Appendix 5

Open Research Indicators: Report of a pilot on Pre-registration

Authors

Mark Kelson 1†, Christopher Tibbs 1, Frederick Breeze 2, Daryl O'Connor 3, Nick Sheppard 3

1 University of Exeter, 2 University of Manchester, 3 University of Leeds
† Correspondence should be addressed to Mark Kelson M.J.Kelson@exeter.ac.uk.

Executive Summary

This report presents findings from ORI Pilot 6, which explored the feasibility of developing an open research indicator for pre-registration. Pre-registration is a key practice in promoting research transparency and integrity, helping to mitigate questionable research practices such as HARKing and selective reporting. Despite its importance, pre-registration remains difficult to detect and evaluate at scale due to inconsistent reporting and lack of standardisation.
The pilot involved collaboration between UK universities and data solution providers - OpenAIRE, Digital Science, and PLOS/DataSeer - who applied automated methods to identify self-reported pre-registration statements in open access articles published in 2023. A subset of outputs was manually coded to benchmark the accuracy of these approaches.

Findings revealed that automated detection of pre-registration is achievable, with all three providers demonstrating high levels of accuracy (81–88%), sensitivity (79–93%), and specificity (78–86%). However, the pilot also highlighted key limitations: current methods cannot reliably assess whether pre-registration occurred before research began, nor can they evaluate the quality or adherence to the registered plans.

The report underscores that self-reported pre-registration alone is not a sufficient indicator of good research practice. It calls for caution in using such indicators for benchmarking, given disciplinary differences, institutional variability, and the lack of consistent policy frameworks. Nonetheless, increasing public declarations of pre-registration may foster transparency and enable future development of more robust indicators.

Recommendations include refining eligibility criteria, improving registry coverage, and exploring integration with timestamped records to assess timing and quality. The pilot demonstrates the value of collaborative, iterative approaches and provides a foundation for future work in developing meaningful indicators of open research practices.

Introduction

Pre-registration is seen as important as it helps minimize questionable research practices (QRPs) by preventing researchers from altering hypotheses after results are known (HARKing), omitting non-significant results, stopping data collection based on statistical significance, selecting outcome variables post hoc, or using multiple analysis strategies to highlight only significant findings. These practices, often described as the "garden of forking paths," can distort research findings and misrepresent evidence. Therefore, there is a strong justification for wanting to have an effective method of gauging how institutions/researchers are increasing use of this aspect of open research.

However, the practice is difficult to automatically identify or evaluate. We note that negative statements are rare (only one of the papers we sampled explicitly stated that it was not pre-registered), and it is more typical for studies to declare that they are pre-registered or not to mention the practice at all.

There are numerous challenges in identifying pre-registration statements. In common With Data Availability Statements, there is no set format or place in the text where the information about pre-registration 'belongs'. Conceptually then, this pilot has several convergences with the DAS indicator pilot, and overlaps in part with the openness and FAIRness of data pilots in the potential ambiguity of unitary indicators, and the range of information challenges posed by the pilot assessment of the representation of research practices in the resulting literature.

So far, attempts to monitor pre-registration in the published literature have been relatively small scale, have tended to focus on specific disciplinary clusters (e.g. a defined subject domain or a group of journals), and have focused on declarations of pre-registrations in published articles.

Hardwicke *et al* (2022)[1] sampled 250 psychology articles that were published between 2014 and 2017, manually checking for the presence of "transparency and reproducibility-related research practices" and tabulating their presence. The practices explored included pre-registration alongside open licensing; sharing of research materials/protocols/data/code; funding and conflict of interest statements; and replication studies. They found that just 3% of the studies reported were declared as pre-registered. Poole *et al* (2024)[2] found a similarly low rate of pre-registration in their review of articles published in six autism research journals between 2011 and 2022. They found that just 2.23% (n=192) of studies reported were pre-registered. Furthermore, they manually assessed the quality of the pre-registrations, finding that "specificity in the pre-registrations was low, particularly in the design and analysis components of the pre-registration. In addition, only 28% of sampled manuscripts adhered to their analysis plan or transparently disclosed all deviations."

These low levels of pre-registration are inconsistent with stated levels of support for the practice. Ferguson *et al* (2023)[3] conducted two surveys of social science researchers to assess levels of 'open science practices' including pre-registration. and found that while 58% of researchers privately expressed support for pre-registration, only 25% reported having pre-registered at least one study. This self-reporting from the survey was tested by randomly

---

[1] Hardwicke, T. E., Thibault, R. T., Kosie, J. E., Wallach, J. D., Kidwell, M. C., & Ioannidis, J. P. A. (2021). Estimating the Prevalence of Transparency and Reproducibility-Related Research Practices in Psychology (2014–2017). Perspectives on Psychological Science, 17(1), 239-251. https://doi.org/10.1177/1745691620979806

[2] Poole, D., Linden, A., Sedgewick, F., Allchin, O., & Hobson, H. (2024). A systematic review of pre-registration in autism research journals. Autism, 0(0). https://doi.org/10.1177/13623613241308312

[3] Ferguson, J., Littman, R., Christensen, G. *et al.* Survey of open science practices and attitudes in the social sciences. *Nat Commun* **14**, 5401 (2023). https://doi.org/10.1038/s41467-023-41111-1

sampling publications by authors who had completed both surveys. The manual audit results were closely aligned with the self-reported behaviours. The authors therefore concluded that their survey findings were likely accurate. Manual sampling of pre-registration practices has consistently found low levels of the practice. Furthermore, so far self-reporting via surveys, while demanding and difficult to replicate consistently or at scale, seems to reinforce the estimates derived from manual sampling.

Low levels of pre-registration should not perhaps be a surprise, as very few journals seem to be systematically requiring it. Spitschan, Schmidt, and Blume (2021)[4] analysed author guidelines in 27 sleep research and chronobiology journals, rating them using the Transparency and Openness Promotion (TOP) Factor guidelines[5] as criteria. They found very low levels of specificity or consistency across journal policies with respect to pre-registration, and a significant minority of the journals did not have a requirement at all. The study also echoed Hardwicke *et al*'s finding that the quality of pre-registration was low "with accountability consequently likewise being very low".

Lee *et al* (2025)[6] evaluated the use of generative artificial intelligence tools (ChatGPT and Claude) compared to manual sampling in detecting pre-registration statements in the published literature. Using a random sample of articles submitted to the UK's Research Excellence Framework evaluation exercise in 2021 by University College London, they found that the Chatbots were close in accuracy both to each other and to human assessors at detecting the presence of a pre-registration statement, with approximately 90% agreement. However, none of the approaches evaluated the quality or level of specificity of the pre-registration.

Simply detecting a statement or citation of a pre-registration in an article is not straightforward and is of limited value in and of itself without a further evaluation of the specificity and accuracy of the pre-registration itself. As such, the presence of a pre-registration statement is not a sufficient indicator of good practice. Gorman's (2020)[7] commentary on the assumed value of pre-registration highlights the issues that this presents, emphasising that the assumption that a pre-registration represents real-life good research practices imbues a study with greater credibility than is warranted. The lack of quality assurance on a pre-registration assertion, according to this argument, means that journals that do not evaluate or enforce rigour and quality "are aiding the authors of these manuscripts in deceiving readers into believing that the results presented have greater validity than those from unregistered studies".

The risks of using a self-reported assertion of pre-registration as an indicator of quality, rigour, or research integrity without a robust framework for monitoring the timing, content, and adherence to or deviation from the registration in place, according to Klonsky (2025)[8], represent a useful tool being "converted into an indicator of strong science and a goal in and of itself.... [which] will distort its use and harm psychological science in unanticipated ways".

---

[4] Spitschan M, Schmidt MH and Blume C. Principles of open, transparent and reproducible science in author guidelines of sleep research and chronobiology journals [version 2; peer review: 2 approved, 1 approved with reservations]. Wellcome Open Res 2021, 5:172 (https://doi.org/10.12688/wellcomeopenres.16111.2)

[5] https://topfactor.org

[6] Lee, D. Y. H., Parker, A. J., Norbury, C., & Shanks, D. (2025, February 25). Validating AI-assisted evaluation of open science practices in brain sciences: ChatGPT, Claude, and human expert comparisons. https://doi.org/10.31234/osf.io/9euw8_v2

[7] Gorman, D. M. (2020) Commentary on Vassar *et al*. (2020): Cautionary observations on the pre-registration revolution. *Addiction*, 115: 1180–1181. https://doi.org/10.1111/add.14945.

[8] Klonsky, E. D. (2025). Campbell's Law Explains the Replication Crisis: Pre-Registration Badges Are History Repeating. Assessment, 32(2), 224-234. https://doi.org/10.1177/10731911241253430

Citing evidence that papers citing pre-registrations "routinely violate the rules and spirit of pre-registration", Klonsky argues that the use of pre-registration as an indicator will encourage questionable research practices, including the new phenomenon of pre-registering after results are known (PRARKing) and enhance the perceived trustworthiness of fragile findings.

Considering the labour intensiveness of checking and validating pre-registration claims, the relatively low instance of pre-registration in the published literature, and the potential pitfalls of relying on pre-registration assertions being taken at face value, we nevertheless believe that there is value in assessing levels of self-reporting. If more researchers publicly claim to have pre-registered and tie those claims to verifiable, timestamped records, then the practice or pre-registration will become more widespread, and the growing preponderance of self-reported pre-registrations will enable us to find the datapoints and data sources to create more robust, in-depth indicators. Even a public declaration of pre-registration that turns out to be incorrect in that it allows future researchers the ability to double check the validity of the pre-registration claim being made. In this context, the value of the openness of research comes from the sense of transparency as much as accessibility.

Given the complexities of evaluating the prevalence and significance of pre-registration as represented in the literature, it was helpful to partner with so many data providers. The mix of approaches and commitment to collegial working enabled us to extract useful lessons and demonstrated the spirit of openness underpinning the collaboration.

Methods

The project team comprised researchers from the universities of Manchester, Leeds and Exeter and representatives from OpenAIRE, Digital Science, and PlosONE/DataSeer.
We met weekly throughout the pilot to discuss and agree our approach.

Together the team developed a specification (Annex 1) to document the requirement being explored by the pilot, drawing from a SCOPE review of the rationale for being interested in monitoring pre-registration.

Each of the universities then contributed a list of research outputs following a common dataset specification:
- output(s) must be only articles
- output(s) must be published in 2023
- output(s) must be published open access
- output(s) must include the mandatory fields: local identifier, DOI, local URL, output type, year of publication, organisation unit

Solutions providers then used (or augmented) their existing automated detection approaches to explore this dataset. A brief description of the approach taken is provided below.

OpenAIRE:

OpenAIRE employed a systematic text mining methodology to identify (pre)registration statements in scholarly publications associated with the UKRN Pilot 6 institutions. The process targeted full-text publications in the OpenAIRE Graph[9] for the year 2023, using a high-recall, multi-stage strategy that was later refined to improve precision.

---

[9] https://graph.openaire.eu/

PDFs were first converted into machine-readable full-text, then analyzed using a custom pipeline powered by the OpenAIRE Information Inference System (IIS)[10]. This system leverages madIS[11] and a suite of user-defined functions (UDFs) to execute SQL-based routines. These routines combine handcrafted rules and regular expressions to detect (pre)registration identifiers and platform-specific patterns (e.g., AsPredicted, OSF, PROSPERO, Clinical Trial Registries). The approach is context-aware, capturing both direct references (e.g., URLs and codes) and indirect mentions—such as platform names near verbs like "pre-registered" or "was registered at." Variants and misspellings were handled to ensure broader coverage. Alongside the extracted identifiers, the system captured contextual snippets—before, during, and after each match—to support verification and further analysis.

Filtering was applied to restrict results to UKRN-supplied DOIs, using DOI and institutional affiliation data from the OpenAIRE Graph. Two types of outputs were generated (originally in JSON, delivered as Excel files):

1. A detailed dataset of matched publications, including the platform, identifier, text snippets, and linked metadata (e.g., funding information).
2. A supplementary dataset listing all UKRN DOIs processed, indicating whether a (pre)registration was found.

These datasets were designed to support institutional review and included metadata such as organization ID, matched platform, snippet context, and associated funding details (e.g., project IDs, acronyms, and funding streams). Multiple lines per DOI were retained where different platforms, institutions, or versions of the same publication were identified.

Digital Science:

In total, **27828** distinct publications were matched to publication records within the Dimensions database (using the DOIs provided).
25821 publications were available for full text analysis - with University of Manchester data added at this point.

In order to extract a fair representation of records for analysis, a GBQ algorithm was written to extract a similar number of publications from each institution, with an even distribution of potential pre-reg vs non pre-reg publications. The dataset was also narrowed down to the 4 participating pilot institutions.

2532 publications were initially identified as having potential Pre-Registration related text.
(A reiteration of the algorithm later in the process resulted in **2146)**.

To avoid overwhelming the final dataset with publications that did not reference pre-registration practices, the algorithm was designed to extract all flagged publications (i.e., those containing potential pre-registration terms) and a random sample of unflagged publications. This approach ensured that the dataset remained balanced, manageable, and suitable for manual review, while still enabling meaningful comparisons across institutions. **Without this step, the much larger number of non-pre-registration publications would have dominated the dataset, making fair analysis more difficult.**

Specifically:

---

[10] https://github.com/openaire/iis. A high-level overview of IIS can be found in the paper "Information Inference in Scholarly Communication Infrastructures: The OpenAIREplus Project Experience", Procedia Computer Science, vol. 38, 2014, 92-99. [sciencedirect.com/science/article]
[11] https://github.com/madgik/madis

- All flagged publications (those containing pre-registration-related keywords - see methodology below) were included in full.

- Unflagged publications (those without any matching keywords) were randomly sampled at three times the number of flagged publications for each institution.

- This ensured an even spread of pre-registration and non-pre-registration outputs across the dataset.

- It prevented the final sample from being overwhelmed by non-pre-registration publications.

- It enabled manageable manual review and fair institutional comparisons.

PLOS-DataSeer:

PLOS and DataSeer developed a new Open Science Indicator for measuring study registration (preregistration). PLOS shared publicly the results of the first version of this indicator applied to the PLOS corpus and its comparator set (~133,000 articles) in June 2024, which includes a small number of articles from UKRN pilot institutions. The UKRN article corpus (=11912 articles) was also analysed using the same method. The method involved developing and detecting regular expressions for 31 study registration databases (registries), including registries for clinical trials, systematic reviews, animal studies, and other types of study (or general purpose registries such as Open Science Framework).

The study registration (preregistration) indicator was created as a preliminary release and optimised for PMC xml. While results for all articles are provided as a proof of concept for these UKRN pilots, the accuracy of results for the 5901 PMC xml articles is currently higher.

More information on the methods and results of the PLOS-DataSeer study registration indicator is available at https://doi.org/10.6084/m9.figshare.21687686.v7

Mapping of the PLOS-DataSeer dataset produced for UKRN to the UKRN requirements for each pilot is here:

UKRN pilots &
PLOS-DataSeer OSI re

General method

The intersection of DOIs from each solution provider was then identified and 113 (the target was 100, but slightly more than that was achieved) of this intersection was double human coded (with discrepancies reconciled by discussion) to produce a human rated judgement (see Overview Report Annex 4). Cross tabulations of this human rating against each of the three solutions providers were produced. Where there were discrepant ratings we explored those DOIs for insights into why differences existed.

Findings

Definitions

Our findings suggested that criteria for measurement are not yet reliably possible. Instead, we identified a key question that precedes the possibility of such metrics: What is an 'ideal' definition of pre-registration, and how does that pose challenges for creating criteria for measurement?

For example, in assessing the temporal aspect of pre-registration: is it feasible to identify/measure when a registration occurred? Could timestamps be used reliably? What other quality measures of a registration are possible? We argue that the ideal definition has three components.

1. The "pre" part. The registration should occur before the work takes place
2. Openness. The registration should be public
3. The "registration" part. The description of what will be done should match the resulting output. In some sense, even a pre-registration that *does not* match is still useful as it can reveal to a careful reader departures from the plan.

This pilot focused on self-reported public registration as a key criterion. We did not attempt to verify the timing or quality of the registration. We note that OSF provided a dataset which was felt to be out of scope for this study, as it did not match up to our approach of identifying references to pre-registration in articles. However, in theory, this could be cross-referenced with self-reported public registration to inform evaluation of timing and quality in future work. Our work focuses, therefore, on the second component.

Characteristics of the data

We created a combined dataset from partner institutions, including all outputs for the year 2023. Both Digital Science and OpenAIRE worked from this dataset. PLOS/DataSeer used their own index to identify the relevant papers.

It is notable that institutions do not typically have complete enumeration of the outputs associated with them. Uploading to institutional repositories is dependent on individuals uploading the relevant work in the right timescale. Institutional repositories are typically augmented with outputs identified by external platforms (e.g. Digital Science). Some combination of these approaches is essential, as in developing a reliable indicator it is critical that the platform being used indexes as much research as possible. The challenge of evaluating the completeness of an output list is by no means solved.

As discussed earlier, identifying whether the registration occurred before the study is an open problem. The date of publication of the output should certainly occur after the date of pre-registration. Otherwise PRARKing or other QRPs become a risk, and the indicator will become susceptible to gaming or distortion – and in fact may incentivise it.

Across disciplines, varying study types, differing methods and conventions of research, there is a significant variation in the applicability of pre-registration as a practice. Many outputs are not eligible for pre-registration and they would ideally be filtered out before an indicator was applied. Eligibility for pre-registration is a continuum, however, and experts may disagree on what should and should not be pre-registered.

We observed "cuckoo" references to pre-registration (i.e. a follow-on study of a clinical trial which references the original registration of the trial, but which contains no information on the follow-on study). From this, we conclude that validity is a key factor to be addressed in

subsequent work, and a reliable way to identify that a pre-registration matches the output that cites it will be important.

Validation

The results of the pilot are provided below. Figure 1 compares the each of the three solutions provider ratings against the Human rating. Overall, 42 out 109 (38.5%) outputs were judged to be pre-registered according to the dual human coding. This is a high prevalence and perhaps is indicative of the fact that the intersection of outputs assessed by all providers is unusual in some sense (perhaps better indexed outputs appear on all three registries and perhaps better indexed outputs are more likely to be pre-registered).

Figure 1: Comparison of human ratings of outputs versus automated approaches

Table 1: Comparison of the three automated approaches taking the human agreed rating as the gold standard.

	True Negative (%)	False Negative (%)	True Positive (%)	False Positive (%)	Sensitivity	Specificity	Accuracy
Digital Science	56 (49.6%)	4 (3.5%)	38 (33.6%)	11 (9.7%)	93	78	86
OpenAIRE	60 (53.1%)	6 (5.3%)	36 (31.9%)	7 (6.2%)	91	84	88
Plos/DataSeer	63 (55.8%)	17 (15%)	25 (22.1%)	4 (3.5%)	79	86	81

The performance of the three approaches is provided in table 1. Overall sensitivity, specificity and accuracy are high across all three approaches.

Interpretation

As observed above, the lack of reliable insight into the timing of a registration and the quality or rigour of its content limit the evaluative potential of pre-registration. As noted in the landscape review, there are therefore good reasons not to regard the presence of a self-reported pre-registration statement on its own as evidence of good research practice.

It is also important to note that the varying use cases for different data platforms could be a source of some confusion. For example, the OSF data was on pre-registrations, while the data

from the other providers was focused on the representation of pre-registration in the selected article sample. Cross-referencing the two was not possible at sufficient scale to be of use in this pilot, but is suggestive of a pathway towards addressing the concerns identified here.

There may be differences between research-intensive vs less research-intensive institutions in frequency/requirement for pre-registration. For example, universities that have large medical schools/biomedical/health sciences departments are likely to be required to pre-register their work which may make benchmarking exercises difficult. Similarly, how could a university benchmark internally? Comparisons between schools or departments may be problematic for all the reasons discussed above.

Related to the issues of applicability and consistency, there is a lack of institutional and national policies on pre-registration. Whether valid or not, concerns would likely be raised around benchmarking without a clear policy framework. This is analogous to the lack of consistently adopted frameworks for pre-registration requirements in journals, but if addressed 'upstream' of publication would enable more consistent approaches to be developed later in the research process and during publication.

As outlined above, current approaches to measuring pre-registration do not capture whether the preregistration actually preceded the research, nor do they evaluate the validity, applicability, or quality of the pre-registration. An open research indicator for pre-registration cannot replace the need for a human to check whether the researchers stuck to their pre-registration plans, or reliably and honestly recorded deviations along the way. With consistent approaches to capturing this data and evidence, with suitable timestamping and openness, there might be opportunities for AI to tackle this problem at scale in the future.

Discussion

This pilot assessed the feasibility of using automated approaches to identifying whether outputs from institutions were pre-registered. All three solutions providers created approaches that identified self-reported pre-registration in institutional outputs. There was only a little variability between the approaches. In general the automated approaches achieved in excess of 80% for accuracy (this of course is influenced by prevalence, which is unusually high in the selected outputs). Sensitivity and specificity for all providers was close to or in excess of 80% also. Less than 10% of the outputs in our combined sample were false negatives. There was a wider range of false positives across providers (3.5%-15%) were false positives.

There are a few key limitations to this work.

1. There remains the possibility that the human agreed rating is incorrect in some way.

2. Pre-registration was operationalised as a self-declaration in this pilot. If an output says it was pre-registered and the pre-registration existed we took this as evidence that the work was pre-registered. We did not check that what was in the pre-registration matched the output.

3. Whether the registration happened before the work described in the output was not assessed in this pilot.

4. Automated processes had better sensitivity than specificity, that is, when the work was pre-registered, the automated approaches were able to detect that was the case (sensitivity 79-93%). When the work was not registered, the automated approaches were less able to reliably detect that (specificity 78-86%).

In addition to these quantitative findings it should be noted that the solutions providers were able to iterate their algorithms on the basis of learning from each other's approaches. The hand coded searches also provided a useful benchmark for the providers to fine tune their algorithms against.

Conclusions

1. Automated identification of self-report of pre-registration is achievable

2. (Much) more work needs to be done to assess the temporal (was it "pre' registered?) and quality components (does the registration match what was delivered?)

3. There is value in encouraging researchers to link to a pre-registration, even if that pre-registration is not associated or high quality as it allows third parties to assess these factors.

Recommendations

1. Pre-registration is discipline and institution specific and comparison between these is unadvisable

2. A key consideration is choosing a registry that indexes lots of outputs.

3. The prevalence of pre-reg is influenced by the denominator. Ideally, studies not eligible for pre-registration (e.g. discussion articles, mathematical theorems, historical reviews) would be excluded from the denominator. The delineation between eligible and not eligible for pre-registration is not entirely clear cut

Data availability

Information relating to this pilot is available on the Open Science Framework https://osf.io/67vrp/

Appendix 5: Pre-registration. Annex 1: Specification for pilot: Pre-registration

DRAFT SPECIFICATION for PILOT 6 - development of an open indicator to summarise the prevalence of pre-registration

Date: 25th June 2024

1. Headline Specification

This pilot will explore the prevalence of preregistration across a broad sample of research articles published by UKRN institutions. The pilot team intends to conduct their own analysis of preregistration prevalence within the publication set. This will provide a comparator dataset to solutions providers, from which we can assess what will constitute an appropriate preregistration indicator.

2. Definition

Following the collation of different definitions of pre-registration we have identified that there are two consistently agreed components to pre-registration.

1. The first is that information about what will be done is recorded before being done. There are minor considerations here reflecting that such documents can be updated and amended over time.

2. The second component is that this information is publicly available.

In this pilot we therefore operationalise pre-registration as evidence of a study plan being publicly published and time-stamped in advance of that study plan being enacted.

We are non-specific about which registries need to be used.

The quality of preregistration has been considered out of scope of the pilot project. It is therefore a requirement to define preregistration in a way that does not consider registration quality.

3. Research Methodology

The pilot will:

Initial Phase
- Conduct a literature review

- Define and build a dataset from participating UKRN institutions.

- Develop a simple methodology for identifying pre-registration based on keyword search to be used as baseline

- Establish a subset of publications for manual checking

    - This will involve manually checking all of the positives (i.e. outputs identified as being pre-registered) and then randomly sampling the same sample size from those categorised as negatives (i.e. outputs not identified as pre-registered). This will allow us to estimate false positive and negative rates.

- Identify baseline % of items at an institutional level that contain a registration

- Apply solutions providers methodologies
  - The potential methodologies include
    - Text searching in the free version of Dimensions
    - Text searching in the paid-for version of Dimensions
    - Bespoke solutions from Dimensions, OpenAIRE, COS, PLoSONE and Open Alex
- Categorise findings against the definition outlined in Section 2 above.
- Data Analysis step, drawing out trends in disciplines, different platforms etc.

Each method will identify a set of outputs deemed positive for pre-registration. We propose manually checking each of these positives to identify false positives (i.e. studies declared as pre-registered that are not). Whichever outputs are not deemed pre-registered will be assumed negative. We will sample from these outputs (the same number of outputs as deemed positive so that the false positive and negative rates will be estimated with similar precision) and manually search these for false negatives.

For example, if a simple text search solution identifies 30 outputs out of 100 are pre-registered for a given institution, we will manually check all 30 outputs for evidence of pre-registration. Then we will randomly select 30 outputs from the 70 deemed not pre-registered to check for evidence of pre-registration.

4. Scope

4.1   Quality of preregistration will not be considered

4.2   This pilot will attempt to discover preregistrations based on a corpus of journal articles. Investigating preregistrations which do not result in a future publication is out of scope

4.3   The results will be broken down by discipline, platform, registration type, institution and funder.

4.4   Registered reports will be included in the analysis

5. Dataset

A dataset of outputs from each participating institution will be collected for the year 2023. A draft reporting template is provided below

Table showing the prevalence of preregistered outputs split by institution and proposed solution. Each cell may also contain information on the false positive and false negative rates from the hand search.

Solution		Institutions			
		UoExeter	UoLeeds	UoManchester	UoReading
Text search in free Dimensions platform	In (%) of pregristered articles				
	False positive rate				

	False negative rate				
	Denominator (N)				
Solution 2					
Solution 3					

6. Deliverables

Project report based on UKRN template

7. Issues

Not all full text documents will be licensed for use by pilot partners. Alternative means of access to author accepted manuscript versions could be explored. Any gaps will be accounted for in analysis.

Each solution will likely work from a different denominator of identified outputs. We need to factor this into our interpretation of the outputs. For example there is potentially a tradeoff to be made between solutions which identify more papers from an institution and the accuracy of the identification of pre-registration.

Appendix 5: Pre-registration. Annex 2: CWTS Review

In this report the authors present the outcomes of a pilot on indicators of preregistration. This pilot is part of a series of pilots carried out in a project coordinated by the UK Reproducibility Network (UKRN).

The report shows that it is technically feasible to produce reasonably accurate indicators of the adoption of preregistration. However, the authors emphasize that such indicators are of somewhat limited value, since they do not indicate whether preregistration took place before the research was carried out and whether the research adhered to the registered plan. Nevertheless, the authors point out that indicators of the adoption of preregistration contribute to increased transparency. In addition, such indicators may enable the future development of more sophisticated indicators of preregistration practices.

I consider this to be a solid and informative report in which the authors offer important and thoughtful reflections on the value of indicators of preregistration. I have provided several minor comments in an annotated version of the report.

I have one more significant comment. In addition to the three recommendations provided by the authors, I would like to add a fourth recommendation: Journals should ask authors whether their study was preregistered, and if this is indeed the case, journals should include a link to the preregistration not only in the full text of an article but also in the metadata of the article. If journals consistently adopt this practice, there is no need for complex text mining analyses to determine whether a study was preregistered or not. This information can then be obtained directly from the metadata of an article, and databases such as OpenAIRE, Dimensions, OpenAlex, and others can ingest this metadata and make it available to their users.

Ludo Waltman
November 23, 2025

Appendix 5: Pre-registration. Annex 3: Team response to CWTS Review

We thank CWTS for providing useful comments on our report. Here is our response.

1. In the first sentence of the introduction section we assert that "Pre-registration is seen as important…" but we agree with that reviewer that pre-registration is valued differently in different disciplines and in different types of research.

2. In our methods we would like to clarify that when we say outputs must be published open access, we did restrict this journal's (thereby potentially excluding green open access, although this is not completely clear in the common dataset).

3. OpenAIRE explored 17,204 unique DOIs

4. In the Digital Science section we did not explore reasons for why only 25,821 publications out of 27,828 were available for full text analysis.

5. In the Digital Science section this phrase "with University of Manchester data added at this point" refers to the fact that there were some manual steps to ensure that information on the correct institutions were included.

6. In the Digital Science section "potential pre-registration related text" refers to a phased approach where outputs without any pre-registration keywords were excluded before a closer check on registration status was made

7. In the Digital Science section, algorithm here refers to the entire workflow to identify publications with pre-registration related text.

8. We note that the numbers of articles identified by the various approaches differs.

9. In the General Method section the number of DOIs in the intersection was 283

10. In Table 1 the following definitions were used.

    Sensitivity = True positive proportion, i.e. the probability of the algorithmic approach flagged a publication as pre-registered given the human raters deemed it was pre-registered

    Specificity = True negative proportion, i.e. the probability of the algorithmic approach flagged a publication as not pre-registered given the human raters deemed it was not pre-registered

    Accuracy = The proportion of publications correctly flagged, i.e. the probability that the algorithm's designation of a publication matches the human rater

11. In our interpretation section we refer to medical schools/biomedical/health sciences departments being likely to be required to pre-register their work. Here we mean a combination of host institutional rules, funder mandates, and publisher requirements.

12. We note the reviewers suggestion for a fourth recommendation of asking journals to force authors to explicitly state whether their study was pre-registered. We acknowledge that this would indeed help to solve the issue, however unless every publisher simultaneously adopted this suggestion and every author similarly complied, we would remain in the situation we currently find ourselves in, i.e. having to produce a system that automatically identifies pre-registration in published outputs where it is not routinely recorded in the meta-data. Moreover, we view our recommendations as being aimed at those who are planning to identify pre-registration in published outputs on how best to go about it. As such, we do not think that the recommendation to journals (while a good idea) is not one that we wish to include here.

13. Our middle recommendation regarding choosing a registry that indexes a lot of outputs refers to the fact that different registries have access to different sets of academic publications. Choosing a registry that indexes more is generally better.