articles are published based on the question, theory, and methods, Registered Reports offer a remedy for a range of reporting and publication biases. Here we reflect on the history, progress and future prospects of the Registered Reports initiative, and also offer practical guidance for authors, reviewers, and editors. We review early evidence that Registered Reports are working as intended, while at the same time acknowledging that they are not a universal solution for irreproducibility. We also consider how the policies and practices surrounding Registered Reports are changing, or must change in the future, to address limitations and adapt to new challenges. We conclude that Registered Reports are promoting reproducibility, transparency and self-correction across disciplines, and may help reshape how society evaluates research and researchers.

After more than a decade of meta-research and debate, the life and social sciences are well in the midst of a credibility revolution[1–3]. Faced with evidence of publication bias[4–7], hindsight bias and selective reporting[8–15], insufficient sample sizes[16], inadequate data sharing[17], and suboptimal rates of both attempted[18,19] and successful replication[20–22], researchers from across a broad range of fields are unifying around a core mission to improve reproducibility and transparency. In doing so, the deeper aim of the openness agenda is to stimulate cultural reform, aligning what is beneficial for individual scientists with what is beneficial for science[23].

As scientists and policymakers grapple with the causes of irreproducibility, it has become clear that one of its main drivers is the so-called *results paradox*. On the one hand, scientists are taught from their earliest years that the one part of the research process they must keep at arms length is the results of their research. The objective investigator – the detective – follows the data with discipline and restraint, never pressuring it to bend to their will, lest they fall prey to Richard Feynman's famous warning that "the first principle is that you must not fool yourself and you are the easiest person to fool"[24] (p. 12). On the other hand, the very same researcher is sent a message from prestigious journals, funding agencies, and evaluation committees, that if you want to succeed in science, be sure to publish a lot of clear, novel, positive findings. Researchers are thus presented with conflicting goals: be a good detective who never indulges in data massaging or cherry picking, but also be a good lawyer who wins arguments and produces a continual supply of beautiful results.[25]

Many observed problems with reproducibility stem from researchers attempting to resolve this paradox while protecting their careers. When publishing in prestigious journals requires confirming one's hypotheses, but the results defy expectations, researchers can resolve this conflict by following the advice of Bem[26,27] and rewriting their hypotheses to "predict" those results — a form of hindsight bias known as "Hypothesizing After Results are Known" (HARKing)[8]. When academic leaders tell authors to "go with strongest studies" because "weak data dilute strong data"[28] (p. 79) and that "what you don't have to do is tell the whole truth...you can select the results you present"[29], the responsive researcher answers by reporting the analyses that tell the best story, diverting negative or inconvenient results to the file-drawer or converting them into publishable (likely false) positives. And when journal editors tell

researchers that, all else being equal, some results are simply more deserving of publication than others, the strategic researcher responds by conducting a large number of small studies and reporting only the most persuasive findings (even if unreliable) rather than gambling on the outcome of larger, more definitive projects that may yield inconclusive data[30].

Registered Reports (RRs) were proposed in 2012 as a way to free researchers from the pressure to engage in these counterproductive practices, thereby breaking the cycle that perpetuates bias and irreproducibility (Figure 1). The RR model originates from the simple philosophy that to defeat the distorting effects of outcome bias on science, we must focus on the *process* and blind the evaluation of science to research outcomes[31]. This blinding is achieved by splitting peer review into two stages. In the first stage, authors submit their research question(s), theory, hypotheses, detailed methods and analysis plans, and any preliminary data as needed. Following detailed review and revision – usually according to specific criteria – proposals that are favourably assessed receive *in principle acceptance* (IPA), which commits the journal to publishing the final paper regardless of whether the hypotheses are supported, provided the authors adhere to their approved protocol and interpret the results in line with the evidence. Following IPA, authors then typically register their approved protocol in a repository, either publicly or under a temporary embargo. Then, after completing the research, they submit a Stage 2 manuscript that includes the approved protocol plus results and discussion, which may include clearly labeled *post hoc* analyses in addition to the preregistered outcomes (i.e., findings from both confirmatory *and* exploratory analyses). The reviewers from Stage 1 and/or newly invited reviewers then assess the completed Stage 2 manuscript, focusing on compliance with protocol and whether the conclusions are justified by the evidence. Crucially, reviewers do not relitigate the theory, hypotheses or methods, thus preventing knowledge of the results from influencing recommendations. RR guidelines specify that editors similarly cannot reject a manuscript on the basis of any new concerns about the methodology or rationale, or on the basis of the results themselves.

[Insert Figure 1 here]

The aim of this modified review process is to reduce as much as possible the potential for biased research practices such as HARKing and selective reporting, while also eliminating the incentive for researchers to employ such practices in the first place. RRs are also designed to mitigate publication bias by journals and outcome bias by reviewers, since the decision to accept or reject is made before results are known. Finally, the format is designed to clearly distinguish the outcomes of pre-planned confirmatory research from exploratory data analysis.
In this review we take stock of the RR initiative, consider its recent history and historical underpinnings, emerging variants, impacts and limitations, and the likely future of the format into the 2020s and beyond. We also offer guidance to authors, reviewers and editors who are becoming familiar with RRs.

Past

Registered Reports as they exist today were first proposed in 2012, independently and simultaneously at two journals: *Cortex* and *Perspectives on Psychological Science*[32,33]. The format was then formally offered at these journals and *Social Psychology* in 2013[34–36]. These first steps precipitated a gradual rise in adoptions, with now over 280 journals across a range of disciplines offering RRs as a new article type (Figure 2). Early launches triggered the rise of RRs into mainstream publishing, but the origins of the format, and of preregistration in general, are much older. As early as 1878, chemist and logician Charles Peirce laid the foundations for preregistration of protocols, writing that "[t]he hypothesis should be distinctly put as a question, before making the observations which are to test its truth" (p. 476)[37]. In the mid-20th century, psychologist Adriaan de Groot further argued that distinguishing exploratory from confirmatory research was vital for scientific progress, and that "it is a serious offense against the social ethics of science to pass off an exploration as a genuine testing procedure" (see Wagenmakers et al.[38] for a detailed historical overview). Embedded in the arguments of Peirce, de Groot, and many others is the maxim that prespecifying predictions and analyses is an important tool for preventing confirmation bias in hypothesis testing.

[Insert Figure 2 here]

Perhaps the earliest proposal for a RR-type review process, in which journal editors would reach editorial decisions based on pre-study or results-blind review, was advanced by psychologist Robert Rosenthal (1966), who wrote: “What we may 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.”[39] (p. 36). Similar ideas were proposed throughout the 1970s and 1980s[40–44] but were not widely implemented. Yet, remarkably and unknown to mainstream science, by 1976 the first RR format had already been launched, albeit in the fringe discipline of parapsychology. For 17 years, the *European Journal of Parapsychology* quietly published RRs alongside regular articles before discontinuing them in 1992[45].

While non-clinical researchers were debating the potential merits of results-blind review, medical researchers were busy weighing up the costs and benefits of public preregistration to address publication bias, particularly in the context of clinical trials. With the US Food and Drug Administration Modernization Act of 1997 (FDAMA) came the first law requiring trial (pre)registration, which in turn led to the launch of the clinicaltrials.gov registry in 2000. By 2005, the International Committee of Medical Journal Editors (ICMJE) was requiring trial registration as a condition of journal publication. An increasing number of journals began offering protocol article types (and in some cases entire journals) with some performing pre-study review of the protocols (e.g. *Trials*, and the *Lancet*'s since abandoned Protocol Reviews format[46]).

Crucially, none of these initiatives, article-types, or journals provided in principle acceptance regardless of the results. Thus, despite decades of debate about preregistration, pre-study review, and results-blind acceptance within isolated channels, it would take until the launch of RRs at *Cortex* and *Perspectives on Psychological Science* in 2013 for the combined model to take hold. From there, the next two years witnessed a gradual increase in the number of adopting psychology journals, followed by the first general STEM journal in 2016 (*Royal Society Open Science*; see Figure 2). After a series of major launches (e.g. *Nature Human Behaviour*, *BMC Medicine, PLOS Biology*) and broader disciplinary expansion throughout 2017-2018, the RR format permanently entered the mainstream.

Present

At the time of writing, Registered Reports are offered by 295 journals, with 483 Stage 2 articles so far published by 81 adopting outlets. With the format becoming more available and associated with a growing published corpus, the first signs of impact are emerging. In this section we review some of the early signs of its effectiveness. We also introduce recent variants of the model, summarise the key ingredients that make a high-quality RR, address some major misconceptions (Table 1) and genuine limitations that have emerged, before offering specific recommendations to authors (**Box 1**), reviewers (**Box 2**) and editors (**Box 3**).

Field spread and author demographics

Since their initial launch within psychology and neuroscience, RRs have spread to specialist journals covering a range of disciplines, primarily in the life and social sciences (Figure S1.1). As this reach has grown, we can begin to explore the demographics of submitting authors to discover the accessibility of the format. The prospect of being pre-accepted at a respectable journal is likely to be appealing to many researchers – and perhaps early career researchers (ECRs) in particular – seeking to eliminate the risk that the results of their research will determine publication and, consequently, their career prospects. On the other hand, the often substantial sample sizes needed for RRs to achieve minimum levels of statistical power required by many journals, combined with the time taken for Stage 1 review (see **Limitations and drawbacks**), could act as a deterrent, especially for researchers with major resources constraints. To provide a preliminary insight on accessibility, we analysed the author demographics of 141 Stage 1 RRs submitted to *Cortex*, *European Journal of Neuroscience*, *NeuroImage* and *Royal Society Open Science*. We found that 77% of submitted Stage 1 manuscripts were first-authored by PhD students or postdocs (Figure S1.2a). At the journal *Cortex*, where a direct comparison between different articles types was possible, we found that 78% of submitted RRs were led by ECRs compared with 67% for a comparison sample of regular articles (Figure S1.2b/c). It would be premature to conclude that RRs present no barriers for researchers[47], but these results at least provide no reason to fear that RRs are beyond the reach of ECRs.

Early impacts

Are RRs working as intended to reduce bias and improve reliability? Although the initiative is too young to answer this question with confidence, metascientific investigations are beginning to reveal signs of bias control, study quality, computational reproducibility, and citation influence.

Bias control. Since reporting and publication biases typically favour positive results, RRs, if successful, should yield a greater proportion of negative results compared with the conventional literature. So far this prediction appears to hold: a recent analysis of 296 hypotheses published across 127 RRs in different fields found that 60% of RRs report null results, approximately five times greater than the rate in regular articles[48]. In psychology this difference is even more striking, with a new study finding that just 4% of regular articles failed to confirm the first hypothesis compared with 56% for RRs[49] (see also Wiseman et al.[45]). It would be tempting to conclude that this increase is caused by the elimination of selective reporting, HARKing and publication bias resulting from pre-study review and IPA – the key ingredients of the RR process. On the other hand, it is possible that authors, knowing that their study will be published regardless of whether their hypotheses are supported, might employ the RR format to test riskier hypotheses. Moreover, RRs themselves might select for authors who are diligent in controlling their own reporting bias, regardless of the article type. To address such confounds, future observational studies could compare the plausibility of hypotheses in RRs compared with non-RRs, as well as indicators of biased reporting in RRs and non-RRs within the same sets of authors (see **Future**).

Computational reproducibility. There are several reasons why RRs might be more computationally reproducible than conventional articles. At many journals, the RR review policy has more stringent expectations concerning open data and code, which is associated with greater accuracy in statistical reporting[50]. In addition, IPA eliminates the incentive for authors to conceal messy or inconvenient elements of their data, and early adopters of the format may also be predisposed to performing research to a higher level of transparency. A recent study indeed suggests that the results of RRs can be more readily reproduced from the acquired data compared with

regular articles[51]. Of 35 RRs published in psychology that made data and code openly available, 57% were computationally reproducible compared with 31% in a previous analysis of regular articles[52]. Although RRs appear to perform better than the status quo, these results clearly show room for improvement – and require substantially more data to be confirmed.

Citation profile. Clinical trials reporting negative results receive between 2 to 10 times fewer citations than trials reporting positive results[53,54]. Given the increased rate of negative results in RRs, authors may therefore be concerned that submitting a RR could be disadvantageous to their careers; similarly one of the immediate reactions to RRs from many journal editors is that the format could risk reducing their outlet's impact factor, a powerful albeit spurious measure of research influence[55–58]. In fact, such concerns may be unwarranted – a recent analysis of 70 RRs reported in a non-peer-reviewed preprint found that RRs are cited the same or slightly higher than comparable regular articles[59].

Study quality. How do expert assessments of RRs compare with non-RRs? In a recent non-peer-reviewed preprint, Soderberg et al. reported an experiment in which 353 scientists rated a sample of published, partially-blinded RRs and non-RRs on 19 study characteristics, including importance, novelty, creativity, innovation and rigour.[122] RRs numerically outperformed RRs on every criterion, showing statistically robust and large improvements in attributes such as methodological rigour and overall article quality, while being statistically indistinguishable from comparison papers in terms of features such as novelty and creativity. These results held even among reviewers who admitted being skeptical or neutral about RRs. .

Emerging variants

As the reach of RRs has grown, several modified versions of the format have arisen to accommodate specific needs. Five major strands in particular have emerged, including results-blind review, accountable replication policies, RRs involving post-publication peer review, publisher-level RRs, and publisher-independent RRs. These variants are summarised briefly below and discussed in more detail in Supplementary Information.

Results-blind review. In this modified workflow, Stage 1 peer review is undertaken after results are known to the authors, but before they are known to the reviewers and editors. Following IPA, the authors then submit the full manuscript containing the data and conclusions. Because authors need not wait until IPA to conduct their research, this format prevents the Stage 1 review time from delaying data collection and analysis. On the other hand, reviewers are unable to improve the study design, and the format does not prevent reporting bias (e.g. *p*-hacking, HARKing) by authors. To date, at least 13 journals, primarily in psychology and management, have adopted results-blind review as an optional article track. Two journals – Cortex and Infant and Child Development – have also launched Verification Reports, a results-blind format dedicated to assessing the computational reproducibility and robustness of previous findings based on reanalysis of the original study data.

Accountable Replications. Conceived by psychologist Sanjay Srivastava[61], this variant emerges from the principle that when a journal publishes a research finding it should commit to publishing all methodologically-sound replications of that finding, regardless of how the results turn out, and regardless of subjective importance or methodological flaws in the original study. Using a modified set of the RR assessment criteria, the journal reaches a Stage 1 IPA decision on the basis of technical validity and the methodological proximity between the replication and target study. To date, Royal Society Open Science is the only journal that implements a complete and fully specified version of this concept, following partial implementations at Clinical Psychological Science, Journal of Research in Personality, and Psychological Science.

Post-publication peer review RRs. RRs usually rely on conventional pre-publication review in which reviewers serve as gatekeepers to IPA and Stage 2 acceptance. In contrast, by combining post-publication peer review (PPPR) with RRs,the Stage 1 manuscript is published almost immediately following initial receipt and is then openly reviewed.[68] If the reviews are positive (with authors having the usual opportunity to revise the protocol) then the article is awarded IPA and, once passing Stage 2 review, the final manuscript with results is badged as a RR. To date, this model has been adopted across ~10 journals, including F1000Research and Wellcome Open Research.

Publisher-level RRs. The review process for RRs is typically managed by the one journal but recently some journals have begun implementing a distributed model in which Stage 1 and 2 manuscripts can be reviewed and published in different journals under the same publisher. In one working model, the completed Stage 2 RR is then cross-linked to the accepted Stage 1 protocol using an International RR Identifier.[49]

Publisher-independent RRs. Can RRs exist beyond journals? As part of the recently created Peer Community in Registered Reports (PCI RR), Stage 1 and 2 preprints are reviewed independently of journals (https://rr.peercommunityin.org). Where the reviews are positive, PCI RR issues a positive recommendation and authors can then choose to publish their recommended preprint in any "PCI RR-friendly" journal without further peer review.

Seven virtues of high-quality Registered Reports

What makes a good RR? In this section we describe seven desirable characteristics that authors should aim to capture in their Stage 1 and Stage 2 manuscripts. Further guidance may be found in **Box 1**, in the RR policies for specific journals (see list at https://cos.io/rr/) and in a recent practical primer by Kiyonaga and Scimeca [70].

The **first** and foremost ingredient is that the proposal tackles a scientifically valid question, and ideally one that other scientists agree is important to answer. The Introduction section of the Stage 1 manuscript should make clear the underlying theory or application from which the question arises, leaving the reader in no doubt as to why the study is being proposed. **Second**, where the study proposes hypotheses, they should be stated as precisely as possible in terms of specific variables to ensure falsifiability. In quantitative hypothesis-driven sciences, we recommend that researchers consider the open theory pathway proposed by Guest and Martin[71] to help ensure that hypotheses are formulated as a natural specification of computational theory rather than emerging loosely – and often with questionable rationale – from a vague conceptual framework. Some of the most effective RRs achieve this by identifying pressure points in competing theories and then devising hypotheses to adjudicate between them.

With the theory, rationale and hypotheses in place, the **third** key ingredient is a study procedure and analysis plan that is as rigorous, transparent, and comprehensive as possible. Data acquisition protocols and analysis plans should be prespecified with sufficient detail to be reproduced by experts in the field, ideally with accompanying code, and with rigorous experimental controls where appropriate – including both negative and positive controls. Where the conclusions will depend on inferential statistics, the procedure should include a detailed sampling plan, such as a statistical power analysis, Bayes factor design analysis [72], or an appropriate alternative, which – crucially – should also make clear the specific hypothesis it interrogates and the rationale for deciding the sensitivity of each statistical test (such as justification of the target effect size or Bayesian prior). When planning the analyses, it is vital to choose the right tools for the job, including assumption checks, detailed consideration of data preprocessing and filtering, and planned contingencies for any data-driven analysis decisions. Where these contingencies would be too numerous (or even impossible) to specify in advance, the inclusion of pilot data at Stage 1 can be used to verify assumptions and narrow the range of possibilities. Alternatively, authors can embrace uncertainty and use blinded analysis methods to control risk of bias.

These three features are the essential building blocks for the **fourth** key ingredient: a seamless link between the research question, theory and its specification, the hypotheses, sampling plan, and contingent interpretation given different outcomes (see **Box 1**). A Stage 1 RR can be thought of as a preparatory chain of inference, leading from the why to the what and how, which as with any chain, is only as strong as its weakest link. Many RRs now include study design tables to elucidate these links as precisely as possible.

Fifth, once the study is completed and results are known, it is essential that the outcomes of the prospective (confirmatory) analyses are clearly distinguished from the outcomes of any post hoc (exploratory) analyses that deviated from the preregistered plans. While RRs are not intended to restrict valid deviation or post hoc exploration, it is vital that at this final hurdle the outcomes that were decided after observing data, and thus potentially bias-prone, are not conflated with those that were protected from bias by prespecification. Clear differentiation of exploratory and confirmatory outcomes, in turn, furnishes the **sixth** key ingredient: ensuring that the

conclusions of the RR are based firmly on the evidence presented and appropriately weighted in favour of the confirmatory outcomes. **Finally**, in line with Level 2 of the Transparency and Openness Promotion Guidelines [73], the seventh key ingredient of a high-quality RR is that study data, code and digital materials are made publicly available to the maximum extent permitted by relevant ethical or legal constraints.

[Insert Table 1]

Limitations and drawbacks

Despite their advantages, RRs are neither a panacea nor a one-size-fits all solution for irreproducibility. As a tool for improving the quality of confirmatory research, they are particularly well suited to hypothesis-driven studies and are not designed to improve the robustness or transparency of purely exploratory science (for which better suited article types are available[74–76]). As the format has evolved, various shortcomings have also been revealed in the workflow and implementation.

Lack of protocol transparency. In 2018, Hardwicke et al. reported that of 70 journals that had adopted RRs permanently, only 50% required that the accepted Stage 1 protocols were publicly registered and available alongside the completed Stage 2 articles[77]. Protocol transparency is an important element of RRs because it enables readers to check whether authors followed the approved protocol, rather than relying on the (typically) closed review process to ensure compliance[78]. One reason for lack of protocol transparency is that, in 2013, the key progenitor of RRs, *Cortex*, did not require accepted Stage 1 protocols to be made public, and this policy omission was then duplicated among subsequent adopters. However, since Hardwicke et al.'s analysis, the recommended author guidelines for RRs at *Cortex* and the Center for Open Science have been updated to include protocol transparency[79], and to date, of 213 permanent adopters with published RR policies, 87% now either publish the Stage 1 protocol as a separate article or require Stage 1 protocols to be registered and made public no later than the point of Stage 2 acceptance (See Supplementary Information for details). It remains an ongoing task to persuade all RR-adopting journals to require protocol transparency, and a key aim of future metascience will be to confirm that journals are enforcing their policies appropriately (**Box 4**).

Lack of standardisation. Previous analyses by Hardwicke et al.[77] and Scheel et al.[49] show that RRs are registered and reported inconsistently and, in many cases, even lack sufficient information to determine the specific hypotheses. This lack of specificity likely arises from the incompatibility between the 17th-century traditional manuscript format – involving discursive and often vague prose documentation – and the demand for precision within RRs. An RR should articulate falsifiable predictions that are linked to specific sampling plans, inferential analyses, and contingent interpretations given different outcomes.

Stage 1 delay and bureaucratic tennis. Despite the fact that RRs might, in aggregate, lead to more efficient knowledge generation compared with regular articles (see Table 1), the fact remains that the Stage 1 review time typically adds a period of several months between submitting a Stage 1 manuscript and the commencement of the research. This downtime can present a significant barrier for researchers on short-term contracts or who hold grants that demand immediate data acquisition. Furthermore, in fields that require very specific ethical approval, such as the clinical sciences, authors can find themselves locked into a time-consuming tennis match between the journal and their ethics committee, both of which can insist on approving a precisely specified protocol (Figure 3).

[Insert Figure 3]

Box 1. Top tips for authors

The planning stage

1. RRs famously champion the rigor of the proposed methodology and analyses, but the specificity and importance of the research question(s) should also be evaluated during the planning stage. Study plans should first target specific, clear research questions and respective hypotheses (if applicable), upon which robust methods and analyses can then be developed.
2. The feasibility of the proposed study should be considered as early as possible during the planning stage. A well-justified sampling plan that meets journal requirements (e.g. 90% statistical power) should be formulated depending on

time and resource constraints. Hypotheses and statistical models may need to be adjusted accordingly.

3. The feasibility and validity of the proposed methods and analyses should be assessed before submission in order to maximise the efficiency of the peer review process and minimise the need for deviations from the Stage 1 protocol following IPA. This can be challenging given that researchers have often not yet acquired data. For this reason, piloting or data simulations are highly recommended.

The submission stage

1. Authors should ensure that there are precise and exhaustive links between each research question, hypothesis, sampling plan, analysis plan and contingent interpretation depending on different outcomes. An RR that minimises researcher degrees of freedom may receive IPA without the need for major corrections, given that the robustness of the methods has been established. The use of pre-study coding protocols can help achieve maximum cohesion (e.g. see https://osf.io/6bv27/).
2. Authors should formulate analysis plans that take into account all the steps in data pre-processing (e.g. exclusions, cleaning, aggregation) and analyses (e.g. statistical model assumptions). If such details are impossible to specify then authors should propose a more general plan in combination with blinded analysis methods[80], for instance as deployed in a recent RR by Dutilh et al.[81].
3. The Stage 1 protocol should, where possible, include outcome-neutral checks for ensuring that the proposed hypotheses are capable of being tested, such as positive/negative controls, tests of intervention fidelity, or data quality checks. These conditions need to be successfully met at Stage 2 and therefore authors should carefully consider which, if any, such conditions can be preregistered at Stage 1.

Box 2. Top tips for reviewers – questions to ask during Stage 1 and Stage 2 assessment

At Stage 1

1. Do the research questions and proposed hypotheses make sense in light of the theory or application? Are they defined precisely? Are the hypotheses capable of answering the research question?
2. Is the protocol sufficiently detailed to enable replication by an expert in the field, and to close off sources of undisclosed procedural or analytic flexibility?
3. Is there an exact mapping between the theory, scientific hypotheses, sampling plan (e.g. power analysis, where applicable), preregistered statistical tests, and possible interpretations given different outcomes?
4. Where relevant, does the power analysis (or alternative sampling plan) reach the minimum threshold required by journal policy (e.g. 90% power, BF>6)?
5. Does the sampling plan for each hypothesis propose a realistic and well justified estimate of the effect size?
6. Have the authors avoided the common pitfall of relying on conventional null hypothesis significance testing to conclude *evidence of absence* from null results? Where the authors intend to interpret a negative result as evidence that an effect is absent, have authors proposed an inferential method that is capable of drawing such a conclusion, such as Bayesian hypothesis testing[72,82] or frequentist equivalence testing[83]?
7. Have the authors minimised all discussion of exploratory analyses, apart from those that must be explained to justify specific design features? Maintaining this clear distinction at Stage 1 can prevent exploratory analyses at Stage 2 being inadvertently presented as pre-planned.
8. Have the authors clearly distinguished work that has already been done (e.g. preliminary studies and data analyses) from work yet to be done?
9. Have the authors prespecified positive controls, manipulation checks or other data quality checks? If not, have they justified why such tests are either infeasible or unnecessary? Is the design sufficiently well controlled in all other respects?
10. When proposing positive controls or other data quality checks that rely on inferential testing, have the authors included a statistical sampling plan that meets the minimum requirement in terms of statistical power or evidence strength (if there is one)?

At Stage 2

1. Did the authors formally preregister their Stage 1 protocol and have they provided a direct URL to the approved protocol in the Stage 2 manuscript? Did they stay true to their protocol? Are any deviations from protocol clearly justified and fully documented?
2. Is the Introduction in the Stage 1 manuscript (including hypotheses) the same as in the Stage 2 manuscript? Are any changes transparently flagged?
3. Did any prespecified data quality checks or positive controls succeed?
4. Are any additional *post hoc* analyses justified, performed appropriately, and clearly distinguished from the preregistered analyses? Are the conclusions appropriately centered on the outcomes of the preregistered analyses?
5. Are the overall conclusions based on the evidence?

Do's and don'ts

- **Do** suggest additional exploratory analyses at Stage 2 but **don't** expect the editor to necessarily require the authors to conduct them. Authors are not obliged to conduct any unregistered analyses unless such tests are necessary to support conclusions that go beyond the preregistered analyses – and these must be conclusions that the *authors* (not reviewers or editors) wish to draw. This protection exists to prevent goal-post shifting and subtle forms of publication bias from affecting the Stage 2 process.
- If you find a flaw in the protocol at Stage 2 that was missed or unaddressed at Stage 1, **do** mention it but **don't** expect the manuscript to be rejected on that basis. It is important to remember that the protocol is not generally subject to re-review at Stage 2, and editors cannot require authors to do extra studies. Barring rare cases in which the editor, authors and reviewers agree that a severe error was made, the most that will likely happen is that the authors will be asked to address potential design limitations in the Discussion.

Box 3. Advice for journal editors considering offering RRs

Are RRs appropriate for my discipline? RRs are potentially useful in any field where at least one of the following problems exist: publication bias, hindsight bias (HARKing), selective reporting of desirable results by authors (including but not limited to *p*-hacking), lack of sufficient sample sizes to draw meaningful conclusions,

or lack of close replication. Not all fields experience these problems to the same extent. For example, many sciences do not employ null hypothesis significance testing or even inferential statistics. Other fields, particularly in the physical sciences, have a strong culture of replication. Publication bias, however, exists across almost all sciences[4]. Therefore, in considering whether to offer RRs, the first question an editor should ask is: Would two studies in my field that addressed the exact same question and were conducted to the exact same methodological standard have the same chance of being published in my field's most prestigious journals if one study showed clear, persuasive results while the other study showed null or ambiguous results? If the answer is no – and the study showing more compelling results would have a greater chance of being accepted – then RRs should be offered as a means to reduce publication bias.

