


BMJ Open Assessing the prevalence, quality and compliance of data-sharing statements in gastroenterology publications: a cross-sectional analysis

Jackson Moore ¹, Khanh Nguyen,¹ Brody Dennis,¹ Mahad Chaudhry,¹ Annes Elfar,¹ Eli Oldham,¹ Eli Paul,¹ Jacob Cohn,² Alicia Ito Ford,^{1,3} Matthew Vassar^{1,3}

To cite: Moore J, Nguyen K, Dennis B, *et al.* Assessing the prevalence, quality and compliance of data-sharing statements in gastroenterology publications: a cross-sectional analysis. *BMJ Open* 2025;**15**:e092490. doi:10.1136/bmjopen-2024-092490

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2024-092490>).

Received 15 August 2024
Accepted 07 March 2025



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

¹Oklahoma State University Center for Health Sciences, Tulsa, Oklahoma, USA

²Department of Internal Medicine, Oklahoma State University Medical Center, Tulsa, Oklahoma, USA

³Department of Psychiatry and Behavioral Sciences, Oklahoma State University Center for Health Sciences, Tulsa, Oklahoma, USA

Correspondence to

Dr Jackson Moore;
jackson.moore.chs@gmail.com

ABSTRACT

Objective To examine the current state of data-sharing practices in gastroenterology literature, focusing on data-sharing statements (DSS) and identifying influential factors on DSS inclusion.

Background High-quality, reproducible research is crucial in addressing the widespread prevalence of gastrointestinal diseases. Data-sharing practices enable researchers to access studies more easily, enhancing reproducibility. Our study aims to analyse the inclusion and influence of DSS in top gastroenterology journals.

Methods We conducted a cross-sectional analysis to examine the use and contents of DSS in gastroenterology clinical trials. Using Clarivate's Journal Citation Reports, we selected five leading gastroenterology journals. Then, we searched MEDLINE (PubMed) for original research articles published between 1 January 2018 and 31 December 2023. In a double-blind, duplicate manner, data were extracted on DSS presence, funding source, study design and open-access status. We then conducted a thematic analysis of all DSS. Additionally, authors were contacted and given 14 days to respond or share data to investigate adherence to their DSS.

Results Of the 953 articles that met inclusion criteria, 400 (400/953; 42.0%) contained a DSS. Open-access articles had a higher likelihood of containing DSS (estimate=0.413; $p<0.05$). *The Lancet Gastroenterology and Hepatology* has the highest percentage of DSS (159/194; 82.0%), while *Clinical Gastroenterology and Hepatology* has the lowest percentage of DSS (33/256; 12.9%). Impact factor is a significant indicator for DSS (estimate=0.138, $p=0.01$). Finally, 'conditional data availability' was the most common data theme in our study (225/303; 74.3%). Over half (153/284; 53.9%) of the authors contacted did not respond to our request for sharing data.

Conclusion Our findings reveal significant variability in DSS inclusion and adherence among top gastroenterology journals. Journals with mandatory data-sharing policies demonstrated higher compliance, while open-access status and journal impact factor were positively associated with data-sharing practices. However, a notable gap remains in authors' follow-through on stated data-sharing commitments.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ This report is strengthened by strict adherence to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.
- ⇒ This analysis is strengthened by double-blind screening and extraction.
- ⇒ The relatively short 14-day response time could be a limitation.

INTRODUCTION

One fundamental aspect of quality research is reproducibility, which allows researchers to access the data necessary to replicate studies more readily.¹ While data sharing is a critical step towards reproducibility, it does not inherently guarantee it. Sharing data provides the foundation for reproducibility by enabling others to validate results, conduct replications and explore secondary analyses. Such practices can greatly improve the clarity and reproducibility of scientific findings.² In an effort to address this issue, the International Committee of Medical Journal Editors (ICMJE) issued a statement promoting the importance of data sharing, mandating investigators to include data-sharing statements (DSS) into their works to enhance transparency across research specialties.³ Adherence to such data-sharing policies allows for reproducibility of research, enabling the verification of results, ensuring proper ethical conduct and upholding policies and practice implications.⁴ Research has shown that reproducibility can greatly strengthen scientific evidence by promoting accountability, reliability and integrity of research findings.⁵ Although data sharing has various proven benefits in medical research, the practice of proper data sharing has been historically underwhelming.⁶ Given the substantial

global burden of gastrointestinal diseases, which account for significant morbidity and healthcare costs worldwide, gastroenterology research plays a critical role in advancing treatment and diagnostic methods. Despite this importance, data-sharing practices in this field have not been well-explored, unlike other medical disciplines.

While data sharing has several positive aspects, existing literature on its importance remains relatively sparse in gastroenterology. This setback is due to several factors, including concerns about data misuse, lack of data sharing incentives and the competitive nature within academic research.⁷ Survey data revealed that while participants recognise the advantages of sharing data, fear of confidentiality breaches and abuse of data persists.⁸ Additionally, research done by Gabelica *et al* demonstrated that regardless of researchers including DSS in their work, over 90% of the authors either did not respond or declined requests for data access when asked.⁹ These data highlight several barriers to data sharing, which hinder reproducibility and damage public perception of research.

Given the frequency of gastrointestinal diseases and the initiative set forth by the ICMJE, there is a need to analyse the effect of data sharing in gastroenterology research.^{3 10} Collaborative efforts between researchers and clinicians are essential to properly ensure reproducibility of treatment outcomes.¹¹ With this in mind, our study aims to analyse the factors concerning the inclusion, content and application of DSS across top gastroenterology journals. We hope to ultimately explore the current state of data sharing in gastroenterology research and derive meaningful recommendations, as well as define key influential factors to DSS incorporation.

METHODS

Study design, reporting and reproducibility

This study used a cross-sectional design to evaluate the prevalence, quality and compliance of DSS in selected gastroenterology journals. The approach involved systematically selecting, screening and extracting data from original research articles published in these journals. Therefore, we deemed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) to be the most suitable reporting guideline over the Strengthening the Reporting of Observational Studies in Epidemiology.^{12 13} To promote open science practices, repeatability and transparency, we uploaded our protocol (and all of its versions), data, analysis scripts and study artefacts to Open Science Framework.¹⁴ Raw data from article screening can be found at <https://osf.io/hm72s>.¹⁴ Raw data from thematic analysis can be found at <https://osf.io/nr8c4>.¹⁴

Patient and public involvement

The persons contacted for the purposes of this project had no input into the design, recruitment or conduct of this study.

Training

Our team underwent thorough training on the study objectives, inclusion/exclusion criteria, data extraction process, general method of the study and Rayyan, an online systematic review platform (<https://www.rayyan.ai/>), before starting the screening and extraction process. After the training, each author in charge of data extraction (JM and KN) took part in a calibration exercise, screening five articles, to ensure reliability and consistency.

Journal selection

We used Clarivate's Journal Citation Reports to identify five of the top journals in the field of gastroenterology based on their impact factors. Journals in this study were required to be indexed in MEDLINE (PubMed). We searched and compiled a list of all original research articles published and/or e-published in these five journals over the 6-year time frame from 2018 to 2023. Additionally, we added the clinical study filter to ensure relevance and of the articles reviewed. In the case that more than one journal under the same subspecialty was included in the top five, we only included the journal with the highest impact factor of that subspecialty to avoid overrepresentation of a subspecialty in order to achieve a diverse balance between comprehensive gastroenterology journals and subspecialty-specific journals. From this list, we assessed all articles from each journal for screening and analysis.

Article selection

Using the International Standard Serial Numbers of the selected journals, we performed a comprehensive literature search in MEDLINE (PubMed) on 5 June 2024. We limited the publication date range to 1 January 2018 through 31 December 2023 to reflect current practices in data sharing during a critical period of transition in open science. Articles e-published before 1 January 2018 or after 31 December 2023 were included if their print publication date fell within our time frame. This time frame aligns with the implementation of the ICMJE's 2018 requirement for DSS in clinical trials, allowing us to assess compliance and trends in response to this policy.

Screening/eligibility criteria

The inclusion criteria of these gastroenterology articles included clinical trials, case series, case-control, cross-sectional analysis, cohort studies and survey studies. Furthermore, these articles must include primary data, either quantitative or qualitative, and must be within the prespecified published time period, from five of the top gastroenterology journals. Articles were excluded if they were review articles, commentaries, editorials, erratum/corrigendum, systematic reviews, case reports of a single person, response to the reader or studies without new primary data. Authors JM and KN performed independent blind screening of each article. One additional author (BD) was available to resolve conflicts.

Primary data extraction

The extraction process was done independently by two authors using a pilot-tested Google Form as the primary method of extraction: bibliographic information (journal's name, publication date, PubMed ID), repository details (database name, URL, data DOI), funding source(s) and its category (eg, government, private, industry, hospital, university, not listed, not funded, self-funded), study design type (see 'Screening/eligibility criteria'), study topic (treatment/intervention or diagnostic/screening), journal data policies, accessibility status of the article, details/presence of DSS (a verbatim extract of the DSS), corresponding author email, format of data (raw, processed or both) and data file type (.sav, .xlsx, or .csv). Results from the extraction were reconciled, and author (BD) was available to resolve conflicts. Our PubMed search was filtered to include articles from 2018 to 2023. However, articles were returned with both e-publication and print dates. As such, we reported the e-publication dates, which included studies from 2016 and 2017.

Primary data analysis

We conducted a descriptive analysis to disclose the frequencies and percentages of our findings across multiple categories: journal, study design, year of publication, study topic, funding source and article accessibility status. Additionally, our analysis explored the proportion of studies with and without DSS from each category. We also examined the trend of DSS presence over time by year for journals and study designs. Study designs with insufficient sample sizes were excluded from this analysis to ensure a comprehensive understanding of DSS trends.

Further, this study employed hierarchical logistic regression to evaluate DSS in gastroenterology research articles, to account for variance at both the journal and publisher levels. Hierarchical logistic regression was chosen to adjust for the nested structure of the data, in which articles are nested within journals, and journals are nested within publishers. This approach helps control the correlations and dependencies within the data, providing more reliable and accurate estimates. Categorical variables were converted to factors, and the journal impact factor was treated as a numeric variable. Sample groups with fewer than 30 observations were excluded to adhere to the central limit theorem and assure reliable statistical approximations.

We chose fixed effects for study design, impact factor, funding source and article access to evaluate their direct influence on the inclusion of DSS. These variables were hypothesised to have a predictable and consistent effect across all articles. Fixed effects allowed us to control for key predictors while evaluating their specific contributions to the outcome. Random effects were included for journals to adjust for the hierarchical structure of the data. This allowed us to model the variance attributed to differences among journals and publishers, which could otherwise confound the fixed effects. By including

random effects, we were able to isolate the influence of the fixed effects and understand the influence of higher-level groupings (journals and publishers) on data-sharing practices. Model diagnostics involved verifying multicollinearity using variance inflation factors to ensure the trustworthiness of the findings. This process confirmed the predictors in our model were not strongly correlated, which could potentially affect the precision of our assessments.

Qualitative analysis of DSS themes

To quantify the variation of DSS provided, we conducted a theme analysis of the extracted data statements. To institute a set of theme categories, we employed ChatGPT-4o (<https://openai.com/chatgpt/>) to analyse all of the verbatim DSS in our pilot study sample. The presence of these themes was assessed across all DSS within our study.

Verifying data accessibility

An option for data sharing is for researchers to provide access to data on reasonable request. We were interested in evaluating the frequency and ease in which authors who issued this phrasing of DSS would actually provide their information. To accomplish this, we emailed the corresponding author of each article and proposed these two questions: (1) *Are you able and willing to share the de-identified data necessary for replication of your primary outcome analysis?* and (2) *If so, approximately when would you be prepared to send it for use in replication?* The researchers were notified this request was a portion of a study assessing the practicality and follow-through of DSS (see online supplemental materials for message template). Researchers were not asked to legitimately gather and send their data. A follow-up request was sent 7 days later if no response was received.

RESULTS

Search results

Initially, our search returned 993 records, of which two duplicates were removed. The remaining 991 articles were screened by title and abstract, with 38 more removed for failing to meet our inclusion criteria. The following 953 articles were assessed for DSS, with 400 being further analysed for data themes. Ultimately, 303 DSS were analysed for specific themes, while 97 were excluded due to ineligibility (figure 1).

General characteristics of the included studies

Our sample consisted of 953 articles from five gastroenterology journals. Non-open-access articles (492/953; 51.6%) and open-access articles (461/953; 48.4%) were nearly equally represented in our study. The number of publications per year remained relatively constant, ranging from 102 (10.7%) in 2023 to 187 (19.6%) in 2020. The most common study designs were clinical trials (734/953; 77.0%) and cohort studies (182/953; 19.1%), with the majority focusing on treatment or intervention

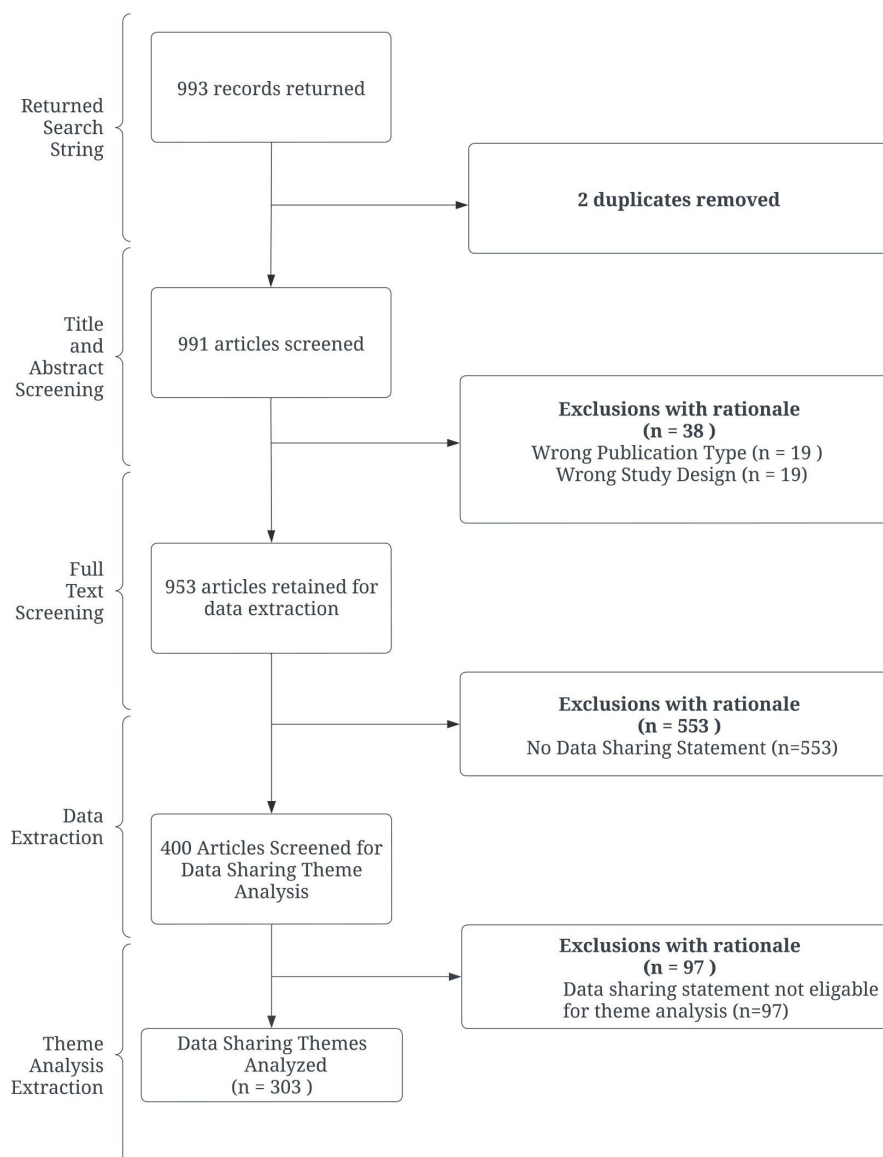


Figure 1 Our exclusion reasonings for the articles that were omitted from our final analysis.

(671/953; 70.4%). Funding sources were fairly evenly distributed between industry (295/953; 31.0%), multiple sources (260/953; 27.3%) and government (220/953; 23.1%). The least common funding sources were hospitals, universities and private entities. Additional information regarding the general characteristics of our included articles can be found in [table 1](#).

Trends in DSS

Significant variability was found among journals, publication years and the likelihood of including a DSS. The top three journals with the highest prevalence of DSS were *The Lancet Gastroenterology and Hepatology* (159/194; 82.0%), *Gut* (90/150; 60.0%) and the *Journal of Hepatology* (71/139; 51.1%). Conversely, *Gastroenterology* (47/214; 22.0%) and *Clinical Gastroenterology and Hepatology* (33/256; 12.9%) were less consistent in their inclusion of DSS. We noted a steady increase in the incorporation of

DSS from 9.2% to 82.8% over the study period in clinical trials from 2018 to 2023 ([figure 2](#)).

Factors influencing the presence of DDS

Our hierarchical logistic regression model detailed the likelihood of DSS across multiple categories. At baseline, our analysis revealed that the articles in our sample were unlikely to include DSS, with an intercept estimate value of -3.57 ($p=0.01$). Among the leading gastroenterological journals, articles within *The Lancet Gastroenterology and Hepatology* had the highest likelihood of including a DSS (mean effect=0.67) compared with the other journals in our sample. Journal impact factor ($p=0.01$) and accessibility ($p<0.05$) were significant positive indicators of DSS presence. Our data showed that neither funding source nor study design were significant indicators of DSS inclusion ($p>0.05$). Additionally, we found no significant variance in DSS incorporation when evaluating publishers ([table 2](#)).

Table 1 General characteristics of the findings within our study

Characteristic	N (%) Total=953	Presence of data-sharing statement
Study design		
Clinical trial	734 (77.0%)	337 (84.3%)
Cohort (prospective or retrospective)	182 (19.1%)	53 (13.3%)
Case-control study	15 (1.6%)	6 (1.5%)
Cross-sectional study	19 (2.0%)	3 (0.8%)
Cost-effective analysis	3 (0.3%)	1 (0.3%)
Year		
2023	102 (10.7%)	83 (20.8%)
2022	127 (13.3%)	94 (23.5%)
2021	151 (15.8%)	83 (20.8%)
2020	187 (19.6%)	77 (19.3%)
2019	171 (17.9%)	44 (11.0%)
2018	215 (22.6%)	19 (4.8%)
Article access		
Not open access	492 (51.6%)	216 (54.0%)
Open access	461 (48.4%)	184 (46.0%)
Study topic		
Treatment/interventional	671 (70.4%)	322 (80.5%)
Screening/diagnostic	282 (29.6%)	78 (19.5%)
Funding source		
Industry	295 (31.0%)	146 (36.5%)
Multiple	260 (27.3%)	103 (25.8%)
Government	220 (23.1%)	91 (22.8%)
Not funded	33 (3.5%)	20 (5.0%)
Private	63 (6.6%)	19 (4.8%)
Not listed	65 (6.8%)	16 (4.0%)
University	13 (1.4%)	4 (1.0%)
Hospital	4 (0.4%)	1 (0.3%)

Data-sharing themes

After completing our thematic analysis of the (303/400) articles containing an eligible DSS, we identified several patterns among the DSS. *Conditional data availability* was the most frequent theme, occurring in 225 entries (225/303, 74.3%). The *gatekeeper role* (134/303, 44.2%) and *privacy concerns* (133/303, 43.9%) were also common, highlighting the emphasis on protecting sensitive data. *Corporate ownership*, *structured access platform* and *access mechanism*, and *conditional timing* appeared in about 20–30% of the entries, indicating significant control over information by corporations or authors. Full dedication to open data was less common in gastroenterology articles, with *open data sharing and public repository use* (39/303, 12.9%) and *immediate data accessibility* (25/303, 8.3%) being less frequent. The remaining themes were rarely mentioned.

Additional information can be found in online supplemental table 3.

Willingness to share data on request

We contacted 284 unique corresponding authors of the 400 articles to assess their willingness to share their data. After the 14-day response period, the majority did not reply (153/284; 53.9%). A number of authors refused to share their data (30/81; 37.0%), while some were willing to share data without conditions (20/81; 24.7%) and others were willing to share data with conditions (31/81; 38.3%). The rest (50/284; 17.6%) were unable to be contacted due to various reasons such as being out of office (19/284; 6.7%) or undeliverable emails (21/284; 7.4%).

DISCUSSION

Data-sharing practices in gastroenterology journals have shown gradual progress yet highlight the need for additional improvements. While there has been a 10-fold increase in DSS inclusion in the last six years across platforms, some journals are still below 50% frequency of DSS inclusion. While *Lancet Gastroenterology and Hepatology* managed to achieve a 100% DSS adherence rate, other journals such as *Clinical Gastroenterology and Hepatology* remained consistently low. Additionally, we found that both the quality and execution of DSS are lacking. An example of this is DSS stating 'data will not be made available' or similar phrasing. While this is considered a DSS by many publishers, it does not lend itself to data sharing or transparency. Our analysis also demonstrated open-access articles were more likely to include DSS than non-open-access articles, potentially revealing broader themes within gastroenterology research. The following paragraphs will expand on these findings.

The stark differences in data-sharing policies across gastroenterology journals emphasise the need for standardisation in order to bolster reproducible research. Our results show a sharp increase in *Lancet Gastroenterology and Hepatology* DSS inclusion beginning in 2018, reaching a perfect 100% beginning in 2021 (figure 3). This is likely due to the *Lancet Gastroenterology and Hepatology* mandating DSS inclusion in clinical trials in 2019 and all clinical research in 2020.¹⁵ On the contrary, *Gastroenterology* and *Clinical Gastroenterology and Hepatology* reportedly follow ICMJE guidelines, which state that any article containing results from clinical trials must contain DSS.³¹⁶¹⁷ Our analysis contradicts this statement, revealing that, although most of the studies in our sample from *Gastroenterology* are clinical trials, fewer than one quarter provided DSS. Additionally, *Clinical Gastroenterology and Hepatology* never achieved DSS adherence rates greater than 31%, further emphasising the gap in adherence between *Lancet Gastroenterology and Hepatology* and these journals. Given that all of these journals are published by Elsevier, it is apparent that strict enforcement of DSS on the journal level results in greater adherence to DSS inclusion.

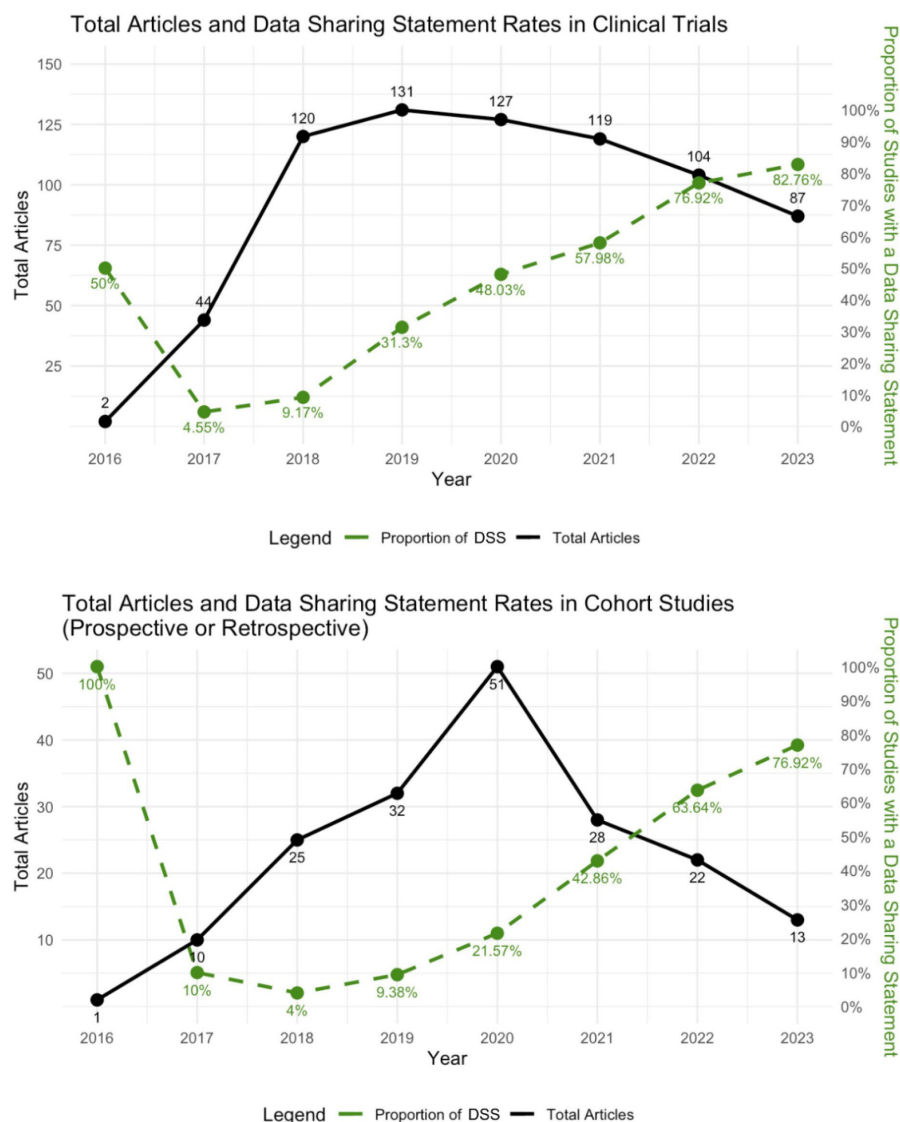


Figure 2 Data-sharing statement (DSS) trends for clinical trials and cohort studies.

During our thematic analysis of all DSS in our sample, we observed various recurring themes. Many of the DSS in our study fell under the category of *gatekeeper role*, requiring investigators to contact a specific person for access to the data. Relying on people to provide access to data represents a barrier in data availability for researchers as it may lead to selective sharing bias and inconsistent standardisation of data-sharing quality.^{18 19} The idea of gatekeeping illustrates the competitive aspect of medical research as researchers may be less likely to publish their findings onto open-access platforms due to fear of losing competitive advantage, lack of incentive and the perception of lower quality data.^{20–22} Moreover, less than 15% of DSS in our sample stated that the data were deposited in an open-access repository. The lack of repository use raises many concerns, as data in repositories are fully available, immediately available and permanently available for secondary analyses.²³ Approximately 20% of DSS in our sample stated a conditional timing for access to data. For example, some authors state that data

will be available after 1–3 years, while some state data will only be accessible for a certain time. Therefore, the use of open data repositories can combat this challenge in order to improve reproducibility and transparency of gastroenterology research.

Our analysis revealed a significant disparity in the inclusion of DSS between open-access and non-open-access journals. Open-access articles were more likely to include DSS than their non-open-access counterparts, which aligns with the broader goals of open access to facilitate greater public engagement and enhance collaboration across different medical fields. The benefits of open access are well-documented, including increased visibility, citation rates and the potential for multidisciplinary collaboration.²⁴ Furthermore, Yoong *et al* discuss the various benefits to open data practices, such as repurposing data to answer secondary questions, examine research gaps not initially explored and allow for rapid reviews that may influence public health decision-making.²⁵ However, the financial burden associated with open-access publication

Table 2 Hierarchical logistic regression analysis with random effects summary

Fixed variable	Estimate	SE	Z value	P value
Intercept	-3.570	1.394	-2.561	0.010*
Study design (compared with clinical trials)				
Cohort (prospective or retrospective)	-0.097	0.234	-0.416	0.677
Impact factor	0.138	0.054	2.585	0.010*
Funding source				
Government	-0.250	0.444	-0.564	0.573
Government, industry	-0.258	0.539	-0.478	0.633
Government, private	-0.255	0.502	-0.509	0.611
Industry	0.182	0.445	0.408	0.683
Not listed	-0.209	0.533	-0.391	0.696
Private	-0.454	0.527	-0.861	0.389
Article access				
Open access	0.413	0.186	2.218	0.027*
Journal sd__(Intercept)	0.751			
Random-effects entity	Mean effect			
Journal (publisher)				
<i>Lancet Gastroenterology and Hepatology</i> (Elsevier)				0.672
<i>Gut</i> (BMJ Publishing Group)				0.646
<i>Clinical Gastroenterology and Hepatology</i> (Elsevier)				0.129
<i>Journal of Hepatology</i> (Elsevier)				-0.130
<i>Gastroenterology</i> (Elsevier)				-1.310

Reference categories: (1) study design: clinical trial; (2) funding source: not funded.
 Interpretation of estimates: (1) positive estimates indicate an increased likelihood of having a data-sharing statement compared with the reference category; (2) negative estimates indicate a decreased likelihood of having a data-sharing statement compared with the reference category.
 Significance levels: *p<0.05.
 Random effects: intercepts for journals and publishers were included as random effects to account for variability across these entities.
 Model details: hierarchical logistic regression model with random effects was used, controlling for clustering at the journal and publisher levels.

fees remains a substantial barrier. For instance, the cost to publish an open-access article in the *Journal of Hepatology* is \$4950, which may be prohibitive for many researchers, particularly those from underfunded institutions or low-income countries. Despite these financial challenges, the correlation between open access and increased data sharing highlights the need for continued efforts to support open-access publishing. Addressing the economic barriers associated with open access will

likely increase the inclusion of DSS, thereby enhancing data transparency and reproducibility. Further research should explore the mechanisms to subsidise open-access fees and investigate the long-term outcome of open access on data-sharing practices.

While many DSS in our sample stated that data would be available, in practice, this was often not the case. After contacting 284 corresponding authors, we encountered 28 who refused to share their data, most often citing proprietary or intellectual constraints. Furthermore, 153 authors did not respond at all. These results highlight significant issues regarding the validity of DSS within gastroenterology research. Similar findings in other literature reveal that researchers frequently do not follow through with their commitments to data sharing.⁹ This trend of failing to honour DSS raises several questions and underscores the need for stricter adherence to data-sharing commitments.

Many considerations can be taken to further increase the inclusion of DSS. One effective measure is the mandatory inclusion of DSS, as seen with the ICMJE.³ This approach had the most significant implication, as evidenced by *The Lancet Gastroenterology and Hepatology* 100% compliance rate with DSS inclusion post-2021.²⁶ However, the presence of a statement alone does not guarantee data availability. Requiring DSS along with the submission of raw data at the time of article submission could address accessibility issues.^{27 28} Additionally, some journals have partnered with data repositories like Figshare, which provide platforms for storing and sharing research data. These partnerships significantly enhance data accessibility. By leveraging such platforms, journals can facilitate compliance with data-sharing requirements and improve the overall quality and effectiveness of scientific research. Other recommendations include incentivising data sharing through grants or recognition programmes, developing standardised templates and guidelines for DSS, and providing training for researchers on the importance and methods of data sharing. Journals could also implement stricter enforcement policies, such as requiring proof of data deposit in a repository before publication. Establishing international collaborations to harmonise data-sharing policies and practices could further streamline efforts and ensure consistency across different journals and disciplines.

While unrestricted data sharing through public repositories is often considered the optimal outcome for promoting transparency and reproducibility, it is not always feasible or appropriate. For data sets containing sensitive information, such as patient data, proprietary information or ethically restricted content, controlled access mechanisms are necessary to ensure compliance with legal and ethical standards. The current infrastructure for data sharing often lacks the capacity to support controlled access effectively, requiring researchers to navigate complex contractual agreements such as Data Use Agreements. In such cases, requesting readers to contact the corresponding author for data access may

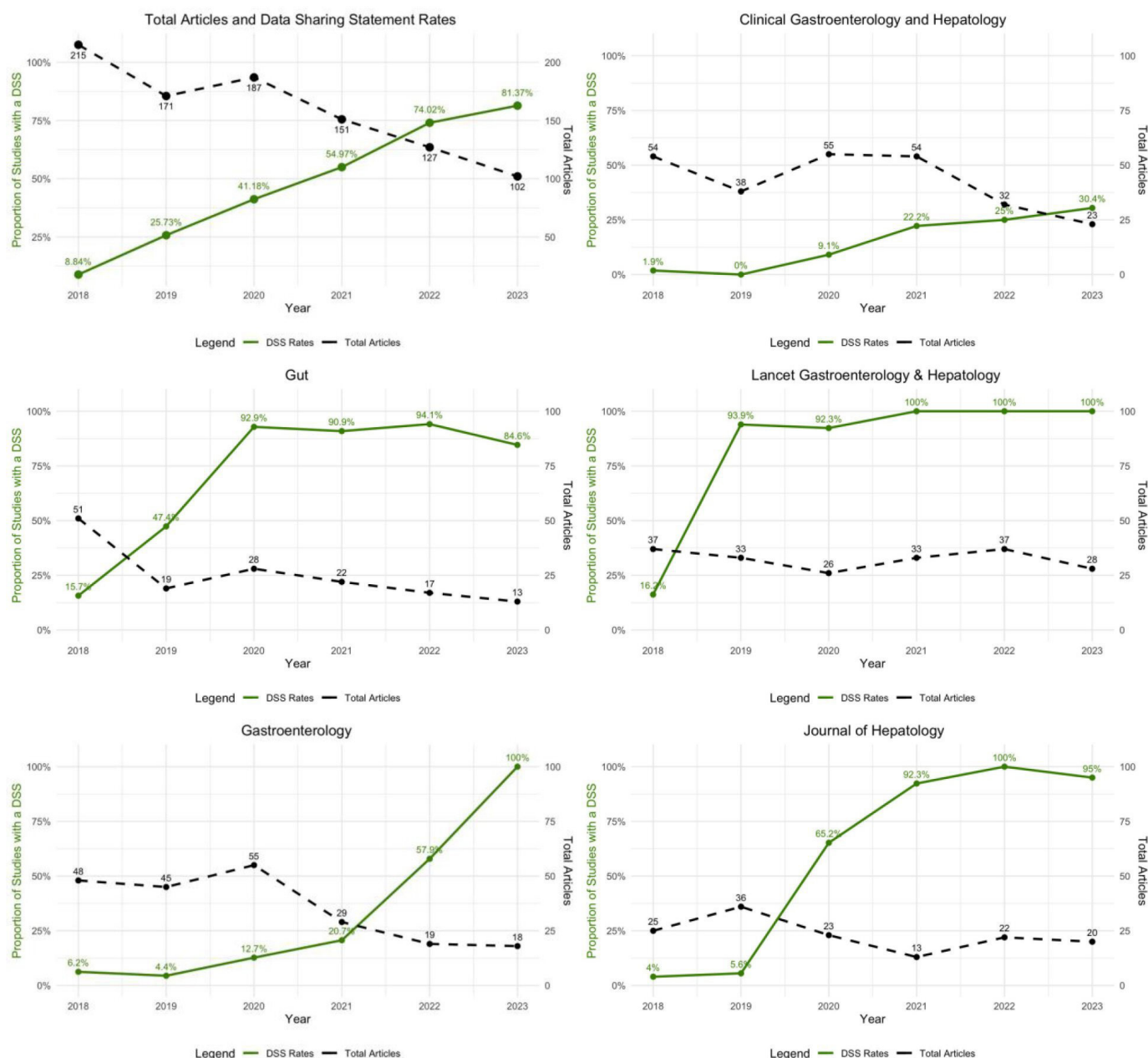


Figure 3 Data-sharing statement (DSS) trends by journals throughout the years.

serve as a practical solution to facilitate data sharing rather than acting as a form of gatekeeping. This underscores the need for improved infrastructure and standardised frameworks to support controlled yet accessible data-sharing practices.

Our study demonstrated multiple strengths. We strictly adhered to the PRISMA guidelines to facilitate transparency and reproducibility of our findings. Moreover, we employed a double-blind data extraction method to ensure the reliability of our data. We also examined data from five of the highest impact factor journals in gastroenterology. However, our study was not without limitations. Our data collection was limited to articles published between 2018 and 2023. Additionally, contacted authors were allotted 14 days to respond to our request for data

availability, and some were out of office, limiting our total responses. We acknowledge that the ICMJE requirement for data availability statements specifically targets clinical trials, and the varying proportion of clinical trials published across the included journals may have influenced our findings. Additionally, while the NIH's Data Management and Sharing Policy was not yet in effect for most of the included articles, its gradual development reflects an increasing expectation for data sharing, which may have indirectly shaped the practices observed in our study. These limitations could have restricted the potential data for our study, and we acknowledge these could be addressed in a follow-up study.

Our results provide up-to-date insight into the current environment of DSS among five leading gastroenterology

journals. While progress has been made following the implementation of data-sharing policies by institutions like the ICMJE, further advancements are needed to emulate the standards set by top journals such as *The Lancet Gastroenterology and Hepatology*. Standardisation and enforcement of data-sharing policies among journals and regulatory agencies have the potential to minimise costs and significantly increase the availability of study data. This, in turn, will enhance the accuracy and reproducibility of research across the field of gastroenterology.^{4 29}

