nature portfolio

Corresponding author(s): Russell J Funk

Last updated by author(s): Oct 26, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We used Python v3.10.6 and MariaDB v.10.6.4 to collect and organize raw XML and CSV files obtained from the relevant data publishers.
Data analysis	Open-source code related to this study is available at https://doi.org/10.5281/zenodo.7258379 and http://www.cdindex.info. We used Python v3.10.6 (pandas v1.4.3, numpy v1.23.1, matplotlib v3.5.2, seaborn v0.11.2, spacy v2.2, jupyterlab v3.4.4) to wrangle, analyze, and visualize data and to conduct statistical analyses. We used MariaDB v.10.6.4 to wrangle data. We used R v4.2.1 (ggplot2 v3.36, ggrepel v0.9.0) to visualize data. We used StataMP v17.0 (reghdfe v.5.7.3) to conduct statistical analyses.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data associated with this study are freely available in a public repository at https://doi.org/10.5281/zenodo.7258379. Our study draws on data from six sources: the American Physical Society, JSTOR, Microsoft Academic Graph, Patents View, PubMed, and Web of Science. Data from Microsoft Academic Graph, Patents View, and PubMed are publicly available, and our repository includes complete data for analyses from these sources. Data from the American Physical Society, JSTOR, and Web of Science are not publicly available, and were used under license from their respective publishers. To facilitate replication, our repository includes limited

versions of the data from these sources, which will enable calculation of basic descriptive statistics. The authors will make full versions of these data available upon request and with permission from their respective publishers.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We chiefly rely on archival, observational, and quantitative citation and textual data on scientific and technological works (e.g., papers and patents), which were obtained from the following sources: Web of Science, Patents View, JSTOR, the American Physical Society, Microsoft Academic Graph, and PubMed. All analyses conducted were retrospective in nature. Our aim was to document the changing nature of innovative activity in science and technology as reflected in the data across time and to closely investigate the large-scale temporal patterns that emerged.
Research sample	There were no human subjects in our sample. We analyzed archival, observational, and quantitative data from Web of Science, Patents View, JSTOR, the American Physical Society, Microsoft Academic Graph, and PubMed. Our sample included approximately 45 million scientific papers and 3.9 million utility patents from 1945-2010. Among the papers, we analyzed their 390 million citations, 25 million titles, and 13 million abstracts. Among the patents, we analyzed their 35 million citations, 3.9 million titles, and 3.9 million abstracts. Our primary sample is composed of Web of Science and Patents View data. Web of Science data from the post-World War II era is among the most accurate, detailed, and comprehensive data available for studying the dynamics of science. Similarly, patent data from Patents View represents the population of machine-readable U.S. patent data available to the public.
Sampling strategy	There was no power analysis used to determine sample size. Rather, we analyzed the population of all available records (with complete data) from the following archival sources: Web of Science, Patents View, JSTOR, the American Physical Society, Microsoft Academic Graph, and PubMed. This approach allowed us to analyze all available data to derive the most robust empirical results.
Data collection	We used Python v3.10.6 and MariaDB v.10.6.4 to initially collect and organize the raw XML and CSV data we obtained from the publishers of Web of Science, Patents View, JSTOR, the American Physical Society, Microsoft Academic Graph, and PubMed.
Timing	Data collection lasted about 30 days during the month indicated in parentheses for each of the dataset: Web of Science (December, 2018), Patents View (June, 2018), JSTOR (November, 2016), American Physical Society (August, 2019), Microsoft Academic Graph (February, 2021), and PubMed (July, 2019).
Data exclusions	N/A: No data was excluded. We analyzed all data that was made available to us from the publishers.
Non-participation	N/A: The study only conducted retrospective analysis of bibliometric data and had no participants (and thus no drop outs or declines).
Randomization	N/A: We used an observational study design on archival data and therefore there was no randomization.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging