International Journal of Advance and Applied Research
www.ijaar.co.in
ISSN – 2347-7075 Impact Factor – 8.141
Peer Reviewed Bi-Monthly
Vol. 7 No. 2 January – February - 2026

Indian Contributions to COVID-19 Research on bioRxiv: A Bibliometric Review (2020–2022)

Mr. Vishnu S. Waghere
Research Scholar, Department of Library & Information Science,
Savitribai Phule Pune University, Pune
Corresponding Author – Mr. Vishnu S. Waghere
DOI - 10.5281/zenodo.18710017

Abstract:

Preprint servers accelerated the dissemination of COVID-19 research, especially during 2020. This paper reviews evidence about India-linked COVID-19 preprints on the bioRxiv preprint server, focusing on volume, temporal dynamics, topical distribution, and institutional concentration. Quantitative figures are drawn primarily from a bibliometric study of Indian COVID-19 preprints deposited in bioRxiv and medRxiv using the NIH iSearch COVID-19 Portfolio (data through November 29, 2022). To contextualize these findings, we also report current (January 2026) counts from the bioRxiv/medRxiv COVID-19 collection page. Results show that India-linked COVID-19 bioRxiv preprints peaked in 2020 and declined thereafter, with a small share of total COVID-19 preprints on bioRxiv at the 2022 snapshot. Topic signals indicate prominence of infectious disease and epidemiology-related work, and institution-level outputs are concentrated among a small set of major national research organizations. The paper discusses implications for rapid, open biomedical communication and the need for regularly updated, affiliation-resolved datasets.

Keywords: COVID-19; SARS-CoV-2; India; bioRxiv; preprints; bibliometrics

Introduction:

The COVID-19 pandemic produced an unprecedented demand for timely biomedical evidence. Preprints—manuscripts posted publicly before formal journal peer review—became a central channel for rapid disclosure, enabling near-real-time access for researchers, clinicians, and policymakers. Evidence from the first months of the pandemic shows preprints influencing global discourse on key parameters such as transmissibility (Majumder & Mandl, 2020). A large-scale analysis of bioRxiv and medRxiv further documented that COVID-19 preprints were accessed, shared, and cited more than non-COVID-19 preprints, and that the pandemic shifted scientific communication norms (Fraser et al., 2021). Within this global surge, understanding the contributions of specific countries is important for assessing research capacity, collaboration patterns, and the diffusion of locally relevant evidence. This paper focuses on India's contribution to COVID-19 research on bioRxiv, the major life-sciences preprint server.

Objectives:

1. Summarize published bibliometric evidence about India-linked COVID-19 preprints on bioRxiv.
2. Quantify annual posting patterns (2020–2022) and India's share of COVID-19 bioRxiv preprints at a late-2022 snapshot.
3. Describe topical and institutional patterns reported for India-linked COVID-19 preprints (bioRxiv and medRxiv).

4. Provide platform-level context using current (January 2026) COVID-19 collection counts.

Methods:

Design: Secondary bibliometric synthesis with simple descriptive statistics.

Primary dataset: Counts for India-linked COVID-19 preprints were taken from Raju, Adakawa, Harinarayana, and Chandrappa (2023), who analyzed Indian COVID-19 preprints deposited in bioRxiv and medRxiv using the NIH iSearch COVID-19 Portfolio (coverage through November 29, 2022). For bioRxiv specifically, the study reports annual posting counts for 2020–2022 and a total number of India-linked COVID-19 bioRxiv preprints at the time of analysis.

Platform context: To provide an up-to-date sense of scale, we extracted the current size of the bioRxiv/medRxiv COVID-19 SARS-CoV-2 collection as displayed on the bioRxiv collection page in early January 2026. This collection also links to a public API for metadata access, but full affiliation-resolved re-analysis is outside the scope of this report.

Analysis. We computed percentage changes across years (2020→2021, 2021→2022) and India's share of total COVID-19 bioRxiv preprints at the study's 2022 snapshot. We visualized annual counts, share, and selected topic and institution distributions (as reported in the source study).

Results:

1. India-linked COVID-19 bioRxiv preprints over time (2020–2022):

The source bibliometric study reports that India-linked COVID-19 bioRxiv preprints were most numerous in 2020 and declined in subsequent years. Using the reported annual counts (118 in 2020; 92 in 2021; 35 in 2022), posting volume decreased by approximately 22.0% from 2020 to 2021 and by 62.0% from 2021 to 2022.

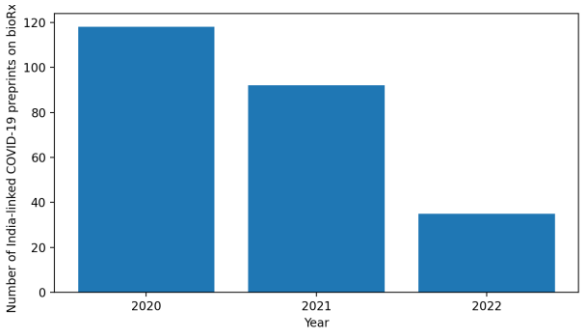

Figure 1. Reported India-linked COVID-19 preprints posted to bioRxiv by year (2020–2022).

2. Share of COVID-19 bioRxiv preprints at the late-2022 snapshot:

At the study's snapshot (November 29, 2022), the NIH iSearch portfolio contained 6,555 COVID-19 bioRxiv preprints. The study reports 240 India-linked COVID-19 bioRxiv preprints at that time, implying an approximate share of 3.66% of the COVID-19 bioRxiv corpus.

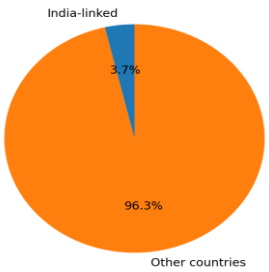

Figure 2. Estimated share of India-linked COVID-19 preprints on bioRxiv at a November 29, 2022 snapshot.

3. Topical signals in India-linked COVID-19 preprints (bioRxiv + medRxiv):

Because many COVID-19 manuscripts with clinical or public-health focus were preferentially deposited on medRxiv, the source study also summarizes topic categories across India-linked COVID-19 preprints on both servers. Infectious diseases (except HIV/AIDS) and epidemiology were the most frequent categories, followed by public and global health and bioinformatics among the top categories reported.

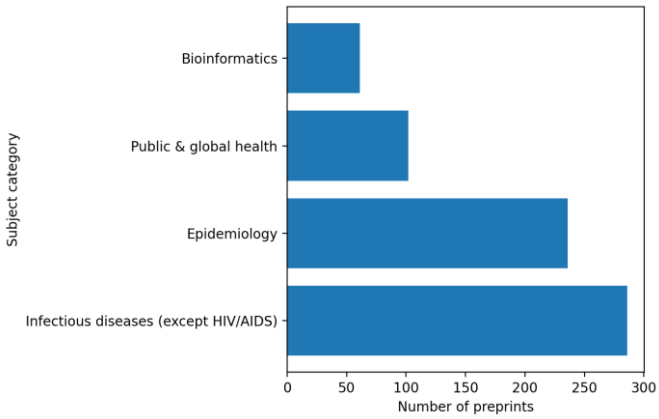

Figure 3. Selected top topic categories for India-linked COVID-19 preprints (bioRxiv + medRxiv) as reported by the source study.

4. Leading India-based institutions:

Institution-level outputs reported in the study show concentration among a small set of major biomedical and science institutions. Among the institutions explicitly highlighted, AIIMS (New Delhi) had the largest number of COVID-19 preprints, followed by the Indian Institute of Science (IISc, Bengaluru) and the ICMR-National Institute of Virology (NIV, Pune).

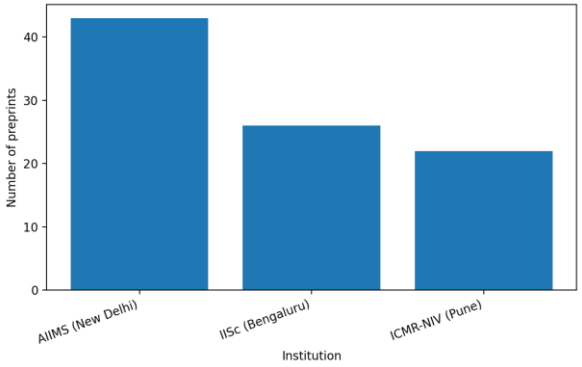

Figure 4. Selected leading India-based institutions by COVID-19 preprint count (bioRxiv + medRxiv) as reported by the source study.

5. Platform context: size of the COVID-19 collection through January 2026:

The bioRxiv/medRxiv COVID-19 SARS-CoV-2 collection continues to grow beyond the study's 2022 cutoff. In early January 2026, the collection page lists 31,198 total COVID-19 preprints across both servers, including 8,485 on bioRxiv and 22,713 on medRxiv. This indicates that the late-2022 snapshot used for India-specific

counts represents an earlier phase of a still-expanding corpus.

Discussion:

India-linked COVID-19 preprints on bioRxiv show a clear early-pandemic peak and a decline by 2022. This pattern matches broader observations that COVID-19 preprint posting surged in the early months and then decreased as the research field matured and journal publication workflows adapted (Fraser et al., 2021). The observed topic mix—dominated by infectious disease and epidemiology—aligns with pandemic-driven priorities such as viral biology, transmission dynamics, and surveillance. The institution concentration suggests that a small set of national research hubs contributed disproportionately to early preprint output, which is consistent with how large-scale laboratory, clinical, and public-health infrastructures shape rapid research production.

A notable interpretation issue is that country attribution depends on how "India-linked" is defined (e.g., corresponding author affiliation, any co-author affiliation, or institution location). Harmonized, open metadata standards are therefore important for robust, repeatable country-level analyses. The public collection API offered by bioRxiv provides a pathway for updated country attribution, but it requires careful affiliation parsing and disambiguation.

Limitations:

- India-specific bioRxiv counts are taken from a published bibliometric study whose data end on November 29, 2022.
- Topic and institution breakdowns are reported for combined bioRxiv+medRxiv Indian preprints, not bioRxiv alone.
- The source study contains small internal inconsistencies between annual counts and the reported total bioRxiv count; we used the published figures as stated and treated the reported total as authoritative for the share estimate.
- We did not re-run a full metadata scrape via the collection API; therefore, post-2022 India-linked bioRxiv trends are not estimated here.

Conclusion:

Published evidence indicates that India contributed a modest but meaningful share of COVID-19 preprints to bioRxiv during 2020–2022, with activity peaking early in the pandemic. Topical signals emphasize infectious disease and epidemiology, and outputs are concentrated in major national research institutions. Because the overall bioRxiv COVID-19 corpus continues to expand through 2026, updated affiliation-resolved analyses using open metadata would be valuable for measuring India's evolving contribution and collaboration patterns.

References (APA):

1. BioRxiv. (2026). COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv (collection page). Retrieved January 4, 2026, from https://connect.biorxiv.org/relate/content/181
2. Fraser, N., Brierley, L., Dey, G., Polka, J. K., Palfy, M., Nanni, F., et al. (2021). The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. PLOS Biology, 19(4), e3000959. https://doi.org/10.1371/journal.pbio.3000959
3. Majumder, M. S., & Mandl, K. D. (2020). Early in the epidemic: Impact of preprints on global discourse about COVID-19 transmissibility. The Lancet Global Health, 8(5), e627-e630. https://doi.org/10.1016/S2214-109X(20)30113-3

4. Raju, N. V., Adakawa, M. I., Harinarayana, N. S., & Chandrappa, H. (2023). Indian COVID-19 preprints submissions in bioRxiv and medRxiv preprint servers. (Preprint). Cold Spring Harbor Laboratory. https://doi.org/10.1101/2023.02.14.23285870
5. Romolak, M., & Homolak, J. (2022). Publishing of COVID-19 preprints in peer-reviewed journals, preprinting trends, public discussion and quality issues. Scientometrics, 127, 1339-1365. https://doi.org/10.1007/s11192-021-04249-7
6. Vlasschaert, C., Giles, C., Hiremath, S., & Lanktree, M. B. (2020). Preprint server use in kidney disease research: A rapid review. American Journal of Kidney Diseases. https://doi.org/10.1101/2020.03.16.20036590