Contributors All authors provided equal contributions to the intellectual development of this protocol and assisted in editing prior to submission. BD and EP performed the drafting and preliminary research needed to formulate this protocol. JM and KN performed data extraction, analysis and drafting of the manuscript. MV provided oversight and critical appraisal of this protocol prior to the onset of the cross-sectional analysis it encompasses. MV acts as the guarantor. To quantify the variation of DSS provided, we conducted a theme analysis of the extracted data statements. To institute a set of theme categories, we employed ChatGPT-4o (<https://openai.com/chatgpt/>) to analyse all of the verbatim DSS in our pilot study sample. The presence of these themes was assessed across all DSS within our study.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests No financial or other sources of support were provided during the development of this manuscript. MV reports receipt of funding from the National Institute on Drug Abuse, the National Institute on Alcohol Abuse and Alcoholism, the US Office of Research Integrity, Oklahoma Center for Advancement of Science and Technology and internal grants from Oklahoma State University Center for Health Sciences—all outside of the present work. AIF reports receipt of funding from the Center for Integrative Research on Childhood Adversity, the Oklahoma Shared Clinical and Translational Resources and internal grants from Oklahoma State University and Oklahoma State University Center for Health Sciences—all outside of the present work. All other authors have no competing interests to declare.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval The Oklahoma State University Center for Health Sciences Institutional Review Board approved the study (study number: 2024096).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. All data are available on Open Science Framework (OSF). Raw data from article screening can be found at <https://osf.io/hm72s>. Raw data from thematic analysis can be found at <https://osf.io/nr8c4>.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Jackson Moore <http://orcid.org/0009-0004-1091-4655>

REFERENCES

- 1 Resnik DB, Shamoo AE. Reproducibility and Research Integrity. *Account Res* 2017;24:116–23.
- 2 Tedersoo L, Küngas R, Oras E, *et al*. Data sharing practices and data availability upon request differ across scientific disciplines. *Sci Data* 2021;8:192:192.
- 3 Taichman DB, Sahni P, Pinborg A, *et al*. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. *Lancet* 2017;389:e12–4.
- 4 Alter G, Gonzalez R. Responsible practices for data sharing. *Am Psychol* 2018;73:146–56.
- 5 Peng RD, Dominici F, Zeger SL. Reproducible epidemiologic research. *Am J Epidemiol* 2006;163:783–9.
- 6 Gomes DGE, Pottier P, Crystal-Ornelas R, *et al*. Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proc R Soc B* 2022;289.
- 7 Hughes LD, Tsung G, DiGiovanna J, *et al*. Addressing barriers in FAIR data practices for biomedical data. *Sci Data* 2023;10:98.
- 8 Kalkman S, van Delden J, Banerjee A, *et al*. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. *J Med Ethics* 2022;48:3–13.
- 9 Gabelica M, Bojčić R, Puljak L. Many researchers were not compliant with their published data sharing statement: a mixed-methods study. *J Clin Epidemiol* 2022;150:33–41.
- 10 Wang R, Li Z, Liu S, *et al*. Global, regional, and national burden of 10 digestive diseases in 204 countries and territories from 1990 to 2019. *Front Public Health* 2023;11:1061453.
- 11 Blevins D, Farmer MS, Edlund C, *et al*. Collaborative research between clinicians and researchers: a multiple case study of implementation. *Implement Sci* 2010;5:76.
- 12 STROBE. Strengthening the reporting of observational studies in epidemiology. Available: <https://www.strobe-statement.org/>
- 13 PRISMA statement. Available: <https://www.prisma-statement.org/>
- 14 Moore J, Dennis B, Nguyen K. Gastroenterology Data Sharing, 2024. Available: https://osf.io/s8hua/?view_only=cad005190b0f4d40b3c312c8790029aa
- 15 The Lancet. Editorial Manager. Available: <https://www.editorialmanager.com/thelancet/default.aspx>
- 16 Elsevier. Clinical Gastroenterology and Hepatology. Available: <https://www.sciencedirect.com/journal/clinical-gastroenterology-and-hepatology/publish/guide-for-authors>
- 17 Elsevier. Gastroenterology. Available: <https://www.sciencedirect.com/journal/gastroenterology/publish/guide-for-authors>
- 18 Stanford Institute for Economic Policy Research (SIEPR). Race to the bottom: How competition to publish first can hurt scientific quality. Available: <https://siepr.stanford.edu/publications/policy-brief/race-bottom-how-competition-publish-first-can-hurt-scientific-quality>
- 19 van den Besselaar P, Hemlin S, van der Weijden I. Collaboration and Competition in Research. *High Educ Policy* 2012;25:263–6.
- 20 Borgman CL. The conundrum of sharing research data. *J Am Soc Inf Sci Tec* 2012;63:1059–78.
- 21 Anger M, Wendelborn C, Winkler EC, *et al*. Neither carrots nor sticks? Challenges surrounding data sharing from the perspective of research funding agencies—A qualitative expert interview study. *PLoS One* 2022;17:e0273259.
- 22 Huang C-KK, Neylon C, Hosking R, *et al*. Evaluating the impact of open access policies on research institutions. *Elife* 2020;9:e57067.
- 23 Pinfield S, Salter J, Bath PA, *et al*. Open-access repositories worldwide, 2005–2012: Past growth, current characteristics, and future possibilities. *Asso for Info Science & Tech* 2014;65:2404–21.
- 24 Springer Nature. Benefits of open research. Available: <https://www.springernature.com/gp/open-research/about/benefits>
- 25 Yoong SL, Turon H, Grady A, *et al*. The benefits of data sharing and ensuring open sources of systematic review data. *J Public Health (Oxf)* 2022;44:e582–7.
- 26 Google Docs. Figure 2. Available: https://docs.google.com/document/d/1hRcRGqQxUBZFcBGPi0kP9N7_PKWqLjvP-KSa7SS04/edit
- 27 Hrynaskiewicz I, Norton ML, Vickers AJ, *et al*. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. *BMJ* 2010;340:c181.
- 28 Federer LM, Belter CW, Joubert DJ, *et al*. Data sharing in PLOS ONE: An analysis of Data Availability Statements. *PLoS One* 2018;13:e0194768.
- 29 Postlethwaite CF, Liss PS. Data sharing: standards are on the rise. *Nature New Biol* 2022;610:257.