Information Research - Vol. 30 No. 3 (2025)

From lockdown to limelight: unexpected gains in information seeking and scholarly communication research

Marek Deja

DOI: https://doi.org/10.47989/ir30357871

Abstract

Introduction. This study investigates the impact of COVID-19 on the impact of research in information seeking and scholarly communication.

Method. An analysis of bibliometric data was performed focusing on publications related to information seeking and scholarly communication, using the SciVal tool to extract documents related to the topic cluster published from 2015 to 2022. To establish causality, the C-ARIMA model was applied to allow a comparison between actual post-pandemic research metrics and counterfactual predictions.

Analysis. Analysing key metrics such as the Field-weighted citation impact (FWCI), the study sought to quantify the causal relationship between the pandemic and changes in research output, while accounting for prior trends and seasonal publishing cycles.

Results. The results showed a significant increase in FWCI after the pandemic. This growth was not a random event; the model also demonstrated that this was a time-dependent outcome of the pandemic's influence on research impact.

Conclusion(s). The pandemic provided a substantial boost to the impact of research related to information seeking and scholarly communication. This may be largely driven by an increased need for accessible information, marking a significant shift in academic priorities after the global crisis.

Introduction

The COVID-19 pandemic, while overwhelmingly negative in its impact on society, did highlight certain shifts, particularly in how we engage with digital environments (Shen et al., 2022). The accelerated digitisation of work and life processes stands out as one of the most pronounced changes, impacting both scholars and students (Fernández et al., 2023; Trevisan et al., 2024), although often with mixed results regarding mental and emotional wellbeing (Newson et al., 2024). Research into human information behaviour has shed light on how these pandemic-induced shifts alter the ways individuals seek and process information, as the need for reliable information becomes critical during crises (Montesi, 2021). The increase in the health information seeking reflects this, especially as it pertains to adjustments in daily life and physical health concerns during the pandemic (Kelly & Sharot, 2021; Ni et al., 2024). However, while seeking health information naturally aligns with pandemic-driven needs related to everyday life or physical condition (Kelly & Sharot, 2021), it is not only health -related information that has changed focus. The growing reliance on digital platforms for academic communication and scholarly discourse has also become a prominent post-pandemic theme, with deeper implications for how we approach the dissemination and sharing of knowledge. The literature points to a significant increase in the importance of scholarly communication practices, particularly those that intersect with open access, open science, and other models of academic openness (Frank et al., 2023; Rowley, 2023; Seguya et al., 2023). The processes of information seeking is evolving with the academic community as issues around predatory publishing (Downes, 2023; Oermann et al., 2023), open science (Chakravorty et al., 2022; Okafor et al., 2022; Ross-Hellauer et al., 2022), open data initiatives (Prince, 2023), and the use of open educational resources (Sousa et al., 2023; Tlili et al., 2023) gain prominence.. Switching to open communication models is closely tied to the use of digital resources, the influences driving academic productivity, and broader transformations in higher education, management, and library and information science (Fitzgerald, 2020; Ma, 2023). This expanding focus on digital scholarly communication promises to shape the future of information seeking research, especially in the light of ongoing digital transformation across various academic disciplines.

However, it remains uncertain whether the COVID-19 pandemic and its restrictions had a causal effect on the observed rise of information seeking and scholarly communication as research topics, especially in terms of how the digital shifts arising from lockdowns might have influenced impact metrics. This paper seeks to provide empirical evidence supporting the idea that these research topics gained greater scientific prominence across various fields in the post-pandemic era. To address this uncertainty, it is essential to establish the causal role of the pandemic in amplifying the impact of research on these topics. This study proposes to illustrate this through a bibliometric analysis, employing a novel causal inference framework with the C-ARIMA model (Menchetti et al., 2023) to establish genuine causation, rather than association, based on bibliometric data drawn from the Scopus database.

Literature review

The COVID-19 pandemic reshaped academic research, creating a complex landscape for academic work. As university buildings closed and researchers transitioned to remote work, a dichotomy emerged in academic publishing. Lee and Haupt (2021) documented a global influx of COVID-19 publications, reflecting the rapid mobilisation of the scientific community. However, Püttmann and Thomsen (2024) revealed a simultaneous decline in non-COVID-related research output. However, this was a global trend that did not capture fluctuations in a specific field or research topic.

The review by Huvila and Gorichanaz (2024) partially supports the assumption of this paper: Information seeking has a strong position in information science studies after pandemics. Searching for health information leads to the importance of the topic. The short-and long-term effects of the pandemic on information seeking were analysed, indicating that information seeking

should present a positive trend in field-specific importance (Huvila & Gorichanaz, 2024). Wilson (2018) presented a similar assumption regarding the diffusion of information seeking behaviour research beyond library and information science just a few years before the global crisis related to the pandemic.

The impact of the pandemic on academic productivity varied. Some researchers found unexpected writing time during lockdowns, while others struggled with remote work constraints. Jackman et al. (2022) found that some UK doctoral students and early career researchers reported increased productivity and improved work-life balance. However, widespread disruptions to research activities were observed, particularly in laboratory-dependent disciplines, causing long term problems related to fewer submissions and low quality published articles (Yang & Li, 2024). While we can presume that some topics related to information seeking might emerge and gain importance, there is a separate trend to such research output related to temporal decrease of academic productivity, which varies field-to-field.

Lowering the quantity of output does not necessarily indicate a decrease in the impact of research or the direction of further research, its pure quality, or a lesser importance of publications in some field or topics. Myers et al. (2020) found that scientists in experimental disciplines experienced severe declines in research time, with reductions of 30-40% below pre-pandemic levels, but fields less dependent on physical infrastructure reported minimal disruptions, underscoring the pandemic's uneven impact on academia. Cui, Ding and Zhu (2022) also showed fluctuations in preprints published in different fields. Education saw an increase in scholarly output during the pandemic, while fields such as anthropology, cognitive sciences, and information systems showed a decline.

Another transformation lies in the disruption of international research collaborations. Gao et al. (2021) observed a significant decrease in new co-authorships of non-COVID-19 papers in 2020, suggesting a contraction of global scientific networks. However, in early 2021, scientists exhibited fewer declines in number of collaborations compared to the pre-pandemic state and observed a smaller workload, indicating time-related combat to work-life normality (Gao et al., 2021).

Although Library and Information Science (LIS) students, researchers, and library professionals experienced some initial barriers related to remote communication (Begum & Habiba, 2023; de Wit & Altbach, 2023), there is no indication that the quality of research in field-specific topics, such as information seeking or scholarly communication, was significantly affected. Moreover, the evolving role of digital forms of scholarly communication, like the popularisation of preprints (Fraser et al., 2021), dealing with misinformation, is also closely aligned to predatory publishing (Otike et al., 2022), and the emerging importance of information seeking (Huvila & Gorichanaz, 2024), support the initial assumption that these processes has gained significant relevance as research topic and might gain impact across many fields of science.

While the pandemic's immediate shocks in academia are documented, the aftermath that emerged during the subsequent period of stabilization remain under-examined. This study aims to verify whether pandemic-related restrictions and the digitization of academic work exerted a causal influence on the growing prominence of research in information seeking and scholarly communication. These topics are situated within the context of Library and Information Science (LIS), and their impact is measured by the cross-disciplinary and field-normalized citation metric. Classical statistical inference alone is insufficient for capturing the complexities of this problem, as it primarily addresses correlations without fully exploring counterfactual scenarios (Pearl, 2010). Instead, a framework grounded in causal inference is necessary, one that surpasses simple association-based reasoning and traditional experimental designs. To investigate how pandemic-induced restrictions shaped information seeking research, a causal inference approach is employed to transcend these limitations. Causal learning integrates methodologies from causal

inference and machine learning, extending beyond correlation-based approaches by incorporating intervention-based modelling (Cheng et al., 2022). Unlike standard machine learning, which primarily seeks predictive accuracy, causal learning emphasizes estimating intervention effects and modelling distributional shifts using both statistical assumptions and data-driven learning.

Causal autoregressive integrated moving average (Causal ARIMA), as applied in this study, exemplifies this causal inference approach. While ARIMA is traditionally a statistical forecasting method, Causal ARIMA extends it by embedding intervention modelling within the potential outcomes framework, allowing for counterfactual estimation by projecting the pre-intervention trend forward to construct a synthetic non-intervened scenario (Menchetti et al., 2023; Yao et al., 2021). Unlike traditional regression-based methods, which assume fixed relationships between variables, causal learning in time series explicitly models dynamic changes in dependencies induced by interventions. This enables counterfactual estimation, such as assessing the absence of the pandemic, while isolating treatment effects and accounting for shifts in publication trends or disciplinary emphasis. As causal inference increasingly integrates representation learning and confounder adjustments, it strengthens time-series causal estimation, mitigating spurious correlations and reinforcing causal learning as a distinct methodological paradigm (Yao et al., 2021). Counterfactual estimation is critical in this context because it allows to differentiate the true effects of the pandemic on research dissemination from general publication trends, thereby isolating the causal impact of pandemic-related disruptions on scholarly communication.

Methodology

This study uses a causal inference framework to model counterfactual scenarios and estimate the effects of interventions, following the principles of potential outcome theory, also known as the Rubin causal model (Imbens & Rubin, 2015). In this paradigm, our aim is to assess the impact of the COVID-19 pandemic on the outcome of research behaviour captured in bibliometric data related to our research topic, as defined by Scopus in the topic cluster: Information seeking; scholarly communication; adoption–TC.905. To do this, the observed post-pandemic data was compared with counterfactual data, using a forecasting ARIMA model to predict paper impact metrics in a scenario where the pandemic did not occur. This process allows us to evaluate the causation through predictive modelling and comparison of real world and counterfactual outcomes (Menchetti et al., 2023).

In line with the causal nature of most social science research questions (Pearl, 2010), this study represents a shift from traditional statistical inference to causal reasoning. These new methods aim to uncover the mechanisms behind observed phenomena and address "what if?" questions (Imai, Kim & Want, 2023), such as: ‘What would have happened if COVID-19 restrictions had not occurred?’ Pearl's (2010) distinction between associational and causal concepts is key here: while associational concepts are derived from data distributions (e.g., the rise in information seeking publications post- 2020), causality requires counterfactual thinking. For example, by comparing pre-pandemic data with machine learning models of bibliometric impact, we explore what impact trends might have been without the pandemic, an approach not possible thirty years ago (Pearl, 2000). Although other attempts have been made to bring causal inference methods into bibliometric research (for instance, Lars Wenaas (2021) used Bayesian structural time series to examine the readership effects of journals converting to open access, and, Riccaboni and Verginer (2022) employed a difference-in-differences model to detect shifts in life science research during the pandemic) the approach presented here advances that effort by focusing on field-normalized impact metrics and broader information seeking and scholarly communication topics.

Data collection

Data for this study were collected in September 2024 using the SciVal tool to extract relevant data from the Scopus database. The query focused on the topic Information seeking; scholarly

communication; adoption–TC.905, and was filtered for publications between 2015 and 2022. This search resulted in a data set of 13,576 publications, of which 5,682 were journal articles with at least one citation. The most frequent key phrases describing the data set are presented in Figure 1. At the top of the list are open access (2281), related to open access publishing (577), open access journal (501); academic libraries (1466) related to academic library (622), digital libraries (270), library collection (353), open educational resources (1282), journals (850) and electronic resources (452). The next important set of key phrases is related to information seeking (519) with information seeking behaviour (251), information behaviour (583), information needs (413), information science (383), information literacy (267).

Figure 1. Keywords in topic cluster 905 for publications between 2015 and 2022

The selected chronological frame from 2015 to 2022 is considered appropriate, as it captures publication trends before and after the onset of pandemic-related restrictions in March 2020. March 2020 marks the point of intervention for this study, as the COVID-19 pandemic began to affect mobility and introduced new demands for digital technologies and communication, particularly in higher education (Cohen et al., 2020; Cullinan et al., 2021; Ruhm, 2024), impacting scholarly communication and well-being (Mok, 2022), and led to substantial global reactions, such as stay-at-home regulations (Alex, 2022). To accommodate the structure of the C-ARIMA model, which allows for the setting of a specific timeframe to observe causal effects, the horizon for this study was set from June 2021 to February 2022. This period was chosen because it likely reflects the more sustained effects of the pandemic, by which time 'constant disruptions in the educational process imposed by repeated lockdowns' were observed (Aristovnik et al., 2023, p. 4). Legal restrictions on social mixing varied slightly, but mostly prevalent between March 2020 and June 2021 and later lifted by February 2022, followed by a shift toward individuals managing their own

risk (Grøsland et al., 2022; Ruhm, 2024; Smith et al., 2022). In general, arguing that the pandemic had an immediate effect on impact metrics will be challenging, given the need to account for the delay in academic research and publishing processes, further compounded by citation delays.

Given the nature of the dataset, with many publications sharing the same date and the irregularity of publication schedules, a temporal aggregation method was employed to aggregate the data into monthly time points (Granger, 1990). This transformation also benefits the forecasting of the trend of impact metrics after intervention, particularly with data starting from 2015, providing a broad context for analysis.

To measure the impact of the pandemic on scholarly output, relying solely on citation counts was deemed insufficient, as citations accumulate over time and are not specific to the publication field. Instead, the Field-weighted citation impact (FWCI) metric was used. FWCI is a field-normalised metric that is stable and reflective of the impact of publications within the first three years after publication (SciVal Support Centre, 2024). For accurate causal assessment in time series, additional unimpacted covariates were required to forecast counterfactual scenarios within the C-ARIMA model. These covariates included the number of publications during the period to capture the lack of long term shifts in research productivity, the average citation count as a measure of peer recognition, and the long term value of research, and the average source normalized impact per paper (SNIP) of journals, indicating the stable influence of journal prestige. Although these metrics are inherently quantitative, taken together they approximate factors beyond raw citation counts (such as disciplinary relevance, visibility, and peer reception),thereby providing a more holistic basis for modelling counterfactual scenarios and evaluating the pandemic's causal effects on research output.

Data analysis

Causal inference in time series analysis integrates counterfactual reasoning with forecasting methodologies, extending the concept of potential outcomes into the temporal domain (Bojinov & Shephard, 2019). This approach enables the estimation of causal effects over time by comparing observed outcomes with counterfactual scenarios, providing a dynamic view of the impacts of interventions (Brodersen et al., 2015). The C-ARIMA model (Menchetti et al., 2023) represents a significant advancement in this area by combining counterfactual outcomes with ARIMA models, creating a robust framework for assessing intervention impacts. Specifically, C-ARIMA is effective in disentangling the effects of interventions such as the COVID-19 pandemic from underlying trends and seasonal patterns, which are common in bibliometric data due to the periodic nature of journal publishing cycles.

In this study, causal inference methods allow for a more nuanced understanding of the impact of the pandemic on information seeking and scholarly communication. Rather than relying on simplistic before-and-after comparisons of publication metrics, this approach estimates the causal impact of pandemic-induced changes while accounting for preexisting trends in research output. This enables the development of a comprehensive overview of how the pandemic has shaped research priorities and influenced the impact of scholarly work across various disciplines.

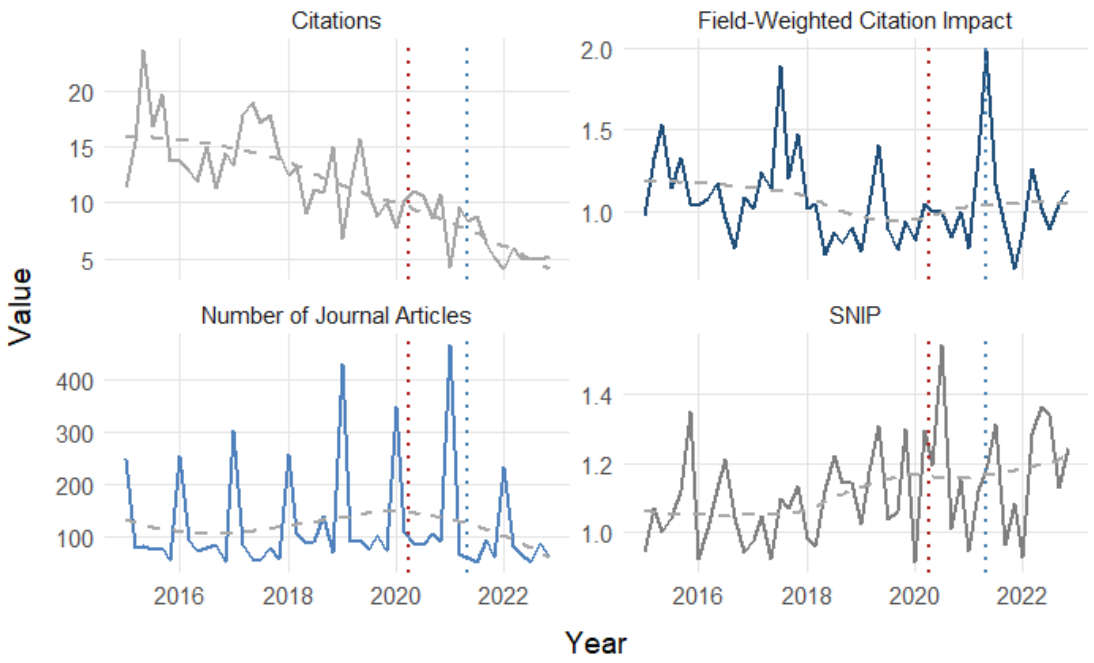

Figure 2. Monthly changes in citations, number of publications, SNIIP, and FWCI metrics in topic cluster TC.905

Regarding the software used, the causalARIMA R library (Menchetti, 2021/2024) was used for this analysis, using a manually selected autoregressive integrated moving average (ARIMA) model, specifically ARIMA (1,1,2). This model was chosen because of its similarity to the damped Holt model, incorporating an autoregressive damping parameter (McKenzie & Gardner, 2010). Such a model is particularly suitable for nonstationary time series where a trend is expected to diminish over time, avoiding unrealistic long term predictions that assume a continued linear trend. This is particularly relevant in the context of citations, where external factors, such as saturation or a slowdown in journal impact factor growth, could decelerate trends.

To account for the characteristics of the time series data (Figure 2), the ARIMA(1,1,2) model incorporates a first-order autoregressive component (AR(1)) to introduce dependency on the values of the previous period. First, differencing (I(1)) was applied to handle non-stationarity and a second order moving average component (MA(2)) was included to address short-term shocks or noise in the data. This component is especially important given the uneven publication periods resulting from periodicity of journals.

Results

Performance and fit of the model

The ARIMA(1,1,2) model performed well in forecasting and estimating causal effects. Significant coefficients AR (1) and MA (2) (Table 1), combined with the minimal bias reflected in the performance metrics (Table 2), indicate that the model effectively captured the temporal dynamics of the FWCI metric. Citations were a significant predictor (t = 8.756), while SNIP demonstrated a trend toward significance (t = 1.854), suggesting a potential contribution of journal quality, though its effect was less robust compared to citations. The model handled the non-stationary nature of the FWCI series with first-order differencing (d=1), removing trends while preserving meaningful fluctuations. The estimated coefficients (Table 1) show AR(1) at -0.862 (t = -4.635) and MA(2) at -0.931 (t = -5.920), both of whom play a key role in modelling the time structure. However, MA (1) was not significant at -0.069 (t = -0.434) and was not helpful for the model to take into account

more complex patterns in the data, and ARIMA used information from two steps back (MA(2)) to handle short-term variations in the data.

Coefficient	Estimate	Standard Error	t-value
AR(1)	-0.862	0.186	-4.635
MA(1)	-0.069	0.158	-0.434
MA(2)	-0.931	0.157	-5.920
Citations	0.056	0.006	8.756
N-publications	0.000	0.001	0.902
SNIP	0.291	0.157	1.854

Table 1. Estimated coefficients and significance levels for the ARIMA model.

Performance metrics confirmed the overall robustness of the forecast: the mean error between predicted and observed values was close to zero (-0.009), the RMSE was 0.231, and the MPE was -3.872%, indicating a slight underestimate of the FWCI by around 3.87%. A MAPE = 13.603, below 15% further suggests strong predictive accuracy. The AIC (13.215), BIC (28.105), and AICC (15.289) all suggest a well-fitted model without overfitting. The Durbin-Watson statistic (1.9183, p=0.3305) indicated that there was no significant autocorrelation, validating the model.

Causal effects of the COVID-19 pandemic

The results of the causal effects analysis, shown in Figure 3 and summarised in Tables 2 and 3, highlight the impact of the COVID-19 pandemic on the research impact of scholarly communication and information seeking behaviour journal articles related to the topic group TC.905. The analysis, which focused on the FWCI metric, compared observed versus forecasted values using C-ARIMA modelling over a horizon from mid-2021 to early 2022. The data reveal a clear cumulative increase in FWCI, with significant point and cumulative causal effects identified across the two key periods of analysis.

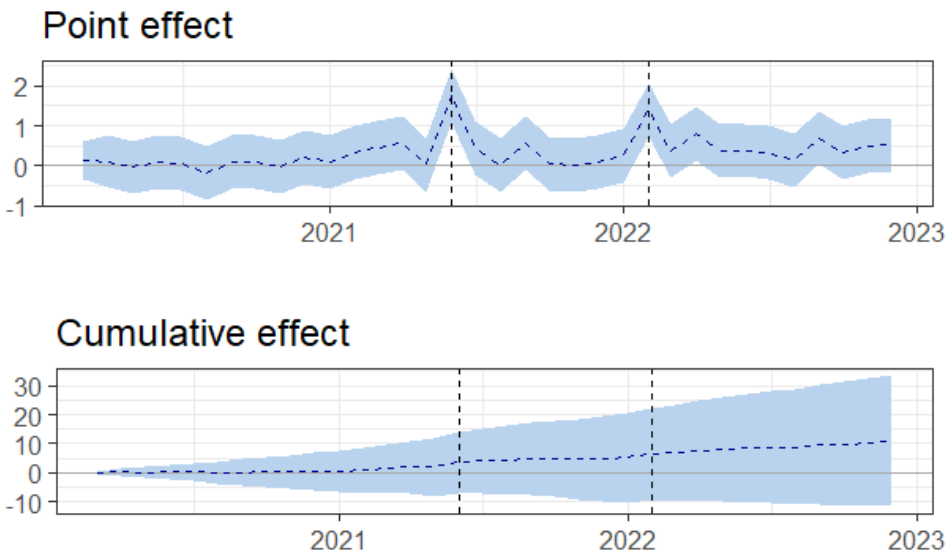

Figure 3. Causal effect horizons.

As illustrated in Table 2, the point causal effects of FWCI increased by 1.764 in June 2021 and 1.400 in February 2022, both with standard errors of 0.34. These results suggest a significant positive impact on the topic cluster during these times, confirmed by right-sided p-values below 0.001. These point effects reflect short-term increases, likely corresponding to a greater focus on

academic communication and information seeking as the academic community adapts to pandemic-driven changes. However, the absence of an immediate effect without horizon (0.536 p =0.118) supports the assumption that the impact of the pandemic on FWCI was not instantaneous but evolved over time.

Estimates (Standard Error)	Without horizons	2021-06	2022-02
Point Causal Effect	0.536 (0.343)[1]	1.764 (0.34)	1.400 (0.34)
Cumulative Causal Effect	10.943 (0.249)	3.482 (0.242)	6.281 (0.243)
Temporal Average Causal Effect	0.322 (0.007)	0.232 (0.16)	0.273 (0.11)
Left-sided p-value	1.000	1.000	1.000
Right-sided p-value	< 0.001	< 0.001	< 0.001
Bidirectional p-value	< 0.001	< 0.001	< 0.001
[1]Left-sided p-value 0.941; Bidirectional p-value 0.118; Right-sided p-value 0.059			

Table 2. Causal effects summary.

By June 2021, the cumulative total of the average monthly FWCI had increased by 3.482, rising to 6.281 by February 2022, with standard errors of 0.242 and 0.243, respectively. This sustained increase indicates a long-term shift in research productivity and impact within the topic cluster, reflecting the cumulative effect of the pandemic on scholarly communication and information-seeking topics. Accounting for the number of publications during these periods, the cumulative FWCI gain corresponds to a total increase of 145.99 by June 2021 and 219.203 by February 2022, representing the whole topic's impact accumulation over time.

Estimates	Without horizons	2021-06	2022-02
Observed Effect	1.082	1.123	1.090
Forecasted Effect	0.760 (95% CI [0.685, 0.841], SD = 0.041)	0.891 (95% CI [0.785, 1.036], SD = 0.063)	0.817 (95% CI [0.722, 0.929], SD = 0.052)
Absolute Effect	0.322 (95% CI [0.241, 0.397], SD = 0.041)	0.232 (95% CI [0.087, 0.338], SD = 0.063)	0.273 (95% CI [0.161, 0.368], SD = 0.052)
Observed Cumulative Effect	36.790	16.840	25.078
Forecasted Cumulative Effect	25.847 (95% CI [23.295, 28.601], SD = 1.384)	13.358 (95% CI [11.771, 15.539], SD = 0.950)	18.798 (95% CI [16.613, 21.372], SD = 1.191)
Absolute Cumulative Effect	10.943 (95% CI [8.189, 13.494], SD = 1.384)	3.482 (95% CI [1.301, 5.070], SD = 0.950)	6.281 (95% CI [3.707, 8.465], SD = 1.191)
Relative Cumulative Effect	0.423 (95% CI [0.317, 0.522], SD = 0.054)	0.261 (95% CI [0.097, 0.380], SD = 0.071)	0.334 (95% CI [0.197, 0.450], SD = 0.063)
alpha = 0.05	p < 0.001	p = 0.002	p < 0.001

Table 3. Observed and forecasted effects.

First, the observed and predicted effects at two key time points, June 2021 and February 2022, indicated that the observed FWCI values were consistently higher than the predicted values. In June 2021, the observed FWCI was 1.123, while the predicted value was 0.891 (95% CI [0.785, 1.036], SD = 0.063). This yielded a temporal average effect of 0.232, with a confidence interval ranging from 0.087 to 0.338 and a small standard deviation of 0.063, suggesting a statistically significant increase in FWCI during this period. Similarly, by February 2022, the observed FWCI was 1.090,

while the predicted value was 0.817 (95% CI [0.722, 0.929], SD = 0.052), resulting in a temporal effect of 0.273 (95% CI [0.161, 0.368], SD = 0.052). The statistical significance ($p < 0.05$) of these effects indicates that the impact of the pandemic on research in these fields was both significant and sustained over the horizon period.

In terms of cumulative effects, by June 2021, the cumulative FWCI reached 16.840, compared to the expected cumulative FWCI of 13.358 (95% CI [11.771, 15.539], SD = 0.950), resulting in an absolute cumulative effect of 3.482 (95% CI [1.301, 5.070]). By February 2022, the cumulative observed FWCI increased to 25.078, while the cumulative forecast value remained lower at 18.798 (95% CI [16.613, 21.372], SD = 1.191), resulting in an absolute cumulative effect of 6.281 (95% CI [3.707, 8.465]). The relative cumulative effects were 26.1% and 33.4%, respectively, indicating that the cumulative impact of the pandemic on FWCI was significantly stronger than anticipated at both time points.

In the non-horizon analysis, covering the period from the intervention in March 2020 to the end of 2022, the cumulative observed FWCI reached 36.790, while the cumulative forecast value was 25.847 (95% CI [23.295, 28.601], SD = 1.384). This resulted in an absolute cumulative effect of 10.943 (95% CI [8.189, 13.494], SD = 1.384), demonstrating a substantial cumulative increase in FWCI since the critical period in academia, during which the rapid digitisation of academic work and communication reshaped the research landscape. The relative cumulative effect of 42.3% (95% CI [0.317, 0.522], SD = 0.054) reflects the considerable influence of the pandemic on the research impact of studies on information seeking and scholarly communication in related fields, far exceeding initial expectations, with effects lasting well beyond the anticipated horizon.

Discussion

The rapid digitisation of academic work and the urgent need for accessible research have reshaped the research landscape in ways that may persist beyond the immediate crisis. However, the present study does not directly confirm widespread changes to research practices or digital tool adoption. Rather, it focuses on whether the pandemic influenced the citation patterns of literature on information seeking and scholarly communication as a topic cluster by shifting research priorities toward digital transformations that address urgent concerns in these areas, thereby heightening scholarly attention. By examining how the pandemic caused the shifts in these research areas, the analysis supports the notion that the global crisis influenced scholarly attention to topics related to information seeking behaviours and scholarly communication, leading to unexpected increased field-specific impact (Riccaboni & Verginer, 2022).

Using the C-ARIMA model, key causal effects were identified, indicating the role of the pandemic in reshaping some research topics' focus. This methodology goes beyond simple correlations, establishing a direct causal link between the pandemic and changes in research impact, an underutilised approach in scientometric studies (Li et al., 2023). By applying causal inference, the study bridges the theory-driven sociology of science with data-driven bibliometrics and scientometrics. Waltman (2016) emphasises that normalised citation impact indicators are critical in research evaluation because they account for disciplinary differences. This perspective is crucial for assessing the impact of the pandemic on research behaviour and academic output in multiple fields.

The observed increase in impact within related areas may reflect growing academic interest in how scholarly communication and information seeking evolved during the pandemic, rather than a direct measure of adoption of digital tools alone. The rise in research impact underscores the importance of these fields in navigating post-pandemic academic discourse, where digital reliance, misinformation, and the demand for open data have increased the value of research (Otike et al., 2022; Ross-Hellauer et al., 2022).

In practical terms, the application of causal inference in scientometrics offers valuable information for both academic institutions and policy makers. Identifying causal relationships between global crises, technological breakthroughs, and academic performance provides essential knowledge for institutional preparedness and policy adjustments. Future research could extend this approach to more granular studies of specific fields, journals, and collaborative practices, further exploring how crises shape knowledge production and dissemination across disciplines (Li et al., 2023).

Empirical studies could involve surveys and interviews with faculty to investigate the social processes driving the observed causal effects on information seeking and scholarly communication post-pandemic. While the pandemic likely triggered a digital transformation in research, communication, and academic collaboration, the present findings mainly suggest increased scholarly interest in these topics (e Rodrigues & Mandrekar, 2021; Elia, 2023). This trend aligns with the broader digitalisation of academia, where open access and efficient information retrieval as a research topic has increased. However, this change also raises ethical and policy concerns, such as data privacy, misinformation, and the commodification of scholarly output (Ross-Hellauer et al., 2022). Causal inference methods, which establish time-related causal links, are better suited to address these concerns than correlational studies, which can overlook the underlying drivers of these changes but still can impose some limitations (Cordero et al., 2018).

Limitations of the study

While this research applies a novel causal inference approach to bibliometric data, several limitations warrant attention. First, although the C-ARIMA model was chosen for its capacity to address time-diminishing trend effects in nonstationary time series, it was not systematically compared to other configurations, such as ARIMA or SARIMA variants, exponential smoothing, or other causal models like Bayesian Structural Time Series (Riccaboni & Verginer, 2022; Wenaas, 2021). Future methodological work could incorporate these alternative models to refine and validate the findings, given the complexities of bibliometric data. In this study, the pragmatic focus on achieving a reliable forecast with minimal errors remains crucial for effectively handling time-series challenges in bibliometric analysis.

Second, the study examines a single topic cluster (TC.905) from 2015–2022 occurring only in Scopus database, aggregated at a monthly level to manage temporal detail and statistical stability. Although this choice reduces noise from publication timing, other aggregation windows (e.g., quarterly) might reveal different patterns or capture disciplinary nuances, but would also reduce data points and necessitate a longer timeframe for analysis. Future investigations could explore specialized clusters to see whether similar trends surface in different publishing contexts.

Third, even though unimpacted covariates were considered, unobserved confounders remain possible (Pearl, 2000). Institutional policies, funding shifts, or other external influences may shape citation trends in ways not fully accounted for here. Incorporating Principal Component Analysis (PCA) to derive latent variables from multiple covariates has proven effective in ARIMA models (Hajirahimi et al., 2023). Hence, while these findings suggest a causal link between the pandemic and shifts in research impact, they should be interpreted with caution considering these constraints.

Conclusions

The adoption of causal inference methodologies in bibliometric analysis marks a significant leap forward in understanding the dynamics of scientific research and knowledge dissemination. These techniques allow for moving beyond descriptive analyses to make causal claims about the factors driving changes in research focus and impact (Li et al., 2023). This shift aligns with the growing recognition of the importance of causal reasoning in data science and machine learning (Pearl & Mackenzie, 2018). Causal reasoning supports more informed decision-making. By understanding the reasons behind observed changes in research practices, rather than simply observing them, academia can develop more effective and ethically sound strategies to address emerging trends.

Based on the growing significance of information seeking research in post-pandemic scholarly communication, this hypothesis is proposed for future studies in both bibliometric and empirical contexts. The increase in research impact related to information seeking behaviour and scholarly communication post-pandemic is driven by an expanding reliance on digital and open science models for knowledge dissemination. This shift, further amplified by the increasing demand for accessible information during global crises, has elevated the prominence and influence of research in these fields. This hypothesis promotes a deeper exploration into how digital transformations and the rise of open science will continue shaping research practices, collaborative networks, and the overall impact of academic work in a post-pandemic world.

About the author

Marek Deja is an Assistant Professor at the Institute of Information Studies, Faculty of Management and Social Communication, Jagiellonian University. His research interests include academic information cultures, digital literacy, information management in higher education, and scientometrics. He can be contacted at marek.deja@uj.edu.pl.

References

Alex, J. K. (2022). Impact of the Covid-19 pandemic on the academic life of higher education students: A rural South African perspective from a global study. *South African Journal of Higher Education*, 36(1), 20–40. https://doi.org/10.20853/36-1-4303

Aristovnik, A., Karampelas, K., Umek, L., & Ravšelj, D. (2023). Impact of the COVID-19 pandemic on online learning in higher education: A bibliometric analysis. *Frontiers in Education*, 8. https://doi.org/10.3389/feduc.2023.1225834

Begum, S. M., & Habiba, U. (2023). Are library professionals overcoming the COVID-19 pandemic challenges in the current reality? *Digital Library Perspectives*, 39(4), 420–443. https://doi.org/10.1108/DLP-01-2023-0009

Bojinov, I., & Shephard, N. (2019). Time series experiments and causal estimands: Exact randomization tests and trading. *Journal of the American Statistical Association*, 114(528), 1665–1682. https://doi.org/10.1080/01621459.2018.1527225

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., & Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics*, 9(1), 247–274. https://doi.org/10.1214/14-AOAS788

Chakravorty, N., Sharma, C. S., Molla, K. A., & Pattanaik, J. K. (2022). Open science: Challenges, possible solutions and the way forward. *Proceedings of the Indian National Science Academy*, 88(3), 456–471. https://doi.org/10.1007/s43538-022-00104-2

Cheng, L., Guo, R., Moraffah, R., Sheth, P., Candan, K. S., & Liu, H. (2022). Evaluation methods and measures for causal learning algorithms. *IEEE Transactions on Artificial Intelligence*, 3(6), 924-943. https://doi.ieeecomputersociety.org/10.1109/TAI.2022.3150264

Cohen, A. K., Hoyt, L. T., & Dull, B. (2020). A descriptive study of COVID-19–related experiences and perspectives of a national sample of college students in spring 2020. *Journal of Adolescent Health*, 67(3), 369–375. https://doi.org/10.1016/j.jadohealth.2020.06.009

Cordero, J. M., Cristóbal, V., & Santín, D. (2018). Causal Inference on education policies: A survey of empirical studies using Pisa, Timss and Pirls. *Journal of Economic Surveys*, 32(3), 878–915. https://doi.org/10.1111/joes.12217

Cui, R., Ding, H., & Zhu, F. (2022). Gender inequality in eesearch productivity during the COVID-19 pandemic. *Manufacturing & Service Operations Management*, 24(2), 707–726. https://doi.org/10.1287/msom.2021.0991

Cullinan, J., Flannery, D., Harold, J., Lyons, S., & Palcic, D. (2021). The disconnected: COVID-19 and disparities in access to quality broadband for higher education students. *International Journal of Educational Technology in Higher Education*, 18(1), 26. https://doi.org/10.1186/s41239-021-00262-1

de Wit, H., & Altbach, P. G. (2023). International higher education for the future: Major crises and post-pandemic challenges. *Change: The Magazine of Higher Learning*, 55(1), 17–23. https://doi.org/10.1080/00091383.2023.2151799

Downes, M. (2023). There is no such thing as a predatory journal. *Learned Publishing*, 36(4), 709–711.. https://doi.org/10.1002/leap.1568

e Rodrigues, M. C., & Mandrekar, B. (2021). Role of the libraries in providing information and changing information-seeking behaviour of students and faculties during the pandemic: A study. *Library Philosophy and Practice*, 1–20. https://digitalcommons.unl.edu/libphilprac/6350/

Elia, E. F. (2023). Information-seeking behaviour of Tanzanians in the time of COVID-19. *Global Knowledge, Memory and Communication*, 72(4/5), 356–375. https://doi.org/10.1108/GKMC-04-2021-0060

Fernández, A., Gómez, B., Binjaku, K., & Meçe, E. K. (2023). Digital transformation initiatives in higher education institutions: A multivocal literature review. *Education and Information Technologies*, 28(10), 12351–12382. https://doi.org/10.1007/s10639-022-11544-0

Fitzgerald, S. R. (2020). Toward a conceptual framework for scholarly information seeking. *The Journal of Academic Librarianship*, 46(6), 102259. https://doi.org/10.1016/j.acalib.2020.102259

Frank, J., Foster, R., & Pagliari, C. (2023). Open access publishing – noble intention, flawed reality. *Social Science and Medicine*, 317, 115592. https://doi.org/10.1016/j.socscimed.2022.115592

Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálfy, M., Nanni, F., & Coates, J. A. (2021). The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLOS Biology*, 19(4), e3000959. https://doi.org/10.1371/journal.pbio.3000959

Gao, J., Yin, Y., Myers, K. R., Lakhani, K. R., & Wang, D. (2021). Potentially long-lasting effects of the pandemic on scientists. *Nature Communications*, 12(1), 6188. https://doi.org/10.1038/s41467-021-26428-z

Granger, C. W. J. (1990). Aggregation of time-series variables: A survey. In T. Barker (Ed.), *Disaggregation in econometric modelling* (Chapter 1). Routledge. https://doi.org/10.4324/9781315822549

Grøsland, M., Larsen, V. B., Reneflot, A., & Hart, R. K. (2022). The COVID-19 pandemic's impact on mental health care use among Norwegian students in higher education: A nation-wide register-based pre-post cohort study. *BMC Health Services Research*, *22*(1), 1506. https://doi.org/10.1186/s12913-022-08816-3

Hajirahimi, Z., Khashei, M., & Hamadani, A. Z. (2023). Principal component-based hybrid model for time series forecasting. *International Journal of Information Technology*, 15(6), 3045–3053. https://doi.org/10.1007/s41870-023-01343-2

Huvila, I., & Gorichanaz, T. (2024). Trends in information behavior research, 2016–2022: An Annual Review of Information Science and Technology paper. *Journal of the Association for Information Science and Technology*, 76(1), 216-237 https://doi.org/10.1002/asi.24943

Imai, K., Kim, I. S., & Wang, E. H. (2023). Matching methods for causal inference with time-series cross-sectional data. *American Journal of Political Science*, *67*(3), 587–605. https://doi.org/10.1111/ajps.12685

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences*. Cambridge University Press.

Jackman, P. C., Sanderson, R., Haughey, T. J., Brett, C. E., White, N., Zile, A., Tyrrell, K., & Byrom, N. C. (2022). The impact of the first COVID-19 lockdown in the UK for doctoral and early career researchers. *Higher Education*, 84(4), 705–722. https://doi.org/10.1007/s10734-021-00795-4

Kelly, C. A., & Sharot, T. (2021). Individual differences in information-seeking. *Nature Communications*, *12*(1), 7062. https://doi.org/10.1038/s41467-021-27046-5

Kiviaho, A., & Einolander, J. (2023). Digital transformation, well-being and shrinking communities: Narrowing the divides between urban and rural. *Heliyon*, *9*(8), e18801. https://doi.org/10.1016/j.heliyon.2023.e18801

Lee, J. J., & Haupt, J. P. (2021). Scientific globalism during a global crisis: Research collaboration and open access publications on COVID-19. *Higher Education*, *81*(5), 949–966. https://doi.org/10.1007/s10734-020-00589-0

Li, J., Liu, X., & Wei, C. (2023). Why do we need causal inference in scientometric research? *Data Science and Informetrics*, *3*(1), 101–110. https://doi.org/10.59494/dsi.2023.1.7

Ma, L. (2023). *The Scholarly communication handbook: From research dissemination to societal impact*. Facet. https://doi.org/10.29085/9781783306268

McKenzie, E., & Gardner, E. S. (2010). Damped trend exponential smoothing: A modelling viewpoint. *International Journal of Forecasting*, 26(4), 661–665. https://doi.org/10.1016/j.ijforecast.2009.07.001

Menchetti, F. (2024). *CausalArima* [R]. https://github.com/FMenchetti/CausalArima (Original work published 2021)

Menchetti, F., Cipollini, F., & Mealli, F. (2023). Combining counterfactual outcomes and ARIMA models for policy evaluation. *The Econometrics Journal*, 26(1), 1–24. https://doi.org/10.1093/ectj/utac024

Mok, K. H. (2022). Impact of COVID-19 on higher education: Critical reflections. *Higher Education Policy*, 35(3), 563–567. https://doi.org/10.1057/s41307-022-00285-x

Montesi, M. (2021). Human information behavior during the Covid-19 health crisis. A literature review. *Library & Information Science Research*, 43(4), 101122. https://doi.org/10.1016/j.lisr.2021.101122

Myers, K. R., Tham, W. Y., Yin, Y., Cohodes, N., Thursby, J. G., Thursby, M. C., Schiffer, P., Walsh, J. T., Lakhani, K. R., & Wang, D. (2020). Unequal effects of the COVID-19 pandemic on scientists. *Nature Human Behaviour*, 4(9), 880–883. https://doi.org/10.1038/s41562-020-0921-y

Newson, M., Zhao, Y., Zein, M. E., Sulik, J., Dezecache, G., Deroy, O., & Tunçgenç, B. (2024). Digital contact does not promote wellbeing, but face-to-face contact does: A cross-national survey during the COVID-19 pandemic. *New Media & Society*, 26(1), 426–449. https://doi.org/10.1177/14614448211062164

Ni, Z., Zhu, L., Li, S., Zhang, Y., & Zhao, R. (2024). Characteristics and associated factors of health information-seeking behaviour among patients with inflammatory bowel disease in the digital era: A scoping review. *BMC Public Health*, 24(1), 307. https://doi.org/10.1186/s12889-024-17758-w

Oermann, M. H., Waldrop, J., Nicoll, L. H., Peterson, G. M., Drabish, K. S., Carter-Templeton, H., Owens, J. K., Moorman, T., Webb, B., & Wrigley, J. (2023). Research on predatory publishing in health care: a scoping review. *Canadian Journal of Nursing Research*, 55(4), 415–424. https://doi.org/10.1177/08445621231172621

Okafor, I. A., Mbagwu, S. I., Chia, T., Hasim, Z., Udokanma, E. E., & Chandran, K. (2022). Institutionalizing open science in Africa: limitations and prospects. *Frontiers in Research Metrics and Analytics* 7, 855198. https://doi.org/10.3389/frma.2022.855198

Otike, F., Bouaamri, A., & Hajdu Barát, Á. (2022). Predatory publishing: A catalyst of misinformation and disinformation amongst academicians and learners in developing countries. *The Serials Librarian*, 83(1), 81–98. https://doi.org/10.1080/0361526X.2022.2078924

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.

Pearl, J. (2010). Causal inference. In I. Guyon, D. Janzing, & B. Schölkopf (Eds.), *Proceedings of Workshop on Causality: Objectives and Assessment at* NIPS 2008 (Vol. 6, pp. 39–58). PMLR. https://proceedings.mlr.press/v6/pearl10a.html

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.

Prince, N. (2023). Continuing education and data training initiatives are needed to positively impact academic librarians providing data services. *Evidence Based Library and Information Practice*, 18(3), 81–83.https://doi.org/10.18438/eblip30382

Püttmann, V., & Thomsen, S. L. (2024). Academics' susceptibility to disruptions of their research productivity: empirical insights from the COVID-19 pandemic. *Higher Education*. https://doi.org/10.1007/s10734-024-01266-2

Renz, A., & Hilbig, R. (2023). Digital transformation of educational institutions accelerated by COVID-19: A digital dynamic capabilities approach. In J. Whalley, V. Stocker, & W. Lehr (Eds.), *Beyond the pandemic? Exploring the impact of COVID-19 on telecommunications and the Internet* (pp. 103–119). Emerald Publishing Limited. https://doi.org/10.1108/978-1-80262-049-820231004

Riccaboni, M., & Verginer, L. (2022). The impact of the COVID-19 pandemic on scientific research in the life sciences. *PLOS ONE*, 17(2), e0263001. https://doi.org/10.1371/journal.pone.0263001

Ross-Hellauer, T., Reichmann, S., Cole, N. L., Fessl, A., Klebel, T., & Pontika, N. (2022). Dynamics of cumulative advantage and threats to equity in open science: A scoping review. *Royal Society Open Science*, 9(1). https://doi.org/10.1098/rsos.211032

Rowley, J. (2023). Getting published in peer reviewed academic journals in business and management: perspectives for doctoral and early career researchers. *Management Research Review*, 46(2) 306–319. https://doi.org/10.1108/MRR-04-2021-0258

Ruhm, C. J. (2024). US state restrictions and excess COVID-19 pandemic deaths. *JAMA Health Forum*, 5(7), e242006. https://doi.org/10.1001/jamahealthforum.2024.2006

SciVal Support Center. (2024, January 15). *SciVal metric: Field-weighted citation impact* (FWCI). https://service.elsevier.com/app/answers/detail/a_id/28192/supporthub/scival/p/10961/

Seguya, A., Salano, V., Okerosi, S., Kim, E. K., Shrime, M. G., Viljoen, G., & Fagan, J. J. (2023). Are open access article processing charges affordable for otolaryngologists in low-income and middle-income countries? *Current Opinion in Otolaryngology and Head and Neck Surgery*, 31(3), 202–207.. https://doi.org/10.1097/MOO.0000000000000892

Shen, Z., Zhong, Z., Xie, J., Zhang, Q., & Li, S. (2022). The effects of information-seeking behaviors on risk perception during the COVID-19 pandemic: A cross-sectional correlational survey. *Psychology Research and Behavior Management*, 15, 1707–1719. https://doi.org/10.2147/PRBM.S368537

Smith, L. E., Potts, H. W. W., Amlôt, R., Fear, N. T., Michie, S., & Rubin, G. J. (2022). Patterns of social mixing in England changed in line with restrictions during the COVID-19 pandemic (September 2020 to April 2022). *Scientific Reports*, 12(1), 10436. https://doi.org/10.1038/s41598-022-14431-3

Sousa, L., Pedro, L., & Santos, C. (2023). A systematic review of systematic reviews on open educational resources: An analysis of the legal and technical openness. *International Review of Research in Open and Distributed Learning*, 24(3), 18–33.. https://doi.org/10.19173/irrodl.v24i3.7196

Tlili, A., Garzón, J., Salha, S., Huang, R., Xu, L., Burgos, D., Denden, M., Farrell, O., Farrow, R., Bozkurt, A., Amiel, T., McGreal, R., López-Serrano, A., & Wiley, D. (2023). Are open educational resources (OER) and practices (OEP) effective in improving learning achievement? A meta-analysis and research synthesis. *International Journal of Educational Technology in Higher Education*, 20(1).. https://doi.org/10.1186/s41239-023-00424-3

Trevisan, L. V., Eustachio, J. H. P. P., Dias, B. G., Filho, W. L., & Pedrozo, E. Á. (2024). Digital transformation towards sustainability in higher education: State-of-the-art and future research insights. *Environment, Development and Sustainability*, 26(2), 2789–2810. https://doi.org/10.1007/s10668-022-02874-7

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365–391. https://doi.org/10.1016/j.joi.2016.02.007

Wenaas, L. (2021). Attracting new users or business as usual? A case study of converting academic subscription-based journals to open access. *Quantitative Science Studies*, 2(2), 474–495. https://doi.org/10.1162/qss_a_00126

Wilson, T.D. (2018). The diffusion of information behaviour research across disciplines In *Proceedings of ISIC, the Information Behaviour Conference, Krakow, Poland, 9-11 October: Part 1. Information Research*, 23(4), isic1801. http://InformationR.net/ir/23-4/isic2018/isic1801.html (Archived by WebCite® at http://www.webcitation.org/74IAnQkY8)

Yang, X., & Li, T. (2024). Does the COVID-19 pandemic affect excellence in academic research? A study of science and engineering faculty members in China. *Higher Education Policy*, 37(3), 532–548. https://doi.org/10.1057/s41307-023-00313-4

Yao, L., Chu, Z., Li, S., Li, Y., Gao, J., & Zhang, A. (2021). A survey on causal inference. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 15(5), 1-46. https://doi.org/10.1145/3444944

Copyright