

Seidlmayer, Eva et al.

**Article — Published Version**

## Research topic displacement and the lack of interdisciplinarity: lessons from the scientific response to COVID-19

Scientometrics

*Suggested Citation:* Seidlmayer, Eva et al. (2024) : Research topic displacement and the lack of interdisciplinarity: lessons from the scientific response to COVID-19, Scientometrics, ISSN 1588-2861, Springer Nature, Berlin, Vol. 129, pp. 5141-5179, <https://doi.org/10.1007/s11192-024-05132-x>

This Version is available at:  
<http://hdl.handle.net/11108/641>

**Kontakt/Contact**

ZBW – Leibniz-Informationszentrum Wirtschaft/Leibniz Information Centre for Economics  
Düsternbrooker Weg 120  
24105 Kiel (Germany)  
E-Mail: [info@zbw.eu](mailto:info@zbw.eu)  
<https://www.zbw.eu/de/ueber-uns/profil-der-zbw/veroeffentlichungen-zbw>

**Standard-Nutzungsbedingungen:**

Dieses Dokument darf zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden. Sie dürfen dieses Dokument nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Sofern für das Dokument eine Open-Content-Lizenz verwendet wurde, so gelten abweichend von diesen Nutzungsbedingungen die in der Lizenz gewährten Nutzungsrechte.



<http://creativecommons.org/licenses/by/4.0/>

**Terms of use:**

*This document may be saved and copied for your personal and scholarly purposes. You are not to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. If the document is made available under a Creative Commons Licence you may exercise further usage rights as specified in the licence.*



# Research topic displacement and the lack of interdisciplinarity: lessons from the scientific response to COVID-19

Eva Seidlmayer<sup>1</sup> · Tetyana Melnychuk<sup>3</sup> · Lukas Galke<sup>4,5,6</sup> · Lisa Kühnel<sup>1</sup> · Klaus Tochtermann<sup>5</sup> · Carsten Schultz<sup>3</sup> · Konrad U. Förstner<sup>1,2</sup>

Received: 17 July 2023 / Accepted: 31 July 2024 / Published online: 22 August 2024  
© The Author(s) 2024

## Abstract

Based on a large-scale computational analysis of scholarly articles, this study investigates the dynamics of interdisciplinary research in the first year of the COVID-19 pandemic. Thereby, the study also analyses the reorientation effects away from other topics that receive less attention due to the high focus on the COVID-19 pandemic. The study aims to examine what can be learned from the (failing) interdisciplinarity of coronavirus research and its displacing effects for managing potential similar crises at the scientific level. To explore our research questions, we run several analyses by using the COVID-19++ dataset, which contains scholarly publications, preprints from the field of life sciences, and their referenced literature including publications from a broad scientific spectrum. Our results show the high impact and topic-wise adoption of research related to the COVID-19 crisis. Based on the similarity analysis of scientific topics, which is grounded on the concept embedding learning in the graph-structured bibliographic data, we measured the degree of interdisciplinarity of COVID-19 research in 2020. Our findings reveal a low degree of research interdisciplinarity. The publications' reference analysis indicates the major role of clinical medicine, but also the growing importance of psychiatry and social sciences in COVID-19 research. A social network analysis shows that the authors' high degree of centrality significantly increases her or his degree of interdisciplinarity.

**Keywords** COVID-19 · Bibliometrics · Interdisciplinarity · Research dynamics · Network analysis · Machine learning

## Introduction

During the COVID-19 pandemic, the world has faced three main crises: the medical crisis, the social crisis, and the economic crisis (Nicola et al., 2020; "syndemic" Ryan, 2021). Dealing with these crises has required a bundling of the strengths of scientists from several very different scientific disciplines (Moradian et al., 2021). Following Rafols and

---

Eva Seidlmayer, Tetyana Melnychuk and Lukas Galke have contributed equally to this work.

Extended author information available on the last page of the article

Meyer (2010), we define interdisciplinary research as research targeting scientific problems by integrating concepts, theories, techniques, instruments, and data from different scientific disciplines. Interdisciplinary collaborations drive breakthrough ideas and innovations (Jung et al., 2021; Schilling & Green, 2011) that were urgently needed in the COVID-19 pandemic (e.g., drugs and vaccines). Despite numerous benefits, interdisciplinary collaborations can be subjected to many cognitive and institutional challenges (Leahey et al., 2017a), are highly risky in terms of the desired outcome, can take many years to become successful, and are particularly difficult to manage (Blackwell et al., 2009). Furthermore, due to multiple organizational and cultural barriers and disciplinary orientations, the willingness and ability for interdisciplinary collaborations are only weakly developed among many actors (Amey & Brown, 2006; Kirby et al., 2019). In the same time, interdisciplinary work became more difficult by the lockdown, canceled conferences and home office (Kastenhofer et al., 2023). Moreover, researchers emphasize the need for international collaboration to tackle COVID-19 (Mohamed et al., 2020). Even though multiple scholars investigated the rapid development of a scientific response to the COVID-19 virus crisis based on a bibliometric analysis (Haghani & Bliemer, 2020; Colavizza et al., 2020; Bonilla-Aldana et al., 2020; Joshua & Sivaprakasam, 2020), their explorative focus did not elaborate on the factors that induce or hinder interdisciplinary collaborations in COVID-19 research. Little is known about the aspect of interdisciplinarity and, in particular, whether the pandemic has increased the degree of interdisciplinarity in scientific collaborations. This is an important issue regarding the guidance for policymakers' decisions on the institutional environment for coping with similar crises in the future. In our study, we aim to focus on the actors and their characteristics in scientific networks to understand them as decision-makers for collaborations with distant disciplines.

Furthermore, there has been a tendency that other important research topics could have been displaced by the intensively growing research field "COVID-19" (Woo Baidal et al., 2020; Watermeyer et al., 2021). Besides a cannibalization of topics, research on quite different other established topics may also have adapted to the new conditions of the pandemic. Topic adjusting effects due to COVID-19 pandemic have not yet been investigated in scientific research. Adding knowledge to this potentially unintended effects may provide important insights for science management and science policy for future global urgent crises and in the course of future mission-oriented research policies.

We go beyond the findings by Zhao et al. (2022) and Coccia (2021) and investigate to what extent COVID-19 has had an impact on how research is conducted in terms of scientific collaborations and interwovenness of scientific topics. In particular, our research questions are:

- Did the emergency of the COVID-19 pandemic lead researchers to focus on topics of COVID-19 and by this to a reduction of efforts in other established scientific topics? How can the scientific response to COVID-19 be compared to the responses to medical emergencies, such as the Zika virus, Ebola, and SARS?
- Did research on COVID-19 increase the degree of interdisciplinarity, as measured by the established representation learning method DeepWalk applied to the network of bibliographic metadata?
- What social network characteristics of the authors of COVID-19-relevant research are important for increasing the degree of interdisciplinarity?

Answering these questions requires a multi-faceted analysis: We first analyze the impact of COVID-19 crisis-driven research and compare it with other disease outbreaks. Then we

conduct a concept (MeSH descriptors<sup>1</sup>) similarity analysis grounded on concept embeddings that are obtained with machine learning from graph-structured bibliographic data, including publications, concept annotations, and citations. Based on the cosine distances between the embeddings of the MeSH descriptors, we measure the (dis)similarity between scientific fields, which is the foundation for our interdisciplinarity indicator. We use this interdisciplinarity indicator to run a fine-grained analysis of the degree of interdisciplinarity of COVID-19 research in 2020. Moreover, we conduct a reference analysis of SARS-CoV-2/COVID-19 relevant scientific articles and preprints. Finally, we investigate the scientific social networks of scholars who have contributed to the COVID-19 research to explore whether an author's betweenness centrality or an author's degree centrality facilitates the degree of interdisciplinarity at the researcher level.

As datasets, we rely on the ZB MED Knowledge Environment<sup>2</sup> (ZB MED KE) and a dedicated COVID-19 dataset (called COVID-19++, see Galke et al. (2021) for details on the dataset), which contains scholarly publications and preprints on COVID-19, enriched with works that are cited by them. The datasets are described in more detail in Appendix A.

In summary, the key contributions of this work are:

- We run four analyses to investigate the dynamics of the scientific response to the COVID-19 outbreak from different angles: research volume, concept similarity, references (i.e., citations), and author networks.
- We show how an incrementally trained DeepWalk model can be utilized to reflect the similarity of concept embeddings and how this can be used to conduct an analysis of research dynamics regarding the interdisciplinary work.
- Our analyses of COVID-19 research dynamics reveal that, while the focus shifted noticeably to the topic of COVID-19 in disregard of other topics, there is no increase in interdisciplinarity.

This article is structured as follows. The literature on COVID-19 research dynamics is reviewed in Sect. 2. Section 3 analyses the shift in research volume by topic and compares this shift with other medical emergencies through a descriptive analysis. Section 4 analyses to what extent research topics change over time and how this change affects their proximity through a concept similarity analysis, carried out with representation learning in citation networks with bibliographic metadata. Subsequently, Sect. 5 investigate what insights researchers use from other fields by an analysis of the relationship between COVID-19 research articles and their references. Lastly, Sect. 6 investigates the factors for interdisciplinarity (quantified as diversity in topics of publications) of authors through a social network analysis, before we discuss our results in Sect. 7.

## State of research and literature gap

In the following, we describe how COVID-19 has affected academic research from multiple angles. During the pandemic, many research articles on COVID-19 have been published. Raynaud et al. (2021) show that the increase in publication volume of COVID-19

<sup>1</sup> The Medical Subject Headings (MeSH) thesaurus is a controlled and hierarchically-organized vocabulary produced by the U.S. National Library of Medicine. See <https://www.nlm.nih.gov/mesh/meshhome.html>.

<sup>2</sup> <https://www.zbmed.de/en/research/completed-projects/zb-med-knowledge-environment/>.

was accompanied by a decrease in the share of non-COVID-19 publications. Dinis-Oliveira labels this as a “paperdemic” and reminds us of the principles of scientific integrity (Dinis-Oliveira, 2020). Within this large volume of research, we compile papers from the scientometric perspective on how the pandemic has impacted research and what new trends can be inferred from bibliographic data.

## Impact of the pandemic on research: a scientometric perspective

From the first months of the COVID-19-pandemic, scholars have attempted to understand the knowledge system behind SARS-COV-2. For instance, Benjamins et al. (2021) analyze geographic sources of publications in medical journals. Although the authors note that the distribution is unexpected, they argue that medical journals should solicit articles from underrepresented countries for a more representative discourse, given the global impact of the crisis. Similar to our research interest, Coccia (2021) is concerned with research dynamics from a scientometric perspective. Using Scopus publication metadata, the author compares scientific output that is driven by a crisis (“crisis driven”; e.g., COVID-19, MERS, ZIKA, HIV, H1N1) with that addressing a continuous problem (“problem-driven”; e.g., lung cancer, chronic obstructive pulmonary disease (COPD)). Crisis-driven research has a high publication rate; whereas problem-driven research-related publication volume is rather linear. While crisis-driven research is concentrated in terms of a few scientific fields (3–5), a few journals, a few funding agents, and countries associated with about 80% of the publication volume; problem-driven research is more widely diffused Coccia (2021). Coccia finds another difference with regard to open-access publications: Crises-driven research is more published in open access (78%), whereas only about 40% of open access publications constitute research in lung cancer and COPD.

Other authors investigate the degree of interdisciplinarity of COVID-19-related research and conclude that COVID-19-related publications are dominated by a few disciplines such as Medicine, Immunology and Microbiology, Biochemistry Genetics, and Molecular Biology (Zhao et al., 2022). The authors found that before 2020 the number of disciplines involved in coronavirus research raised, while the balance and diversity of disciplines showed a falling trend from 1990 to 2019. They conclude that coronavirus research has not become more interdisciplinary, but it has been catalyzed by the COVID-19 pandemic (Zhao et al., 2022). Also, Fassin (2021) takes a bibliometric approach and investigates the impact of the COVID-19 publication explosion on bibliometric indicators. After providing an analysis of  $h_a$  and  $h$  indices, the paper stresses the salience of the topic, the magnitude of the problem, and urgency, as the key drivers for citations. Aviv-Reuven and Rosenfeld (2021) investigate changes in publishing patterns, particularly in the volume and in average time to acceptance of both preprints and journal articles. The authors find a sharp increase in publication volume and a significantly faster mean time of acceptance for COVID-19 papers.

Liu et al. (2022) argue that the pandemic is a catalyst of scientific novelty. The authors have applied a BioBERT model, pretrained on 29 million PubMed articles, on 98,981 COVID-19 related papers to find that scientific novelty has increased, along with an increase in first-time collaborations, whereas international collaboration experienced a sudden decrease. Harsanto (2020) analyzes the bibliographic metadata of publications in the first three months after COVID-19 hit. The study finds that, on average, 150.33 documents were published every

month, with medicine as the main field of research (62.04%). Shan et al. (2020) argue that it becomes increasingly difficult to publish non-COVID-19 research.

### **Need for interdisciplinary collaboration during the COVID-19 crisis?**

Due to the effective results, such as the development and successful testing of the coronavirus vaccine (Li et al., 2021), we expect to find interdisciplinary research dynamics in the scientific activities on COVID-19 since the multidimensional challenge of the pandemic requires collaborative efforts of multiple disciplines. The advantages of an interdisciplinary approach in the fight against the pandemic are seen by around 88% of respondents to a delphi study conducted between February and April 2022, "The incorporation of research paradigms from diverse disciplines has greater potential to end COVID-19 as a public health threat [...]" (Lazarus et al., 2022). In particular, the need for cooperation between the social sciences and medicine is repeatedly emphasized (Corsi & Ryan, 2022). We adopt the definition of interdisciplinary research of the US National Academy of Sciences, Engineering, and Medicine which defines it as follows: "(...) a mode of research by teams or individuals that integrates information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice." (Sciences et al., 2005) The definition also describes interdisciplinarity as interdisciplinarity between subgroups of disciplines ("bodies of knowledge").

A new combination of technological knowledge and components may result in breakthroughs (Fleming, 2001), and a wide breadth of knowledge enables the development of radical innovations (Xu, 2015). A high level of interdisciplinarity can drive the creativity and innovativeness of researchers in academia (Leahey et al., 2017b), increase publication impact (Yegros-Yegros et al., 2015), and enhance the R & D performance of public research centers (Jung et al., 2021). A high degree of interdisciplinarity of researchers is positively related to their intention and ability to launch a company and license their patents to private firms (D'Este et al., 2019). Moreover, a solution-driven R & D approach also forces organizations to employ professionals from different science and technology fields who must work with their colleagues from other disciplines and, therefore, go beyond the boundaries of their disciplines. In this regard, the distance of knowledge fields in interdisciplinary R & D is especially large during early phases of radical innovation, for instance, developing a vaccine against the SARS-CoV-2 virus, and when new techno-scientific fields emerge (Meyer-Krahmer, 1997). Therefore, interdisciplinary collaborations could have facilitated the development of urgently required vaccines and medicine against the novel coronavirus. Nevertheless, studies for the period before the COVID-19 pandemic show that there is little openness among scientists to implementing interdisciplinary approaches. And this is also in contrast to the broad-based funding opportunities offered by third-party founders who want to support interdisciplinary approaches between STEM (science, technology, engineering and math) and Non-STEM research (Uddin et al., 2021).

**Table 1** MeSH terms for the comparison of the novel coronavirus to other medical emergencies' publication volume

Medical emergency	Mesh term	Concept number
AIDS	HIV-1	D015497
SARS	Severe Acute Respiratory Syndrome (SARS)	D045169
Swine Flue	Influenza A Virus, H1N1	D053118
MERS	Middle East Respiratory Syndrome (MERS)	D065207
Ebola	Ebolavirus	D029043
Zika Virus	Zika Virus	D000071244
Marburgvirus	Marburgvirus	D029024

## Research volume analysis

### Methods

*Comparison to other medical emergencies* To get a general sense of the publication volume related to the COVID-19 pandemic, we conducted a comparison to the occurrence of MeSH terms related to other similar medical emergencies in the Medline database (He & Chen, 2018; Xiang et al., 2020). Therefore, we retrieved amounts of publications annotated with the specific MeSH terms related to the AIDS crisis (since the early 1980ies), the SARS pandemic (2002-2003), the Swine Flue (pandemic in 2009), the MERS epidemic (2014), Ebola epidemic (2014-2016), Zika virus outbreak (2015–2016), as well as Marburgvirus with its last major outbreak in 2017 (Table 1 and Fig. 1).

*Concept Drift Magnitude* To quantify the change over time in the concept annotations, we evaluate the drift magnitude of the concept distribution in our datasets in order to estimate the shift in content over time. We follow Webb et al. (2018) and use total variation distance. Formally, it is defined as:

$$\sigma_{t-1,t} = \frac{1}{2} \sum_{y \in \text{dom}(Y_{t-1} \cup Y_t)} |P_{t-1}(y) - P_t(y)|$$

where  $t-1$  and  $t$  may be time points or time intervals that define the respective observed distribution. Figure 2 shows the results for the class drift magnitude on our COVID19++ dataset. We use the total variation distance in two ways. First, we compare the drift magnitude of the data over time. Second, we quantify the topic shift between primary COVID-19 articles and referenced literature.

We use the ZB MED Knowledge Environment (ZB MED KE) and a derived COVID-19-specific subset of the ZB MED KE, called COVID-19++ Galke et al. (2021). The ZB MED KE is a heterogeneous database environment containing over 70 million records from over 70 specialized databases in the life sciences field. We describe the datasets in more detail in Appendix A.

In order to get a broader impression of the publication amount in general, the publication volume on the following MeSH terms of broader topics has been included (Fig. 1, upper panel): *Betacoronavirus* (D000073640), *Heart Diseases* (D006331), *Cardiovascular Diseases* (D002318), *Myocardial Ischemia* (D017202), *Respiratory Tract Diseases* (D012140), *Betacoronavirus* (D000073640), and *Heart Diseases* (D006331).



## Results

Although there has not been a specific MeSH term for the SARS-CoV-2 pathogen nor for the SARS-CoV-2 illness in 2020, we can assess the increasing number of publications on the topic from the development of the term *Betacoronavirus* (D000073640) (Fig. 1). Moreover, the steep increasing curve progression of SARS, which is related to SARS-CoV-2, can be taken as an indication. Notably, the term *COVID-19* (D000086382) was only introduced in the MeSH edition 2021 but was not available before that date. Fig. 1 shows that the research on respiratory tract diseases has also drastically increased in 2020 reflecting that the novel coronavirus predominantly attacks the human respiratory tract (Ali & Alharbi, 2020).

We took the increasing amount of publications during the last decades into account by normalizing the data as a percentage of the actual volume of research output. The outbreak of HIV in the early 1980s and the research on it seem to put research on other virus-related analyzed topics in shade in the data. However, we can still observe the decline of research on HIV that appears to be related to other outbreaks, especially on SARS, Influenza H1N1 Virus, and Zika Virus.

Similarly, in 2002–2003, when the SARS pandemic occurred, the data in Fig. 1 (lower panel) shows an increase in research activities on the topic. Analogously, in 2009, we know the swine flu pandemic took place. It was mainly caused by the H1N1 virus and resulted in an increase in the amount of research on the virus. The Ebola outbreak in 2014 induced a comparatively slight rise in scholarly publications on the topic. It might be also overlapped with the Zika Virus epidemic in 2015–2016, which caused a lot of publications annotated with the respective MeSH term. All these curves appear to decline at the moment when betacoronavirus and SARS research enter the scene again in 2020. Both terms are related to SARS-CoV-2 research, which lacked a specific term at that time. Compared to the other curves related to medical crises, the rise of scientific output on the MeSH term *Betacoronavirus* as well as on *SARS* during 2020 is much stronger in terms of the amount of research output. If we combine both curves of *Betacoronavirus* and *SARS*—since some publications will be annotated with both MeSH terms, and this is why we cannot simply add the numbers—the research dynamic can be compared to the crisis of HIV rather than to other crises in the recent decades. Notably, we observe a decline in the share of publications in other topics such as heart diseases and cardiovascular diseases.

Figure 2 shows the drift magnitude in the datasets, i.e. the change in topics over time of the COVID-19 subset and a 15M subset of the ZB MED Knowledge Environment (ZB MED KE) (filters: English and MeSH descriptors availability). Both COVID19++ (N=122,886) and ZB MED KE (N=15,196,066) exhibit an unusually strong drift in 2020. The drift in the COVID19++ data set is consistently larger than the drift of ZB MED KE.

To compare the publication process in relation to other outbreaks of viruses, namely the SARS outbreak in 2002–2003, the Influenza pandemic caused by the H1N1 virus in 2009–2010, and most recently the Zika virus epidemic in 2015–2016, publications from these very periods of the acute health crises had been extracted from ZB MED Knowledge Environment. Referring to the open CrossRef Database<sup>3</sup>, the referenced publications had been supplemented. For identified references, we extracted MeSH descriptors if available. Afterwards, the MeSH descriptors had been mapped to tree number level 1 of the MeSH hierarchy. We then use the MeSH descriptors to calculate the shift in topics between primary and referenced publications. The results are shown in Table 2. COVID-19 shows a similar topic shift as Zika Virus. SARS and H1N1 have a higher topic shift (Table 2).

<sup>3</sup> <https://www.crossref.org>.



**Fig. 1** Trending MeSH terms within health crisis publications in Medline. The upper picture shows the broad context of publication topics and their publication volume; the lower image concentrates on individual pathogens with a more precise focus

Moreover, we observe significant differences in the distribution of publications and references over the MeSH thesaurus level 1 domain between MeSH descriptors *Respiratory Tract Diseases* and *Cardiovascular Diseases*, *Heart Diseases*, and *Myocardial Ischemia* (Fig. 3, upper figure). These differences are salient for MeSH domains M (Named Groups) and N (Health Care). Further, the analysis reveals considerable differences in the distribution of publications and references over the MeSH thesaurus level 1 domain between MeSH descriptors *SARS* and *Influenza A Virus, H1N1* as well as *Zika Virus* for MeSH domains A (Anatomy), B (Organisms), C (Diseases), D (Chemicals and Drugs), G (Phenomena and Processes), and N (Health Care) (Fig. 3, lower panel).

Consequently, our analyses show that COVID-19 research in the analysis period of 2020 has grown unprecedentedly fast and at least partially has displaced research efforts on other viruses and diseases.

## Concept similarity analysis

### Methods

#### Dynamic concept embedding space learned by DeepWalk

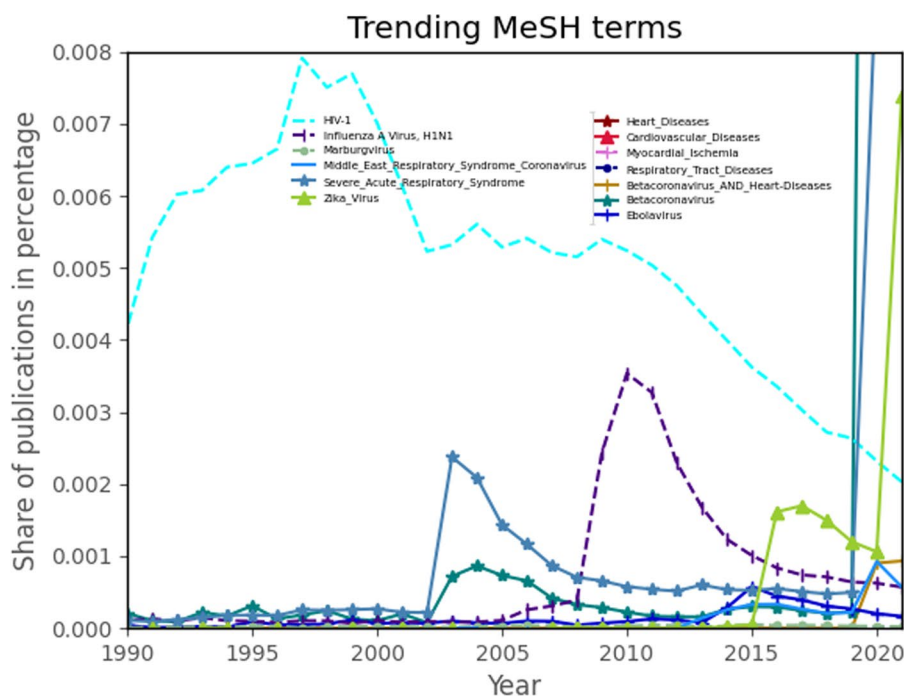
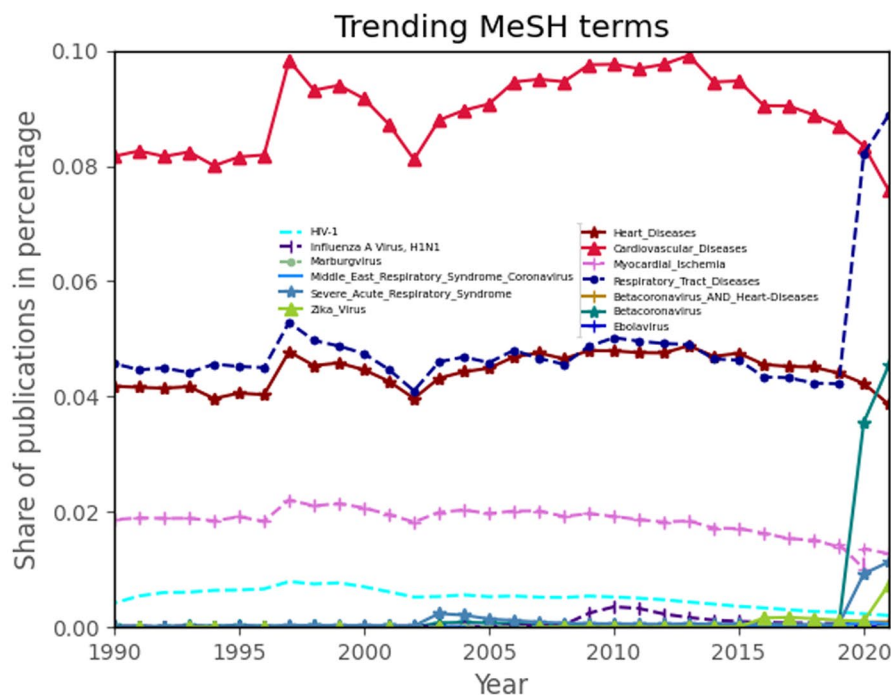
Counting how often a concept is assigned to a document together with another concept is a common proxy for the similarity of these two concepts (Tijssen, 1992). However, such co-occurrence counts are insufficient to derive a notion of similarity between concepts. That is because two concepts may be very similar, which is not reflected in co-occurrence counts unless the articles are consistently annotated with all similar concepts. Moreover, the annotation of publications with concepts from a controlled vocabulary rather reflects the diversity of aspects of that particular publication. Thus, co-occurrence does not imply that the concepts are similar (Galke et al., 2018).

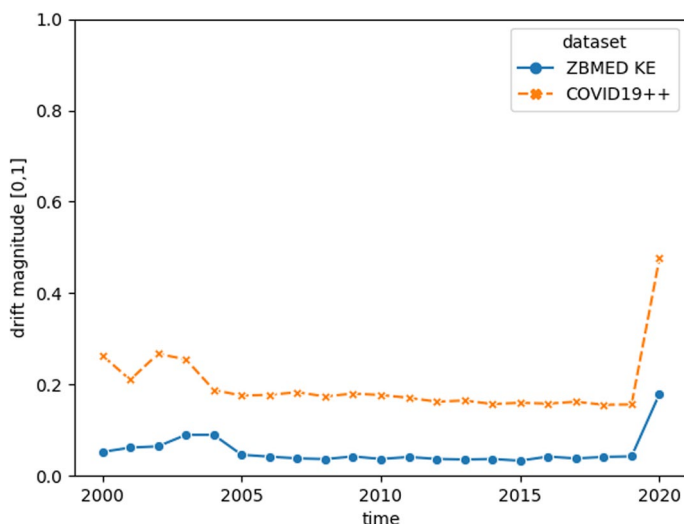
A crucial example is that the MeSH descriptor *COVID-19* only appeared on 7 July 2020<sup>4</sup>. Thus, research articles published before July had to resort to other terms of MeSH, such as *Betacoronavirus* or *Pneumonia, Viral*. With our machine learning model, the similarity of the two concepts (or between any other concepts) is learned and taken into account. We pursue an embedding approach (Perozzi et al., 2014; Grover & Leskovec, 2016) that exploits bibliographic metadata such as the authors of the publications, its journal, and the annotation with concepts from a controlled vocabulary.

In the following, we outline how concept embeddings can be learned from the graph of bibliographic metadata and how these embeddings facilitate a measure of concept similarity over time. For training this model, we use the COVID-19++ dataset (Galke et al., 2021). The dataset is described in more detail in Appendix A.

To transform bibliographic metadata into graphs, we consider each publication and each concept to be a node in the graph. The publication node is connected with the concept node

<sup>4</sup> <https://meshb.nlm.nih.gov/record/ui?ui=D000086382>.





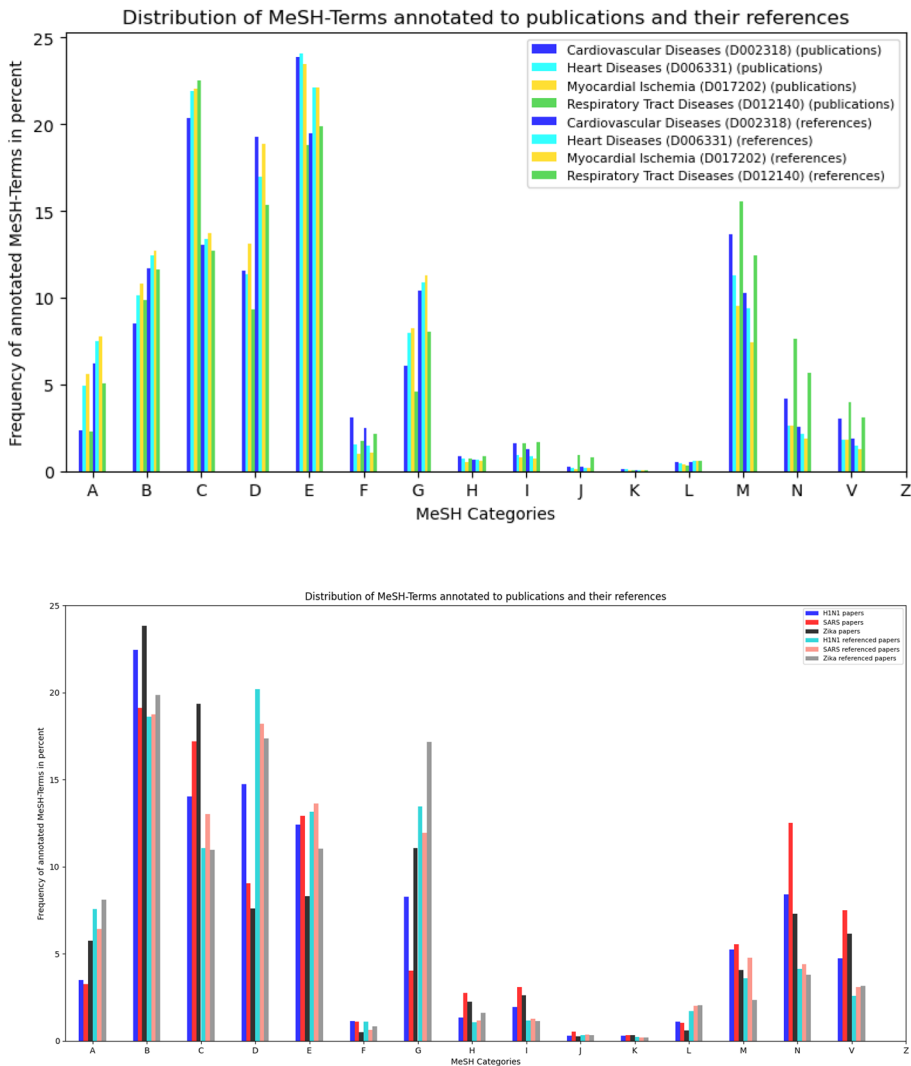
**Fig. 2** Changes in the distribution of MeSH terms. Class Drift Magnitude between 2000 and 2020 and the 15M documents from the ZB MED KE compared to the COVID-19 specific dataset COVID-19++. COVID19++ is a subset of ZB MED KE that contains articles on COVID-19, along with preprints, and the publications cited by those

**Table 2** Shift in Topics between primary and referenced literature: A control group of MeSH terms related to virus diseases outbreaks

MeSH term	Publ. year(s)	#Primary	#Referenced	Topic Shift
Zika virus	2015/16	693	4534	0.1605
SARS	2002/3	3164	45657	0.2198
H1N1	2009/10	843	15098	0.2280
COVID-19	2020	50572	307448	0.1541

via an edge when the publication is annotated with that particular concept. Similarly, we insert a node for each author and each journal into the graph linked to the publication by an edge if the author has written the publication or if the publication is published in the journal, respectively. As a result, we obtain a graph  $\mathcal{G} = (V, E)$  constructed based on bibliographic metadata. The set of nodes  $V$  is comprised of documents (original publications and their references) nodes  $\mathbb{P}$ , concept nodes  $\mathbb{C}$ , author nodes  $\mathbb{A}$ , and journal nodes  $\mathbb{J}$ . The undirected edges  $E$  may resemble authorship relations between authors and publications, annotation relations between publications and concepts, and the journal. We use this graph as a basis for learning concept embeddings. Based on these concept embeddings, we compute the similarity between concepts.

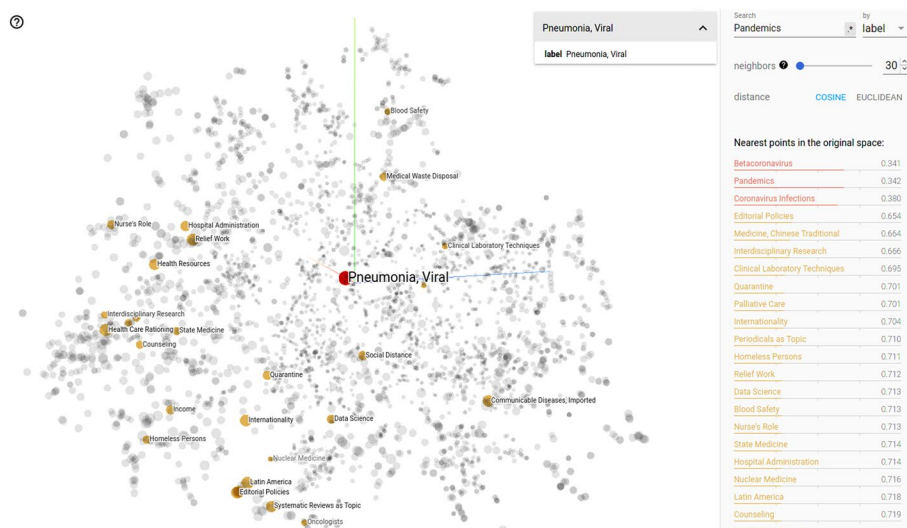
To learn the embedding function  $f$ , we use the DeepWalk algorithm (Perozzi et al., 2014; Grover & Leskovec, 2016). DeepWalk is an established approach for learning node embeddings in a graph. The Deepwalk algorithm randomly initializes a node embedding  $X \in \mathbb{R}^{n \times s}$ , where  $s$  is the embedding size. To update the embeddings of the



**Fig. 3** Shift of topics in primary publications and referenced publications for control groups

nodes, DeepWalk samples random walks  $(u_1, u_2, \dots, u_l)$  through the graph such that  $\forall i \in 1, 2, \dots, l : u_i \in V$  and  $\forall i \in 1, 2, \dots, l-1 : (u_i, u_{i+1}) \in E$ , where  $l$  is the length of the random walk. For each node on each random walk, its current embedding is used to predict neighboring nodes within a fixed context window size  $n_{\text{cwnd}}$  along the random walk. The respective node embedding is updated according to the error signal from the prediction. DeepWalk is connected to the Word2vec model from natural language processing (Mikolov et al., 2013) by considering random walks as sentences.

In the present work, we intend to learn node embeddings not only on a single graph but also on multiple graph snapshots over time. DeepWalk relies on random initialization. Therefore, different runs may converge to different node embeddings. This is a critical factor as the node embedding at time  $t+1$  should be based on the embedding of



**Fig. 4** A t-SNE visualization of the final embedding space produced by the DeepWalk algorithm after incrementally training on all available data. The 30 nearest neighbors (in original high-dimensional space) of the concept *Pneumonia, Viral* are highlighted. The three closest neighbors are *Betacoronavirus*, *Pandemics*, and *Coronavirus Infections*

the node at time  $t$ . In the original work on DeepWalk (Perozzi et al., 2014), the authors have suggested the possibility of online learning. In our snapshot-based setting, we run multiple epochs for time step  $t$  before advancing to time step  $t + 1$ . To stabilize training across time steps, we reuse the node embeddings from the previous time step to initialize the embeddings for the next time step. This is similar to the method for deep autoencoders proposed by Goyal et al. (2018, 2020). Thus, we expect the embeddings to remain similar but still account for the changes in the graph.

Regarding hyperparameter choices, we use a walk length of  $l = 80$ , a context size of  $c = 10$ , and an embedding size of  $s = 128$ . We sample 10 random walks per concept per month. We optimize the embedding with a learning rate of 0.025 that linearly decays to zero at each time step. The resulting embedding space is visualized by t-Stochastic Neighbor Embedding (t-SNE) (van der Maaten & Hinton, 2008) in Fig. 4.

To compare similarities, we first normalize the resulting concept embedding  $X_{|\text{concepts}|}$ , whose rows  $i$  hold concept embedding  $f(c_i)$  in two steps. First, we center the embedding by subtracting the centroid. Then, we normalize the columns to unit  $L_2$ -norm. As a result, we have normalized and centered vectors for each concept  $c^{(t)}$  in each time step  $t$ .

To compare a pair of concepts at time  $t$ , we compute the cosine distance on normalized and centered concept vectors  $a, b$ .

$$d_{\cos}^{(t)}(a, b) = 1 - \frac{a^{(t)} \cdot b^{(t)}}{\|a^{(t)}\| \cdot \|b^{(t)}\|}$$

The cosine distance ranges between 0 and 2 and a lower cosine distance indicates a higher degree of similarity between the two concepts.

We are interested in the relationship and dynamics of whole research fields rather than merely two concepts. For that purpose, we assume that a research field can be described by a set of concepts. We define a metric to compare two sets of concepts, given the learned concept embedding. This is connected to the average linkage metric in the context of clustering methods (Chapter 17: Hierarchical Clustering Manning Raghavan, & Schütze 2008). We consider two concept sets  $\mathbb{A}$  and  $\mathbb{B}$ . Then we compute the pairwise cosine distances between all the concepts of  $\mathbb{A}$  and all the concepts of  $\mathbb{B}$ . We regard the mean of these pairwise distances as the distance between the two concept sets.

$$\overline{d_{\cos}^{(t)}(\mathbb{A}, \mathbb{B})} = \frac{1}{|\mathbb{A}| \cdot |\mathbb{B}|} \sum_{a \in \mathbb{A}, b \in \mathbb{B}} d_{\cos}^{(t)}(a, b)$$

We consider the prevalence of increasing similarity (convergence) between concept sets  $\mathbb{A}$  and  $\mathbb{B}$  during the time span  $t_{\text{start}}, t_{\text{end}}$  if the distances are monotonously decreasing, i. e., if the following equation holds:

$$\forall t \in [t_{\text{start}}, t_{\text{end}}) : \overline{d_{\cos}^{(t+\Delta t)}(\mathbb{A}, \mathbb{B})} \leq \overline{d_{\cos}^{(t)}(\mathbb{A}, \mathbb{B})}$$

With this model, we have generated temporal embeddings on our newly introduced COVID19++ dataset that we use for the analyses of research dynamics, which we will describe next.

## Identification of relevant scientific fields on the basis of MeSH terms

Further, we conducted a specific analysis of MeSH terms (descriptors) in our data set. After the exclusion of particular common and COVID-19-related MeSH descriptors, we identified the most frequent MeSH descriptors over the analysis period (January–December 2020) and allocated it according to the MeSH thesaurus hierarchy at the second level (which we define as MeSH subdomains). Afterward, we determined 12 MeSH subdomains with the largest numbers of the most frequent MeSH descriptors. These MeSH subdomains and especially MeSH descriptors represent the most important medical and social spheres related to or affected by the COVID-19 pandemic.

For instance, the MeSH descriptors *Pneumonia*, *Pneumonia-Viral*, *Severe Acute Respiratory Syndrome*, *Asymptomatic Infections* of the subdomain Infections (C01) reflect diseases caused by the novel coronavirus (Huang et al., 2020). The efforts of scientists to develop efficient medicine and vaccine are displayed in the subdomain Amino Acids, Peptides, and Proteins (D12), since its most frequent MeSH descriptors *Antibodies*, *Protein S*, *C-Reactive Protein*, *Cytokines* are indicative of the scientists' considerations regarding the triggers of the immune responses to the novel coronavirus and describe different approaches for immune response mechanisms at medical and biochemical levels (Watanabe et al., 2020; Rogers et al., 2020; Kayser et al., 2020).

The MeSH descriptors *Diagnosis*, *Viral Load*, *Tomography*, *Disk Diffusion Antimicrobial Tests*, *Respiratory Rate*, *Diagnostic Imaging* of the subdomain Diagnosis (E01) and *Hospitalization*, *Therapeutics*, *Respiration-Artificial*, *Intubation*, *Vaccination* MeSH descriptors of the subdomain Therapeutics (E02) reveal that scientists have intensively investigated and employed the proper diagnostic instruments and therapy approaches for the novel COVID-19 disease which were that time utilized in hospitals and health care (Huang et al., 2020).

Furthermore, the scope of the pandemic is reflected in the MeSH descriptors *Incidence, Prevalence, Sensitivity and Specificity, Mortality, In Vitro Techniques, Polymerase Chain Reaction* of the subdomain Investigative Techniques (E05) which represents an identification method of the SARS-CoV-2 RNA and an analysis of the dissemination and transmission of SARS-CoV-2 infections.

The frequent usage of MeSH descriptors *Epidemics, Quarantine, Infection Control, Disinfection* of the subdomain Environment and Public Health (N06) underlines that SARS-CoV-2 is highly infectious and can be widespread very fast without the appropriate infection control measures (Ali & Alharbi, 2020).

Moreover, the frequent occurrence of the MeSH descriptors *Anxiety, Stress-Psychological, Social Distance, Depression, Fear* of the subdomain Behavior and Behavior Mechanisms (F01) reveal that the COVID-19 triggers also severe psychological issues (Cullen et al., 2020).

The geographical dissemination of COVID-19 can be followed by the Geographic Location (Z01) MeSH descriptors *China, United States, Italy, United Kingdom, and India* as countries that were most affected by or related to the coronavirus topic at the beginning of the pandemic (Ali & Alharbi, 2020).

To sum up, our results show that MeSH descriptors are a reliable instrument to reflect multiple facets of an emerging scientific topic in the medical field as well as societal issues related to this topic. Even though the prior literature uses a keyword co-occurrence analysis to cluster publications on coronavirus and COVID-19 (Haghani & Bliemer, 2020; Radanliev et al., 2020), our proposed method allows for a more fine-grained and validated analysis of the topics, since we use the established MeSH thesaurus for the identification of topics.

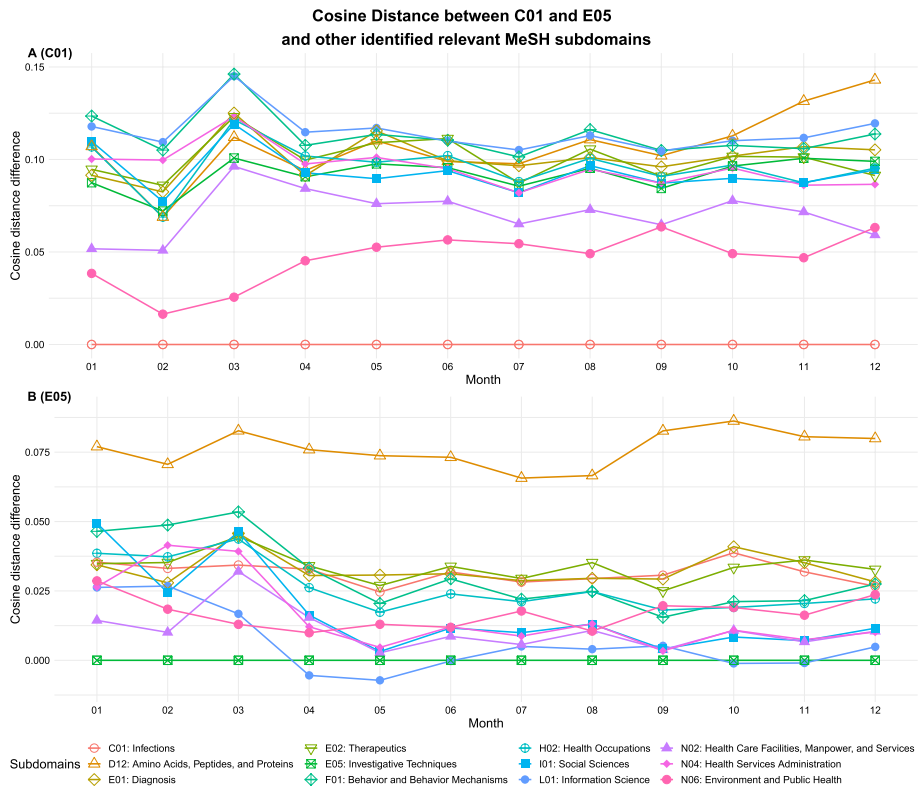
## Results

We used all MeSH descriptors that we identified in our dataset and that were allocated to the twelve relevant above-mentioned subdomains. We analyzed whether the topics represented by MeSH descriptors approach each other (converge) to unveil interdisciplinary collaborations. Since scientific convergence can be revealed by the usage of research results, methods, and techniques of one separate discipline by another one (Curran, C-S. & Leker, J., 2009; Curran & Leker, 2011), we spanned bibliographic graphs of scientific publications' metadata containing information on publications, authors, MeSH descriptors as annotated concepts and applied machine-learning algorithms on these graphs to train a model for the evaluation of the degree of interdisciplinary collaborations. Furthermore, we applied our developed measurement based on the topic (concept) similarity, as described in Sect. 4.1. Fig. 5 shows the degree of convergence (growing interdisciplinarity) or divergence (growing specialization) of relevant MeSH subdomains, measured as the normalized mean cosine distance.

The curves in the Fig. 5 should be interpreted in the following way: a decrease of cosine distance over time implies that two relevant fields approach each other, and they start to converge—i.e. the degree of research interdisciplinarity increases; an increase in cosine distance over time means that the scientific fields depart from each other and suggest a scientific divergence pattern—i.e. a specialization within scientific fields grows. In general, the lower a cosine distance is, the nearer two scientific fields are to each other.

To calculate the normalized cosine distances difference to other subdomains, we subtracted the normalized cosine distance of the MeSH descriptors of the MeSH subdomain





**Fig. 5** Cosine distance difference in DeepWalk embedding space between MeSH subdomain Infections C01 (a) and Investigative Techniques E05 (b) and other identified relevant MeSH subdomains in 2020. The values of the cosine distance curves of C01 and E05 were set to zero as baselines

Infections (C01) from the values of the normalized cosine distance of other subdomains. In this way, the normalized mean cosine distance of the MeSH descriptors of the MeSH subdomain Infections (C01) was set to zero value as a baseline.<sup>5</sup>

As Fig. 5a shows in the first month of 2020 the contextual distance between the MeSH subdomain Infections (C01)) and other selected subdomains decreased implying that the research on COVID-19 in other subdomains was primarily associated with the infections and diseases that the coronavirus triggered. As such, the degree of research interdisciplinarity increased. From February to March 2020, other subdomains were deepened in their own research topics, thus, the cosine distance between Infections (C01) and other MeSH subdomains increased—a specialization prevailed. From March 2020, the cosine distance between the MeSH subdomain Infections (C01) again slightly decreased and remained at a stable level for most of the other analyzed subdomains, except the subdomain Environment and Public Health (N06). The subdomain

<sup>5</sup> The values below zero imply that the normalized cosine distances between the subdomain Infections (C01) and other subdomain are lower than the mean of normalized cosine distances between all MeSH descriptors within subdomain Infections (C01).

N06 containing such MeSH descriptors as *Epidemics*, *Quarantine*, *Infection Control*, *Ventilation*, and *Hand Disinfection* had in February 2020 the lowest cosine distance to the subdomain C01 (Infections), revealing the urgent necessity of preventing measures against COVID-19 spreading. Thus, we observe an increase in research interdisciplinarity between subdomains C01 and N06. However, after February 2020 this subdomain N06 was evolving more independently of the field marked as Infections (subdomain C01). As such, the cosine distance between these two subdomains gradually increased, reaching the highest distance difference in September 2020. Meanwhile, the contextual closeness and relatedness of the subdomains Infections (C01) and Health Care Facilities, Manpower, and Services (N02), containing MeSH descriptors *Hospitals*, *Intensive Care Units*, and *Triage*, started to grow from March 2020, since the cosine distance was falling from March 2020, reaching the lowest distance between two subdomains in December 2020. This interrelatedness of these two subdomains can be explained by the drastically increasing need for hospitalization and intensive care for COVID-19 patients. Surprisingly, the cosine distance between Infections (C01) and Amino Acids, Peptides, and Proteins (D12) was even incrementally rising from the mid of the year till December 2020.

Figure 5b reveals that the cosine distance between the subdomain Investigative Techniques (E05) and other subdomains was substantially lower in comparison to the Infections (C01) subdomain. In the first month of 2020, the cosine distance between Investigative Techniques (E05) and other subdomains decreased, implying rising interdisciplinarity. In March 2020, it raised for the subdomain Social Science (I01) and Behavior and Behavioral Mechanisms (F01), underlying a weak relationship between E05 MeSH concepts and *Socioeconomic Factors*, *Demography*, *Culture*, *Policy*, *Economics* MeSH concepts of I01. Similarly, the cosine distance between E05 MeSH concepts and concepts of Health Care Facilities, Manpower, and Services (N02) and Amino Acids, Peptides, and Proteins (D12) subdomains increased from February to March 2020, revealing the disciplines' specialization. Afterward, cosine distance fell for all pairs between E05 and other subdomains. The decreasing cosine distance means that despite the discipline and subdomain differences, the contextual interrelatedness between E05 and other subdomains' MeSH concepts increased. Thus, the degree of interdisciplinarity slightly increased. Strikingly, the cosine distance between E05 and Information Science (L01) subdomain even dropped under zero (baseline), highlighting the higher importance of L01 MeSH concepts for the Investigative Techniques subdomain. L01 subdomain contains, among others *Computer Simulation*, *Software*, *Social Media*, *Information Storage and Retrieval* MeSH concepts. The low cosine distance difference between E05 and L01 implies an amplified impact of information science technologies on conventional and novel medical analysis methods. Hence, we observe a high degree of interdisciplinarity between subdomains E05 and L01 in April and May 2020.

In total, our analysis of scientific topics related to COVID-19 (defined as MeSH subdomains) shows that most of the scientific subdomains and fields rely on their own research insights, techniques, and methods. This finding suggests that interdisciplinarity is less prevalent in COVID-19 research. However, the fine-grained analysis of pairwise topics shows that researchers produce knowledge on particular topics by integrating insights from both topics. We performed a similar analysis for the MeSH descriptor pairs denoting vaccine candidates. Appendix C shows the results.

## Reference analysis

### Methods

The analysis of research interdisciplinarity is usually performed by consulting the citation data of scientific publications (Yegros-Yegros et al., 2015; Rafols & Meyer, 2010). Hence, we conduct an analysis of research flows embedded in SARS-CoV-2/COVID-19 publications and preprints to evaluate the level of interdisciplinarity in the knowledge flows.

To determine what knowledge has been used to conduct research on COVID-19, we compared the annotations to publications on COVID-19 with those annotations to papers cited in them. We examined the distribution of MeSH terms annotated to the primary journal publications on SARS-CoV-2 as well as preprints and the referenced publications.

Specifically, we used the total variation distance to quantify the differences in the distributions of MeSH descriptors over the 15 MeSH thesaurus domains for the COVID++ dataset between primary publications and the respective references. Section 3.1 contains details of the total variation distance calculation.

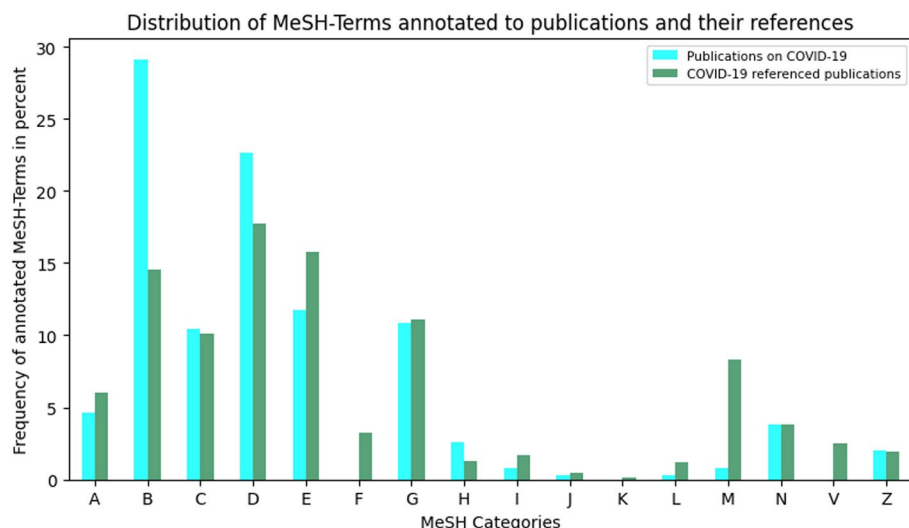
Additionally, we contrasted the differences with a random sample of publications and their references, which we call control datasets. For these control datasets, we sampled 25,000 papers from the database ZB MED Knowledge Environment that appeared in 2020 and had MeSH terms. These publications were randomly selected from the ZB MED Knowledge Environment database. Afterward, we extracted the referenced works from the CrossRef database. We repeated this procedure 100 times.

### Results

The results show that SARS-CoV-2 publications use knowledge from slightly differently oriented life science fields. In Fig. 6, the assigned MeSH terms of primary articles published on SARS-CoV-2 and their references are allocated to MeSH categories. On average, a lower share of MeSH terms is allocated to the referenced publications than to the primary articles and preprints. One reason for this could be that much of the literature cited is not in the field of life sciences. These articles from other fields may not carry MeSH terms. This explanation is further examined in Appendix B by the subject area of the journals in which the articles had been published.

A slight shift in topics can be observed in Fig. 6. The shift is indicated by the categories of MeSH terms assigned to primary publications and preprints compared to those categories of MeSH terms assigned to the cited papers. This way, we identify the subject areas on which the SARS-CoV-2 papers build. Both primary articles and referenced articles share a focus on topics that are annotated with MeSH concepts from categories Organism (B), Diseases (C), Chemicals and Drugs (D), Analytical, Diagnostic and Therapeutic, and Equipment (E), and Phenomena and Process (G). Both publication types correspond to the low numbers of annotations from the fields: Disciplines and Occupations (H), Anthropology, Education, Sociology, and Social Phenomena (I), as well as Technology, Industry, and Agriculture (J).

For categories Phenomena and Process (G) and Health Care (N), a more or less equal distribution can be observed. In contrast, in the categories Analytical, Diagnostic and Therapeutic, and Equipment (E), Psychiatry and Psychology (F), Named Groups (M), and Publication Characteristics (V), we observe a surplus of MeSH terms annotated to references over primary literature. A clear preponderance in the distribution of MeSH

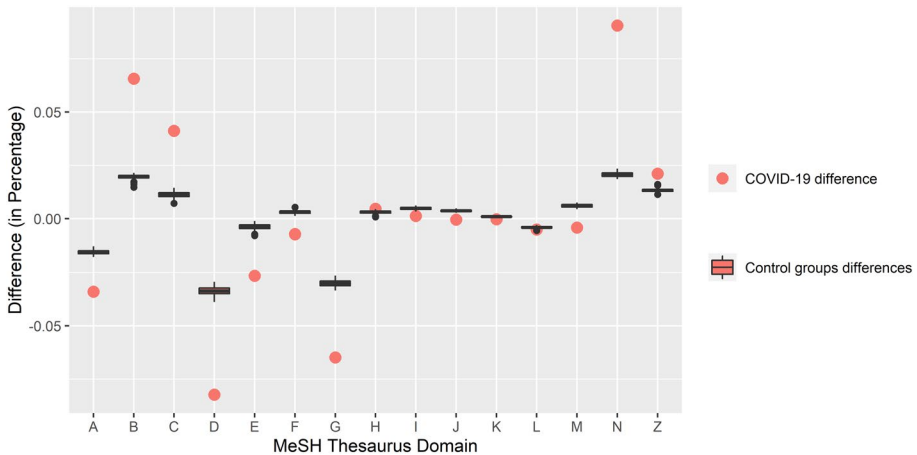


**Fig. 6** Shift of topics. Distribution of MeSH terms related to primary publications on COVID-19 and the cited work. Legend MeSH research fields see Appendix A.3

terms to primary papers can be observed in the categories Organism (B) and Chemicals and Drugs (D).

The biggest differences between subject fields of primary publication and references were determined in categories Organism (B; primary literature: 29.11%, references: 14.57%), Chemicals and Drugs (D; primary literature: 22.66%, references: 17.71%), Analytical, Diagnostic, and Therapeutic Techniques, and Equipment (E; primary literature: 11.79%, references: 15.78%), and Psychiatry and Psychology (F, primary literature: 0%, references: 3.28%), Named Groups (M; primary literature: 0.79%, references: 8.3%), and Publication characteristics (V; primary literature: 0%, references: 2.54%). The high value for categories Organism (B) and Chemicals and Drugs (D) reflect the intensive search for a vaccine for this phase of research and publishing. The knowledge used for this research comes mainly from these fields and, in addition, from the fields that deal with therapeutic techniques (E), as well as with certain population groups (M). With SARS-CoV-2, the named groups discussed here include obese people, the elderly, and men as particular risk groups. The named groups of people, like the content on psychology, are far more prevalent in the received works than in the publications on SARS-CoV-2 itself. This shows that conditional interdisciplinary exploitation of knowledge is taking place here.

Thus, our results show definite discrepancies between the primary publications on SARS-CoV-2 and the references of these publications in particular MeSH categories. The surplus of cited works in topic areas A, E, F, M, and V shows that publications on SARS-CoV-2 exploit information from these fields. The knowledge is used to research COVID-19, mainly in Organism (B), and Chemicals and Drugs (D). This implies that scholars also applied research results from categories other than their research field, which is one of the signs of interdisciplinarity. The results of further analysis of primary publications and their references are presented in Appendix B.



**Fig. 7** Differences of distributions of MeSH descriptors over the 15 MeSH thesaurus domains between primary publications (citing) and references (cited) of the COVID-19 data and 100 control groups. The results of control groups are displayed as boxplots. COVID-19 difference is shown as a dot

Regarding the comparison with control datasets, Fig. 7 shows that differences in the distributions between primary publications and references of 100 control groups are in the range of  $\pm 3\%$  in all MeSH domains. The respective differences of distributions between primary publications and references of the COVID-19 dataset have a range of  $\pm 9\%$ . The conducted one-way two-tailed *t*-test confirms that different values of the COVID-19 dataset are significantly different from the control group data.

## Network analysis

Since our presented results slightly revealed particular characteristics of scientific collaboration or at least citation of knowledge from other research fields, we aimed to determine specific factors that can have an impact on the interdisciplinarity of COVID-19 research. Thus, we investigated whether and how social network characteristics affect the interdisciplinarity of authors of publications at the individual level.

## Methods

In our dataset of COVID-19 journal articles and preprints, we were able to identify particular authors unambiguously with the help of their ORCIDs. In order to focus on the authors who rapidly responded to the COVID-19 crisis in their research, we included only those authors that have published at least two or more journal articles or preprints. Our dataset contained 6,283 authors in total. For these authors, we identified all their COVID-19 journal articles and preprints and their co-authors in these journal articles and preprints. This data enabled us to create a scientific social network for the focal authors. In our scientific social network, authors represent nodes and joint publications (journal articles or preprints) are edges.

**Table 3** Descriptive statistics

Variables	Mean	Std. dev.	Min.	Max.
Level of interdisciplinarity	0.51	0.07	0.12	0.69
Betweenness centrality	72.1	255	0	2080.7
Degree centrality	5.5	5.6	0	27
Number of COVID-19 publications	1.9	0.9	1	5
Number of MeSH domains	6.6	1.3	4	9
Number of co-authors	3.9	4.3	0	19

n = 4670

**Table 4** Correlation matrix

Variables	1	2	3	4	5	6
1 Level of interdisciplinarity	1					
2 Betweenness centrality	0.01	1				
3 Degree centrality	0.14***	0.34***	1			
4 Number of publications	−0.10***	0.13***	0.14***	1		
5 Number of MeSH domains	0.46***	0.02	0.05**	−0.02	1	
6 Number of co-authors	0.18***	0.14***	0.88***	−0.04**	0.04***	1

\*\*\*p &lt; .001; \*\*p &lt; .01, \*p &lt; .05; n = 4670

Each edge implies the knowledge exchange and information linkage between two authors (Aboelela et al., 2007). Having the information on the nodes' connectedness over the available edges, we calculated the degree centrality and betweenness centrality for each author (node) according to Freeman (1979). Afterward, for each author of the scientific social network, we identified all MeSH descriptors of their respective publications. To measure an author's level of interdisciplinarity, we calculated the mean cosine distance between all MeSH descriptors of all author's COVID-19 publications and preprints. The level of interdisciplinarity of an author is our dependent variable. Degree centrality and betweenness centrality are independent variables. Then we calculated an average number of different MeSH domains per publication and per author to control for the diversity of the focal author's research. For the estimated statistic model, we also determined the number of total COVID-19 publications of an author as well as the number of co-authors per publication at the author level as control variables. The resulting dataset contained 4,814 observations. Since the data set contained publications with too many authors, whose contribution to the paper might be limited, we excluded observations with extreme values under the 1% percentile and over the 99% percentile for the dependent variable and all explanatory and control variables. Thus, the analyzed data set contained 4,760 observations.

*Estimations.* Since our dependent variable is a continuous nonnegative variable, we used ordinary least squares (OLS) estimations. After the specification of the model, we found that an assumption on heteroscedasticity of residuals was not satisfied according to the Breusch-Pagan test (Breusch & Pagan, 1979). Hence, we used the Huber-White sandwich estimator to obtain heteroscedasticity-consistent standard errors (Wooldridge, 2010).

Descriptive statistics and correlations for the variables appear in Tables 3 and 4. Due to the high correlation of the variables' degree centrality and number of co-authors per

**Table 5** OLS results for the level of interdisciplinarity of authors

Regressors	Coefficients	Std. Errors	p-value
Intercept	0.364	0.006	$< 2 \times 10^{-16}***$
Betweenness centrality	$-8.28 \times 10^{-06}$	$3.34 \times 10^{-06}$	0.013*
Degree centrality	$1.72 \times 10^{-03}$	$1.39 \times 10^{-04}$	$< 2 \times 10^{-16}***$
# COVID-19 publications	$-8.27 \times 10^{-03}$	$9.75 \times 10^{-04}$	$< 2 \times 10^{-16}***$
# MeSH domains	$2.38 \times 10^{-02}$	$7.23 \times 10^{-04}$	$< 2 \times 10^{-16}***$

Robust standard errors; \*\*\* $p < .001$ ; \*\* $p < .01$ , \* $p < .05$ ; F-statistic: 361.3 (4; 4,786),  $p < 2.2 \times 10^{-16}***$ ;  $n = 4,670$

publication (0.88) (Table 4) as well as high VIF-values of these variables (over 6) (Allison, 1999), we excluded the number of co-authors per publication from the model estimation. Correlation values among other variables are low. The VIF-values for other variables are under 1.2 and don't reveal any multicollinearity issues (Allison, 1999).

## Results

The results of OLS estimations are displayed in Table 5. The results reveal that the position of the author in the scientific social network of COVID-19 research indeed determines the level of the authors' interdisciplinarity. The betweenness centrality determines the potential control of knowledge flows and communication within the network (Freeman, 1979). Thus, authors with a central position can control the information that they obtain from different and possibly not connected parts of the network. However, even though this information can provide valuable insights for the authors, our results (Table 5) suggest that these insights are difficult to internalize completely in the short term since the coefficient of betweenness centrality is negative and significant ( $\beta = -8.28 \cdot 10^{-6}$ ,  $p < 0.05$ ), if holding the degree centrality constant. Thus, these authors cannot profit from the betweenness centrality position in terms of increasing the level of interdisciplinarity of their publications. On the other hand, OLS estimation results show that the degree centrality in the scientific social network is beneficial for authors if they want to increase their level of interdisciplinarity (Table 5), since degree centrality is positive significant ( $\beta = 1.72 \cdot 10^{-3}$ ,  $p < 0.001$ ). The central position in the network with regard to the degree of centrality is an indicator of the enhanced communication activity and direct interactions of the focal author with other authors in the network. Through efficient, fast, and direct linkages (edges) to their peers, the focal authors gain valuable knowledge of their network colleagues at a lower cost in terms of time and access to specific information. This knowledge and information might come from different research fields and through the direct communication that the focal authors have with their peers, the authors can absorb this knowledge easier even if the knowledge is outside their subject area.

Further, the number of publications is disadvantageous for the promotion of the authors' level of interdisciplinarity (Table 5,  $\beta = -8.27 \times 10^{-3}$ ,  $p < 0.001$ ). Authors who publish many articles might be very specialized and more efficient exclusively in their subject. In contrast, the higher number of research fields, measured as a number of MeSH domains per author, raises the level of interdisciplinarity (Table 5,  $\beta = 2.38 \times 10^{-2}$ ,  $p < 0.001$ ).



## Discussion

We comprehensively analyzed the dynamics of the scientific response to SARS-CoV-2 and COVID-19 with a focus on research interdisciplinarity and topic adjustment effects. For this purpose, we investigated the total research volume on SARS-CoV-2 and COVID-19, MeSH thesaurus descriptors as research topics, MeSH thesaurus domains as well as authors' networks with regard to research interdisciplinarity and increasing focus on COVID-19 topics at cost of activities in other established topics. We ascertained that the research on novel coronavirus and COVID-19 has drastically risen in the first year of the pandemic and has partially displaced other research topics. However, we did not find a complete displacement of particular research topics by the research on COVID-19. Consequently, our study suggests that crisis-driven research, such as research on SARS-CoV-2 and COVID-19, occupies valuable research resources – which is in line with the finding by Coccia (2021) but does not suppress other research topics completely, which scientists usually elaborate on in times without extreme emergencies.

In consent with the study by Zhao et al. (2022), we found that the COVID-19-related research was only to some extent interdisciplinary. The finding that interdisciplinarity did not increase during the first year of the COVID-19 pandemic therefore stands in contrast to the expectation and the benefits declared by scientists in the large-scale Delphi study (Lazarus et al., 2022) or the smaller study by Kastenhofer (Kastenhofer et al., 2023) of a cooperation across scientific boundaries. Similar to our results, Coccia (2021) concluded that crisis-driven research is less interdisciplinary. In pursuit of an explanation, we have conducted an additional analysis (see Appendix B) to confirm that the pandemic-induced research is indeed only partially interdisciplinary *because* COVID-19 related research is dominated by only a few disciplines (Clinical Medicine, Immunology, Molecular Biology, and Genetics as well as Microbiology) – in line with Zhao et al. (2022). Har-santo (2020) also concluded that Medicine was the prevailing scientific field in the first months of the pandemic.

Additionally, we examined the factors that facilitate interdisciplinary collaborations. Although attempts to analyze the research development on SARS-CoV-2 and COVID-19 by the usage of MeSH descriptors were made (Stegmann, 2020), our paper provides a deeper explication of knowledge development on these topics. In particular, beyond the confirmation of the prior research (Stegmann, 2020; Colavizza et al., 2020; Haghani & Bliemer, 2020) that ascertained the rapid growth of scientific knowledge on SARS-CoV-2 and COVID-19, we investigated which patterns (scientific convergence or divergence) were prevailing in the research on COVID-19 and whether researchers from different scientific fields have been working together to cope with the global pandemic, implying an increasing level of research interdisciplinarity. For this purpose, we have applied our developed method for measuring research dynamics based on machine learning (Galke et al., 2019; Perozzi et al., 2014; Grover & Leskovec, 2016).

First, our analysis shows that due to the dramatic impact of the worldwide spread of COVID-19 and the related disease, scientific research on these topics has rapidly grown by exploiting the strengths of researchers from multiple research fields. This finding is in line with the study by Shan et al. (2020), which suggests that medical journals preferred to publish COVID-19-relevant research compared to other topics. Thus, owing to the emergency of the COVID-19 pandemic and journal preference for COVID-19 topics, many scientists contributed to COVID-19-related matters in their fields, potentially at the expense of their main research interests and topics (assuming limited research capacity).

Notably, the journal preference for COVID-19 topics could be regarded as a bias to our bibliographic analysis. However, the preference for COVID-19 topics is deeply intertwined with the crisis that needed to be addressed, such that we consider the effect as a *part* of the scientific response to the global pandemic – leading to a temporary displacement of topics. In particular, we also expect that there will be bounce-back effects in a way that the share of COVID-19 publications will be again decreasing and other topics will be increasing their share in publications. In our analysis on comparing COVID-19 with other medical emergencies, we have observed an annealing of the publication share of the emergencies' topics. For instance, even though the research on HIV-1 and related AIDS topics has decreased in the analyzed period, the actual disease and 38 million people across the globe who lived with HIV-1/AIDS in 2019 (according to the Joint United Nations Program on HIV and AIDS (UNAIDS, 2021)), have not disappeared with the emergence of the novel coronavirus.

Second, MeSH descriptors provide an accurate evaluation of an emerging scientific topic “COVID-19”, even before the descriptor “COVID-19” is officially introduced into the classification system. The most frequent MeSH descriptors allocated to their domains or subdomains at the higher level provide a fast and precise overview of the topics related to COVID-19. Ranging from medical and biochemical research domains to social, psychological, and behavioral topics to medical device and engineering scientific domains, MeSH descriptors in the respective MeSH thesaurus subdomains reproduce all facets of the COVID-19 pandemic. These insights are in line with the prior literature on the research and development on coronavirus, investigating the distribution of papers on coronavirus over research categories (Haghani & Bliemer, 2020). Since our results overlap with the findings from the literature, we conclude that our developed measurement of research dynamics based on the topic (concept) similarity is a valid method to determine scientific evolution trajectories and to measure the degree of interdisciplinary of research. Our results suggest that at the beginning of the pandemic, scholars put multiple efforts to comprehensively analyze diseases caused by SARS-CoV-2 to quickly respond with medication approaches. The SARS-CoV-2 infection triggers except mild symptoms and respiratory problems such as sore throat, fatigue, dry cough, and high fever, also much more serious complications and diseases such as severe pneumonia, kidney failure, cardiac injury, and even death (Huang et al., 2020). For people having other diseases such as heart diseases, lung diseases, cancer, or diabetes, the risk for complications and severe symptoms induced by the SARS-CoV-2 infection is higher (Ali & Alharbi, 2020). Our findings show that in the first period (January–March 2020), researchers have worked within their field of research, which led to the agglomeration of specific discipline knowledge (specialization of knowledge in a particular medical indication). Thereby, researchers may profit from their prior experiences in related scientific topics like other viruses, as also indicated by Laufs et al. 2024 for the development of SARS-CoV-2 vaccines and medications (Laufs et al., 2024). This specialization resulted in a divergence pattern of research dynamics reflected in higher cosine distance between the MeSH subdomain Infections (C01) and other MeSH subdomains since multiple organ systems were subjected to the SARS-CoV-2 infection and had to be profoundly investigated. However, after the identification of major diseases and critical medical conditions, scientists started to apply the knowledge, research methods, and techniques of their colleagues from other fields of medical research. The bundling of knowledge has helped scientists to make advancements beyond the boundaries of their field of study. Moreover, scientists have started employing the methods and techniques of their colleagues from other fields, which resulted in the decrease of normalized cosine distance between the MeSH subdomain Investigative Techniques (E05) and Information Science

(L01) or Social Sciences (I01) in our study. Thus, we observe patterns of interdisciplinary research. Such dependence on knowledge from other scientific fields underlines the importance of joint interdisciplinary research in times of crises and emergencies.

Third, our analysis of references in COVID-19 publications provides further insights into research dynamics. Although clinical medicine research has been most cited by COVID-19 scholars, also the knowledge from other MeSH categories than Diseases (C) has constituted the research basis for medical publications. Insights from the research fields of microbiology, immunology, molecular biology, and genetics were important for research on the COVID-19 as well as on vaccines and medicine against the SARS-CoV-2 infection. Furthermore, the research on health care (MeSH category N) builds on publications from diverse MeSH categories.

Fourth, the author network analysis reveals that the network position of authors determines whether they can contribute to interdisciplinary research, and, as such, to interdisciplinary inventions and innovations. Our results are in line with prior findings (Aboelela et al., 2007). Aboelela et al. (2007) found that researchers working in interdisciplinary research centers having a high degree and betweenness centrality in their local research networks are more productive since they obtain more information from their colleagues and can collaborate more efficiently. However, our results show that only a high degree of centrality position in the collaborative network positively affects the level of research interdisciplinarity, suggesting that authors who can make use of their interdisciplinary scientific network profit in terms of a higher degree of interdisciplinarity of COVID-19 research. This might be a positive effect of short communication between authors in the scientific networks, which is characteristic of the dynamics of science under crisis (Coccia, 2021). In contrast to the degree centrality, the authors' betweenness centrality position hinders the high level of interdisciplinarity, implying that though authors gain diverse knowledge from eventually dispersed unconnected subnetworks, they are not able to rapidly internalize the obtained knowledge and exploit it for interdisciplinary research.

We have conducted additional analyses to trace the research on critical vaccine development (see Appendix C). Our results suggest that conventional viral vector-based vaccines dominated the research on the vaccine against SARS-CoV-2 compared to novel vaccine development approaches (i.e. mRNA-based vaccines). Our results indicate that novel mRNA-based vaccine technologies have reached the late development stage faster compared to conventional vector-based vaccines, which is in line with prior literature (Li et al., 2021).

Our study has several limitations. First, we use MeSH descriptors as research topics, which can be broadly or fine-grained defined. Also, we make use of preprints that contained the most current knowledge and the latest research results on SARS-CoV-2 in the first year of the pandemic. However, since preprints do not undergo a strict review process, they are not published in journals at the initial stage and as such are not indexed with MeSH descriptors. Though we have applied the dictionary lookup-based annotation tool ConceptMapper to index preprints, which provides comparable results with the *MeSH on Demand* system, human-conducted indexing might have differences from our machine-conducted indexing. However, we have conducted a comparison between manual annotation and our ConceptMapper-based annotation in previous work (Galke et al., 2021). Second, our temporal embedding similarly measure provides quantitative insights about research dynamics between two or more research

topics. However, our method does not provide any insights into how interdisciplinary research alliances have been formed or how scientists have chosen their collaborative partners from other research fields. These issues can be explored by future research.

## Conclusion

To summarize, compared with other medical emergencies, the volume of publications related to the COVID-19 pandemic has developed unusually strongly. The publication volume is most comparable with the volume of publications related to HIV emerging in the early 1980s. However, the crisis-driven research on SARS-CoV-2/COVID-19 did not displace other research topics (problem-driven research) completely. For the first year of the COVID-19 pandemic, the initial hypothesis that different scientific disciplines would increasingly collaborate on an interdisciplinary basis was not confirmed. Nevertheless, our results show that particular research topics rely on knowledge, methods, and techniques from other research fields, which is an indicator of interdisciplinary research. Our findings suggest that clinical medicine accompanied by immunology, microbiology as well as molecular biology, and genetics constitute the foundation for SARS-CoV-2/COVID-19 research. These disciplines predominantly cite each other, revealing further limited indications of interdisciplinary research. Further, our results suggest that researchers having a high degree centrality in scientific networks can develop a higher degree of interdisciplinarity.

In total, our results on research dynamics in the first COVID-19 pandemic year recommend research policymakers:

- To foster research on the emerging crisis topic, but to leave enough room for problem-driven research.
- To facilitate interdisciplinary collaborations of researchers from different scientific fields.
- To encourage researchers to produce less, but more high-quality research on the emerging topic, since a high quantity of research publications hinders interdisciplinary research.
- To develop scientific community networks to enable fast and efficient communication opportunities for researchers in order to promote scientists' degree centrality in such networks.
- To encourage the development of novel medical technologies rather than rely solely on conventional technologies since novel technologies lead to breakthrough innovations helping to overcome the crisis.

## Appendix

### Dataset description

#### Data basis

Our investigation is based on an expanded data set COVID-19++ combining scholarly publications and preprints as well as its cited works (Galke et al., 2021). The primary data are scholarly publications on COVID-19 annotated with MeSH terms of the Medical Subject Heading, a vocabulary provided by the National Library of Medicine. All primary publications that we analyzed had been published between January 1 and December 31, 2020.

Compiling the primary data set, we had to deal with difficulty: Since the new coronavirus was named in February 2020 as SARS-CoV-2 by the World Health Organization, the MeSH classification of the year 2020 did not contain a specific term. It was first introduced in the edition in 2021. Publications on SARS-CoV-2 had to be determined by other terms until the exact term was introduced.

Accordingly, the set of primary publications was derived mainly from the database ZB MED Knowledge Environment<sup>6</sup>. Those papers had been annotated with the “COVID” keyword provided by ZB MED Information Centre for Life Sciences. The keyword “COVID” was assigned to scholarly data referring to the content of Semantic Scholar “COVID-19 Open Research Dataset” (CORD-19). It was supplemented by publications listed in Bielefeld Academic Search Engine (BASE) determined by keywords “covid-19”, “SARS-CoV-2”, “covid19”. The annotation was performed by bioinformaticians. Additionally, preprints had been made available by the ZB MED Knowledge Environment and tagged with “COVID”.

Since the relevant research on COVID-19 has developed quite rapidly, scientific preprints had been included as the second source of the data set. The preprint set contains more than 13k articles from bioRxiv, medRxiv, chemRxiv, and preprints.org and was fetched from the semantic search engine *preVIEW COVID-19* (Langnickel et al., 2021a, 2021b, 2022).

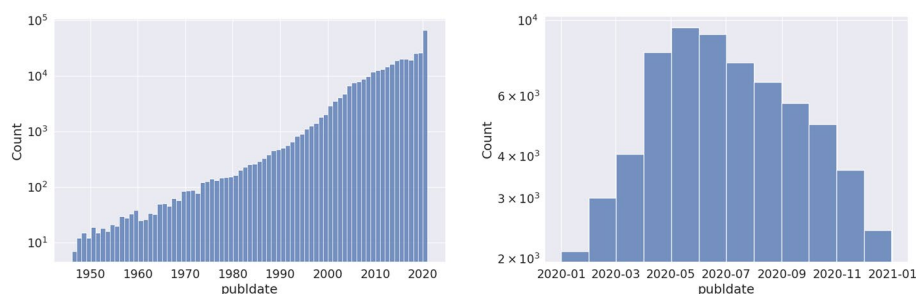
Two difficulties go along with the decision to include preprints: With regard to the possible interdisciplinarity of research on SARS-CoV-2: while the life sciences produce preprints in a relatively short period of time which gain high relevance in the discussion, the text genre of preprints is very much underrepresented in other scientific fields such as Social Sciences and Humanities. Here, we also find much longer periods in the publishing process. Although, the analysis is also open to interdisciplinary collaborations.

The second difficulty in introducing preprints is the absence of official MeSH terms because they are not fully published in a journal. The subsequent annotation with MeSH terms is needed for preparing the data for our analysis. As our analysis runs on MeSH terms, we performed an annotation approach based on string matching (see Paragraph A.2).

<sup>6</sup> The ZB MED Knowledge Environment is a heterogeneous database environment containing over 70 million records from over 70 specialized databases in the life sciences field. See <https://www.zbmed.de/en/research/completed-projects/zb-med-knowledge-environment/>.

**Table 6** COVID-19++ dataset characteristics

Characteristic	Value
# Original Publications	47,242
# Added Preprints	11,154
# Cited Publications	279,027
# Total Publications	337,423
# Citations within Dataset	906,110
# Authorships	1,421,320
# Unique Authors	367,743
# Concept Annotations	5,255,479
# Unique Concepts	13,253
Avg. #Author per Publ.	4.2
Avg. #Concepts per Publ.	15.6



**Fig. 8** Non-cumulative count of publications (log-scale) by publication date in our COVID-19++ dataset (*Left*). Non-cumulative count of publications (log-scale) since 2020 by publication date in our COVID-19++ dataset (*Right*)

Our final data set COVID19++ consists of more than 337k publications made up of more than 47k primary publications dealing with SARS-CoV-2 (Galke et al., 2021) (see Table 6). Considering the investigation on interdisciplinary collaborations, we also compiled a third data set on the cited works as a resource. Querying the open CrossRef Database for citation, we had been able to identify more than 279k publications that influence the primary sources on SARS-CoV-2. The average of references is 15.6 citations for each publication or preprint of the primary data set. The referenced works represent about 86% of the COVID++ data set. On average, the publications in the COVID++ dataset are annotated by seven MeSH terms each. In total, more than 1421k authorships had been identified. 25.8% of the authors can be distinguished by an ORCID ID.

Table 6 shows the characteristics of the resulting dataset. Figure 8 shows the count of publications by month within our COVID-19++ dataset. Note that publications before 2020 are included when they have been cited by publications on COVID-19.

## Retrieval of preprints and annotation with MeSH

As shown in Galke et al. (2021), we fetched preprints from five different preprint servers via the application programming interface (API) of *COVID-19 preVIEW*. *COVID-19 preVIEW* is a semantic search engine available under <https://preview.zbmed.de> and hosted by

**Table 7** Sample-based precision, recall, and  $F_1$  score times 100 of ConceptMapper applied to an evaluation set of 1001 MeSH-annotated articles from 2020

#Mesh terms	#Docs	Method	Precision	Recall	$F_1$
3,781 (all)	1001	ConceptMapper	46.69	27.62	31.64
3,781 (all)	1001	MeSH on Demand*	60.83	48.03	51.14
3,722 (cleaned)	1001	ConceptMapper	46.69	34.00	35.43
3,722 (cleaned)	1001	MeSH on Demand*	59.77	48.87	49.59
567 (diseases)	550	ConceptMapper	62.45	42.41	38.71
567 (diseases)	550	MeSH on Demand*	78.64	61.61	57.97

\*The results for MeSH on Demand are listed for comparison, even though the MeSH on Demand system has had access to our test documents (as observed by seeing the original document as a top-1 similar article to its title and abstract)

**Table 8** Sample-based precision, recall, and  $F_1$  score times 100 of ConceptMapper applied to an evaluation set of 100 COVID-19-related preprints

#Mesh terms	#Docs	Method	Precision	Recall	$F_1$
648 (all)	99	ConceptMapper	82.89	55.21	65.11
601 (cleaned)	99	ConceptMapper	82.89	63.31	71.69
80 (diseases)	64	ConceptMapper	85.74	38.90	38.86

As there is no ground truth for these articles, we use the predictions of MeSH On Demand as a reference to evaluate ConceptMapper

ZB MED (Langnickel et al., 2021a, b, 2022). The preprints are not yet published and are also not indexed with MeSH terms, which is why we indexed them automatically. For this reason, we used the dictionary lookup-based annotation tool ConceptMapper (Tanenblatt et al., 2010), which is based on the Apache Unstructured Information Management Architecture (UIMA) (Ferrucci & Lally, 2004) environment (see Table 7). ConceptMapper is a highly adaptable tool that allows advanced string matching in a reasonable amount of time. We used the implementation of Funk et al. (2014) which is available under <https://github.com/UCDenver-ccp/ccp-nlp-pipelines>. For evaluation reasons, we compared the results with those of the National Library of Medicine service MeSH on Demand<sup>7</sup> which suggests MeSH Terms for a text using the Medical Text Indexer (MTI) program (see Table 8).

## Mesh categories

The MeSH thesaurus which we use for our analyses is organized in a hierarchical tree structure with sixteen main categories (Table 9). All MeSH terms are allocated to these categories by tree numbers. Some of the terms have more than one field they belong to. The subthesauri of the mesh thesaurus are shown in Table 9.

<sup>7</sup> <https://meshb.nlm.nih.gov/MeSHonDemand>.



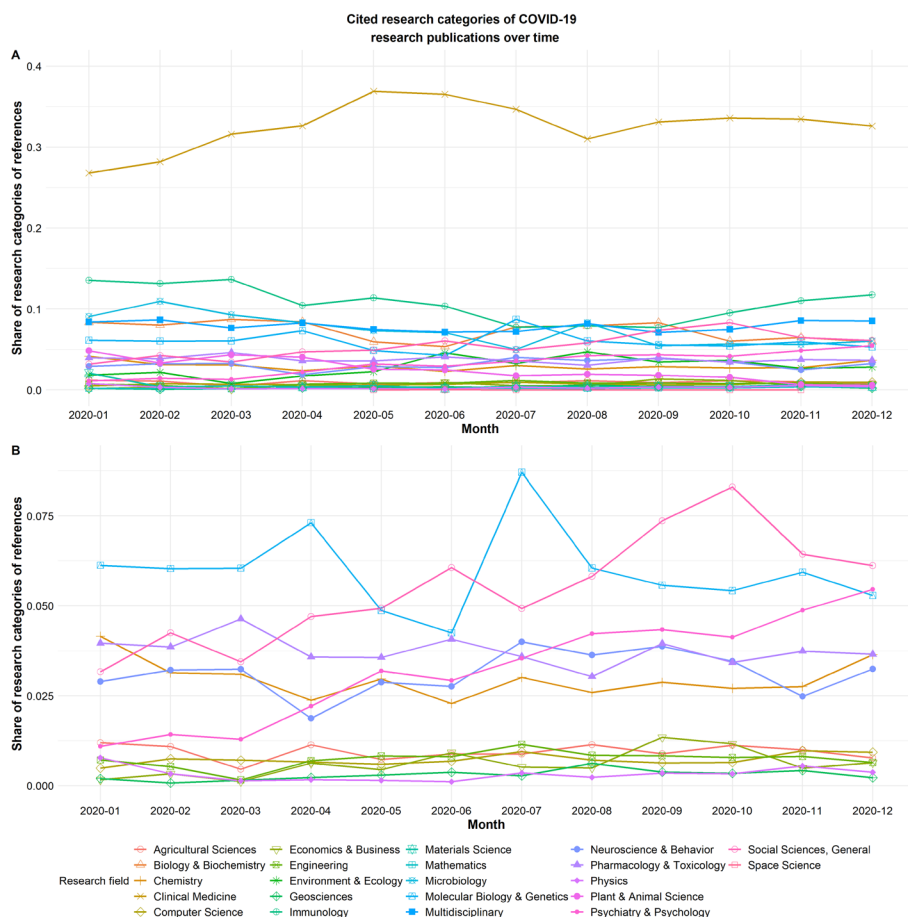
**Table 9** Top-level concepts (subthesauri) of the MeSH hierarchy

Prefix	MeSH category name
A	Anatomy
B	Organism
C	Diseases
D	Chemicals and Drugs
E	Analytical, Diagnostic and Therapeutic Techniques, and Equipment
F	Psychiatry and Psychology
G	Phenomena and Processes
H	Disciplines and Occupations
I	Anthropology, Education, Sociology, and Social Phenomena
J	Technology, Industry, and Agriculture
K	Humanities
L	Information Science
M	Named Groups
N	Health Care
V	Publication Characteristics
Z	Geographicals

## Analysis of references' research fields

To have a deeper insight into whether COVID-19 scientists indeed employed the results of their colleagues from other scientific disciplines or they were reluctant to integrate the results of “foreign” research fields in their own research, we conducted another analysis of the research fields of the references used in the journal articles and preprints in our data set. For this purpose, we analyzed the research fields of referenced journals.

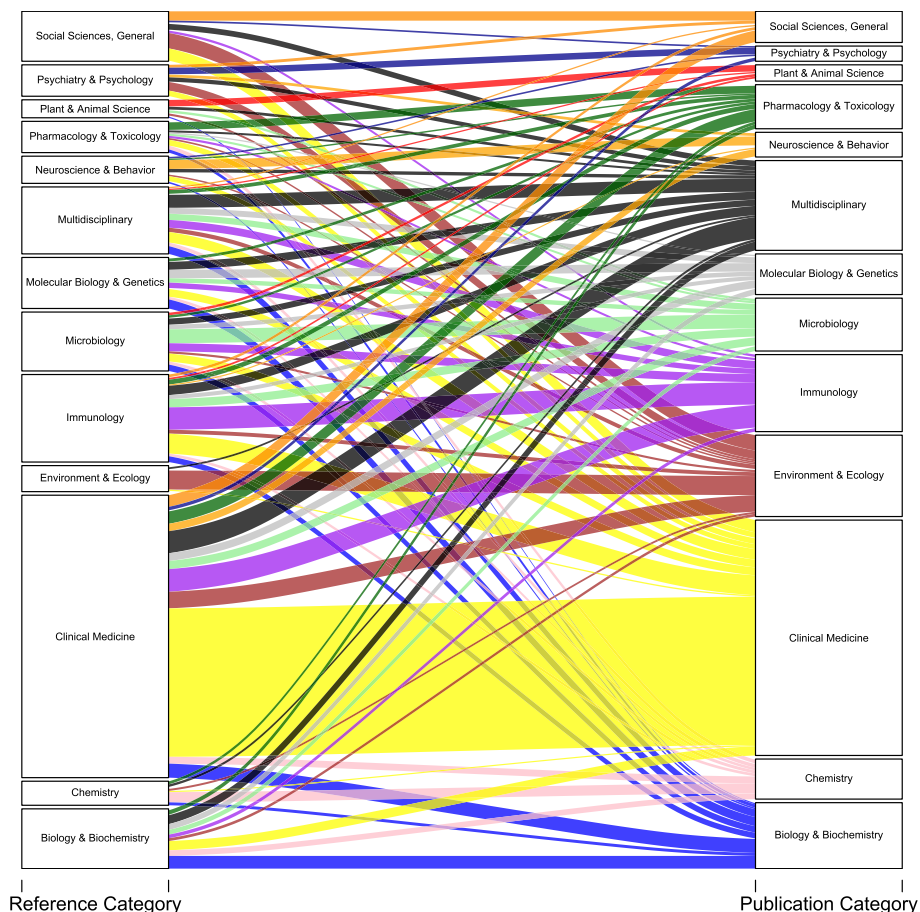
For this investigation, we used 22 research fields that were defined and employed by the Web of Science (WoS) Database (for calculating Essential Science Indicators) for the thematic allocation of research journals. Each scientific journal has a definite unique International Standard Serial Number (ISSN) allocated to one of the 22 research fields according to WoS. Figure 9 shows the temporal analysis results of the usage of knowledge from different WoS research fields. Fig. 9a reveals that during the observation period (January–December 2020), most journal articles on COVID-19 cited prior works in the research field of Clinical Medicine (amber line). The share of Clinical Medicine references raised from 27% in January to approx. 37% in May and then fell to approximately 31% level in August. Owing to the medical relevance of COVID-19, this high share of medical journals cited in the focal COVID-19 research is not surprising. Meanwhile, the shares of natural sciences such as Microbiology, Molecular Biology & Genetics, and Biology & Biochemistry (Fig. 9a) fell in the period of January–June. This insight again confirms the focus on clinical medicine in the research topic COVID-19, since due to the severity of COVID-19, knowledge about similar infectious diseases and therapy approaches was paramount in the first months of the pandemic (Huang et al., 2020). The share of Immunology (green line with circles) decreased from February to September, afterward, it was gradually increasing, suggesting that the research on SARS-COV-2 and COVID-19 was focusing on the human immune responses to the coronavirus and the development of vaccines against it. The share of the field



**Fig. 9** Research fields of references of COVID-19 journal articles

Pharmacology & Toxicology (Fig. 9b, purple line with triangles) reveals a different trajectory: first, the share rose from about 4% in January to almost 5% peak in March, after that, the share was decreasing hardly reaching 3% in August. After a slight increase in September, the share of Pharmacology & Toxicology remained at about 3.8% level. However, having the range of around 2.8% to ca. 4.5%, the share of Pharmacology & Toxicology is considerably lower compared to the 8–13% range of Immunology and 27–37% of Clinical Medicine. Nevertheless, the growing shares of Pharmacology & Toxicology as well as Immunology represent the increasing research on the development of vaccines and effective medicine against SARS-CoV-2.

In addition, Figure 9b shows that the shares of Neuroscience & Behavior rose from around 2.5% in January to almost 4% in July as well as Psychiatry & Psychology was rapidly increasing from approx. 1% in January to almost 6% in December. This is striking evidence for increasing research on the neurological implications of the disease itself and the psychological consequences of COVID-19 protection and control measures taken by a number of governments across the globe. Moreover, the impact of Social Science enormously increased, since its share grew from around 3% in January to almost 8% in October.



**Fig. 10** Knowledge flows between COVID-19 journal articles and their references visualized according to their research fields. Legend: The reference category is the research category of references; Publication Category is the research category of the COVID-19 journal articles

This finding reveals the colossal scope of the COVID-19 pandemic, covering multiple spheres of life, including social and political discourses.

Furthermore, a share of multidisciplinary cited journals (Fig. 9a, blue line with filled squares) decreases from January (ca. 8%) to June (less than 7%). Then it raised to 8% in August and at the end of the year. Thus, interdisciplinary research became less important during the first months of the pandemic.

The dependence of COVID-19 research on different scientific fields can be visualized by the knowledge flows between the COVID-19 journal publications and their references (Fig. 10). Journal publications in the research field of Clinical Medicine were predominantly based on Clinical Medicine prior knowledge. However, they also employed knowledge from the fields of Immunology, Biology & Biochemistry, Molecular Biology & Genetics, Psychiatry & Psychology as well as Social Sciences. The journal publications on Immunology were applying the research results from the following fields: Clinical Medicine, Immunology, Microbiology, Molecular Biology & Genetics as well

as Multidisciplinary Research. The research publications in the field of Pharmacology & Toxicology were grounded on knowledge from Clinical Medicine, Pharmacology & Toxicology, Immunology, Chemistry, Biology & Biochemistry, etc.

In total, our findings are in line with the results from Section 5: Clinical Medicine has been a prevailing research field that COVID-19 scholars build upon. However, clinical researchers have needed to obtain a deep understanding of SARS-CoV-2 composition and the mechanisms of its action in the human body. This is why such research categories as Microbiology, Immunology, Molecular Biology, and Genetics have been playing an important role in the research on SARS-CoV-2 as well. Moreover, due to the global character of the pandemic leading to the global movement and traveling restrictions, and social isolation, such research categories as Psychiatry & Psychology, Social Sciences, Neuroscience & Behavior have gained in importance.

## Research on vaccine development

We analyzed the research dynamics of the research topics related to the development of vaccines against SARS-CoV-2. For this purpose, we identified the MeSH descriptors, which are used for the FDA<sup>8</sup> or EMA<sup>9</sup> approved vaccines against the new coronavirus (*COVID-19 Vaccines* [D000086663]; *ChAdOx1 nCoV-19* (a vector-based vaccine developed by AstraZeneca<sup>10</sup>), *Ad26COVS1* (a vector vaccine developed by Johnson & Johnson<sup>11</sup>), *2019-nCoV Vaccine mRNA-1273* (an mRNA-based vaccine developed by Moderna<sup>12</sup>), and *BNT162 Vaccine* (an mRNA vaccine developed by Pfizer and BioNTech<sup>13</sup>)). For these MeSH descriptors, which were added in 2021 to the MeSH thesaurus, we collected articles published in 2020 in PubMed (the articles were annotated with these MeSH descriptors a posteriori) and extracted other MeSH descriptors in those articles. Since no article annotated with the MeSH descriptor *Ad26COVS1* was found in 2020, we extracted articles published in 2021 with this descriptor to obtain other MeSH descriptors related to the COVID-19 vaccine descriptors. In the collected articles, we extracted MeSH descriptors that were used before the SARS-CoV-2 outbreak in 2020. These vaccine-related MeSH descriptors can show the vaccine development trends even before the introduction of specific MeSH descriptors such as *COVID-19 Vaccines*. We counted the most frequently used MeSH descriptors related to or having a stem “vaccine” for each of the four different vaccines. Then we distinguished between mRNA and vector-based vaccine types for the analysis. We found that the most frequently used MeSH descriptors for mRNA-based vaccine articles were *Viral Vaccines*; *RNA*, *Messenger*, and *Vaccines*, *Synthetic*. The most frequently applied MeSH descriptors for vector-based vaccine articles were *Viral Vaccines*; *Vaccination*; *Immunogenicity*, *Vaccine*, and *Vaccines*.

We plotted the normalized cosine distances of these MeSH descriptors to the descriptor *Spike Glycoprotein*, *Coronavirus*, which plays a crucial role in SARS-CoV-2 attachment to

<sup>8</sup> US. Food and Drug Administration

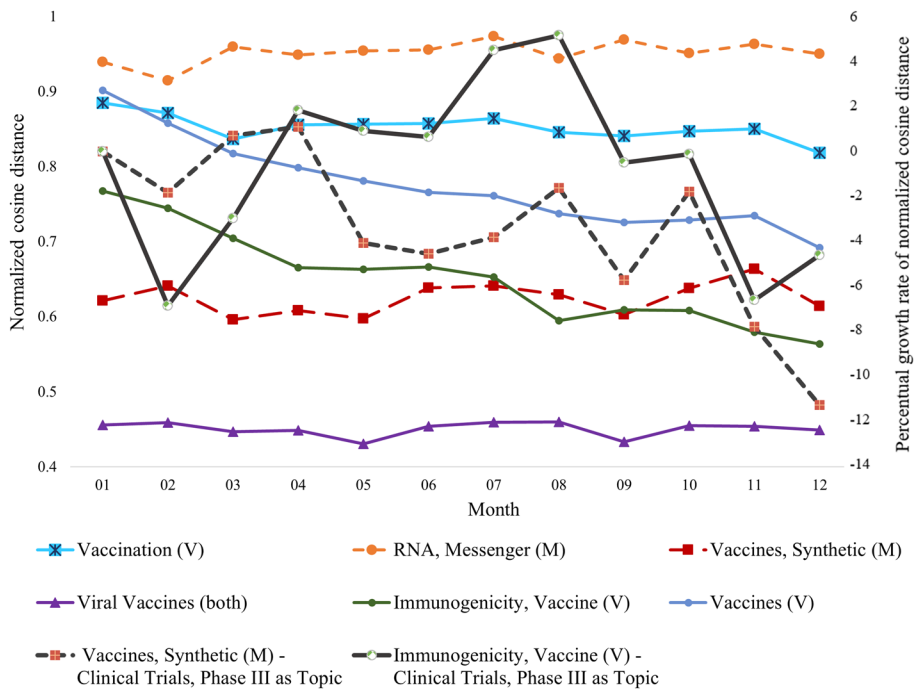
<sup>9</sup> European Medicines Agency

<sup>10</sup> <https://meshb.nlm.nih.gov/record/ui?ui=D000090985>.

<sup>11</sup> <https://meshb.nlm.nih.gov/record/ui?ui=D000090984>.

<sup>12</sup> <https://meshb.nlm.nih.gov/record/ui?ui=D000090983>.

<sup>13</sup> <https://meshb.nlm.nih.gov/record/ui?ui=D000090982>.



**Fig. 11** Left Axis: Normalized cosine distance between selected MeSH descriptors related to vaccine development of four FDA and/or EMA approved vaccines against SARS-CoV-2 and MeSH descriptor *Spike Glycoprotein, Coronavirus*. “M” stands for mRNA-based vaccine development publications; “V” stands for vector-based vaccine development publications; “both” means that the respective MeSH descriptor was identified in both vaccine types publications. Right Axis: Percentual growth rate of normalized cosine distance of MeSH descriptor pairs *Vaccines, Synthetic (M)—Clinical Trials, Phase III as Topic* and *Immunogenicity, Vaccine (V)—Clinical Trials, Phase III as Topic*. The reference point is the value of the normalized cosine distance of January 2020

and intrusion into the host cell and—as such—becomes a target for host immune reactions like neutralizing antibodies (Duan et al., 2020; Sahin et al., 2020; Li et al., 2021) (Fig. 11).

Thus, this Spike glycoprotein is a target for the majority of vaccine development strategies (Duan et al., 2020; Li et al., 2021), and therefore, this MeSH descriptor appropriately qualifies for the exploration of vaccine development trends in the first year of the pandemic. In Fig. 11 (left axis), we observe that a conventional vaccine technology such as vector-based vaccines, in which a viral vector encodes the SARS-CoV-2 Spike protein represented by very common vaccine-related MeSH descriptors *Vaccination*, *Immunogenicity*, *Vaccine*, and *Vaccines* is characterized by the normalized cosine distance ranging from 0.56 to 0.90. The trajectories of cosine distance of pairs *Immunogenicity, Vaccine*, *Vaccines* and *Spike Glycoprotein, Coronavirus* respectively, are almost parallel to each other and show a clear declining trend from January to December 2020. This implies that the research and development of vaccines against SARS-CoV-2 were indeed focused on the SARS-CoV-2 Spike surface protein, after the genome structure of SARS-CoV-2 was identified in January 2020 (Chen et al., 2020) and specific Spike proteins which constitute spikes on the viral surface and facilitate the attachment to host receptors were detected (Chen et al., 2020). In contrast, a more broad and general MeSH descriptor *Vaccination* belonging to multiple

MeSH thesaurus subdomains (Therapeutics (E02), Investigative Techniques (E05), Health Care Facilities, Manpower, and Services (N02) as well as Environment and Public Health (N06)) shows a higher normalized cosine distance ranging from 0.81 to 0.88. Nevertheless, the cosine distance gradually decreases from January to December 2020. The more novel RNA-based vaccine technologies represented by the MeSH descriptors *RNA*, *Messenger* and *Vaccines*, *Synthetic* display rather constant development trajectories of normalized cosine distances (Fig. 11). The novelty of these technologies is especially reflected by the high cosine distance of the MeSH descriptor *RNA*, *Messenger*, ranging from 0.91 to 0.97. This indicates that at the beginning, a new narrowly defined technology has particular difficulties permeating traditional research on vaccines, underlying that these novel vaccines have been developed with rather specialized disciplinary than conventional methods. However, if this new technology is mentioned in the broader context by the usage of the MeSH descriptor *Vaccines*, *Synthetic* having a much lower cosine distance of 0.59–0.66, it has a high potential to permeate the boundaries of conventional vaccines research and to get established as one of the important vaccines research and development topics regarding SARS-CoV-2. Specifically, if mRNA-based vaccines were discussed or compared with other vaccine types—reflected in our study setting by the MeSH descriptor *Viral Vaccines* having the lowest values of cosine distance (0.43–0.46) during the period of investigation (Fig. 11). The novelty of these technologies is especially reflected by the high cosine distance of the MeSH descri— they show also the potential for gaining high acceptance in the vaccine research and development community.

### Vaccine development technologies

Additionally, using MeSH descriptors, we investigated how rapidly two different vaccine technologies (mRNA-based and viral vector-based vaccines) have reached late development stages. For this purpose, we plotted the trajectories of the percentual growth rate of normalized cosine distances of the MeSH descriptors *Vaccines*, *Synthetic* (for mRNA-based vaccines) or *Immunogenicity*, *Vaccine* (for vector-based vaccines), and the MeSH descriptor *Clinical Trials, Phase III as Topic*, indicating a late development stage of the vaccine product development (Banerjee & Siebert, 2017) (Fig. 11, right axis). Figure 11 The novelty of these technologies is especially reflected by the high cosine distance of the MeSH descri shows compared to January 2020 (the reference value for the calculations of percentual growth rate), the growth rate of normalized cosine distance of the MeSH descriptor pair *Vaccines*, *Synthetic*—*Clinical Trials, Phase III as Topic* was declining to reach the value -11.25% in December 2020, whereas that of the MeSH descriptor pair *Immunogenicity*, *Vaccine*—*Clinical Trials, Phase III as Topic* was dropping to -4.6% in December 2020. These results imply the higher pace of decrease of normalized cosine distance for mRNA-based vaccines in comparison to vector-based vaccines. Therefore, mRNA-based vaccines (according to the MeSH descriptor *Vaccines*, *Synthetic*) have been reaching the late development stage (clinical trials phase III) faster than vector-based vaccines (represented by the MeSH descriptor *Immunogenicity*, *Vaccine*). In the future, these technologies might have the potential to at least partially displace conventional vaccine development technologies.

To sum up, the revealed research dynamics of the research topics related to the development of vaccines against SARS-CoV-2 indicates that the research on traditional vaccine development technologies was prevailing in the first year of the COVID-19 pandemic. But the rate of diffusion of novel technologies—such as mRNA-based vaccines—into their



application fields (development of vaccine products in late development stages such as clinical trials phase III) is higher than that of conventional technologies (viral vector-based technologies).

**Author contributions** All authors contributed to the study's conception and design. Material preparation, data collection, and analysis were performed by Eva Seidlmayer, Lukas Galke, Tetyana Melnychuk, and Lisa Kühnel. The first draft of the manuscript was written by Eva Seidlmayer, Lukas Galke, and Tetyana Melnychuk. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Declarations

**Competing interest** The authors have no competing interests to declare.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aboelela, S., Merrill, J., Carley, K. M., & Larson, E. (2007). Social network analysis to evaluate an interdisciplinary research center. *The Journal of Research Administration*, 38, 61–75.
- Ali, I., & Alharbi, O. M. (2020). COVID-19: Disease, management, treatment, and social impact. *Science of The Total Environment*, 728, 138861. <https://doi.org/10.1016/j.scitotenv.2020.138861>
- Allison, P. D. (1999). *Multiple Regression: A Primer*. Thousand Oak: Calif. Pine Forge Press.
- Amey, M.J. & Brown, D. (2006). Breaking out of the box: Interdisciplinary collaboration and faculty work.
- Aviv-Reuven, S., & Rosenfeld, A. (2021). Publication patterns' changes due to the COVID-19 pandemic: A longitudinal and short-term scientometric analysis. *Scientometrics*, 126(8), 6761–6784. <https://doi.org/10.1007/s11192-021-04059-x>
- Banerjee, T., & Siebert, R. (2017). Dynamic impact of uncertainty on R & D cooperation formation and research performance: Evidence from the bio-pharmaceutical industry. *Research Policy*, 46(7), 1255–1271. <https://doi.org/10.1016/j.respol.2017.05.009>
- Benjamins, S., de Meijer, V. E., Pol, R. A., & Haring, M. P. D. (2021). Are all voices heard in the COVID-19 debate? *Scientometrics*, 126(1), 859–862. <https://doi.org/10.1007/s11192-020-03730-z>
- Blackwell, A.F., Wilson, L., Street, A., Boulton, C., & Knell, J. (2009). Radical innovation: crossing knowledge boundaries with interdisciplinary teams (Tech. Rep. No. UCAM-CL-TR-760). University of Cambridge, Computer Laboratory.
- Bonilla-Aldana, D. K., Quintero-Rada, K., Montoya-Posada, J. P., Ramírez-Ocampo, S., Paniz-Mondolfi, A., Rabaan, A. A., & Rodríguez-Morales, A. J. (2020). SARS-CoV, MERS-CoV and now the 2019-novel CoV: Have we investigated enough about coronaviruses? A bibliometric analysis. *Travel Medicine and Infectious Disease*, 33, 101566. <https://doi.org/10.1016/j.tmaid.2020.101566>
- Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47(5), 1287–1294.
- Chen, Y., Liu, Q., & Guo, D. (2020). Emerging coronaviruses: Genome structure, replication, and pathogenesis. *Journal of Medical Virology*, 92(4), 418–423. <https://doi.org/10.1002/jmv.25681>
- Coccia, M. (2021). Evolution and structure of research fields driven by crises and environmental threats: The COVID-19 research. *Scientometrics*, 126(12), 9405–9429. <https://doi.org/10.1007/s11192-021-04172-x>



- Colavizza, G., Costas, R., Traag, V.A., van Eck, N.J., van Leeuwen, T., Waltman, L. (2020). A scientometric overview of COVID-19. *bioRxiv*. <https://doi.org/10.1101/2020.04.20.046144>
- Corsi, M., & Ryan, J. M. (2022). What does the Covid-19 crisis reveal about interdisciplinarity in social sciences? *International Review of Sociology*, 32(1), 1–9. <https://doi.org/10.1080/03906701.2022.2064695>
- Cullen, W., Gulati, G., & Kelly, B. D. (2020). Mental health in the COVID-19 pandemic. *QJM: An International Journal of Medicine*, 113(5), 311–312. <https://doi.org/10.1093/qjmed/hcaa110>
- Curran, C. S., & Leker, C. S. (2009). Employing stn anavist to forecast converging industries. *International Journal of Innovation Management*, 13(4), 637–664. <https://doi.org/10.1142/s13639196090002455>
- Curran, C.-S., & Leker, J. (2011). Patent indicators for monitoring convergence—examples from nff and ict. *Technological Forecasting and Social Change*, 78(2), 256–273. <https://doi.org/10.1016/j.techfore.2010.06.021>
- Dinis-Oliveira, R. J. (2020). COVID-19 research: Pandemic versus paperdemic, integrity, values and risks of the speed science. *Forensic Sciences Research*, 5(2), 174–187. <https://doi.org/10.1080/20961790.2020.1767754>
- Duan, L., Zheng, Q., Zhang, H., Niu, Y., Lou, Y., & Wang, H. (2020). The sars-cov-2 spike glycoprotein biosynthesis, structure, function, and antigenicity: Implications for the design of spike-based vaccine immunogens. *Frontiers in Immunology*, 89, 57662211. <https://doi.org/10.3389/fimmu.2020.576622>
- D'Este, P., Llopis, O., Rentocchini, F., & Yegros, A. (2019). The relationship between interdisciplinarity and distinct modes of university-industry interaction. *Research Policy*, 48(9), 103799. <https://doi.org/10.1016/j.respol.2019.05.008>
- Fassin, Y. (2021). Research on Covid-19: A disruptive phenomenon for bibliometrics. *Scientometrics* 126(6), 5305–5319. Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 6 Publisher: Springer International Publishing <https://doi.org/10.1007/s11192-021-03989-w>
- Ferrucci, D., & Lally, A. (2004). Building an example application with the unstructured information management architecture. *IBM Systems Journal*, 43(3), 455–475. <https://doi.org/10.1147/sj.433.0455>
- Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science*, 47(1), 117–132. <https://doi.org/10.1287/mnsc.47.1.117.10671>
- Freeman, L.C. (1979). Centrality in social networks 1: Conceptual clarification. *Social Networks*.
- Funk, C., Baumgartner, W., Garcia, B., Roeder, C., Bada, M., Cohen, K. B., & Verspoor, K. (2014). Large-scale biomedical concept recognition: an evaluation of current automatic annotators and their parameters. *BMC Bioinformatics*, 15(1), 59. <https://doi.org/10.1186/1471-2105-15-59>
- Galke, L., Mai, F., Vagliano, I., Scherp, A. (2018). Multi-modal adversarial autoencoders for recommendations of citations and subject labels. *UMAP* (pp. 197–205). ACM.
- Galke, L., Melnychuk, T., Seidlmayer, E., Trog, S., Förstner, K.U., Schultz, C., Tochtermann, K. (2019). Inductive learning of concept representations from library-scale bibliographic corpora. *Gi-jahrestagung* (pp. 294, 219–232). GI.
- Galke, L., Seidlmayer, E., Ludemann, G., Langnickel, L., Melnychuk, T., Forstner, K.U., Schultz, C. (2021). COVID-19+4: A citation-aware Covid-19 dataset for the analysis of research dynamics. In 2021 IEEE international conference on big data (big data) (pp. 4350–4355). IEEE. <https://doi.org/10.1109/BigData52589.2021.9671730>
- Goyal, P., Chhetri, S. R., & Canedo, A. (2020). dyngraph2vec: Capturing network dynamics using dynamic graph representation learning. *Knowledge-Based Systems*, 187, 104816. <https://doi.org/10.1016/j.knsys.2019.06.024>
- Goyal, P., Kamra, N., He, X., Liu, Y. (2018). DynGEM: Deep embedding method for dynamic graphs. [arXiv:1805.11273](https://arxiv.org/abs/1805.11273) [cs].
- Grover, A. & Leskovec, J. (2016). node2vec: Scalable feature learning for networks. *KDD* (pp. 855–864). ACM.
- Haghani, M., & Bliemer, M. C. J. (2020). Covid-19 pandemic and the unprecedented mobilisation of scholarly efforts prompted by a health crisis: Scientometric comparisons across sars, mers and 2019-ncov literature. *Scientometrics*, 125(3), 2695–2726. <https://doi.org/10.1007/s11192-020-03706-z>
- Harsanto, B. (2020). The first-three-month review of research on Covid-19: A scientometrics analysis. 2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC) (pp. 1–6). <https://doi.org/10.1109/ICE/ITMC49519.2020.9198316>
- He, J., & Chen, C. (2018). Predictive effects of novelty measured by temporal embeddings on the growth of scientific literature. *Frontiers in Research Metrics and Analytics*, 3, 9. <https://doi.org/10.3389/frma.2018.00009>
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., & Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, china. *The Lancet*, 395(10223), 497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)


- Joshua, V., & Sivaprakasam, S. (2020). Coronavirus: Bibliometric analysis of scientific publications from 1968 to 2020. *Medical journal of the Islamic Republic of Iran*, 34, 64. <https://doi.org/10.34171/mjiri.34.64>
- Jung, Y., Kim, E., & Kim, W. (2021). The scientific and technological interdisciplinary research of government research institutes: Network analysis of the innovation cluster in south korea. *Policy Studies*, 42(2), 132–151. <https://doi.org/10.1080/01442872.2019.1593343>
- Kastenhofer, K., Friesache, H. R., Reich, A., & Capari, L. (2023). (Re-)connecting academia during a sudden, global crisis. *TATuP—Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis*, 32, 17–23. <https://doi.org/10.14512/tatup.32.2.17>
- Kayser, S., Brunner, P., Althaus, K., Dorst, J., & Sheriff, A. (2020). Selective apheresis of c-reactive protein for treatment of indications with elevated crp concentrations. *Journal of Clinical Medicine*, 9, 2947. <https://doi.org/10.3390/jcm9092947>
- Kirby, C. K., Jaimes, P., Lorenz-Reaves, A. R., & Libarkin, J. C. (2019). Development of a measure to evaluate competence perceptions of natural and social science. *PLoS ONE*, 14(1), 1–15. <https://doi.org/10.1371/journal.pone.0209311>
- Langnickel, L., Baum, R., Darms, J., Madan, S., & Fluck, J. (2021). COVID-19 preVIEW: Semantic search to explore COVID-19 research preprints. *Public Health and Informatics*, 89, 78–82. <https://doi.org/10.3233/SHTI210124>
- Langnickel, L., Darm, J., Baum, R., & Fluck, J. (2021). preVIEW: From a fast prototype towards a sustainable semantic search system for central access to COVID-19 preprints. *Journal of EAHIL*, 17, 8–14. <https://doi.org/10.32384/jeahil17484>
- Langnickel, L., Darms, J., Heldt, K., Ducks, D., Fluck, J. (2022). Continuous development of the semantic search engine preVIEW: from COVID-19 to long COVID. Database2022baac048. <https://doi.org/10.1093/database/baac048>
- Laufs, D., Melnychuk, T., & Schultz, C. (2024). Effects of prior knowledge and collaborations on r & d performance in times of urgency: The case of covid-19 vaccine development. *R & D Management/an/a*. <https://doi.org/10.1111/radm.12670>
- Lazarus, J.V., Romero, D., Kopka, C.J., Karim, S.A., Abu-Raddad, L.J., Almeida, G., El-Mohandes, A. (2022). A multinational Delphi consensus to end the COVID-19 public health threat. *Nature* 611(7935)332–345. [2024-02-15]<https://www.nature.com/articles/s41586-022-05398-2> Number: 7935 Publisher: Nature Publishing Group <https://doi.org/10.1038/s41586-022-05398-2>
- Leahey, E., Beckman, C. M., & Stanko, T. L. (2017). Prominent but less productive: The impact of interdisciplinarity on scientists' research. *Administrative Science Quarterly*, 62(1), 105–139. <https://doi.org/10.1177/0001839216665364>
- Leahey, E., Beckman, C. M., & Stanko, T. L. (2017). Prominent but less productive: The impact of interdisciplinarity on scientists' research\*. *Administrative Science Quarterly*, 62(1), 105–139. <https://doi.org/10.1177/0001839216665364>
- Li, Y., Tenchov, R., Smoot, J., Liu, C., Watkins, S., & Zhou, Q. (2021). A comprehensive review of the global efforts on COVID-19 vaccine development. *ACS Central Science*, 7(4), 512–533. <https://doi.org/10.1021/acscentsci.1c00120>
- Liu, M., Bu, Y., Chen, C., Xu, J., Li, D., Leng, Y., & Ding, Y. (2022). Pandemics are catalysts of scientific novelty: Evidence from COVID-19. *Journal of the Association for Information Science and Technology*, 73(8), 1065–1078. <https://doi.org/10.1002/asi.24612>
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.
- Meyer-Krahmer, F. (1997). Systems of innovation : Technologies, institutions, and organizations Systems of innovation: Technologies, institutions, and organizations. In C. Edquist (Ed.), *Science-Based Technologies and Interdisciplinarity* (pp. 298–317). Pinter London.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J. (2013). Distributed representations of words and phrases and their compositionality. NIPS NIPS (pp. 3111–3119).
- Mohamed, K., Rodríguez-Román, E., Rahmani, F., Zhang, H., Ivanovska, M., Makka, S. A., et al. (2020). Borderless collaboration is needed for COVID-19-a disease that knows no borders. *Infection Control & Hospital Epidemiology*, 41(10), 1245–1246. <https://doi.org/10.1017/ice.2020.162>
- Moradian, N., Moallemian, M., Delavari, F., Sedikides, C., Camargo, C. A., Torres, P. J., & Rezaei, N. (2021). Interdisciplinary approaches to COVID-19. In N. Rezaei (Ed.), *Coronavirus disease—COVID-19*. Springer International Publishing.
- Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., & Agha, R. (2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *International Journal of Surgery*, 78, 185–193. <https://doi.org/10.1016/j.ijsu.2020.04.018>

- Perozzi, B., Al-Rfou, R., Skiena, S. (2014). Deepwalk: online learning of social representations. *KDD KDD (701-710)*. ACM.
- Radanliev, P., De Roure, D., & Walton, R. (2020). Data mining and analysis of scientific research data records on Covid-19 mortality, immunity, and vaccine development—in the first wave of the Covid-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(5), 1121–1132. <https://doi.org/10.1016/j.dsx.2020.06.063>
- Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, 82(2), 263–287. <https://doi.org/10.1007/s11192-009-0041-y>
- Raynaud, M., Goutaudier, V., Louis, K., Al-Awadhi, S., Dubourg, Q., & Truchot, A. (2021). Impact of the COVID-19 pandemic on publication dynamics and non-COVID-19 research production. *BMC Medical Research Methodology*, 21(1), 255. <https://doi.org/10.1186/s12874-021-01404-9>
- Rogers, T. F., Zhao, F., Huang, D., Beutler, N., Burns, A., He, W.-t., & Burton, D. R. (2020). Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science*, 369(6506), 956–963. <https://doi.org/10.1126/science.abc7520>
- Ryan, J.M. (2021). 1. Introduction: COVID-19: Global pandemic, societal responses, ideological solutions. COVID-19: Volume I: Global Pandemic, Societal Responses, Ideological Solutions (1st edition Ed, 1). Abingdon
- Sahin, U., Muik, A., Derhovanessian, E., Vogler, I., Kranz, L. M., Vormehr, M., & Türeci, Ö. (2020). COVID-19 vaccine bnt162b1 elicits human antibody and th1 t cell responses. *Nature*, 586(7830), 594–599. <https://doi.org/10.1038/s41586-020-2814-7>
- Schilling, M. A., & Green, E. (2011). Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences. *Research Policy*, 40(10), 1321–1331. <https://doi.org/10.1016/j.respol.2011.06.009>
- National Academy of Sciences, National Academy of Engineering, Institute of Medicine (2005). Facilitating Interdisciplinary Research. Washington, DC The National Academies Press. [https://doi.org/10.17226/11153](https://nap.nationalacademies.org/catalog/11153/facilitating-interdisciplinary-research)
- Shan, J., Ballard, D., & Vinson, D. R. (2020). Publication non grata: The challenge of publishing non-COVID-19 research in the COVID Era. *Cureus*, 12(11), e11403. <https://doi.org/10.7759/cureus.11403>
- Stegmann, J. (2020). Mesh descriptors indicate the knowledge growth in the SARS-CoV-2/COVID-19 pandemic. <https://doi.org/10.48550/ARXIV.2005.06259>
- Tanenblatt, M., Coden, A., Sominsky, I. (2010). The ConceptMapper approach to named entity recognition
- Tijssen, R. J. (1992). A quantitative assessment of interdisciplinary structures in science and technology: Co-classification analysis of energy research. *Research Policy*, 21(1), 27–44. [https://doi.org/10.1016/0048-7333\(92\)90025-Y](https://doi.org/10.1016/0048-7333(92)90025-Y)
- Uddin, S., & Imam, T. (2021). Research interdisciplinarity: STEM versus non-STEM. *Scientometrics*, 126, 603–618.
- UNAIDS. (2021). Global HIV & AIDS statistics—2020 fact sheet. [2022-12-12] <https://erna.redcrossredcrescent.com/en/english-unaids-global-hiv-aids-statistics-2020-fact-sheet>/Joint United Nations Program on HIV and AIDS (UNAIDS)
- van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86), 2579–2605.
- Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., & Crispin, M. (2020). Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*, 369(6501), 330–333. <https://doi.org/10.1126/science.abb9983>
- Watermeyer, R., Crick, T., Knight, C., & Goodall, J. (2021). COVID-19 and digital disruption in uk universities: afflictions and affordances of emergency online migration. *Higher Education*, 81, 623–641. <https://doi.org/10.1007/s10734-020-00561-y>
- Webb, G. I., Lee, L. K., Goethals, B., & Petitjean, F. (2018). Analyzing concept drift and shift from sample data. *Data Mining and Knowledge Discovery*, 32(5), 1179–1199.
- Woo Baidal, J. A., Chang, J., Hulse, E., Turetsky, R., Parkinson, K., & Rausch, J. C. (2020). Zooming toward a telehealth solution for vulnerable children with obesity during coronavirus disease 2019. *Obesity*, 28(7), 1184–1186. <https://doi.org/10.1002/oby.22860>
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data (0262232588)*. The MIT Press.
- Xiang, Y., Zhang, Z., Zeng, C., Hu, Z., Liu, Y., Chen, W., Yang, G. (2020). COVID-19, SARS, MERS and Ebola: A comparative analysis registration of intervention clinical trials preprint. In Review. <https://doi.org/10.21203/rs.3.rs-21604/v1>
- Xu, S. (2015). Balancing the two knowledge dimensions in innovation efforts: An empirical examination among pharmaceutical firms. *Journal of Product Innovation Management*, 32(4), 610–621. <https://doi.org/10.1111/jpim.12234>

- Yegros-Yegros, A., Rafols, I., & D'Este, P. (2015). Does interdisciplinary research lead to higher citation impact? the different effect of proximal and distal interdisciplinarity. *PLoS ONE*, 10(8), 1–21. <https://doi.org/10.1371/journal.pone.0135095>
- Zhao, Y., Liu, L., & Zhang, C. (2022). Is coronavirus-related research becoming more interdisciplinary? A perspective of co-occurrence analysis and diversity measure of scientific articles. *Technological Forecasting and Social Change*, 175, 121344. <https://doi.org/10.1016/j.techfore.2021.121344>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

**Eva Seidlmayer**<sup>1</sup>  · **Tetyana Melnychuk**<sup>3</sup> · **Lukas Galke**<sup>4,5,6</sup> · **Lisa Kühnel**<sup>1</sup> · **Klaus Tochtermann**<sup>5</sup> · **Carsten Schultz**<sup>3</sup> · **Konrad U. Förstner**<sup>1,2</sup>

✉ Eva Seidlmayer  
seidlmayer@zbmed.de

Tetyana Melnychuk  
melnychuk@bwl.uni-kiel.de

Lukas Galke  
galke@imada.sdu.dk

Lisa Kühnel  
kuehnel@zbmed.de

Klaus Tochtermann  
k.tochtermann@zbw.eu

Carsten Schultz  
schultz@bwl.uni-kiel.de

Konrad U. Förstner  
foerstner@zbmed.de

<sup>1</sup> Data Science and Services, ZB MED - Information Centre for Life Sciences, Gleueler Str. 60, 50931 Cologne, Germany

<sup>2</sup> Institut für Informationswissenschaft, TH Köln – University of Applied Sciences, Claudiusstraße 1, 50678 Cologne, Germany

<sup>3</sup> Kiel Institute for Responsible Innovation - Department of Technology Management, Kiel University, Westring 425, 24118 Kiel, Germany

<sup>4</sup> Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, Gelderland, The Netherlands

<sup>5</sup> ZBW - Leibniz Information Centre for Economics, Düsternbrooker Weg 120, 24105 Kiel, Germany

<sup>6</sup> Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark