Forensic science needs registered reports*

**Abstract**

The registered report (RR) format is rapidly being adopted by scientific researchers and journals. RRs flip the peer review process, with reviewers evaluating proposed methods, rather than the data and findings. Editors then accept or reject articles largely based on the pre-data collection review. Accordingly, RRs reduce the incentive for researchers to exaggerate their findings, and they make any data-driven changes to the methods and analysis more conspicuous. They also reduce publication bias, ensuring studies with null or otherwise unfavorable results are published. RRs are being used in many fields to improve research practices and increase confidence in study findings. The authors suggest RRs ought to be the default way in which validation studies are conducted and reported in the forensic sciences. They produce more reliable findings, advance criminal justice values, and will lead to several efficiencies in the research process.

© 2019 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Science is an ongoing race between our inventing ways to fool ourselves, and our inventing ways to avoid fooling ourselves [1].

Not all testing is equal — and some forms of testing can produce misleading results. Clinical medical researchers learned this lesson decades ago when published studies in their field concluded that an unrealistically high percentage of new drugs were effective [2,3]. Since then, researchers in many other fields have found a startling number of false positive findings attributable to the fact that earlier tests were not sufficiently controlled and demanding [4–9]. One particular reform aimed at curbing these problems (i.e., a new way to avoid fooling ourselves), registered reports (RRs), is rapidly being adopted by researchers and journals (see Fig. 1) [10]. RRs flip the peer review process, with reviewers evaluating proposed methods, rather than the data and findings. Editors then accept or reject articles primarily on the basis of the pre-data collection review [10,11]. Accordingly, RRs reduce the incentive for researchers to exaggerate their findings, and they make any data-driven changes to the methods and analysis more conspicuous. In this article, we suggest that an RR is the most appropriate way to conduct and report a validation study in forensic feature-comparison disciplines.

After a brief description of RRs, we will consider their expanding use across the sciences and the weaknesses in the research process they respond to. From there, we turn to forensic science, particularly recent calls for validation testing of forensic feature-comparison disciplines. Foundational tests in these disciplines are confirmatory, have a clear research question, and demand a high level of rigor and precision. In consequence, they are well suited to the RR format. Moreover, the criminal justice system requires accurate and transparent forensic feature-comparison methods. Traditional closed testing practices thwart these system aspirations. Registered reports advance them.

1. What are registered reports?

There are many excellent descriptions of RRs and guides to implementing them, so we will provide only the necessary background [10,12,13]. The RR format encompasses two key innovations: pre-specification of hypotheses, methods, and analysis; and, division of the peer review process into two stages.

At the first stage of the review process, the authors pose a research question, create materials, devise a methodology, and formulate an analysis plan. The stage 1 “manuscript” is then submitted to a journal for peer review focused on the soundness of the question and the validity of the methods. If the stage 1 manuscript is accepted, publication is virtually guaranteed at stage 2 if that plan is followed. An editor at the neuroscientific journal Cortex recently reported that the stage 2 acceptance rate at that journal is 100% [10].

RRs are quickly gaining acceptance in a range of scientific disciplines. The number of journals that accept RRs has expanded from just a handful in 2014 to over 200 at the most recent count (see Fig. 1) [10]. These include influential biomedical, psychological, and multidisciplinary journals. We will now turn to the reasons for this increased interest in RRs and the advantages they confer.

2. The rise of registered reports

What part of a research study — hypotheses, methods, results, or discussion — should remain beyond a scientist’s control? The
answer, of course, is the results: the part that matters most for publishing in prestigious journals and advancing careers. This paradox means that the careful scepticism required to avoid massaging data or skewing analysis is pitted against the drive to identify eye-catching outcomes. Unbiased, negative and complicated findings lose out to cherry-picked highlights that can bring prominent articles, grant funding, promotion and esteem [10].

RRs respond to the “results paradox” described in this quote. Put simply, researchers naturally care about the results of studies [14]. This creates a motivation (unconscious or otherwise) to use practices that help ensure the results tell a clean story that fits their hypotheses. These practices include dropping observations for ad hoc reasons and strategically stopping data collection as soon as the desired effect is found [15]. Metascientific researchers have termed these methods questionable research practices (QRPs) [16,17]. They are widely used in many fields [16–18]. QRPs demonstrably bias research results, making it possible to portray even a randomly generated data set as probative of the researchers’ desired result [15]. Beyond QRPs, researchers may simply not publish studies that do not support their hypothesis or desired outcome (i.e., publication bias) [19].

Scientific concern with QRPs — and the reforms that have been responsive instituted — came to a head recently with the failure of several largescale efforts to confirm findings published in leading scientific journals [4–9]. Across several efforts at replication, only about 50% of studies confirmed previous findings, and even those that were confirmed reported considerably smaller effects. Some refer to these failures as a “crisis.” [20] But such worries are not new.

Clinical medical research was confronted with similar problems in the 1980s and 90s. During this time, attentive researchers were becoming concerned that drug company-funded clinical trials were rarely finding that the drug under examination was ineffective [2,3,21]. These concerns prompted legal reform mandating trial preregistration on a government online registry to counter publication bias [21]. Preregistration creates a record of the study so it can be found even if it is not published. It also makes it possible to determine if the researchers made major changes to their research after seeing initial results (i.e., engaged in QRPs).

The road to acceptance of preregistration in clinical medicine was not straightforward. One challenge has been researchers’ hesitation to preregister, despite the legal requirement [21,22]. This reluctance diminished somewhat after 2004 when the New York State Attorney General sued GlaxoSmithKline (GSK), alleging it failed to disclose null results [23]. GSK then announced it would begin preregistering its studies. Shortly after the GSK affair, the International Committee of Medical Journal Editors (ICMJE) issued a new policy stating that its member journals would only publish preregistered clinical medical trials [24]. The number of trials being registered online increased significantly following the implementation of the policy in 2005, likely due to the prestigious status of the participating journals — which include the Journal of the American Medical Association, The New England Journal of Medicine and The Lancet [25].

Despite these early bumps in the road, preregistration has had beneficial effects in medicine. For example, one recent analysis of large trials funded by the National Heart, Lung, and Blood Institute found that prior to 2000, 43% reported negative results. This number rose dramatically to 92% negative results after 2000 [2].

RRs extend medicine’s preregistration reforms, effectively subjecting the materials that will be preregistered to peer review. Splitting peer review into two stages decreases incentives to engage in QRPs because publication does not depend on the data. And, as with preregistration, RRs make it possible to assess data-contingent changes to the methods and analysis. Furthermore, by engaging peer review at an earlier stage, RRs facilitate better research by avoiding mistakes and strengthening designs.

Early returns on RRs are encouraging. One recent analysis found that studies conducted as RRs support the main hypothesis just 45% of the time (see Fig. 2) [26]. This contrasts with the traditional scientific literature, which contains an overwhelming and unlikely percentage of positive findings (80–95%) [26]. Anecdotally, one editor has reported that early stage peer review has also been a useful part of the process: “One of the most striking characteristics of RRs is that they can help authors to improve the protocol or rationale while it is still possible to make changes. I have overseen numerous cases in which reviewers have intervened to prevent a serious flaw in a study design.” [10].

![Fig. 1. Number of journals adopting RRs by year, with several key dates.](image)
3. Validation studies in the forensic sciences

One important type of study in forensic feature-comparison disciplines — perhaps the most important — is the validation study [27]. These studies test whether a forensic feature comparison procedure works — if it yields a result that accords with ground truth [28]. In 2009, the National Academy of Sciences famously declared that for many practices, forensic science professionals “have yet to establish either the validity of their approach or the accuracy of their conclusions” [29]. Subsequently, in 2016, the U.S. President’s Council of Advisors on Science and Technology reaffirmed the importance of validation studies and found that only DNA analysis from single source samples and fingerprint analysis had been appropriately validated [28].

Although validation studies are ongoing in the forensic feature-comparison disciplines (in response to the NAS and PCAST Reports), there are serious risks that prospective studies will be poorly designed, strongly oriented to confirming existing practices of forensic scientists, and rarely disclosed to defendants and attentive scientists [30]. Indeed, issues of study design and analysis were conspicuous in the PCAST report [28]. Consider, for instance, studies assessing the accuracy of firearms analysis. PCAST found the available studies — those routinely relied upon by those providing firearms evidence — to be poorly designed and to suggest misleadingly low error rates [28]. The RR format can help to avoid overclaiming and ensure an accurate and transparent record of all the validation studies conducted across the forensic sciences.

4. The benefits and importance of registered validation studies

RRs ought to be the default way in which validation studies are conducted and reported. In this section we provide several arguments in favor of RRs, namely: (1) they help to produce more reliable findings; (2) they are more susceptible to comprehension and confrontation because they make the strengths and weaknesses of the research more apparent; (3) they encourage early stage fixes to methodological weaknesses in studies; (4) validation studies are not exploratory, a feature that aligns with the strengths of RRs; and (5) RRs help to manage researchers’ biases, producing several downstream efficiencies.

(1) More reliable findings

First, RRs of studies supporting forensic identification methods advance legal goals that depend on the reliability of scientific evidence. Access to the results of methodologically rigorous studies helps users — investigators, courts, factfinders, and parties — make decisions about whether to pursue other lines of investigation, consider the value of expert opinions, determine whether to make or encourage a guilty plea, and decide whether to formally challenge the evidence [31]. Indeed, even studies with null findings (which are guaranteed publication under the RR format) are still informative and important — it is important to know what works and what does not work [26]. In many jurisdictions, such as the United States, Canada and to some extent England and Wales, and New Zealand, evidence rules require judges to consider reliability in their admissibility determinations [32–35]. Without rigorous methods, forensic science risks repeating the mistakes encountered in other fields [4–9].

(2) More susceptible to comprehension and confrontation

RRs also improve the defendant’s ability to comprehend and, where appropriate, confront inculpatory state evidence [36]. RRs may, for instance, suggest a study’s findings should be viewed cautiously because the researchers did not follow the registered analysis plan and did not have good reason for deviating from it. Either way, deviations will be on the record. Similarly, registration helps to ensure that all validation studies are traceable — not just those with results that favor the group reporting the research. On the other hand, forensic scientists should be aware that RRs may support the probative value of a forensic science procedure in ways that were not previously available. An example may assist.

In its review of fingerprint studies, PCAST was complimentary of an FBI-sponsored study by Ulery and colleagues [37] and critical of a NIJ-sponsored study by Pacheco, Cerchiai and Stoiloff (the Miami-Dade Study) [38]. The review identified several methodological problems with the Miami-Dade study, including a possibly results-driven approach to determining if the examiner made an error:

The paper observes that “in 35 of the erroneous identifications the participants appeared to have made a clerical error, but the authors could not determine this with certainty.” In validation studies, it is inappropriate to exclude errors in a post hoc manner (see Box 4) [28].

As PCAST implies, without preregistration we cannot readily ascertain whether the examiners simply recorded their answers carelessly (i.e., a clerical error) or whether they were actually mistaken in their identification [39].

Had the Miami-Dade study been conducted as an RR, the issue of clerical errors would have been clearer. It may have been that the erroneous identifications were indeed clerical errors and that their exclusion was not influenced by a motivation to demonstrate the accuracy of fingerprint analysis. Prespecified methodologies and analysis plans typically incorporate criteria defining what qualifies as an outlier or clerical error that guide subsequent analysis [40]. In any event, PCAST’s criticism provides a glimpse of future cases where litigants may seek to challenge methodologically ambiguous (or inadequate) studies supporting forensic feature-comparison disciplines.

(3) Early stage methodological improvements

The pre-data collection peer review aspect of RRs may also be

Fig. 2. This early analysis of 113 RRs (the first three data points) compiled by the Center for Open Science found that RRs (including both replications and novel studies) were more likely to report null results than the rest of the scientific literature (the final dot). The RR studies were concentrated in the psychological and biological literature [26].

J.M. Chin et al. / Forensic Science International: Synergy 2 (2020) 41–45

particularly beneficial in the design of studies for the forensic feature-comparison disciplines. One historic critique of some forensic science research is a lack of alignment with academic scientific norms [41]. Another, is the limited methodological expertise of many forensic practitioners [42]. Constructive stage 1 reviews of forensic science RRs may include suggestions from independent scientists with expertise in methodology. In the case of the criticized Miami-Dade study, for instance, methodological flaws may have been identified during the review process. Such review would help ensure that validation tests are performed as rigorously as feasible, and that the results are less vulnerable to challenge.

(4) Validation studies align with the strengths of RRs

Another reason validation studies are particularly apt for the RR format is that they are not exploratory and instead seek to confirm an existing suspicion that a method works. This aligns their aims with the strengths of RRs [10]. For instance, one common qualm among researchers considering submitting an RR is that it may stifle their ability to conduct exploratory analyses on their data. This hesitation is somewhat misplaced because RRs do not prohibit exploratory work; they simply make clear which parts of the analysis are exploratory. In any case, no such misgivings should arise with validation studies. They are conducted once a practice has been developed. The goal of a validation test is to confirm that the method works and to provide an estimate of error. In other words, validation tests should never be exploratory and if there is an exploratory component, that component must be clearly specified in advance.

(5) RRs control biases and improve efficiency

Finally, preregistration’s genesis in pharmaceutical and biomedical research reinforces its importance in forensic science research. As we explained, a pressing problem with biomedical research was the biasing impact of the funders (which is still a concern to some extent), who sought published studies supporting the effectiveness of their products. Similarly, in forensic science, independent reviews of methods and findings will lend the imprimatur of mainstream science. It will help to insulate researchers and individual forensic scientists from allegations of bias because RRs represent a substantial scientific safeguard. These advances contrast with opaque in-house research and reviews, which does little to protect the actual and perceived legitimacy of test results [43]. RRs enable forensic scientists to rely on validation studies that were conducted transparently and subjected to robust scrutiny. Further, just as there is clear public importance in producing demonstrably reliable medicines and therapeutics, society needs demonstrably reliable forensic science procedures.

Unlike Pharma and biomedicine — with access to billions of dollars of private and public capital to invest in research and development — it is all the more important that the resource-starved forensic sciences make sure that they do the most meaningful and methodologically rigorous studies possible on their comparatively modest budgets. Registration may also encourage collaboration with academic researchers as well as across agencies and jurisdictions.

In the long run, RRs may also reduce the burden on the courts that unreliable forensic practices present. With RRs, forensic science communities, prosecutors, and defense counsel will all have access to studies that were critically reviewed. One day, RRs may contribute to largescale systematic reviews of forensic science fields, like those produced by the Cochrane Collaboration in healthcare and diagnostic tests. Such work will not completely eliminate scope for interpretative differences, particularly on the applicability of foundational research to different laboratories or the facts of a specific case. But, as the forensic sciences align themselves with the best scientific practices, the state’s evidence will grow resistant to speculative challenges.

5. Bringing registered reports to forensic science research

So how can we move forward with RRs in forensic science research? The first step, as has been the case in other fields, is for journals to begin accepting them. Currently, no forensic science journal accepts RRs [13,44]. This may be due to barriers that could easily be overcome. Metascientific researchers in other fields have suggested such barriers may include: “Lack of knowledge about the format; A presumption that there is no demand among researchers to pursue this format; Misconceptions about the need or utility of the format.” [45].

To overcome these hurdles, researchers from a variety of fields have established a campaign (Registered Reports Now!) to lobby journals to begin accepting RRs [45]. This involves circulating a list of journals that one wishes to see adopt RRs, drafting a letter based on a template available online, adding those who agree to sign, and sending it to the journal editors. Information about the status of the project is available online [46]. We encourage you to contact the authors of this article if you are interested in adding your name to letters sent to forensic science journals.

Declaration of competing interest

The authors declare no conflicts of interest.

CRediT authorship contribution statement

Jason M. Chin: Writing - original draft. Rory McFadden: Writing - original draft. Gary Edmond: Writing - original draft.

References

[1] R. Nuzzo, Fooling ourselves, Nature 626 (2015) 182–185.
[2] R.M. Kaplan, V.L. Irvin, Likelihood of null effects of large NHLBI clinical trials has increased over time, PLoS One 10 (2015), e0123832, https://doi.org/10.1371/journal.pone.0123832.
[3] K. Dickerson, I. Chalmers, Recognizing, investigating and dealing with incomplete and biased reporting of clinical research: from Francis Bacon to the WHO, J. R. Soc. Med. 104 (2011) 532–538, https://doi.org/10.1258/jrsm.2011.110402.
[4] Open Science Collaboration, Estimating the reproducibility of psychological science, Science 349 (2015) aac4716, https://doi.org/10.1126/science.aac4716.
[5] R.A. Klein, et al., Investigating variation in replicability A “Many labs” replication project, Soc. Psych. 45 (2014) 142–152, https://doi.org/10.1037/s0033-2909.14535.a000178.
[6] R.A. Klein, Many labs 2: investigating variation in replicability across samples and settings, Adv. Meth. Pract. Psych. Sci. 1 (2018) 443–490, https://doi.org/10.1177/2515245918810225.
[7] C.R. Ebersole, et al., Many Labs 3: evaluating participant pool quality across the academic semester via replication, J. Exp. Soc. Psychol. 67 (2016) 68–82, https://doi.org/10.1016/j.jesp.2015.10.012.
[8] C.F. Camerer, et al., Evaluating replicability of laboratory experiments in economics, Science 351 (2016) 1433–1436, https://doi.org/10.1126/science.aaf918.
[9] C.F. Camerer, et al., Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015, Nat. Hum. Behav. 2 (2018) 637–644, https://doi.org/10.1038/s41598-018-0359-2.
[10] C. Chambers, What’s next for registered reports? Nature 573 (2019) 187–189.
[11] National Academies of Sciences, Engineering, and medicine, in, in: Open Science by Design Realizing a Vision for a 21st Century Research, 2018, https://doi.org/10.17226/25116.
[12] A. Kyonoga, J.M. Scimeca, Practical considerations for navigating registered reports, Trends Neurosci. 42 (2019) 568–572, https://doi.org/10.1016/j.tins.2019.07.003.
[13] Center for Open Science, Registered Reports: Peer Review before Results Are Known to Align Scientific Values and Practices. https://cos.io/rr/; (Accessed 10 October 2019).
[14] U. Mitroff, Norms and counter-norms in a select group of the apollo moon
scientists: a case study in the ambivalence of scientists, Am. Sociol. Rev. 39 (1974) 579–595, https://doi.org/10.2307/2094423.

J.P. Simmons, L.D. Nelson, U. Simonsohn, False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant, Psychol. Sci. 22 (2011) 1359–1366, https://doi.org/10.1177/0956797611417632.

J.M. Chin, G. Ribeiro, A. Rairden, Open forensic science, J. Law Biosci. (2019), 10.1145/3340220.

K. Dickersin, D. Rennie, The evolution of trial registries and their use to assess the clinical trial enterprise, J. Am. Med. Assoc. 307 (2012) 1861–1864, https://doi.org/10.1001/jama.2012.4230.

B. Goldacre, et al., Compliance with requirement to report results on the EU Clinical Trials Register: cohort study and web resource, BMJ 362 (2018) k3218, https://doi.org/10.1136/bmj.k3218.

M. Wadman, Spitzer sues drug giant for deceiving doctor, Nature 429 (2004) 2095–2096, https://doi.org/10.1038/4292095a.

C. De Angelis, et al., Questionable research practices in ecology and evolution, PLoS One 11 (2016), e0200303, https://doi.org/10.1371/journal.pone.0200303.

K. Fiedler, N. Schwarz, Questionable research practices revisited, Soc. Psychol. Pers. Sci. 7 (2016) 45–52, https://doi.org/10.1177/1948550615612150.

K. Dickersin, The existence of publication bias and risk factors for its occurrence, J. Am. Med. Assoc. 263 (1990) 1385–1389, https://doi.org/10.1001/jama.1990.03440100109014.

M. Baker, 1,500 scientists lift the lid on reproducibility, Nature 533 (2016) 452–454.

K. Dickersin, D. Bennie, The evolution of trial registries and their use to assess the clinical trial enterprise, J. Am. Med. Assoc. 307 (2012) 1861–1864, https://doi.org/10.1001/jama.2012.4230.

B. Goldacre, et al., Compliance with requirement to report results on the EU Clinical Trials Register: cohort study and web resource, BMJ 362 (2018) k3218, https://doi.org/10.1136/bmj.k3218.

M. Wadman, Spitzer sues drug giant for deceiving doctor, Nature 429 (2004) 2095–2096, https://doi.org/10.1038/4292095a.

C. De Angelis, et al., Questionable research practices in ecology and evolution, PLoS One 11 (2016), e0200303, https://doi.org/10.1371/journal.pone.0200303.

K. Fiedler, N. Schwarz, Questionable research practices revisited, Soc. Psychol. Pers. Sci. 7 (2016) 45–52, https://doi.org/10.1177/1948550615612150.

K. Dickersin, The existence of publication bias and risk factors for its occurrence, J. Am. Med. Assoc. 263 (1990) 1385–1389, https://doi.org/10.1001/jama.1990.03440100109014.

M. Baker, 1,500 scientists lift the lid on reproducibility, Nature 533 (2016) 452–454.

K. Dickersin, D. Bennie, The evolution of trial registries and their use to assess the clinical trial enterprise, J. Am. Med. Assoc. 307 (2012) 1861–1864, https://doi.org/10.1001/jama.2012.4230.

B. Goldacre, et al., Compliance with requirement to report results on the EU Clinical Trials Register: cohort study and web resource, BMJ 362 (2018) k3218, https://doi.org/10.1136/bmj.k3218.

M. Wadman, Spitzer sues drug giant for deceiving doctor, Nature 429 (2004) 2095–2096, https://doi.org/10.1038/4292095a.

C. De Angelis, et al., Questionable research practices in ecology and evolution, PLoS One 11 (2016), e0200303, https://doi.org/10.1371/journal.pone.0200303.

K. Fiedler, N. Schwarz, Questionable research practices revisited, Soc. Psychol. Pers. Sci. 7 (2016) 45–52, https://doi.org/10.1177/1948550615612150.

K. Dickersin, The existence of publication bias and risk factors for its occurrence, J. Am. Med. Assoc. 263 (1990) 1385–1389, https://doi.org/10.1001/jama.1990.03440100109014.

M. Baker, 1,500 scientists lift the lid on reproducibility, Nature 533 (2016) 452–454.

K. Dickersin, D. Bennie, The evolution of trial registries and their use to assess the clinical trial enterprise, J. Am. Med. Assoc. 307 (2012) 1861–1864, https://doi.org/10.1001/jama.2012.4230.