First Monday
PEER-REVIEWED JOURNAL ON THE INTERNET

Interventions in scholarly communication: Design lessons from public health

by Micah Altman, Philip N. Cohen, and Jessica Polka

Abstract

Despite widespread perception that changes to scholarly communication are needed, there are substantial disagreements over the short- and long-term benefits of most proposed approaches to changing this system, and the lack of systematic, empirical research in this area makes these controversies difficult to resolve. We argue that experience in public health can be usefully applied to scholarly communication. Starting with the history of DDT (dichlorodiphenyltrichloroethane) application, we illustrate four ways complex human systems threaten reliable predictions and blunt *ad hoc* interventions. We then show how these apply to interventions in scholarly publication — open access based on the article processing charges (APC) and the adoption of preprints, with an emphasis on the period of the COVID-19 pandemic. Finally, we offer approaches to help guide the design of future interventions: identifying measures and outcomes, developing infrastructure, incorporating assessment, and contributing to theories of systemic change.

Contents

1. Introduction

The scholarly ecosystem is evolving rapidly, and many argue that fundamental changes are needed. A wide spectrum of governmental and non-governmental organizations, scholars, and corporate entrepreneurs have proposed changes to practically every part of the system. These include proposals to change the way that research is funded and research outputs evaluated (Committee on Toward an Open Science Enterprise, *et al.*, 2018); how researchers are trained and credited (Allen, *et al.*, 2014); how research is planned and designed (Committee on Toward an Open Science Enterprise, *et al.*, 2018); how evidence is collected, managed, and shared (Wilkinson, *et al.*, 2016); and how research findings are published, cited, and replicated (Nosek, *et al.*, 2015).

In this paper, we explore how scientific communication can improve by systematically assessing itself. We describe how a systems approach can help identify and anticipate the consequences of changes to scholarly communication, using two examples: the article-processing charge model and preprints. We suggest greater efforts to anticipate categories of potential consequences, share data on interventions, and integrate systemic assessment into the design of interventions. In so doing, those who would introduce interventions into the system may increase both the chance that they succeed and their capacity to learn whether and how they have.

This analysis takes place in the shadow of the COVID-19 pandemic, which put enormous pressure on the system of scholarly communication in science in particular. Progress in understanding and mitigating COVID-19 has built on a scientific and empirical foundation that describes the dynamics of global health at the systems level. That system includes standards, processes, and infrastructure to measure and compare outcomes, an evidence base that supports comparisons across time and between groups, and a methodology for interventions that support statistically reliable inferences about their effects. The system of scholarly communication — which plays a vital role in public health and all other scientific fields — substantially lacks these features.

The urgent demand for timely science during the pandemic led to an increased role for preprints, as scientists sought to push their research forward more rapidly (Fraser, *et al.*, 2021), and to a dramatic increase in media and public attention to preprints (Yan, 2020). The growth of preprints, although not a new mechanism, is a signal of substantial shifts in how scientific scholarly communication works.

The long history of public health intervention and the (shorter) history of scholarly communications suggest that we approach these shifts with both hope and caution. Hope because these shifts are consistent with the goals of major initiatives to promote open access with the goal of social benefit. Caution because of the long history of unintended consequences in public health, and because of the unintended consequences of the previous revolution in scholarly communication: article processing charges. The preprint trend story highlights the need for data to inform hopes and cautions, and the incomplete evidentiary base for understanding scholarly communications and other scientific outputs in general.

Rather than coming from real-time indicators, much of our knowledge comes from individual analyses focused on specific research questions published in scholarly journals. As a result, our systemic knowledge lags years behind the data and contains many gaps. Moreover, many analyses rely on databases such as Clarivate's Web of Science, Google Scholar, Elsevier's Scopus, and Digital Science Dimensions — which are proprietary, use opaque methods, and are subject to undocumented revision over time. On the other hand, an open resource, the database of retracted scientific papers maintained by Retraction Watch — the only systematic data collection of its kind — is produced by a small nonprofit organization with a budget of less than US$100,000 per year (Retraction Watch, n.d.). There are some notable initiatives in progress to share open data on the scholarly ecosystem and to produce standardized indicators and the volume and types of science output systematically over time, using existing open data sources (Altman, 2022; Priem, *et al.*, 2022; Tsuji, 2018; Nishikawa-Pacher, *et al.*, 2022). However, most of these are still in the early stages, are funded through grants and contributed efforts, and lack a sustainability model.

In contrast to scholarly communication, the evidence base for the field of public health is much more open, current, systematic, and standardized. For example, public health systems run on accepted protocols for surveillance (Centers for Disease Control and Prevention, 2022), with standardized measurement, shared data, and norms of accountability (Hammonds, *et al.*, 2019). One program, the Demographic, and Health Surveys, has facilitated more than 400 surveys in 90 countries, which are largely implemented to common standards and collect comparable data (Demographic and Health Surveys, 2022).

2. Learning from public health interventions

The history of public health is replete with complex cases of interventions with myriad effects. In some cases, selecting an intervention to pursue is difficult because the same problem may have multiple causes. Assessing interventions may be difficult because they slip within the stream of changes already underway. And indirect effects flowing from an intervention often defeat its intended aims. We can use experiences from public health as exemplars for the system of scholarly communication.

Consider, for example, the use of household dichlorodiphenyltrichloroethane (DDT). DDT application to control mosquitoes, starting in the 1940s, was initially a successful intervention to combat malaria in affected regions in Africa, Asia, and the Americas. The treatments resulted in reductions in relevant mosquito populations and in subsequent malaria cases and deaths, with beneficial economic effects (Mabaso, *et al.*, 2004). However, DDT would later emerge as a paradigmatic case of unintended consequences for public health interventions (Carson, 1962). Some aspects of the cautionary tale of DDT for malaria control are illustrated in Figure 1.

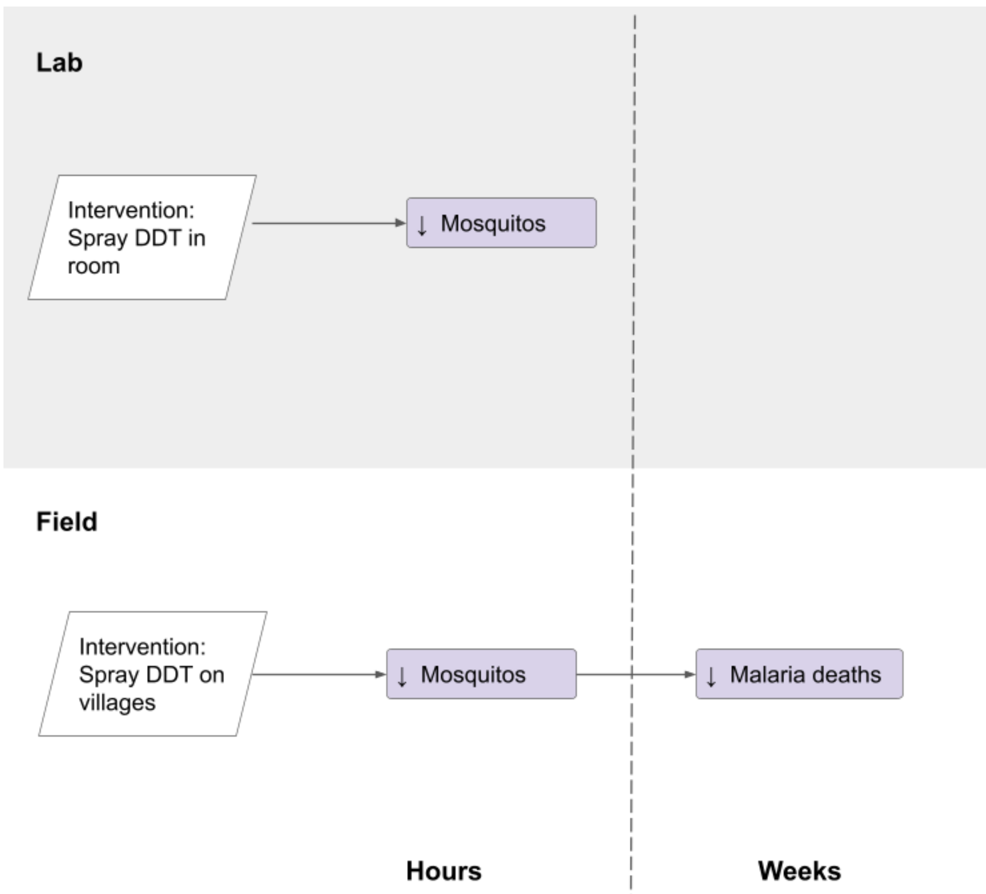


Figure 1: Measured local effects of DDT administration.

Based on successful lab results, the initial use of DDT in the field focused on its short-term efficacy to produce the desired outcomes (shown in gray in Figure 1): mosquito use and malaria reduction within targeted areas. Measuring these outcomes showed the intervention to be highly successful.

Figure 2 shows a more complete picture. Side effects (in white boxes) were not systematically monitored and contributed to a range of adverse effects that emerged only later. This example focuses on the effects of the pesticide on the biological ecosystem; human interactions also complicated the effects of DDT application programs, including migration from affected areas, the lack of trained pesticide applicators, and people replastering their walls to cover up DDT stains (Mabaso, *et al.*, 2004). One common effect was the decimation of helpful species. For example, DDT was fatal to a parasitic wasp in Malaysia but not to its prey, a caterpillar that fed on thatched roof material. As a result of DDT application, therefore, the local human population experienced deterioration of their houses as the caterpillar was freed from predation (O’Shaughnessy, 2008).

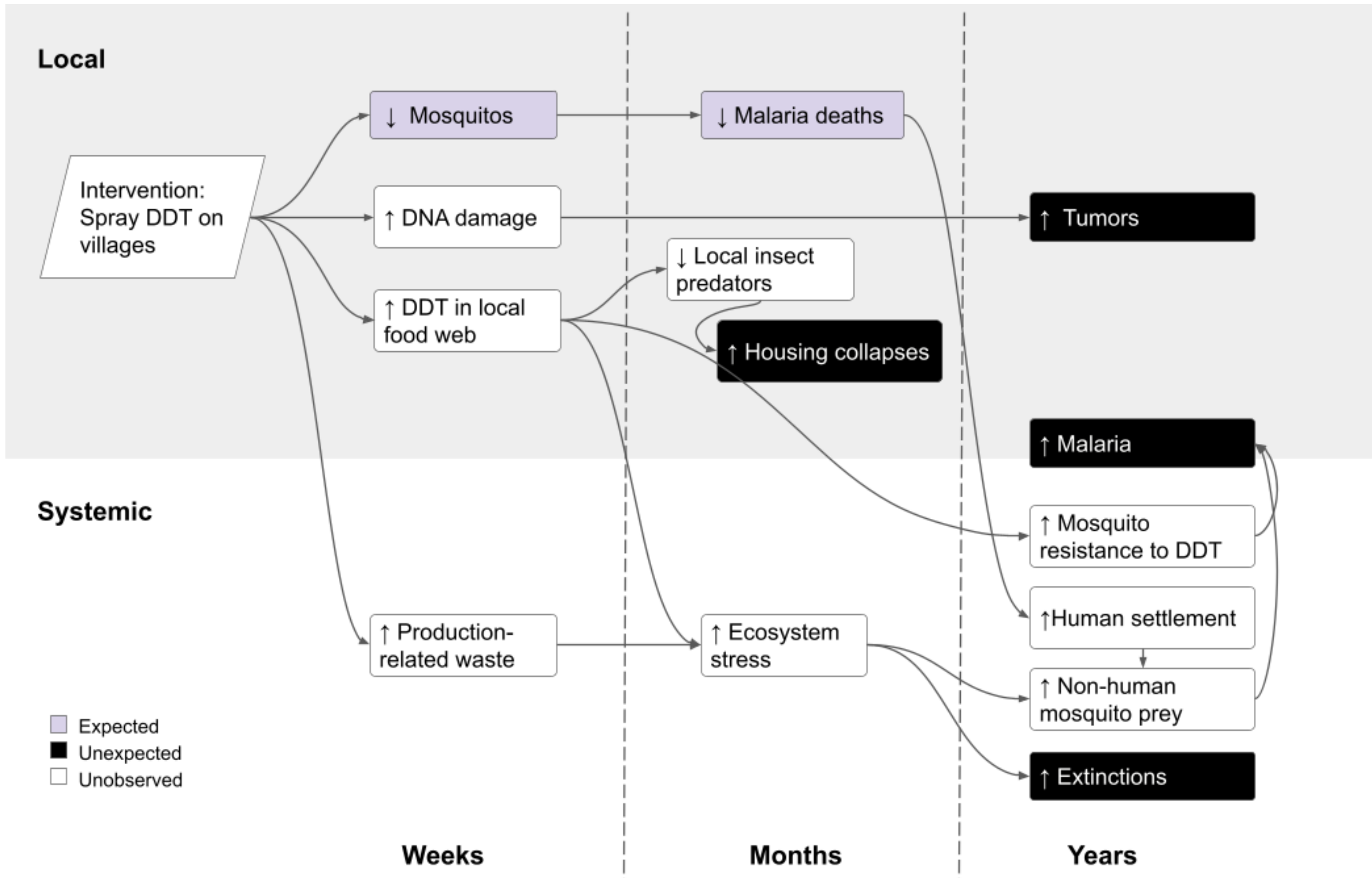


Figure 2: Measured and unmeasured effects of local DDT application.

Years later, it became apparent that DDT had stressed the entire biosphere, leading eventually both to global extinctions and to local increases in malaria, partly due to the emergence of pesticide-resistant mosquitoes (van den Berg, 2009). Another ecosystem adaptation was the spread of resistant species following the use of DDT and later generations of insecticides, leading to the resurgences of mosquitoes and even malaria (Green, 2018; Kupferschmidt, 2016). And, of course, there were the human health effects of DDT, and its effects on species up the food chain from its intended targets (Beard, 2006; van den Berg, 2009). Nevertheless, because of its advantages relative to existing alternatives, DDT still has the endorsement for limited use by some public health agencies, including the World Health Organization (WHO) (van den Berg, *et al.*, 2017)

It may have been impossible to predict the impact of DDT on the eggshells of California condors before the pesticide was deployed (Tubbs, 2016), but the experience of that intervention has surely helped inform the need for wider monitoring and assessment in subsequent efforts. DDT is a prototypical example, but public health interventions provide a panoply of cases in which targeted interventions produced or contributed to undetected adverse events, degradation of unmeasured characteristics of the examined outcome, systemic effects beyond the subjects treated, or long-run dynamics that were beyond the initial scope of design and assessment. Other instructive cases include, for example insecticide-treated mosquito bed nets to control malaria (World Health Organization, 2020; Short, *et al.*, 2018); the fluoridation of public drinking water, starting in the mid-1970s (Ho and Neidell, 2009); many vehicle safety innovations, and other preventive health measures with complex effects (Hedlund, 2000; Cassell, *et al.*, 2006); and also cases of abuse by institutional actors, such as in the infamous Tuskegee Study (Alsan and Wanamaker, 2018). A key lesson to be drawn is: effecting durable and important changes in complex systems is rarely possible absent an understanding of the scope and dynamics of the ecosystem in which the change is made.

Our intention in using these public health examples is to spur consideration of innovations to anticipate potential problems and plan to assess a diverse set of intervention outcomes. We categorize these into four threats to reliable prediction, each representing a targeted outcome for intervention and the potential for unanticipated real-world effects.

1. *Unobserved events*. Assessing only proximate effects, or events intended to be affected (such as the number of people who use a service, or the market share of a product), can leave the researcher biased toward pronouncing the success of an intervention.

2. *Unmeasured characteristics*. An intervention may improve a measured outcome even as it worsens unmeasured outcomes. For example, a treatment that yields a positive outcome when applied by one set of practitioners may have negative effects when applied by others.

3. *System effects*. Unobserved or unmeasured factors may cohere in systemic ways, in the process altering contextual features of the system. As Figure 2 illustrates, systemic effects may be temporally distant from the initial intervention.

4. *Long-run dynamics*. Organizations implementing interventions are likely to have shorter-term ambitions, partly out of necessity (such as the need to generate income), and thus fail to assess those unintended, unmeasured, or systemic effects that take longer to emerge. Over the long run, others may adapt their strategies based on new incentives, so the system as a whole re-equilibrates.

The discipline of public health has long recognized such threats to inference associated with assessing interventions, resulting in systematic attempts to mitigate them (as well as analyses of the system's shortcomings [AbouZahr and Boerma, 2005]). We see four major mechanisms deployed in the field of public health information to address these threats:

1. *Standardized frameworks of measurement*. As described earlier, standardized measurement frameworks and protocols are used widely in public health, emerging from the historic use of surveillance data to respond to crises (Thacker and Stroup, 1994; Choi, 2012).

2. *Data collection & sharing infrastructure*. Public health data is systematically collected as part of program administration. Public health infrastructure supports data aggregation and sharing, facilitated by requirements (Kiley, *et al.*, 2017; U.S. Institute of Medicine, 2015; Jorgenson, *et al.*, 2021), data repositories [1], and surveillance and reporting programs (*e.g.*, Koo and Wetterhall, 1996).

3. *Research design and intervention approaches*. Substantial public health programs are expected to include plans for standardized summative evaluation (Centers for Disease Control and Prevention, 1999). Public health research designs further include mechanisms such as clinical trials registration [2], meta-analysis (Moher, *et al.*, 2009), protocol deposition (Teytelman, *et al.*, 2016), field experiments (Banerjee and Duflo, 2017), and (group-) randomized control trials (Victora, *et al.*, 2004).

4. *Systems theory*. Public health is guided by systemized theory encompassing epidemiology (Rothman, *et al.*, 2021), public health behaviors (Glanz and Bishop, 2010; Valente and Pitts, 2017), and population demography (Birkhead, *et al.*, 2022). Theory suggests how interventions may spill over, where critical parts of the system affect each other, and how patterns of unwanted dynamics could arise.

In summary, standardization reduces the risk of unwanted effects going undetected and supports generalization. Data infrastructure enables observation beyond the local- and short-term effects, including system dynamics. Research designs support the community as a whole in accumulating reliable evidence across studies and linking programs with larger patterns of outcomes. And, although theory may not be not precise enough to predict or avoid all adverse dynamics, it aids in their identification, guides the development of monitoring systems, and is improved through empirical tests.

In contrast to this suite of mechanisms, we are only starting to recognize the dynamics of the scholarly knowledge production system. Theory and research describing the scholarly ecosystem are insufficiently well-articulated to generate testable predictions about the generalizable effects of standardized interventions within local communities of practice. Further, the absence of standard measurement, comprehensive data collection and sharing infrastructure, and causal research design makes building systematic theories challenging. This scarcity weakens the ability to identify and prevent adverse system dynamics that pose the most risk to long-term interventions.

The comparison between the public health system and the scholarly communication ecosystem is most useful for epistemological analysis, as these systems differ fundamentally in their substantive scope and mission. And the public health system suffers from its own structural shortcomings, involving underfunding and political conflicts, among many other issues [3]. Notwithstanding, a lesson to be extracted from this comparison is the key role that a robust and accessible infrastructure for assessment and evaluation plays in understanding and intervening in complex human systems.

3. Interventions in the scholarly ecosystem

3.1. Article processing charges

The article processing charge (APC) model of open access publishing, represents a larger share of scholarly publishing than preprints (to which we turn below), having reached 26 percent of articles in Journal Citation Reports journals by 2019 (Kim and Park, 2021), and 20 percent of Elsevier articles published in 2021 (Elsevier, 2022). Clearly, preprints and APCs — both rapidly growing innovations in the development of scholarly communication — comprise a substantial portion of scientific and other scholarly output, with the former representing a more dramatic increase during the COVID-19 pandemic [4].

Early proponents of open access publishing were convinced that scientific journals should not charge readers to access their articles. Still, they left open the means by which publishing expenses would be covered (Budapest Open Access Initiative, 2002). Open access publishing paid for with article processing charges (APCs) was popularized by the PLOS mega-journals beginning in 2003. In the implementation, the first PLOS journal (*PLoS Biology*) charged an APC of US$1,500, which was intended to be "less than 1 percent of the cost of conducting the research itself" (Public Library of Science, 2003). Despite initial concerns that commercial publishers would stand in the way of open-access publishing (Delamothe, *et al.*, 2003), APC-based open access rapidly gained market share among both non-profit and for-profit publishers, with payments increasingly being bundled into large contracts between universities or research funders and publishers (McKenzie, 2021). Consistent with the original intent of the initiative, this business model appears to have increased the share of research reports that are free to read online. However, once the model was accepted, it became possible (some would say necessary) for APCs to rise. Now, high-prestige journals (including hybrid journals that charge to unlock individual articles) demand higher APCs, contributing to disparities in author access to top-tier publications (Brainard, 2020; Schönfelder, 2020). And an industry of low-quality publishers has adopted the business model as well, extracting profits from poorer institutions and authors (Siler, 2020). Figure 3 illustrates some selected spillover effects of this launch using the same causal graph notation as above.

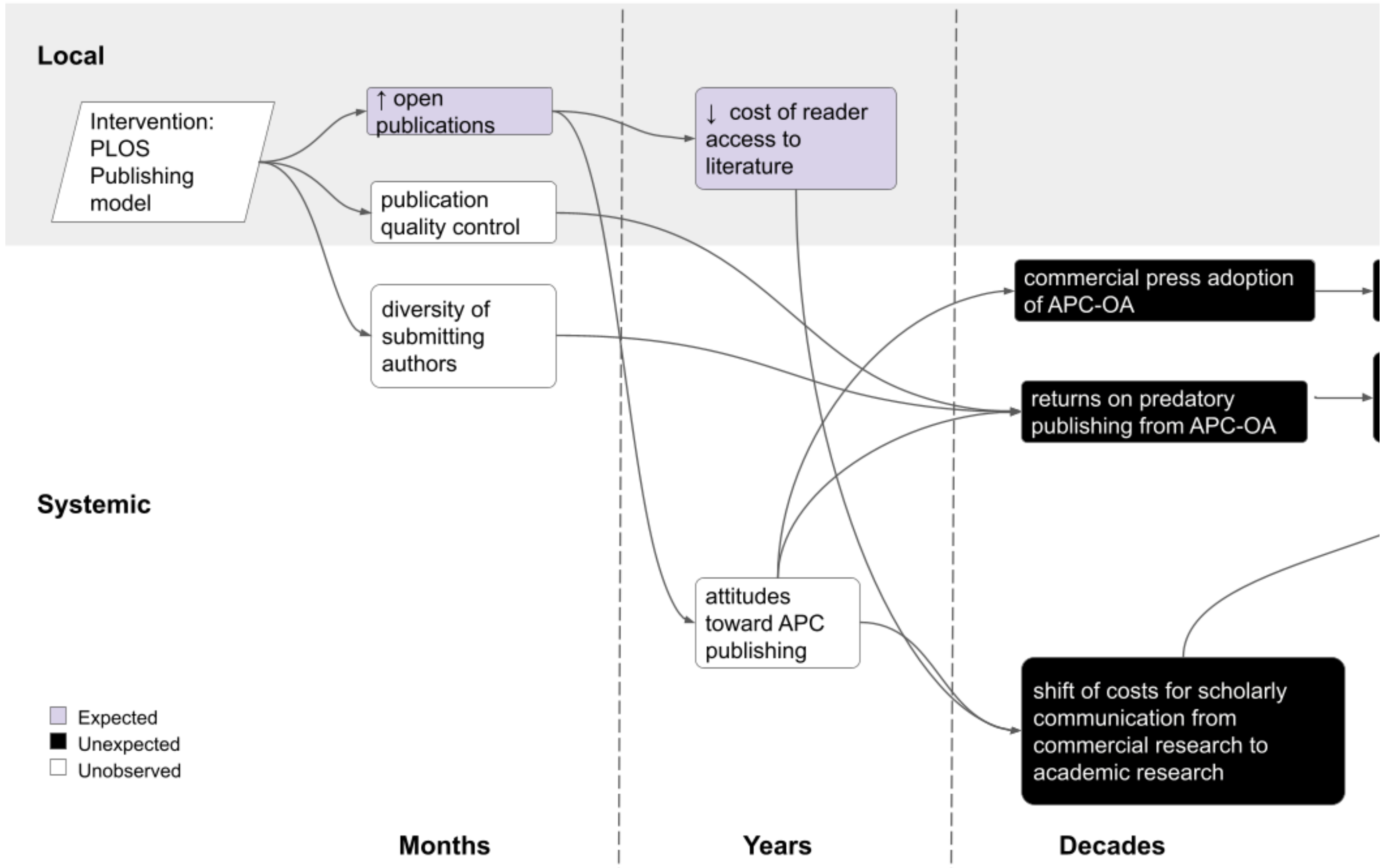


Figure 3: Some unexpected effects of article processing charges.

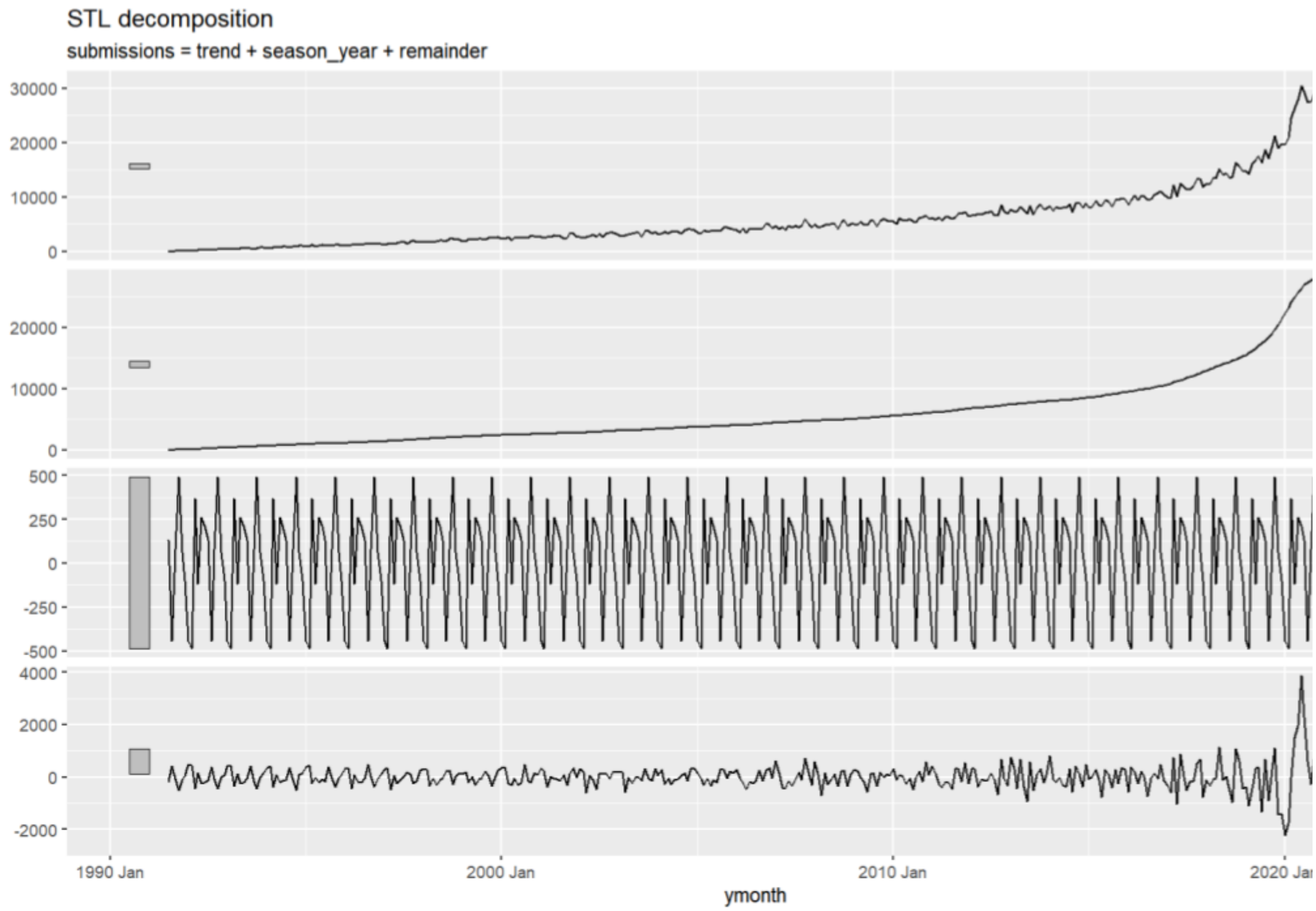
As illustrated in Figure 3, the introduction of APCs had unanticipated consequences. *Unobserved events* made assessing the PLOS' initial success difficult. Submissions and published papers were a natural target, but the choice not to submit was unobserved. This blind spot underlies concerns that the APC model could deter scientists with fewer resources, a problem that some publishers have only partially addressed with fee waivers and discounts (Momeni, *et al.*, 2022). The quality of review in PLOS' innovation is one of several *unmeasured characteristics* of the intervention. The decision to prioritize technical correctness over novelty or impact can lead to "quality censoring," decreasing the ability to use the fact of publication alone as a measure of novelty or impact (García, *et al.*, 2021). As the APC model becomes increasingly popular, systemic costs are shifted from the reader to the author (or their sponsoring organizations). Part of this *system effect* is intended. However, the wider adoption of the APC model also shifts costs within the research sector — in particular from commercial research organizations (some of which are well-funded) and less research-intensive academic institutions to research-intensive universities (Smith, *et al.*, 2016; Parsons, *et al.*, 2011).

Finally, APCs may have *long-run dynamics*. These include making predatory publishing models more profitable (Shen and Björk, 2015). By the mid-2010s, commercial publishers succeeded in capturing the majority of revenue generated by the APC publishing model (Laakso and Björk, 2012). In the long run, comprehensive licensing agreements between large research organizations and commercial publishers — which lower the costs to both read and publish for those covered — raise further equity concerns for independent researchers, those at poorer institutions, and those in the global south (Price and Chaytor, 2015; Poynder, 2019).

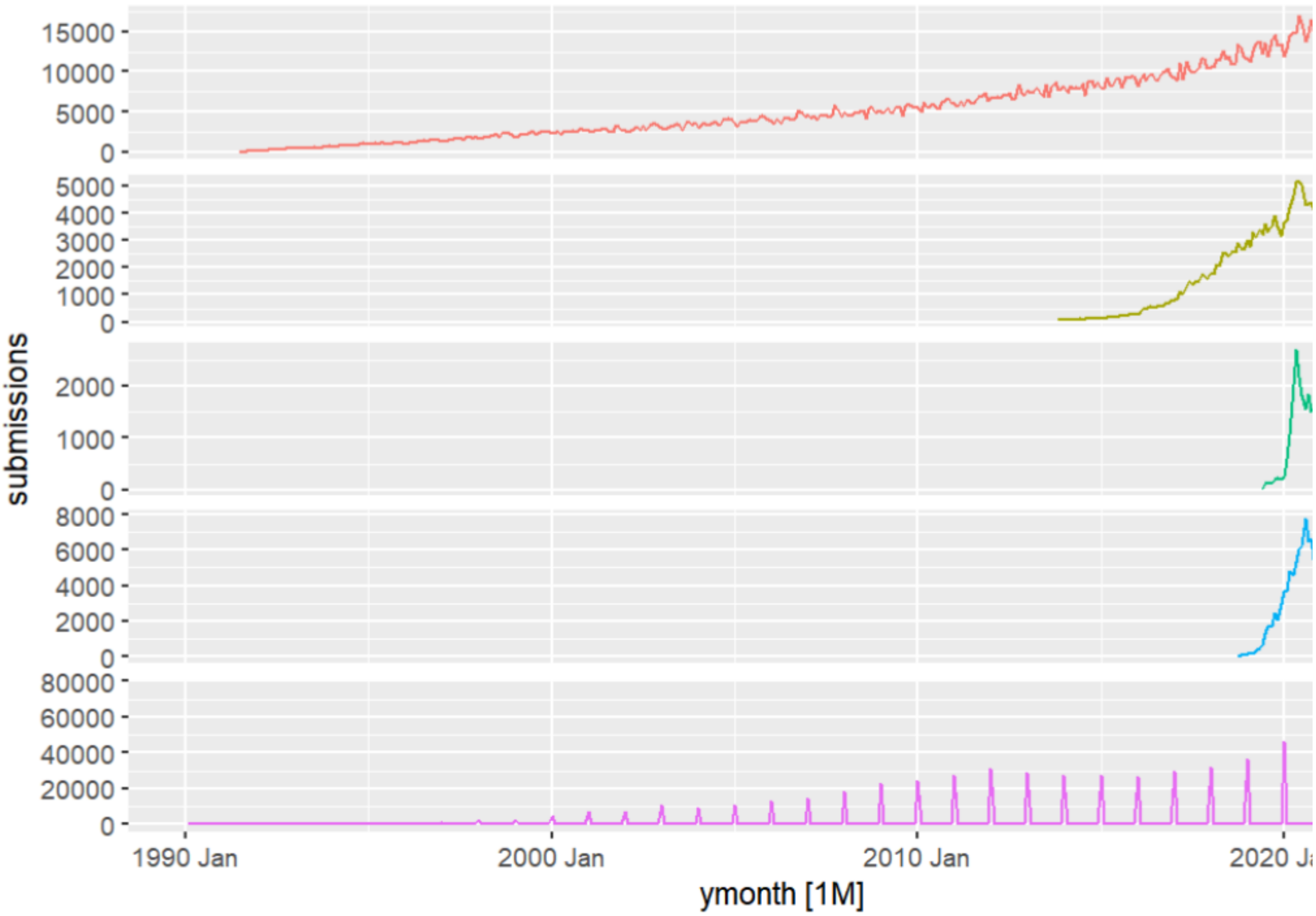
3.2. Preprints

The model of using preprints to disseminate research findings is commonly dated to the founding of arXiv in 1991 (Ginsparg, 2021). Their use was increasing exponentially before the pandemic, and in 2021 there were approximately 360,000 preprints posted, across arXiv, bioRxiv, medRxiv, OSF Preprints, PsyArXiv, SocArXiv, and additional services indexed by Europe PMC — an increase of 70 percent from the number posted in 2019. For comparison, relative to the number of articles listed in the *Web of Science*, preprints rose from 10 percent in 2019 to 13 percent in 2021, with 2020 representing the greatest one-year increase since 2013 [5].

The future exponential nature of this growth cannot be taken for granted as a permanent feature of the scholarly communication landscape, however. As Figure 4 shows, there was a large number of papers posted from March through May 2020, followed by an unprecedented decline in the rate of increase. This change is most dramatic in the data for bioRxiv and medRxiv (panel B of Figure 4), but is apparent in the combined data as well (panel A of Figure 4). Seeing a change such as this prompts questions about the extent to which current trends reflect cyclicity, a one-time increase, or an acceleration of production. Without an evidentiary base from which to study and compare these descriptive trends, the relationship between dramatic shocks such as the pandemic and ongoing changes, such as the increasing use of preprints, is difficult to analyze.



(a) Monthly preprint volume over time: Totals and time series decomposition.
Decomposition of total volume into overall trend, seasonal cycle, and random variation estimated through standard Loess Seasonal Trend Decomposition (13). Gr
component show the relative magnitude of the variations in each of the components: smaller bars indicate that the component trend contributes a larger share of th
above, the time-trend component contributes most to the overall pattern of data, whereas seasonality has a small effect. (Note: Excludes SSRN, which records i



(b) Preprint volume by source.
Note: SSRN provides yearly totals only.

Figure 4: Monthly total preprint issued from leading preprint servers since inception.
Source: Integrated and aggregated from CrossRef DOI metadata API, ColdSpring Harbor server API, and arXiv APIs.

By some interpretations, the pandemic has marked a turning point in the history of scientific publishing (Callaway, 2020); Tingley, 2020; Wellcome, 2020). Although it seemed the long standing arguments for preprints (and open access) had increased resonance during the pandemic, the complexity of the moment means the net effect of preprints on the system is an open question. To put this rapid evolution in context, we review the example of bioRxiv.

At the time of bioRxiv's launch in 2013, the most prominent preprint server was arXiv, and PeerJ Preprints had only recently launched (PeerJ Staff, 2013). Now there are more than 40 preprint servers that host medical and biomedical content (Kirkham, *et al.*, 2020), and many others outside the life sciences. bioRxiv is widely considered a success (Kirkham, *et al.*, 2020) with policy changes by funders and universities having created incentives to post preprints (ASAPbio, 2021), and the growth of new submissions following a geometric pattern (Sever, *et al.*, 2019) (at least until 2020). Many journals have linked with bioRxiv to facilitate the simultaneous submission of manuscripts to the preprint server. Recently, eLife announced it would become a system of "reviewed preprints" rather than a journal that only publishes accepted works (eLife, 2022). At the same time, major institutions such as Plan S promoted the use of institutional repositories to archive and disseminate peer-review publications, often in the form of preprints that were in fact pre-publication versions of an accepted manuscript (Rumsey, 2021). bioRxiv thus emerged as a leader in a rapidly developing field in which preprints were seen as a leading edge.

An increase in use and acceptance is not the only goal — or the only outcome — of the movement toward preprints in the life sciences. Any reduction in time to public disclosure can translate into an acceleration of scientific progress. In the case of the pandemic, more than 10,000 papers about COVID-19 were published in preprint form before appearing in journals in 2020 (in bioRxiv and its medical sibling, medRxiv), and they were shorter and had fewer references then other papers posted in the same period (Fraser, *et al.*, 2021). The vast majority of COVID-19 research published in journals after appearing on bioRxiv or medRxiv did not substantially change in terms of results and conclusions (Brierley, *et al.*, 2022; Nelson, *et al.*, 2022). One interpretation of these results is that preprints were used to disseminate findings at an earlier stage in response to the crisis, reducing the time from discovery to reporting of reliable results — a hypothesis that deserves additional attention.

Preprints increase the potential for visibility earlier in the research cycle. In the case of the COVID-19 pandemic, preprints opened a voluminous conduit for transmitting early research to the public through the news media. Consider three prominent examples. First, a preprint posted on medRxiv on 7 April 2020, showed that the vast majority of COVID-19 infections were transmitted indoors. The report was picked up by hundreds of news outlets (Levenson, *et al.*, 2020) and tweeted almost 20,000 times (according to Altmetrics). It was published in the journal *Indoor Air* months later (Qian, *et al.*, 2021), but by then it had already played a key role in shifting public understanding of how the virus is transmitted (Tufekci, 2020).

In another case, a counterfactual analysis of early non-pharmaceutical interventions to slow the spread of the virus, posted as a preprint on medRxiv, showed that tens of thousands of lives in the U.S. would have been saved by implementing non-pharmaceutical interventions one or two weeks earlier (Pei, *et al.*, 2020a). Posted on the preprint server in May, with data less than one month old, it was reported prominently by the *New York Times*, with dramatic graphics

(Glanz and Robertson, 2020), and hundreds of other outlets (according to Altmetrics). The version that was subsequently published by *Science Advances* (the relatively rapid outlet for *Science*), appeared more than six months later (Pei, *et al.*, 2020b).

Finally, in a notable case of incorrect findings having a potential negative impact when released without peer review, a preprint claiming to find "uncanny similarity" between SARS-CoV-2 and HIV was posted to bioRxiv on 31 January 2020. However, it was withdrawn two days later after other scientists posted public comments identifying errors in its analysis (Pradhan, *et al.*, 2020).

Although many servers subject preprints to screening before posting, this process is usually much less selective than journal peer review. Posting preprints is generally free, and, because of the light screening process, it is inexpensive in terms of time and energy as well. If researchers compete at least in part on volume of outputs, it is likely that the availability of preprint services and the legitimacy of preprints would increase the fraction of research made public. Indeed, as of 2019 approximately 30 percent of papers on bioRxiv did not subsequently appear in a journal (Abdill and Blekhman, 2019). Just under 25 percent of papers related to COVID-19 published during the first six months of the pandemic appeared in a journal within four months (Kodvanj, *et al.*, 2022). This implies that preprints are leading to a greater volume of research being made public than the journal system alone, which is seen as an advantage by those who view the gatekeeping of journals as biased or limiting in undesirable ways (for example by filtering out negative results and contradictory findings). Again, however, we lack the capacity — especially with regard to measuring research output in a consistent way across time — that would be required to evaluate this conclusion.

3.3. Potential unanticipated effects of preprints

As the use of preprints has mushroomed, especially within the life sciences, a wide range of concerns and problems have been raised (Puebla, *et al.*, 2022). Here we categorize these issues according to the scheme used for APCs above: unobserved outcomes, unmeasured characteristics, systemic effects, and long-run dynamics. A sketch of potential unanticipated effects is presented in Figure 5.

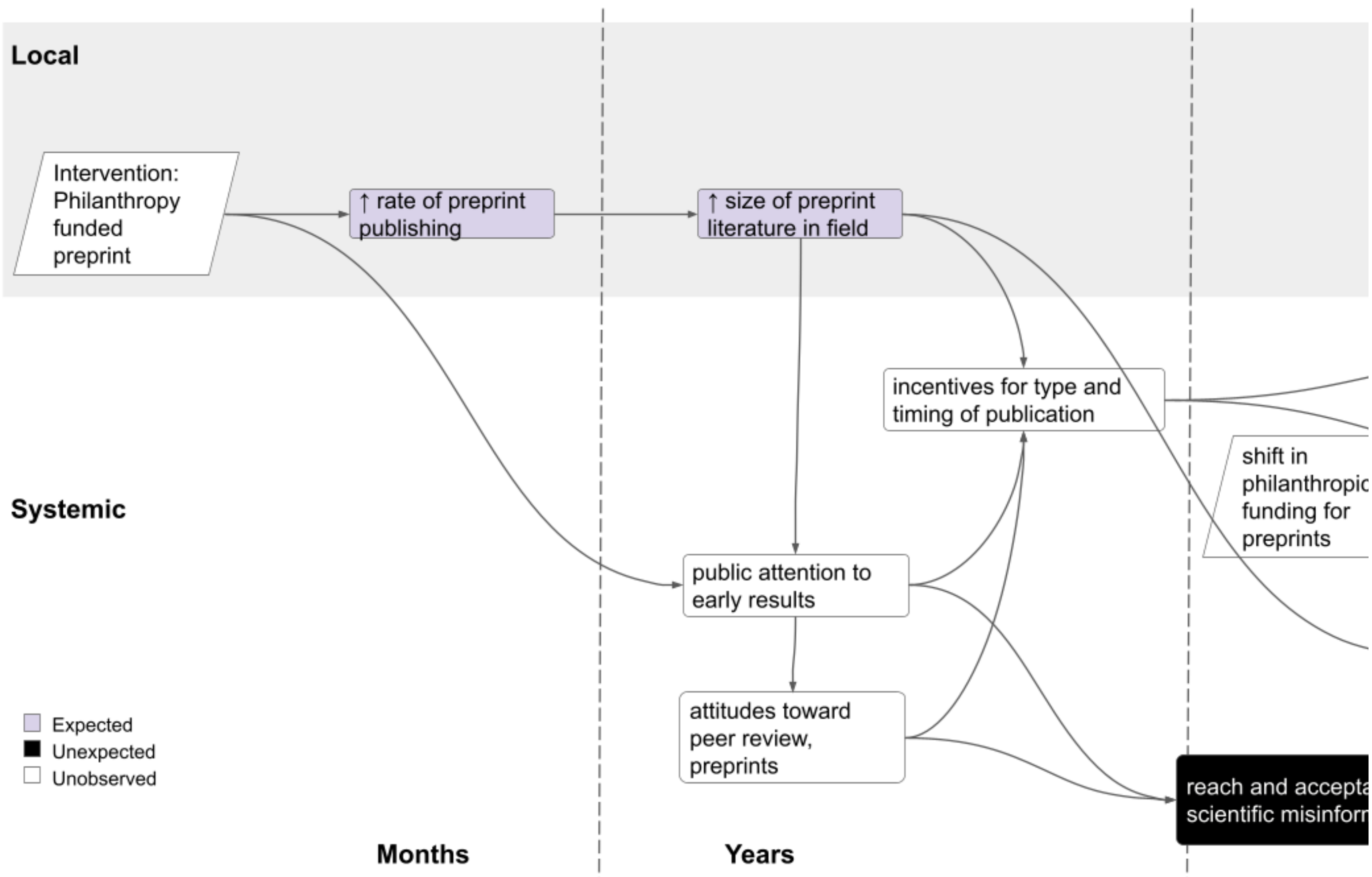


Figure 5: Possible unanticipated effects of preprint servic

While the establishment of preprint servers has increased the rate of preprint publishing and the overall share of scholarly communication available as preprints, it is challenging to assess potential indirect effects of their use. For example, preprint public access raises the possibility of an *unobserved outcome*, that readers who rely on preprints because they're freely available won't read the subsequent journal versions that are behind paywalls (what some publishers label as the "version of record"), which are at least sometimes improved over the preprint version. We might ask, then, whether preprints contribute to the spread of scientific misformation by propagating results that are later rejected through peer review, or reinforce a knowledge gap between those who have access to paywalled journal articles and those who do not.

Assessing the short and long-term implications of preprints is complicated by the likelihood of selection bias among the researchers who post them, which is also unobserved. On one hand, better-resourced scientists may post preprints because they have higher confidence in their early work. But on the other hand, those with fewer resources may use preprints as a way to reach readers with work they can't publish in peer-reviewed journals. The balance of such preferences — which itself may vary across disciplines and fields — remains unknown. It is possible to compare preprint and journal publications of the same paper (*e.g.*, Polka, *et al.*, 2021), but we have no capacity to gather and analyze this systematically.

Assessing any quality of preprints and their potential systemic impact on the body of published science is difficult because they represent a different form of science communication from journal articles. This introduces a set of *unmeasured characteristics* in the system. Thus, if the number of unreliable preprints is added to the numerator in the ratio of unreliable science to all science, and all preprints go in the denominator, what does that rate tell us? If a high proportion of preprints are unreliable that might be a good outcome, if it leads to success in identifying errors and problems early in the publication process, but it exacerbates the problem of making flawed science public. The problem of evaluating the overall quality of science is not new or unique to preprints. Of course, peer-reviewed journal articles have a broad range of reliability despite their categorical status as trustworthy. Nevertheless, preprints present a flood of new information with uncertain implications for the question.

When researchers' behavior with regard to preprints affects others, the result could be unanticipated *systemic effects*. The greater visibility of early work distributed in preprint form may have negative consequences for those who post them. For example, a preprint might alert competitors to the presence of the work, inducing them to change their own research and publication plans. Just as public health interventions can overlook or exploit disadvantaged groups, open science interventions may offer benefits to those who are already well-positioned within the research ecosystem. Preprints place authors in a position of vulnerability because their work is shared early, but in ways for which they might not be institutionally rewarded — and status recognition practices that rely on journal publications exacerbate this risk. In addition, well-established investigators may use preprints to garner visibility as a result of their name recognition, while junior scholars' preprints remain undiscovered (Faulkes, 2016). Thus, preprint posting has the potential to exacerbate systemic inequalities.

Further, the *long-run dynamics* of preprints have not yet fully emerged. Like dentists driven out of business by fluoridation, the sudden availability of public goods (or public information) can create downstream effects that disrupt existing business models. In the case of preprints, they might help challenge the dominance of corporate publishers. If preprint versions of papers are available, the subscription or APC models of generating profits for publishers might be undermined. However, leading publishers seem well prepared for such an eventuality, as they are creating their own preprint and preprint-adjacent services (Springer Nature, n.d.; Cell Press, n.d.; Sage, n.d.).