Can RRs be suitable even when authors are not testing hypotheses? Yes. It is true that almost all RRs test hypotheses, however the format can also have value in an observational setting where publication bias could either suppress certain outcomes or encourage biased analysis or interpretation by authors. For example, a marine biologist may be interested in measuring the concentration of fish, plankton and particulate organic matter to assess the overall health of an ecosystem. Even though the researcher has no specific hypotheses, by detailing the data acquisition method, analysis techniques, and which outcomes will lead to which conclusions, the researcher can control their own analytic and interpretive bias, and by incorporating pre-study review and IPA, they eliminate risk of publication bias.

Are RRs suitable only for replication studies? No. While RRs are well suited for replications (especially when the review criteria are modified to ensure accountability; see Supplementary Information), approximately 50% of the hypotheses in published RRs arise from original studies[48].

If a journal offers RRs, does the RR format need to replace the traditional research article format at the journal? No. So far, virtually all new adopters have added RRs as a new article type, with the exception of *Comprehensive Results in Social Psychology* and a small number of additional journals/platforms which were created as RR-only outlets. Given current limitations in the pace of Stage 1 review

(see **Limitations and drawbacks**, and **Future**) there will always be a place for hypothesis-driven research outside the RR format, in addition to pure exploratory research.

Will RRs lower my journal's impact factor? Probably not. Preliminary evidence suggests that RRs are cited the same or slightly higher than comparable non-RRs[59]. Although impact factor carries a currency with many journals and publishers, the prevailing evidence suggests that it does not contain information about scientific quality[55–58].

How should editors triage initial Stage 1 submissions? Many journals assess Stage 1 manuscripts according to specific criteria (see Supplementary Information). To maximise the efficiency of the review process, we recommend that editors always perform a desk assessment against these criteria (and see also **Box 2**) to ensure that a submitted manuscript avoids common pitfalls before being sent for specialist peer review.

Would editors be required to accept any methodologically sound protocol, regardless of the importance or relevance of the research question? No. Many journals assess the subjective value of the research as part of Stage 1 assessment. For instance, at PLOS Biology, Stage 1 Criterion 1 is "The importance of the research question(s)"[84], while *Nature Human Behaviour* judges "The importance of the research question(s) and relevance for a broad, multidisciplinary audience"[85]. Other journals place less emphasis on such judgements; for example, at Cortex and *Royal Society Open Science*, the "importance" of the research question is replaced with an assessment of the "scientific validity of the research question"[86,87]. Each journal is free to determine how selective it wishes to be at Stage 1. The main requirement is that any such selectivity is applied before the results are known and communicated transparently to authors in the journal policy.

Would a journal be obligated to publish the results of a RR that appeared promising at Stage 1 but was conducted to a low standard? No. The Stage 1 review process allows reviewers and editors to prespecify positive controls, manipulation checks and data checks for assessing quality of implementation (e.g.

data verifying that a particular intervention or measure was administered appropriately). To prevent publication bias, the only requirement is that such tests are prespecified at Stage 1 before results are known, and that they are independent of the primary outcome measures and main study hypotheses. Stage 2 rejection will be very rare, e.g at *Cortex*, *European Journal of Neuroscience* and *Royal Society Open Science*, where C.D.C. is RR editor, the Stage 2 rejection rate is currently zero.

How complicated and arduous is the process of adding RRs to a journal? Installing RRs has become increasingly easy over time. With around 300 journals now offering them, all major publishers have at least one adopter under their umbrella. In many cases, the format and workflow can be imported very easily between journals. The Center for Open Science hosts dedicated Resources for Editors to assist in implementation[88]. Journals can also offer RRs by joining the Peer Community in Registered Reports (PCI RR), and, if they wish, use the review infrastructure of PCI RR in place of establishing an internal infrastructure (https://rr.peercommunityin.org/).

Do RRs create more work for editors compared with regular articles? Editing RRs requires careful attention to study rationale and methodology, ensuring that authors adhere to the journal's RR policy, guiding them on how to best address (the sometimes competing) recommendations from reviewers, and making certain that authors do not feel compelled to comply with requests from reviewers that violate RR policy. An RR editor must therefore be highly engaged with each submission and read the article. Insofar as these practices define the minimum requirements for competent editing, the RR process will produce a comparable editorial workload. At the same time, it is important to note that RR submissions undergo a minimum of two phases of review, at Stage 1 and Stage 2, whereas a regular article is subjected to a single review process. Whether this bifurcation of review for RRs translates to increased editorial workload is unclear, and is likely to depend on the extent to which each RR replaces a regular article that the authors would otherwise have submitted, and the comparative rejection rate of RRs vs regular articles at the journal. If, as observed informally (see Table 1), the rejection rate for RRs is lower than for regular articles, and if each RR replaces at least one regular article that the authors would have submitted, then RRs may still reduce overall editorial workload by reducing the

number sequential journal submissions and thus the duplication of the review process across journals.

Are RRs possible in fields where researchers analyse existing datasets? Yes. The majority of adopting journals invite RRs proposing secondary analyses of existing datasets, provided authors take sufficient steps to minimise risk of bias and analytic overfitting. Such measures could include avoiding all prior observation of the data or key variables, proposing key analyses in an unseen holdout sample, or recruiting a blinded analyst.

Could RRs be useful in qualitative research? Yes. To the extent that publication bias is a concern in both qualitative and quantitative research[89], RRs hold promise for improving reproducibility. To date, however, only a handful journals currently offer qualitative RRs (e.g. *BMC Medicine* and 13 more; see column 20 at https://tinyurl.com/RRpolicylist) and none include specific RR policy text tailored for qualitative methods. The development of qualitative protocols for preregistration and RRs is a key area for future innovation[90–93].

Future

What does the next decade and beyond hold for RRs? The gradual rise of the format has unlocked a range of possibilities for expansion and innovation, while also posing challenges for implementation and quality control. Here we consider some of the major possible developments as RRs scale up. We also reflect on the key outstanding questions for metascience (**Box 4**) and consider how RRs may influence systems for evaluating research and researchers.

Improving efficiency

Perhaps the greatest limitation of the RR format is the time taken for submissions to be reviewed at Stage 1 and receive IPA, thus delaying the commencement of research (see **Limitations and drawbacks**). While it can be argued that elimination of publication bias offsets this cost at a community level, and furthermore that the Stage 1 delay could improve quality by increasing startup costs[94], this downtime nevertheless reduces accessibility of the format to individual researchers and can

make it prohibitive for short-term projects. Here we consider four innovations that could significantly accelerate Stage 1 review without reducing quality.

Rapid review. One way to accelerate RR review is to create a network in which reviewers agree to evaluate submissions within a short time frame. In 2020, Royal Society Open Science became the first journal to launch such a network for RRs related to the COVID-19 pandemic[95]. As part of this special initiative, the journal calls for submissions that are relevant to any aspect of COVID-19 in any field, including biological, medical, economic, and psychological research, while also seeking specialist reviewers who are willing and able to evaluate Stage 1 RRs within 24-48 hours of accepting a review request. To date, nearly 900 scientists across a range of disciplines have joined the reviewer network, which is also accessible to other journals. To gain access, the journal must commit to rapid peer review of COVID-19 RRs – striving for 7 days for the initial Stage 1 review round – and waive all article processing charges. Since then, 11 additional journals have joined the network, including *Nature Human Behaviour*, *Nature Communications* and *PLOS Biology*. To date, *Royal Society Open Science* has published six Stage 2 RRs arising from the initiative, and additional submission statistics are available in Supplementary Information[96–101].

Scheduled review. To date, RRs at all journals are subjected to the same serial review process as regular articles. At Stage 1, the manuscript is received and undergoes editorial triage; if it meets minimal requirements, editors then seek and obtain specialist reviews, ideally leading to IPA following revision and, in many cases, re-review (Figure 4a). Despite prompt engagement by editors and reviewers, this process can take several months to achieve resolution. An alternative approach is to parallelise key elements of the initial Stage 1 assessment (Figure 4b). Under this model, authors initially submit a short, structured protocol for consideration prior to writing the Stage 1 manuscript. If this passes editorial triage then reviewers are invited to assess a full Stage 1 manuscript at a fixed date in the future (e.g. 6 weeks ahead). During this time, the authors write and submit their complete Stage 1 manuscript, which is then reviewed on the scheduled date or during a short range of dates. With sufficient contingencies in place, this modified review process could

reduce the *initial* Stage 1 review time (but not re-review time) from weeks/months down to a matter of days.

[Insert Figure 4]

Observer-evaluator review and "rolling IPA". A more radical alternative to accelerated review and scheduled review would be to abolish the current peer review system altogether, replacing the serial assessment of documents – arguably a throwback to the 17th century exchange of letters – with a more dynamic "observer-evaluator" mechanism. Authors could use existing infrastructure such as the Open Science Framework (and associated add-ons) to create a virtual lab space containing the project rationale, study protocol(s), code, and data as applicable. Reviewers could then be parachuted in as virtual observers, monitoring, commenting and approving specific components as the research unfolds in real-time, with the editor providing guidance, monitoring and oversight. This system could make RRs more compatible with the rapid sequential workflow that is common in fields such as chemical biology, virology and psychophysics, where the results of one experiment often lead within days to the design and implementation of the next experiment. As each update to the protocol is approved, IPA could be rolled over and extended.

RR funding models. One of the major barriers to RRs in clinical research, especially, is the additional bureaucracy imposed by Stage 1 review. Many researchers already face multiple pre-study hurdles, including grant review, ethics review, and in some cases regulatory review, and all before even contemplating a Stage 1 RR submission. One promising solution to this problem is for journals and funders to perform concurrent or near concurrent review of RR proposals. Under this partnership model, which was first trialled in 2017[102,103], authors submit a Stage 1 proposal to a journal and funder, either simultaneously or in succession. Following assessment by both parties (either separately or as part of a joint process), if expert reviews are favourable then IPA and funding are awarded in synchrony, reducing two pre-study review phases into one. This mechanism could be further enhanced by incorporating ethics and regulatory review, which would be particularly useful for clinical trials where these phases of review often require the assessment of a detailed and precise protocol.

Improving quality, accountability and rewards

Alongside improvements in efficiency, the next decade is likely to see a range of measures to optimise the openness and reliability of both RRs and the assessment of RRs. In this final section we consider six key innovations including computationally-generated RRs, mandatory RRs for clinical trials, the development of an RR training and accreditation process for journal editors, tools for monitoring the speed and quality of journal assessment, ways to improve the recognition of reviewer contributions, and emerging steps to incorporate RRs into formal research evaluation.

Computationally-generated RRs. Quality control of accepted protocols is limited by the subjectivity of the review process and variable implementation of RR criteria across journals and fields. As noted previously (see **Lack of standardisation**), the specificity of hypotheses, statistical tests and interpretations may not always be achievable with the current format. It may therefore be unrealistic to expect authors to achieve the required level of precision without tools that can guide them through this process, such as RR study design templates or software (e.g. the ScienceVerse project developed by Lisa DeBruine and Daniel Lakëns https://scienceverse.github.io/scienceverse/). For RRs that test specific confirmatory hypotheses, all preregistered hypotheses and statistical predictions could also become machine-readable[104]. A machine-readable output can be used to extract key metadata from the RR, as the results either confirm or disconfirm the prespecified hypotheses. Evaluating a Stage 2 manuscript on the basis of adherence to preregistered statistical predictions would then become a more efficient and standardised process for the reviewers and editors.

Several preregistration templates are available for assisting researchers in communicating their study plans in a concise and structured format[105], and some journals also implement their own protocol-coding checklists with strict criteria for IPA[106] (e.g. https://osf.io/6bv27/). These template and checklists and can be time-consuming to complete on top of manuscript preparation, therefore an essential innovation for RRs will be the creation of a user-friendly web-based RR generation tool that guides the authors through the implementation of stringent criteria, including

precise specification of hypotheses and linking with sampling plans, analysis plans, inference criteria, and contingent interpretation given different outcomes. The tool would ideally accommodate a wide range of disciplines, similar to the Experimental Design Assistant offered by NC3Rs (https://www.nc3rs.org.uk/experimental-design-assistant-eda), producing a standardised, submission-ready Stage 1 protocol.

Mandatory RRs for clinical trials. In basic (non-clinical) research, RRs are usually proposed as an additional option for authors rather than a requirement, which makes sense given the vital importance of exploratory science. However, we believe a strong case can be made for all clinical trials to be conducted and reported exclusively as RRs. Even though trial registration is now the norm, registration does not guarantee that trials are preregistered rather than "post-registered"[107,108], that trial results will be reported free from bias[109,110], or that the results will be published at all[111,112]. With trials being vulnerable to all the same publication and reporting biases that afflict basic research, and with the first RR model for clinical trials now available at *BMC Medicine*[113], the next decade will hopefully see mounting pressure on clinical trial funders and major medical journals to embrace the format, ideally via RR funding models to maximise efficiency.

Training and accreditation for editors. As a general rule, the standard with which a journal reviews and administers RRs can never exceed its standard of RR editing, and for this reason it is crucial that editors have the required skills and training. Guidelines to assist editors in evaluating Stage 1 RR submissions are available in most RR-adopting journals or at the Center for Open Science (https://cos.io/rr/), and the Peer Community in Registered Reports initiative requires new editors (called "recommenders") to pass a 2 hour entrance test. The expansion of RRs into new disciplines and less familiar terrain is bound to introduce variability in the standard of editing. Therefore, an important future step for increasing and standardising the quality of RR editing will be to provide editors with training materials that are tailored for the busy schedules of academics and possibly using a massive open online course (MOOC). In this way, editors, and perhaps entire journals, could receive accreditation for their knowledge and understanding of the RR process, including criteria for manuscript acceptance/rejection at Stage 1 (IPA) and Stage 2. This

system would equip authors with data to inform their choice of journal, and readers with the confidence in the standard of RR editing across journals.

Community monitoring and feedback. Related to the issue of editorial training is the issue of journal accountability. At present, there is little information available for authors to judge the quality of editing and review at an RR-adopting journal, apart from word of mouth and conventional (likely uninformative) indicators such as journal prestige[55]. We believe journals should regularly publish all data on the number of RR submissions received, rejection rates at the different stages, and time spent under review. In addition, open review policies and a Yelp-style website in which authors and reviewers could leave anonymous feedback ratings on the quality of the editorial process would provide an incentive for journals to maintain high standards.

Reviewer recognition. Reviewers often make significant contributions to RRs that are not transparently recognised. During Stage 1, reviewers can recommend major changes in the study design, hypotheses, methods and analyses - contributions that would readily justify authorship if made outside the review process. One way to ensure that reviewers are properly credited for their contribution to RRs is for a journal to adopt an open, signed review policy. Although most journals employ closed, anonymous peer review, a minority of RR-adopting journals such as *Royal Society Open Science* and *Meta-Psychology* publish the accepted article alongside the reviews, which reviewers can sign to increase transparency, credit and accountability. Reviewers can further list their contributions in online platforms, such as Publons (https://publons.com/). We believe it is important for reviewers, and especially ECR reviewers, to have publicly available evidence demonstrating the quality of their reviewing. To further recognise this contribution, one possibility would be to create a “reviewer contributor” role that formally acknowledges the intellectual input of reviewers to the final RR, without being named as an author. This role could also be recognised through use of the CRediT taxonomy (https://casrai.org/credit/).

Research evaluation. To become normative in the long term, RRs will need to be recognised within formal systems for evaluating research quality. In the UK there are already promising moves in this direction. In the 2021 Research Excellence Framework – the regular national exercise for assessing research quality and

apportioning public funds – RRs are specifically noted as an indicator of research rigour[114], which in turn means that authors who publish RRs could attract increased funding for their institutions. Following the recent formation of the UK Reproducibility Network (http://www.ukrn.org), institutions are also signalling their support for RRs[115]. University College London "strongly encourages" researchers to use RRs where appropriate[116], an approach echoed by learned societies including the British Neuroscience Association[117] and British Psychological Society[118]. The Norwegian funder, Stiftelsen Dam, also recommends that grantees consider publishing their research in the form of RRs[107] while the Templeton World Charity Foundation goes so far as to mandate RRs for certain funding schemes[120]. The next five years will hopefully see international expansion in the recognition of RRs at all stages of evaluation, from research outputs and grant applications to the criteria for employment and promotions. It is crucial that such judgments are applied cautiously with continual reference to ongoing metascience that will establish evidence of the costs and benefits of the format (see **Box 4**).

Box 4. The big questions for metascience

1. How exactly do RRs differ from regular articles, both quantitatively and qualitatively?[121] Are the method sections longer, more detailed, and more easily repeatable? Are the sample sizes larger, as expected by the high statistical power requirements at many adopting journals? Following Soderberg et al., do blinded expert raters judge RRs to be of higher quality than comparable regular articles[122]? Are RRs more likely to include open data and materials?

2. Are the results reported in RRs more likely to replicate when the studies are independently repeated? Are the outcomes of the confirmatory preregistered analyses in RRs more likely to replicate than the outcomes of *post hoc* exploratory analyses?

3. To what extent do Stage 2 RRs deviate from their accepted Stage 1 protocols? Are such deviations always explicit?

4. Is the review process for RRs, on average, more or less efficient than the review process for regular articles? Across all journals, is the acceptance rate higher or lower for RRs?

5. Using a randomised controlled trial, can we conduct a definitive causal test of the hypothesis that the key ingredients of RRs - pre-study review and IPA - reduce publication bias and reporting bias?

6. Are RRs more likely to include "single-shot" studies (*vs.* sequences of studies) compared with regular articles?

7. Are RRs having beneficial collateral effects on open research policies or practices? For example, is the adoption of RRs at a journal associated with rising standards of reproducibility and transparency for regular articles within the same journal? Are there any detrimental effects?

8. Are journal editors adhering to their own RR policies? Are Stage 1 or Stage 2 manuscripts ever rejected for reasons that violate RR policy? Are journals requiring that Stage 1 protocols are formally registered and linked to the Stage 2 manuscript?

9. What are the career costs and benefits of pursuing RRs for individual researchers? Does publishing a RR lead to changes in a researcher's attitude to their own work or the work of their field? Are there any signs of RRs influencing career opportunities? Once authors have published their first RR, do the *regular* articles they subsequently publish demonstrate reduced risk of bias compared with regular articles they published in the past?

10. How do the wider author demographics of RRs compare with the normative demographics in their field? Do any discrepancies vary between fields?

11. How does the tone and content of peer review differ between RRs and regular articles? Does the fact that Stage 1 review offers reviewers the opportunity to improve a study design lead to a more constructive and collegial process, as

suggested by preliminary survey evidence?[123] Are the reviews themselves higher quality? This question could be answered by performing qualitative content analysis of the RR reviews *vs.* regular reviews published by journals that employ an open review policy. The comparison would be especially powerful for journals such as *Meta-Psychology* that also publish the reviews of rejected articles.

12. In the long term, what contribution are RRs making to theory or applications? Can we measure their costs and benefits to health, society and the economy?[94]

Conclusion

In this review we have reflected on the history, preliminary impacts and future potential of the Registered Reports (RR) initiative. For the past 8 years, the life and social sciences have embarked on a journey into the unknown — one that has been mooted for decades but has only now reached open waters. Early suggestions of impact are promising, with RRs more likely to disconfirm *a priori* hypotheses and to be computationally reproducible, while also receiving higher quality ratings and the same or higher attention through citations. The prospects of the initiative now hinge on more detailed metascience, while addressing limitations and maintaining quality control as the format scales up and into new disciplines. As we look into the next decade, we believe RRs are showing all the signs of becoming a powerful antidote to reporting and publication bias, realigning incentives to ensure that the practices which are best for science — transparent, reproducible, accurate reporting — also serve the interests of individual scientists.

Acknowledgements

We are grateful to David Mellor, Brian Nosek and the Center for Open Science for their ongoing collaboration and discussion, to Aoife O'Mahony for the collation of key statistics concerning published RRs, and to the many authors, reviewers and editors who have supported the RR initiative.

Competing Interests

C.D.C. is a member of the Advisory Board of *Nature Human Behaviour*, chair of the Registered Reports committee supported by the Center for Open Science, is a co-founder of *Peer Community in Registered Reports*, and currently serves as RR editor at *BMJ Open Science*, *Cortex*, *European Journal of Neuroscience*, *NeuroImage*, Neuroimage: Reports, *PLOS Biology*, and *Royal Society Open Science.*

References

1. Vazire, S. Implications of the Credibility Revolution for Productivity, Creativity, and Progress: *Perspect. Psychol. Sci.* (2018) doi:10.1177/1745691617751884.
2. Munafò, M. R. *et al.* A manifesto for reproducible science. *Nat Hum Behav* **1**, 0021 (2017).
3. The Academy of Medical Sciences. *Reproducibility and reliability of biomedical research: improving research practice*. https://acmedsci.ac.uk/file-download/38189-56531416e2949.pdf (2015).
4. Fanelli, D. "Positive" Results Increase Down the Hierarchy of the Sciences. *PLOS ONE* **5**, e10068 (2010).
5. Franco, A., Malhotra, N. & Simonovits, G. Publication bias in the social sciences: Unlocking the file drawer. *Science* **345**, 1502–1505 (2014).
6. Harrison, J. S., Banks, G. C., Pollack, J. M., O'Boyle, E. H. & Short, J. Publication Bias in Strategic Management Research: *J. Manag.* (2014) doi:10.1177/0149206314535438.
7. Jennions, M. D. & Møller, A. P. Publication bias in ecology and evolution: an empirical assessment using the 'trim and fill' method. *Biol. Rev.* **77**, 211–222 (2002).
8. Kerr, N. L. HARKing: Hypothesizing After the Results are Known. *Personal. Soc. Psychol. Rev.* **2**, 196–217 (1998).
9. Bruns, S. B. & Ioannidis, J. P. A. p-Curve and p-Hacking in Observational Research. *PLOS ONE* **11**, e0149144 (2016).
10. Khan, M. J. & Trønnes, P. C. p-Hacking in Experimental Audit Research. *Behav. Res.*

Account. **31**, 119–131 (2018).

11. Holman, L., Head, M. L., Lanfear, R. & Jennions, M. D. Evidence of Experimental Bias in the Life Sciences: Why We Need Blind Data Recording. *PLOS Biol.* **13**, e1002190 (2015).
12. Fiedler, K. & Schwarz, N. Questionable Research Practices Revisited: *Soc. Psychol. Personal. Sci.* **7**, 45–52 (2016).
13. Rabelo, A. L. A. *et al.* Questionable research practices among Brazilian psychological researchers: Results from a replication study and an international comparison. *Int. J. Psychol.* (2019) doi:10.1002/ijop.12632.
14. Fraser, H., Parker, T., Nakagawa, S., Barnett, A. & Fidler, F. Questionable research practices in ecology and evolution. *PLOS ONE* **13**, e0200303 (2018).
15. John, L. K., Loewenstein, G. & Prelec, D. Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychol. Sci.* **23**, 524–532 (2012).
16. Button, K. S. *et al.* Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* **14**, 365–376 (2013).
17. Wicherts, J. M., Borsboom, D., Kats, J. & Molenaar, D. The poor availability of psychological research data for reanalysis. *Am. Psychol.* **61**, 726–728 (2006).
18. Mueller-Langer, F., Fecher, B., Harhoff, D. & Wagner, G. G. Replication studies in economics—How many and which papers are chosen for replication, and why? *Res. Policy* **48**, 62–83 (2019).
19. Makel, M. C., Plucker, J. A. & Hegarty, B. Replications in Psychology Research: How Often Do They Really Occur? *Perspect. Psychol. Sci. J. Assoc. Psychol. Sci.* **7**, 537–542 (2012).
20. Camerer, C. F. *et al.* Evaluating replicability of laboratory experiments in economics. *Science* **351**, 1433–1436 (2016).
21. Reproducibility Project: Cancer Biology. *eLife* https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-biology (2014).

22. Open Science Collaboration. Estimating the reproducibility of psychological science. *Science* **349**, aac4716 (2015).
23. Nosek, B. A., Spies, J. R. & Motyl, M. Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspect. Psychol. Sci.* **7**, 615–631 (2012).
24. Feynman, R. P. Cargo Cult Science. *Eng. Sci.* **June**, 10–13 (1974).
25. Johnson, J. A. Are Research Psychologists More Like Detectives or Lawyers? *Psychology Today* http://www.psychologytoday.com/blog/cui-bono/201307/are-research-psychologists-more-detectives-or-lawyers-0 (2013).
26. Bem, D. J. Writing the Empirical Journal Article. in *The Compleat Academic: A Practical Guide for the Beginning Social Scientist* (eds. Zanna, M. P. & Darley, J. M.) (Lawerence Erlbaum, 1987).
27. Bem, D. J. Writing the Empirical Journal Article. in *The Compleat Academic: A Practical Guide for the Beginning Social Scientist* (eds. Darley, J. M., Zanna, M. P. & Roediger III, H. L.) (American Psychological Association, 2003).
28. Developing a Program of Research. in *The Sage Handbook of Methods in Social Psychology* (eds. Sansone, C., Morf, C. C., Panter, A. T. & Fiske, S. T.) (SAGE Publications, 2003).
29. Sanes, J. R. Tell me a story. *eLife* **8**, e50527 (2019).
30. Bakker, M., van Dijk, A. & Wicherts, J. M. The Rules of the Game Called Psychological Science. *Perspect. Psychol. Sci.* **7**, 543–554 (2012).
31. Grand, J. A., Rogelberg, S. G., Banks, G. C., Landis, R. S. & Tonidandel, S. From Outcome to Process Focus: Fostering a More Robust Psychological Science Through Registered Reports and Results-Blind Reviewing. *Perspect. Psychol. Sci.* **13**, 448–456 (2018).
32. Eich, E. PSCI Initiatives for 2013. https://docs.google.com/viewer?a=v&pid=forums&srcid=MDM4NDU5NjQ2ODQzODQ0OTU4OTYBMDA2MTUwNDQ3NTQ0ODMwMDY5NjYBNVVLMmhkT0tVWTRKATAuMQE

BdjI&authuser=0 (2012).

33. Chambers, C. Changing the culture of scientific publishing from within. *NeuroChambers* https://neurochambers.blogspot.com/2012/10/changing-culture-of-scientific.html (2012).
34. Simons, D. J. Registered Replication Reports - Stay Tuned! http://blog.dansimons.com/2013/05/registered-replication-reports-stay.html (2013).
35. Call for proposals: Special issue of Social Psychology on “Replications of important results in social psychology”. *Soc. Psychol.* **44**, 59–60 (2013).
36. Chambers, C. D. Registered reports: a new publishing initiative at Cortex. *Cortex J. Devoted Study Nerv. Syst. Behav.* **49**, 609–610 (2013).
37. Peirce, C. S. Illustrations of the Logic of Science VI: Deduction, Induction, and Hypothesis. *Popular Science Monthly* vol. 13 470–482 (1878).
38. Wagenmakers, E.-J., Dutilh, G. & Sarafoglou, A. The Creativity-Verification Cycle in Psychological Science: New Methods to Combat Old Idols. *Perspect. Psychol. Sci.* **13**, 418–427 (2018).
39. Rosenthal, R. *Experimenter effects in behavioral research.* (Appleton-Century-Crofts, 1966).
40. Weiss, D. J. An Experiment in Publication: Advance Publication Review: *Appl. Psychol. Meas.* **13**, (1989).
41. Kupfersmid, J. Improving what is published: A model in search of an editor. *Am. Psychol.* **43**, 635–642 (1988).
42. Newcombe, R. G. Towards a reduction in publication bias. *Br Med J Clin Res Ed* **295**, 656–659 (1987).
43. Mahoney, M. J. Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cogn. Ther. Res.* **1**, 161–175 (1977).
44. Walster, G. W. & Cleary, T. A. A Proposal for a New Editorial Policy in the Social Sciences. *Am. Stat.* **24**, 16–19 (1970).
45. Wiseman, R., Watt, C. & Kornbrot, D. Registered reports: an early example and analysis. *PeerJ* **7**, e6232 (2019).

46. Lancet, T. E. of T. Protocol review at The Lancet: 1997–2015. *The Lancet* **386**, 2456–2457 (2015).
47. Maizey, L. & Tzavella, L. Barriers and solutions for early career researchers in tackling the reproducibility crisis in cognitive neuroscience. *Cortex* **113**, 357–359 (2019).
48. Allen, C. & Mehler, D. M. A. Open science challenges, benefits and tips in early career and beyond. *PLOS Biol.* **17**, e3000246 (2019).
49. Scheel, A. M., Schijen, M. & Lakens, D. An excess of positive results: Comparing the standard Psychology literature with Registered Reports. *Adv. Methods Pract. Psychol. Sci.* (in press). Preprint at https://psyarxiv.com/p6e9c/
50. Wicherts, J. M., Bakker, M. & Molenaar, D. Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results. *PLOS ONE* **6**, e26828 (2011).
51. Obels, P., Lakens, D., Coles, N. A., Gottfried, J. & Green, S. A. *Analysis of Open Data and Computational Reproducibility in Registered Reports in Psychology*. Preprint at https://osf.io/fk8vh (2019) doi:10.31234/osf.io/fk8vh.
52. Hardwicke, T. E. *et al.* Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal Cognition. *R. Soc. Open Sci.* **5**, 180448 (2018).
53. Jannot, A.-S., Agoritsas, T., Gayet-Ageron, A. & Perneger, T. V. Citation bias favoring statistically significant studies was present in medical research. *J. Clin. Epidemiol.* **66**, 296–301 (2013).
54. Misemer, B. S., Platts-Mills, T. F. & Jones, C. W. Citation bias favoring positive clinical trials of thrombolytics for acute ischemic stroke: a cross-sectional analysis. *Trials* **17**, 473 (2016).
55. Brembs, B., Button, K. & Munafò, M. Deep impact: unintended consequences of journal rank. *Front. Hum. Neurosci.* **7**, (2013).
56. Fang, F. C., Steen, R. G. & Casadevall, A. Misconduct accounts for the majority of retracted scientific publications. *Proc. Natl. Acad. Sci.* **109**, 17028–17033 (2012).

57. Lozano, G. A., Larivière, V. & Gingras, Y. The weakening relationship between the impact factor and papers' citations in the digital age. *J. Am. Soc. Inf. Sci. Technol.* **63**, 2140–2145 (2012).
58. Seglen, P. O. Why the impact factor of journals should not be used for evaluating research. *BMJ* **314**, 497 (1997).
59. Hummer, L., Thorn, F. S., Nosek, B. A. & Errington, T. Evaluating Registered Reports: A Naturalistic Comparative Study of Article Impact. (2017) doi:10.31219/osf.io/5y8w7.
60. Button, K. S., Bal, L., Clark, A. & Shipley, T. Preventing the ends from justifying the means: withholding results to address publication bias in peer-review. *BMC Psychol.* **4**, 59 (2016).
61. Srivastava, S. A Pottery Barn rule for scientific journals. *The Hardest Science* https://thehardestscience.com/2012/09/27/a-pottery-barn-rule-for-scientific-journals/ (2012).
62. Lilienfeld, S. O. Clinical Psychological Science: Then and Now. *Clin. Psychol. Sci.* **5**, 3–13 (2017).
63. Lucas, R. E. & Donnellan, M. B. Enhancing transparency and openness at the Journal of Research in Personality. *J. Res. Personal.* **68**, 1–4 (2017).
64. Preregistered Direct Replications: A New Article Type in Psychological Science. *APS Obs.* **30**, (2017).
65. Royal Society Open Science. Replication studies. https://royalsocietypublishing.org/rsos/replication-studies.
66. The Royal Society. Reproducibility and transparency collection. https://royalsocietypublishing.org/topic/special-collections/rsos-reproducibility.
67. Kriegeskorte, N. Open Evaluation: A Vision for Entirely Transparent Post-Publication Peer Review and Rating for Science. *Front. Comput. Neurosci.* **6**, 79 (2012).
68. Murray, H. Transparency meets transparency. *F1000 Blogs* https://blog.f1000.com/2017/10/12/transparency-meets-transparency/ (2017).
69. Carlsson, R. *et al.* Inaugural Editorial of Meta-Psychology. *Meta-Psychol.* **1**, a1001–

a1001 (2017).

70. Kiyonaga, A. & Scimeca, J. M. Practical Considerations for Navigating Registered Reports. *Trends Neurosci.* **42**, 568–572 (2019).
71. Guest, O. & Martin, A. E. How Computational Modeling Can Force Theory Building in Psychological Science. *Perspect. Psychol. Sci.* 1745691620970585 (2021) doi:10.1177/1745691620970585.
72. Schönbrodt, F. D. & Wagenmakers, E.-J. Bayes factor design analysis: Planning for compelling evidence. *Psychon. Bull. Rev.* **25**, 128–142 (2018).
73. Nosek, B. A. *et al.* Promoting an open research culture. *Science* **348**, 1422–1425 (2015).
74. BMJ Open Science. https://openscience.bmj.com/pages/authors/.
75. International Review of Social Psychology. Exploratory Reports at IRSP: Guidelines for Authors. http://www.rips-irsp.com/about/exploratory-reports/.
76. McIntosh, R. D. Exploratory reports: A new article type for Cortex. *Cortex* **96**, A1–A4 (2017).
77. Hardwicke, T. E. & Ioannidis, J. P. A. Mapping the universe of registered reports. *Nat. Hum. Behav.* **2**, 793–796 (2018).
78. Chambers, C. D. & Mellor, D. T. Protocol transparency is vital for registered reports. *Nat. Hum. Behav.* **2**, 791–792 (2018).
79. Center for Open Science. Template Reviewer and Author Guidelines. *Open Science Framework* https://osf.io/8mpji/ (2018).
80. MacCoun, R. & Perlmutter, S. Blind analysis: Hide results to seek the truth. *Nat. News* **526**, 187–189 (2015).
81. Dutilh, G. *et al.* A test of the diffusion model explanation for the worst performance rule using preregistration and blinding. *Atten. Percept. Psychophys.* **79**, 713–725 (2017).
82. Dienes, Z. Using Bayes to get the most out of non-significant results. *Front Psychol* **5**, 1–17 (2014).
83. Lakens, D. Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses. *Soc. Psychol. Personal. Sci.* **8**, 355–362 (2017).

84. PLOS Biology. *Preregistered Research Article Guidelines for Authors*. https://plos-marketing.s3.amazonaws.com/Marketing/Biology+Preregistered+Articles+Guidelines+for+Authors.pdf (2020).
85. Nature Human Behaviour. *Registered Reports: Author and Reviewer Guidelines*. https://media.nature.com/original/nature-cms/uploads/ckeditor/attachments/4127/RegisteredReportsGuidelines_NatureHumanBehaviour.pdf.
86. Cortex. *Guidelines for reviewers*. https://cdn.elsevier.com/promis_misc/PROMIS%20pub_idt_CORTEX%20Guidelines_RR_29_04_2013.pdf (2013).
87. Royal Society Open Science. Registered Reports. *The Royal Society* https://royalsocietypublishing.org/rsos/registered-reports#ReviewerGuideRegRep.
88. Center for Open Science. Registered Reports: Resources for Editors. https://cos.io/rr/.
89. Petticrew, M. *et al.* Publication bias in qualitative research: what becomes of qualitative research presented at conferences? *J. Epidemiol. Community Health* **62**, 552–554 (2008).
90. Piñeiro, R. & Rosenblatt, F. Pre-Analysis Plans for Qualitative Research. *Rev. Cienc. Política* **36**, 785–796 (2016).
91. Kern, F. G. & Gleditsch, K. S. *Exploring Pre-registration and Pre-analysis Plans for Qualitative Inference*. Preprint (2017) doi:10.13140/RG.2.2.14428.69769.
92. Haven, T. L. & Grootel, D. L. V. Preregistering qualitative research. *Account. Res.* **26**, 229–244 (2019).
93. Hartman, A., Kern, F. & Mellor, D. *Preregistration for Qualitative Research Template*. https://osf.io/j7ghv/ (2018).
94. Tiokhin, L., Morgan, T. & Yan, M. Competition for priority and the cultural evolution of research strategies. (2020) doi:10.31222/osf.io/x4t7q.
95. Chambers, C. NeuroChambers: CALLING ALL SCIENTISTS: Rapid evaluation of COVID19-related Registered Reports at Royal Society Open Science. *NeuroChambers*

http://neurochambers.blogspot.com/2020/03/calling-all-scientists-rapid-evaluation.html (2020).

96. Zhou, T., Nguyen, T. T., Zhong, J. & Liu, J. A COVID-19 descriptive study of life after lockdown in Wuhan, China. *R. Soc. Open Sci.* **7**, 200705 (2020).
97. Weinstein, N. & Nguyen, T.-V. Motivation and preference in isolation: a test of their different influences on responses to self-isolation during the COVID-19 outbreak. *R. Soc. Open Sci.* **7**, 200458 (2020).
98. Khan, K. A. & Cheung, P. Presence of mismatches between diagnostic PCR assays and coronavirus SARS-CoV-2 genome. *R. Soc. Open Sci.* **7**, 200636 (2020).
99. Riello, M., Purgato, M., Bove, C., MacTaggart, D. & Rusconi, E. Prevalence of post-traumatic symptomatology and anxiety among residential nursing and care home workers following the first COVID-19 outbreak in Northern Italy. *R. Soc. Open Sci.* **7**, 200880 (2020).
100. Lieberoth, A. *et al.* Stress and worry in the 2020 coronavirus pandemic: relationships to trust and compliance with preventive measures across 48 countries in the COVIDiSTRESS global survey. *R. Soc. Open Sci.* **8**, 200589 (2021).
101. Yonemitsu, F. *et al.* Warning 'Don't spread' versus 'Don't be a spreader' to prevent the COVID-19 pandemic. *R. Soc. Open Sci.* **7**, 200793 (2020).
102. PLOS ONE Editors. PLOS ONE partners with the Children's Tumor Foundation to trial Registered Reports. *EveryONE: The PLOS ONE blog* https://blogs.plos.org/everyone/2017/09/26/registered-reports-with-ctf/, http://blogs.plos.org/everyone/?p=21545 (2017).
103. Munafò, M. R. Improving the Efficiency of Grant and Journal Peer Review: Registered Reports Funding. *Nicotine Tob. Res.* **19**, 773–773 (2017).
104. Lakens, D. & DeBruine, L. *Improving Transparency, Falsifiability, and Rigour by Making Hypothesis Tests Machine Readable*. Preprint at https://psyarxiv.com/5xcda/ (2020) doi:10.31234/osf.io/5xcda.
105. Mellor, D. & DeHaven, A. Templates of OSF Registration Forms. https://osf.io/zab38/

(2016).

106. Wicherts, J. M. *et al.* Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. *Front. Psychol.* **7**, (2016).

107. Mathieu, S., Boutron, I., Moher, D., Altman, D. G. & Ravaud, P. Comparison of Registered and Published Primary Outcomes in Randomized Controlled Trials. *JAMA* **302**, 977–984 (2009).

108. Gopal, A. D. *et al.* Adherence to the International Committee of Medical Journal Editors' (ICMJE) prospective registration policy and implications for outcome integrity: a cross-sectional analysis of trials published in high-impact specialty society journals. *Trials* **19**, 448 (2018).

109. Tracking switched outcomes in clinical trials. *COMPare* http://compare-trials.org.

110. Ramagopalan, S. V. *et al.* Funding source and primary outcome changes in clinical trials registered on ClinicalTrials.gov are associated with the reporting of a statistically significant primary outcome: a cross-sectional study. *F1000Research* **4**, 80 (2015).

111. Goldacre, B. *et al.* Compliance with requirement to report results on the EU Clinical Trials Register: cohort study and web resource. *BMJ* **362**, k3218 (2018).

112. Chen, R. *et al.* Publication and reporting of clinical trial results: cross sectional analysis across academic medical centers. *BMJ* **352**, i637 (2016).

113. The BMC Medicine Team. BMC Medicine becomes the first medical journal to accept Registered Reports. *Research in progress blog* http://blogs.biomedcentral.com/bmcblog/2017/08/24/bmc-medicine-becomes-the-first-medical-journal-to-accept-registered-reports/ (2017).

114. Research Excellence Framework. Panel criteria and working methods. https://www.ref.ac.uk/media/1084/ref-2019_02-panel-criteria-and-working-methods.pdf (2019).

115. Munafò, M. Raising research quality will require collective action. *Nature* **576**, 183–183 (2019).

116. University College London. UCL Statement on Transparency in Research. https://www.ucl.ac.uk/research/sites/research/files/ucl_statement_on_transparency_in_research_november_20191.pdf (2019).

117. Rousselet, G. A., Hazell, G., Cooke, A. & Dalley, J. W. Promoting and supporting credibility in neuroscience. *Brain Neurosci. Adv.* **3**, 2398212819844167 (2019).

118. British Psychological Society. We're offering Registered Reports across all eleven of our academic journals. *BPS News* https://www.bps.org.uk/news-and-policy/were-offering-registered-reports-across-all-eleven-our-academic-journals (2018).

119. Stiftelsen Dam. Krav om forhåndregistrering av studier finansiert av Stiftelsen Dam. https://www.dam.no/krav-om-forhandregistrering-av-studier-finansiert-av-extrastiftelsen/ (2018).

120. Accelerating Research on Consciousness. *Templeton World Charity Foundation* https://www.templetonworldcharity.org/our-priorities/accelerating-research-consciousness.

121. Mehlenbacher, A. R. Registered Reports: Genre Evolution and the Research Article. *Writ. Commun.* **36**, 38–67 (2019).

122. Soderberg, C. K. *et al.* Initial evidence of research quality of registered reports compared with the standard publishing model. Nat. Hum. Behav. (2021).

123. DeHaven, A. C. *et al. Registered Reports: views from editors, reviewers and authors*. Preprint at https://osf.io/ndvek (2019) doi:10.31222/osf.io/ndvek.

124. Heycke, T., Aust, F. & Stahl, C. Subliminal influence on preferences? A test of evaluative conditioning for brief visual conditioned stimuli using auditory unconditioned stimuli. *R. Soc. Open Sci.* **4**, 160935 (2017).

125. Ait Ouares, K., Beurrier, C., Canepari, M., Laverne, G. & Kuczewski, N. Opto nongenetics inhibition of neuronal firing. *Eur. J. Neurosci.* **49**, 6–26 (2019).

126. Sassenhagen, J. & Bornkessel-Schlesewsky, I. The P600 as a correlate of ventral attention network reorientation. *Cortex* **66**, A3–A20 (2015).

127. Allinson, M. Royal Society Open Science launches Registered Reports.

https://blogs.royalsociety.org/publishing/registered-reports/ (2015).

128. Nosek, B. A. & Errington, T. M. Reproducibility in Cancer Biology: Making sense of replications. *eLife* **6**, e23383 (2017).

129. Guo, W., Del Vecchio, M. & Pogrebna, G. Global network centrality of university rankings. *R. Soc. Open Sci.* **4**, 171172 (2017).

130. Promoting reproducibility with registered reports. *Nat. Hum. Behav.* **1**, 1–1 (2017).

131. OSF Registries. https://osf.io/registries/discover?provider=OSF&type=Registered%20Report%20Protocol%20Preregistration.

Table 1. Misconceptions and realities of Registered Reports (RRs)

The misconception	The reality
1. RRs hamper exploratory research	RRs can include clearly labelled *post hoc* analyses at Stage 2, and almost all published RRs do.
2. RRs limit creativity or serendipity	The opposite is more likely. Assessing study design and rationale in the absence of results encourages a perceptive evaluation of why a study is worth conducting in the first place, and IPA† protects serendipitous discoveries from publication bias.
3. RRs are only for "single shot" studies	RRs in many journals can include preliminary studies (e.g. Heycke et al.[124]), prespecified sequences of studies (e.g. Ait Ouares et al.[125]) or incremental registrations after Stage 2 acceptance.
4. RRs are particularly vulnerable to scooping	In 7 years there have been no reports of scooping linked to RRs, and Stage 1 protocols can be kept under private embargo until Stage 2 submission or acceptance. In the rare event of an idea being appropriated and implemented by a reviewer faster than the authors, doing so would have no effect on publication of the RR because "novelty" is not assessed at Stage 2.
5. RRs can be easily gamed by "post-registering" a study that the authors have already conducted	When submitting a RR, authors must certify that no data collection or analysis – other than that undertaken as part of any preliminary studies – has yet been undertaken. Post-registration would therefore constitute fraud, and would in any case be ineffective because Stage 1 reviewers and editors usually ask for at least minor changes to the proposed study design.
6. RRs require authors to prespecify a fixed plan that cannot be changed	RRs can be changed *en route* in consultation with the journal editor and (where changes are substantial) the reviewers, as long as all changes are transparently reported at Stage 2.
7. RRs are not suitable for studying hard-to-reach samples and/or for researchers with limited resources	Some journals do set high requirements for statistical power or Bayes factors (e.g. *Nature Human Behaviour*, *PLOS Biology*, *Cortex*), but others such as *Royal Society Open Science* set no requirements at all. Authors who are constrained by data availability and/or resources need only justify why their proposed study would make an important contribution to the field.
8. RRs are not suitable for early career researchers (ECRs)	Since RRs eliminate publication bias, Stage 1 IPA can be time-saving and anxiety-relieving. It is perhaps not surprising that most RRs are first-authored by ECRs (see Supplementary Information). ECRs should carefully consider journal requirements when there is a constraint on time and/or resources[47].
9. RRs will slow the pace of research	Even if the RR review process at any one journal does take longer (which is currently unknown), the rejection rate is probably lower (e.g. at *Cortex* the rejection rate is ~10% at Stage 1 and 0% at Stage 2, *cf.* ~90% for regular articles). In contrast, regular articles are often resubmitted to multiple journals and rejected because of unfixable flaws or unattractive results. RRs may therefore accelerate rather than hinder discovery.
10. RRs give reviewers too much power to decide what research is undertaken	RRs redirect the power of reviewers to provide input when it is most valuable, and major changes are limited to Stage 1. With regular submissions, reviewers can request extra studies or recommend rejection based on the results, neither of which is possible for RRs.

† IPA = in principle acceptance

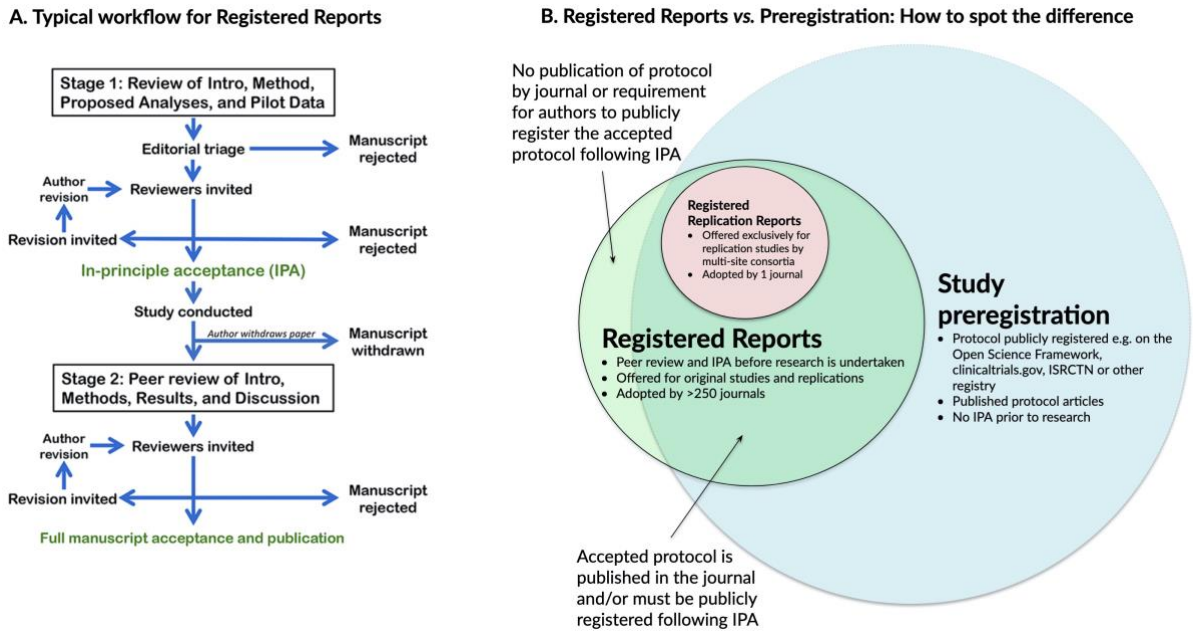


Figure 1. The basics of Registered Reports (RRs). (A) The typical RR workflow involves pre-study review of study rationale, design and proposed analyses, including preliminary data as needed to provide proof of concept, effect size estimation for a sampling plan, or hypothesis-generation. Following in principle acceptance (IPA), authors conduct the research before submitting a complete manuscript with results. The original reviewers then return at Stage 2 to assess compliance with protocol and ensure that the conclusions are appropriately evidence-based. Figure reproduced from https://cos.io/rr/. **(B)** The distinction between RRs, Registered Replications Reports (RRRs) and study preregistration is a common source of confusion. Aside from the minority of RR formats which do not require public preregistration (see **Limitations and drawbacks**), RRs are mostly a subset of the wider family of preregistration methods, but with the additional features of pre-study review and in principle acceptance (IPA) regardless of results. RRRs, offered by one psychology journal, are a subset of RRs.

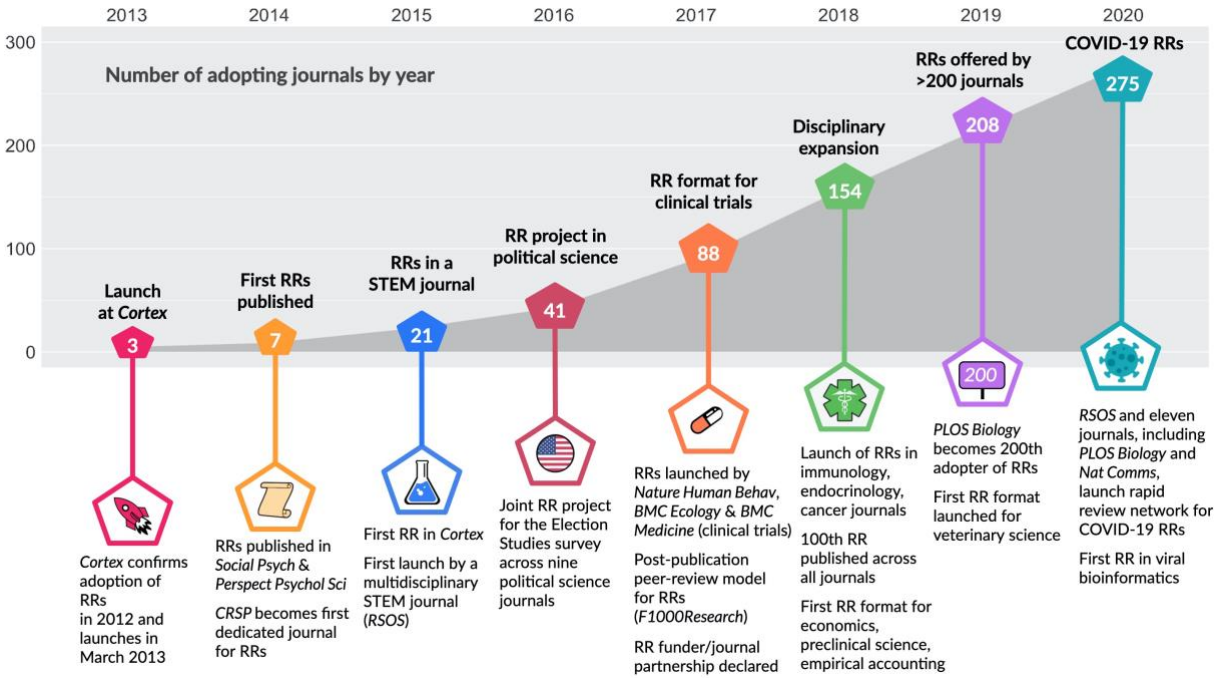


Figure 2. Key milestones in the evolution of Registered Reports. Registered Reports (RRs) in their current form were first introduced in 2012 and *Cortex* was the first journal to officially offer RRs as an article type in 2013. The first RRs were published in the journals *Social Psychology* and *Perspectives on Psychological Science*. The first journal to be exclusively dedicated to RRs was also launched in 2014 (*Comprehensive Results in Social Psychology*). In 2015, *Cortex* published its first RR[126] and *Royal Society Open Science* (*RSOS*) became the first multidisciplinary journal covering all STEM to offer RRs[127]. In 2015, nine political science journals launched a joint RR project for the 2016 American National Election Studies survey (https://www.erpc2016.com/), marking the first application of RRs in political science. The number of RR-adopting journals further increased in 2017, which was a key year on several fronts. As part of the Reproducibility Project: Cancer Biology, *eLife* published the first of many RRs[128] and the first RR format for clinical trials was launched by *BMC Medicine*. The first RR in the field of computer science was also published in *RSOS*[129] and the format was introduced for the first time in a specialist ecology journal (*BMC Ecology*). In the same year, *Nature Human Behaviour* launched RRs[130] and *F1000Research* and *Meta-Psychology* paved the way for the post-publication-peer-review model for RRs. The first RR funder/journal partnership was also announced in 2017[103]. By the end of 2018 the number of adopting journals had risen to 150 and the 100th Stage 2 RR was published across all journals. This increase in the number of adopters paralleled a significant disciplinary expansion, with the format being applied to preclinical science (*BMJ Open Science*), economics (*Journal of Development Economics*), empirical accounting (*Journal of Accounting Research*), animal neurophysiology (*European Journal of Neuroscience*), cancer research (*Cancer Medicine*), immunology, endocrinology, gastroenterology, herpetology, and

agricultural/soil sciences. In 2018, the British Psychological Society became the first society to launch RRs concurrently across all of its journals[118]. In 2019, *PLOS Biology* became the 200th adopter of RRs, *Nature Human Behaviour* published its first two RRs and the format was launched for the first time in the field of veterinary science (*Equine Veterinary Journal*). In 2020, RSOS and 11 journals launched the COVID-19 RR rapid review network[95]. As part of this ongoing initiative, participating journals strive to review Stage 1 RRs related to COVID-19 in 7 days and commit to open access publication with no article processing charge. As a result of this initiative, the past year also marked the first published RR in viral bioinformatics[98].

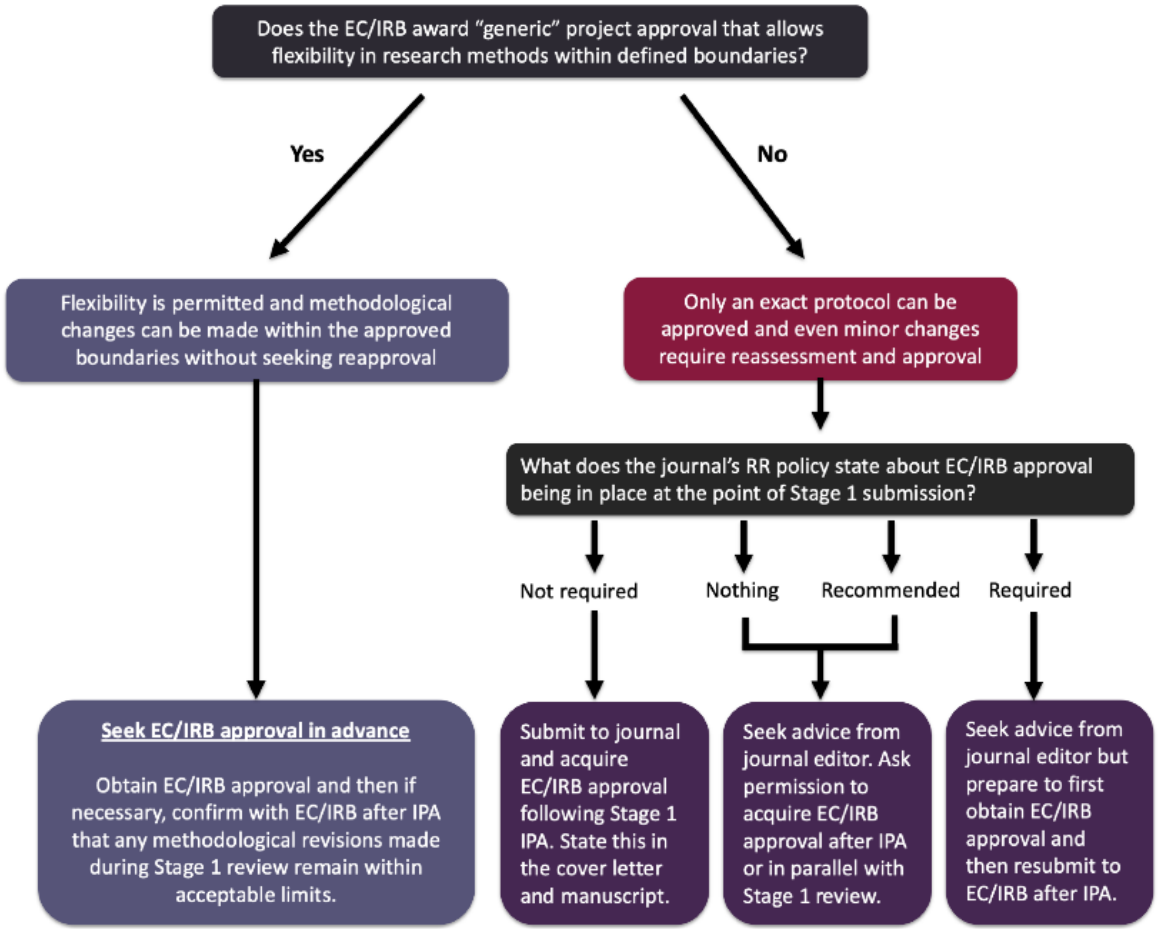


Figure 3. Registered Reports (RRs) and ethics approval. One frequently-asked question is when authors should obtain ethics committee (EC) or institutional review board (IRB) approval for their Stage 1 RR. The answer depends on the tolerance of the EC/IRB for methodological flexibility and the requirements of the journal policy. Where the EC/IRB permits flexibility (left track), it is usually most efficient to obtain a generic EC/IRB approval prior to manuscript submission. Where the EC/IRB will instead only approve a precise protocol, and any deviations to the protocol must be submitted for reapproval, the most efficient course of action depends on the specific journal requirements (right track). Most RRs proceed via the left track, but most RR policies also leave the door open for authors to discuss barriers arising from EC/IRB rigidity. For example, *Cortex* requires that *"all necessary...approvals (e.g. ethics) are in place for the proposed research. Note that manuscripts will be generally considered only for studies that are able to commence immediately;* ***however, authors with alternative plans are encouraged to contact the journal office for advice****."* (emphasis added)

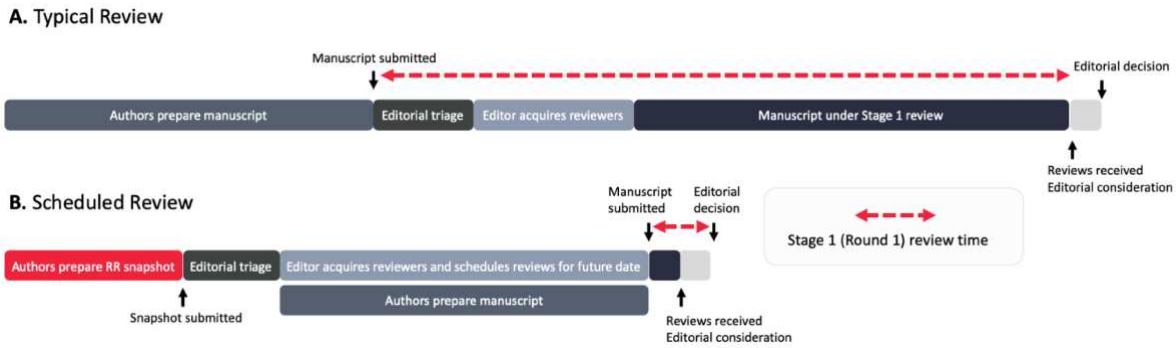
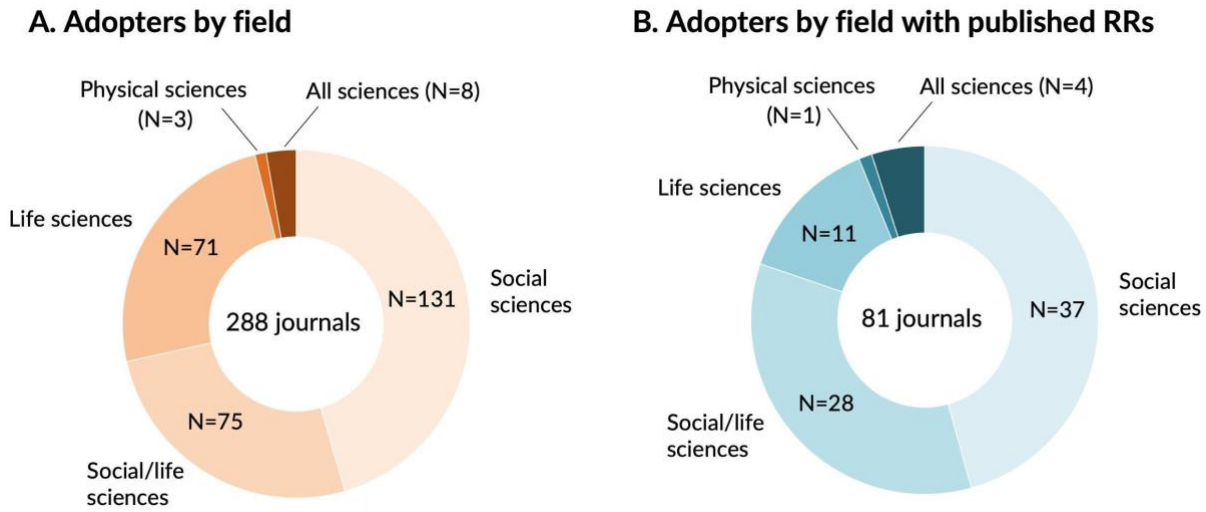


Figure 4. Proposed scheduled review workflow for Registered Reports (RRs). Restructuring the RR submission workflow could significantly reduce the duration of Stage 1 peer review (dashed red arrows). In the typical RR chronology **(A)**, the total time taken for editors to triage submissions, acquire reviewers, obtain reviews, and reach an editorial decision accumulates serially and only after authors have prepared and submitted a full manuscript. Scheduled review **(B)** could significantly accelerate this process by performing key tasks in parallel. Rather than submitting a full manuscript, authors would initially submit a one-page, template-based RR "snapshot" which undergoes editorial triage. If deemed suitable, the editor would then organise the review process for a fixed future date (or short range of dates) while the authors prepare the full manuscript. Although this process could only feasibly expedite the first round of Stage 1 review (and not the re-review of a revised Stage 1 submission), the overall time-saving could be substantial since the first round of assessment is usually the most onerous. Although no journals currently offer Scheduled Review, the workflow has been recently introduced as part of the *Peer Community in Registered Reports* initiative (https://rr.peercommunityin.org/) .

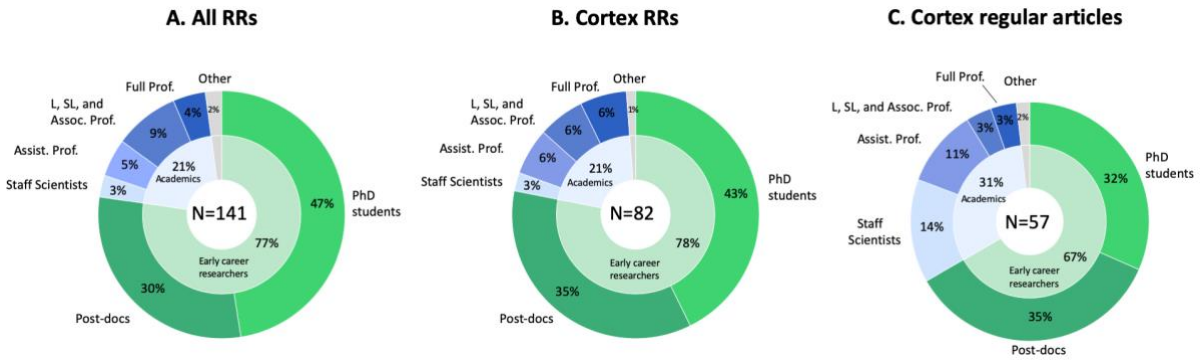
Supplementary Information

A. Field spread and author demographics

Which fields are offering Registered Reports (RRs) and who is using them? Supplementary Figure S1 below shows the broad field distribution of 288 adopting journals (panel a) and the subset of those adopters that have published at least one completed Stage 2 RR (panel b). Supplementary Figure S2 reports an analysis of author career demographics in a sample of RRs submitted to several journals.



Supplementary Figure S1. The general field spread of Registered Reports (RRs). (A) Over 70% of adopting journals publish research in the social sciences or at the intersection of the social and life sciences, alongside a smaller number of specialist life science journals and an even smaller number of STEM-wide and physical sciences journals. **(B)** Within the 81 adopting journals that have published at least one completed Stage 2 RR, the greatest representation is again in the social and social/life sciences. Of the four included STEM-wide journals, all published RRs have been in the social or life sciences except for one RR in computer science[1]. Journals situated in the social/life sciences category are typically those that interface between psychology and neuroscience, either by publishing at least some articles in cognitive/behavioural neuroscience or by publishing psychological research that can have implications for understanding biological systems (e.g. psychophysics, computational modelling).



Supplementary Figure S2. Who is submitting Registered Reports (RRs)? (A) Of 141 RRs submitted to *Cortex*, *European Journal of Neuroscience*, *NeuroImage* and *Royal Society Open Science*, 77% were first-authored by early career researchers (ECRs). A comparison of first-author career demographics between RRs **(B)** and regular articles **(C)** submitted to *Cortex* between 2013-2019 reveals a similar trend. The sample of regular articles at *Cortex* consists of comparable (non-clinical[†]) empirical reports drawn from adjacent manuscript submission numbers to the RRs. This comparison should be treated with caution as factors other than the status of the article as a RR could drive any differences. The results, however, suggest that RRs are unlikely to be less popular among ECRs compared with regular articles, at least at *Cortex*. Data were obtained and analysed by C.D.C as RR editor at these journals. Assist. Prof = Assistant Professor; L = Lecturer; SL = Senior Lecturer; Assoc. Prof = Associate Professor.

[†]Clinical articles were excluded from the sample of regular articles because the great majority of RRs at *Cortex* are non-clinical submissions, and preliminary observation revealed that regular clinical articles were far more likely to be first-authored by senior academics (including medical doctors and neurologists) compared with regular non-clinical articles. Including clinical articles in the sample of regular articles would therefore have skewed the comparison even more in favour of demonstrating a predominance of ECR authors for RRs.

B. Emerging Variants of Registered Reports

Results-blind review. Also known as "results-free review", "results-masked review" or "Hybrid RRs", this initiative modifies the standard RR workflow by performing Stage 1 peer review after results are known to the authors, but before they are known to the reviewers and editors. Following IPA, the authors then submit the full manuscript containing the data and conclusions. This format has the advantage that authors need not wait until IPA to conduct their research, thus preventing the Stage 1 review time from delaying data collection and analysis. On the other hand, the fact that the study is already complete and authors know the results limits the extent to which reviewers can improve the study design, and the format also does nothing to prevent reporting bias (e.g. *p*-hacking, HARKing) by authors. Results-blind review thus fits more easily into the workflow of regular articles but at the cost of greater risk of bias. To date, at least 13 journals, primarily in psychology and management, have adopted results-blind review as an optional article track, sometimes alongside new policies such as the Verification Reports initiative[2,3], and often while also offering standard RRs. In 2016, *BMC Psychology* launched the format in the form of a randomised controlled trial to assess feasibility and compare the rate of article acceptance and positive results with regular articles[4]. At the time of writing, the results of this trial are pending.

Accountable Replications. The standard RR format provides a popular route for replication studies by ensuring that peer review and IPA occur before authors invest substantial resources into performing the research, and regardless of the eventual results. However, at many adopting journals an RR can still be rejected at Stage 1 if the editor or reviewers consider the original study to be flawed, or the replication attempt to be a low priority. Moreover, the tendency for Stage 1 review to improve the proposed study design – which for original studies is a key strength of the process – can conflict with the need to keep the design of a replication study as similar as possible to the original work, even when the original study is suboptimal. Thus, while standard RRs are certainly a preferable route for replication studies compared with regular articles, they are still not an ideal fit for the requirements of replication.

The solution to this puzzle is a modified form of RRs and results-blind review conceived by psychologist Sanjay Srivastava[5]. Srivastava's "Pottery Barn rule" proposes that

when a journal publishes a research finding it should commit to publishing all methodologically-sound replications of that finding, regardless of how the results turn out, and regardless of subjective importance or methodological flaws in the original study. This policy would ensure that journals are publicly accountable for the replicability of the work they accept, providing an incentive for them to publish replicable studies. In 2018, *Royal Society Open Science* became the first journal to implement a complete version of this concept, following partial implementations at *Clinical Psychological Science*[6], *Journal of Research in Personality*[7], and *Psychological Science*[8]. Under the journal's Replications policy, authors are invited to submit close replication studies either before results exist using a workflow similar to RRs, or after the study is finished using results-blind review. Using a modified set of the RR assessment criteria (Supplementary Figure S3), the journal then reaches a Stage 1 IPA decision on the basis of technical validity and the methodological proximity between the replication and target study. In addition, the journal commits to reviewing replication attempts not only of studies published in *Royal Society Open Science*, but also in dozens of other major journals[9]. The central aim of this policy is to unlock the file drawer of unpublished replications while also minimizing as much as possible the barriers to conducting new replication studies[10]. Since then, a similar policy has been launched at the journal *Infant and Child Development*[3].

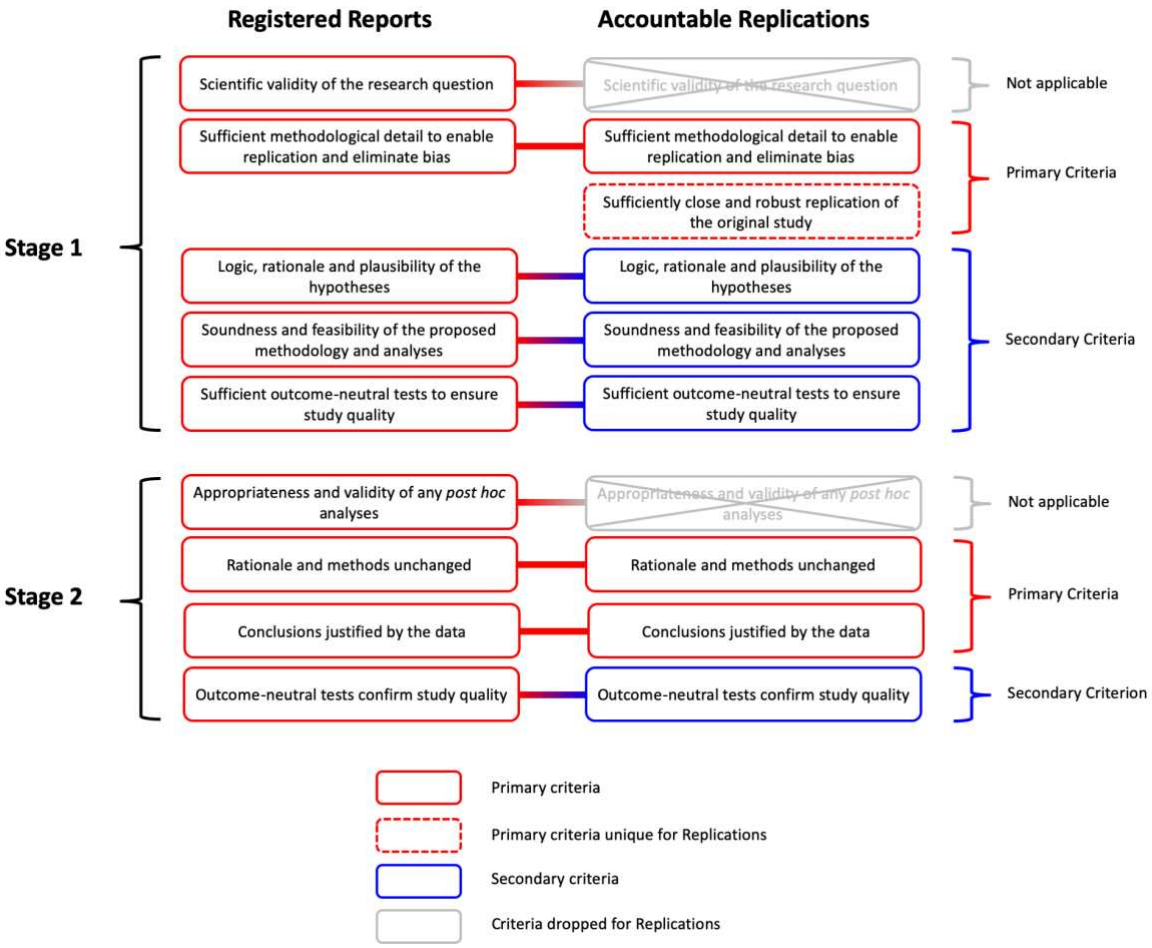
Verification Reports. Launched in 2020 at the journal *Cortex*, Verification Reports provide a dedicated format for assessing the computational reproducibility and robustness of previous findings based on reanalysis of the original study data.[2] Like RRs, submissions are assessed over two stages – before and after the presentation of results – and, as with the Accountable Replications initiative, researchers can also submit Stage 1 submissions in which the outcomes are already known to the authors but temporarily redacted to ensure results-blind peer review.

Post-publication peer review RRs. RRs usually rely on conventional pre-publication review in which reviewers serve as gatekeepers to IPA and Stage 2 acceptance. Recognising the inherent limitations of pre-publication review[11], in 2017 the journal *F1000Research* launched a variant of RRs involving post-publication peer review (PPPR)[12]. With PPPR-RRs, the Stage 1 manuscript is published almost immediately following initial receipt and is then openly reviewed. If the reviews are positive (with

authors having the usual opportunity to revise the protocol) then the article is awarded IPA and, once passing Stage 2 review, the final manuscript with results is badged as a RR. Unlike pre-publication review RRs, if authors fail to obtain IPA, or fail Stage 2 review, the journal will still publish the protocol and results, but the final article will not receive the RR badge. The PPPR model thus replaces the gatekeeping role of traditional peer review with a certification role, increasing the transparency and immediacy of RRs. To date, this workflow has been duplicated by several more journals, including *Wellcome Open Research*, and a similar route was independently launched in 2017 at the journal *Meta-Psychology*[13].

Publisher-level RRs. Under the standard RR model, the Stage 1 and Stage 2 manuscripts are reviewed by the same journal. Recently, however, some journals have begun implementing a distributed model in which they are reviewed and published in different journals. For example, in the system created by JMIR Publications – which publishes in medicine, engineering, technology and bioinformatics – authors submit their Stage 1 protocols to the journal *JMIR Protocols*. Following IPA, the authors can then choose which outlet among 26 JMIR journals to submit the Stage 2 manuscript. Once accepted, the completed Stage 2 RR is cross-linked to the accepted Stage 1 protocol using an International RR Identifier.[49]

Publisher-independent RRs. Can RRs exist beyond journals? Under the recently created *Peer Community in Registered Reports* (PCI RR) platform, Stage 1 and 2 preprints are reviewed independently of journals (https://rr.peercommunityin.org). PCI RR is a non-profit, non-commercial platform that coordinates the peer review of Registered Report preprints. Once a submission is accepted following peer review (or in PCI terms “recommended”), the completed Stage 2 Registered Report is posted at the preprint server where the preprint is hosted, and the peer reviews and recommendation of the preprint are posted at the PCI RR website. Following the completion of peer review, authors then have the option to publish their articles in a growing list of “PCI RR-friendly” journals that have committed to accepting PCI Registered Report recommendations without further peer review. This initiative separates RRs from mainstream publishers and gives authors the power to choose their destination journal (if any).



Supplementary Figure S3. The review criteria for RRs and Accountable Replications at *Royal Society Open Science*, following Srivastava's "Pottery Barn Rule"[5]. The left side of the figure shows the primary criteria (red) for RRs, all of which must be met to achieve Stage 1 and Stage 2 acceptance. The right side of the figure shows how these criteria are relaxed (blue) or eliminated (grey) for Replications to ensure that replication studies are accepted or rejected solely based on methodological detail, methodological robustness (including statistical power) and methodological similarity to the original research. Note that judgments about the validity of the research question, which can lead to Stage 1 rejection or major revisions for a RR, are not applicable for a Replication because publication of the original study is considered to have already settled that question. A journal that adopts this policy thus becomes accountable for its previous acceptance decisions. Other conditions that are essential for RRs are downgraded to secondary criteria (blue), meaning that weaknesses and failures in these areas, including methodological flaws in the original research, will require discussion in the manuscript but will not lead to outright rejection. Only the primary criteria must be met for a Replication submission to be accepted at each stage. By softening these criteria, the aim of this policy is to eliminate as many barriers as possible to publishing close, methodologically robust replications.

C. Protocol transparency requirements across the family of RR journals

Of 288 journals that have adopted RRs, 75 either launched the initiative as part of a short-term special issue or have yet to formally publish their RR policy. The remaining 213 journals have launched RRs as a permanent option, of which 186 (87.3%) support protocol transparency by either publishing the accepted Stage 1 manuscript as a separate article or by requiring authors to formally register the Stage 1 manuscript in a time-stamped repository at the point of IPA. The remaining 27 journals do not require protocol transparency. The fact that a journal nominally requires protocol transparency does not guarantee that it will happen[14]; equally, the lack of such a policy does not prohibit authors from voluntarily registering their Stage 1 protocols. Authors can register their accepted Stage 1 protocols either publicly or under a temporary private embargo using a dedicated RR registration portal established by the Center for Open Science (https://osf.io/rr/). Entries in this registry become discoverable once made public[15].

D. Rapid review network for COVID-19 Registered Reports

The COVID-19 rapid review network for RRs was established in March 2020 at the journal *Royal Society Open Science* (see original call for submissions). As part of this special initiative, the journal calls for submissions that are relevant to any aspect of COVID-19 in any field, including biological, medical, economic, and psychological research, while also seeking specialist reviewers who are willing and able to evaluate Stage 1 RRs within 24-48 hours of accepting a review request. To date, nearly 900 scientists across a range of disciplines have joined the reviewer network, which is also accessible to other journals. To gain access, the journal must commit to rapid peer review of COVID-19 RRs – striving for 7 days for the initial Stage 1 review round – and waive all article processing charges. Since then, 11 additional journals including *Nature Human Behaviour*, *Nature Communications*, *PLOS Biology*, *Collabra: Psychology*, and all seven *PeerJ* journals. Submission statistics for *Royal Society Open Science* are shown in Supplementary Table S1.

Supplementary Table S1. Summary statistics for Stage 1 and Stage 2 RRs received by *Royal Society Open Science* as part of the COVID-19 rapid review network between March 2020 – April 2021. At the time of writing, 17 submissions have been received. Of these, 14 manuscripts have been awarded Stage 1 in-principle

acceptance, including 8 studies currently in progress and 6 fully-published Stage 2 RRs. Note that *review duration* in the table includes the time taken from submission to an editorial decision (or interim decision), and thus includes the time taken by reviewers as well as the editorial handling time; it does not include time taken by authors to revise or resubmit following an interim decision. So far, the average review duration per round is consistently less than the 7 days, with an average of 4 reviewers recruited per manuscript, which together point to the feasibility of the rapid review process in achieving a sufficient number of specialist reviews in a short time frame. Peer reviews and editorial decisions are published alongside the final Stage 2 RRs (see here).

	Stage 1			**Stage 2**			**Stage 1 & 2**
	No. reviewers	**Review duration per round (days)**	**Total review time (days)**	**No. reviewers**	**Review duration per round (days)**	**Total review time (days)**	**Total review time (days)**
Average	4.1	6.8	15.4	2.8	5.2	8.7	17.7
Median	4	5.4	10.8	3	4.2	7.3	17
Minimum	2	1.1	3.3	1	1.5	2.9	6.2
Maximum	6	13.1	44.3	4	11	17.9	28.7
Std. Dev.	1.1	4.2	13.5	1	4.2	6.3	9.4
No. submissions	17	17	17	6	6	6	6

References

1. Guo, W., Del Vecchio, M. & Pogrebna, G. Global network centrality of university rankings. *R. Soc. Open Sci.* **4**, 171172 (2017).
2. Chambers, C. D. Verification Reports: A new article type at Cortex. *Cortex* **129**, A1–A3 (2020).
3. Syed, M. Infant and Child Development: A journal for open, transparent and inclusive science from prenatal through emerging adulthood. *Infant Child Dev.* **30**, e2215 (2021).
4. Button, K. S., Bal, L., Clark, A. & Shipley, T. Preventing the ends from justifying the means: withholding results to address publication bias in peer-review. *BMC Psychol.* **4**, 59 (2016).

5. Srivastava, S. A Pottery Barn rule for scientific journals. *The Hardest Science* https://thehardestscience.com/2012/09/27/a-pottery-barn-rule-for-scientific-journals/ (2012).
6. Lilienfeld, S. O. Clinical Psychological Science: Then and Now. *Clin. Psychol. Sci.* **5**, 3–13 (2017).
7. Lucas, R. E. & Donnellan, M. B. Enhancing transparency and openness at the Journal of Research in Personality. *J. Res. Personal.* **68**, 1–4 (2017).
8. Preregistered Direct Replications: A New Article Type in Psychological Science. *APS Obs.* **30**, (2017).
9. Royal Society Open Science. Replication studies. https://royalsocietypublishing.org/rsos/replication-studies.
10. The Royal Society. Reproducibility and transparency collection. https://royalsocietypublishing.org/topic/special-collections/rsos-reproducibility.
11. Kriegeskorte, N. Open Evaluation: A Vision for Entirely Transparent Post-Publication Peer Review and Rating for Science. *Front. Comput. Neurosci.* **6**, 79 (2012).
12. Murray, H. Transparency meets transparency. *F1000 Blogs* https://blog.f1000.com/2017/10/12/transparency-meets-transparency/ (2017).
13. Carlsson, R. *et al.* Inaugural Editorial of Meta-Psychology. *Meta-Psychol.* **1**, a1001–a1001 (2017).
14. Hardwicke, T. E. & Ioannidis, J. P. A. Mapping the universe of registered reports. *Nat. Hum. Behav.* **2**, 793–796 (2018).
15. OSF Registries. https://osf.io/registries/discover?provider=OSF&type=Registered%20Report%20Protocol%20Preregistration.