In the long run, preprint services — operating as free for both authors and readers — do not have a clear path to sustainability beyond the philanthropic funding streams that have allowed them to thrive thus far (Puebla, *et al.*, 2022). Where preprint services do succeed in adapting their business models to recover costs, this will necessarily change the incentives of preprint services. For example, to the extent that preprint services follow the Social Science Research Network (SSRN) model — now owned by Elsevier — they may contribute to the further commercial consolidation of the publishing industry.

3.4. Designing scholarly communications interventions to promote collective learning

Analyses of scholarly communication have generated calls for systemic reforms (*Nature*, 2018; Altman and Bourg, 2018). However, the majority of interventions in open science are launched as iterative, exploratory activities without a systematic treatment or measurement framework. Over the last two decades, the practice of science communication has advanced through the efforts of individual communities and stakeholders to make science more open, transparent, and reliable. Thus, many interventions have been designed and deployed with a primary focus on operational and advocacy concerns — and evaluation of these efforts are primarily observational and retrospective. While this is not a formal research design, it appears as a common design pattern (Alexander, 1979; Pree, 1995), which we label "Do it now, check sometime" (DOINCS).

In the case of preprints, a simple assessment based on the flow of research through major servers signals success of the specific initiatives, but if there are unanticipated systemic consequences, that success might not imply overall benefits to production and dissemination of knowledge. Critics already complain that disseminating research without prior peer review threatens public health (Majumder and Mandl, 2020), in which case the success of the preprint intervention would perversely imply negative effects on a growing scale. Developers of any one intervention aren't responsible for assessing all possible system effects of their work. But their interests are probably broader than a narrow uptake or revenue analysis implies. We need to develop mechanisms for asking and answering questions on a scale somewhere in between the immediate effects of a given intervention and the universe of possible system effects. Funders may not be enthusiastic to support assessments of possible effects that are far removed from an individual project, but the research community has an interest in wider understanding. Successfully addressing such questions will require community-based assessments that different researchers can use to study the impact of specific interventions. Done right, this may also mitigate the problem of self-interested assessments, and allow better documentation of failures (for example, after the staff of an unsuccessful project are dismissed).

Actors in the throes of innovation frequently are called upon to make in-stream modifications of their systems without incorporating comprehensive assessment plans or threat analyses. For example, as the pandemic unfolded, bioRxiv created a collection of papers related to COVID-19 (available at COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv). At first it was a simple list of related papers. However, after accumulating approximately 2,000 articles, they added a disclaimer at the top of the page clarifying that the preprints are not peer reviewed and should not be considered conclusive, a disclaimer that was subsequently further revised [6]. During the pandemic such modifications occurred in the middle of what is also a large, uncontrolled experiment in scientific dissemination, in which the perceptions of reliability and peer review around preprints are being closely scrutinized. Although it is impossible to anticipate all possibilities in complex systems (by their nature), without design principles in place to capture and assess data from an evolving intervention, the opportunity for systematic analysis and understanding (difficult in the best of circumstances) may be lost.

In the case of bioRxiv, fortunately, the service's papers and their metadata are available for systematic analysis, allowing them to be compared to subsequently published versions, and linked to other sources for analysis, such as Retraction Watch's COVID-19 database, Altmetrics, and news media databases. Thus, it may be possible to study the trace effects of specific preprint reporting practices. However, to begin to describe the full dynamics of preprints within the system of scholarly communication would require systematic data from outside the preprint workflow as well, involving, for example, attitudes and behaviors of researchers, institutional responses, the impact of preprints in the public sphere, the effects of preprinting on academic careers, and so on. Such research design and data needs are beyond the scope of this paper.

The history of public health interventions demonstrates that retrospective descriptive analysis is insufficient to yield the depth of causal understanding needed to effect desirable change. Understanding the causal impact of different preprint initiatives will require more than describing the dynamics of the scholarly communications ecosystem over time — it will also require that we be able to link changes in these dynamics to initiatives and interventions. Interventions in public health are designed not only to inform short-term local evaluation needs (*e.g.*, by measuring disease reduction in the treated sample), but also to be comparable with other interventions (*e.g.*, by providing standardized measures of risk reduction, and demographic correlates), to contribute to a global evidence base (*e.g.*, through reporting to a global indicators program), and to provide reliable evidence for evaluating causal effects (*e.g.*, through field experiments and randomization).

4. Discussion and conclusions

The COVID-19 pandemic interrupted many aspects of science and social life simultaneously. By the time it arrived we were already decades into the intervention of preprints in scholarly communication, and still without a systematic understanding of how they have contributed to the changing ecosystem.

Like DDT and the other public health examples mentioned earlier, the APC model of open-access publishing, and preprints, clearly have generated direct effects — as can be measured by their growing presence in the scholarly record — but also indirect and unintended effects that we may not be able to systematically assess. Our conclusion from this review is that although we can learn by doing, if we want deeper understanding we have to design our interventions to be studied, and then study the implementation of those designs. There are recent examples of systemic assessments of open science practices, but these are usually *post hoc* in nature (*e.g.*, Piwowar, *et al.*, 2018; Momeni, *et al.*, 2019). Other proposals for large-scale experimental studies are in the developmental stage (*e.g.*, Nosek, *et al.*, 2020).

Despite these exceptions, interventions in the scholarly ecosystem predominantly follow the pattern of DOINCS. Under favorable conditions, these interventions can provide substantial information about what works in science communication. If eventual analyses are shared, they may offer opportunities for future innovators to learn from past efforts. However, as research in public health suggests, in complex interdependent ecosystems the outcomes from DOINCS often do not generalize, are difficult to compare, and may be misleading.

Complex systems are inherently challenging to manage. As APCs and preprints illustrate, even when there is general agreement that an open science intervention works — for example, through growing uptake — there may be no systematic measurement of its impact, nor a robust understanding of the reasons for its success. The history of public health implies that we can only mitigate unwanted and potentially severe downstream consequences in scholarly communication by developing better systemic understanding of open science interventions. We suggest several broad approaches drawing from this history.

First, identify key measures of practices and outcomes — beyond short-term local program measures such as uptake and revenue — to be used across interventions and assessments. This should include characteristics of the research producers and consumers (such as demographic data and career-status), outputs or outcomes that can be tracked or compared to others over time (such as publication DOIs, measures of readership, and total production cost), and other events or effects (such as retractions, participant effort, and user satisfaction).

Second, develop the infrastructure needed for data collection and sharing these measures across the system, and condition support for new interventions on its use. Many new interventions will benefit from common features — such as linked bibliographic data, usage statistics, or other metadata — and funding for new projects should include contributions to public goods infrastructure to manage and share this information.

Third, build research assessment into the foundation of new initiatives, using the standards of scientific research. For example, those proposing interventions in scholarly communication may be expected to preregister a study of their project even without an experimental design, as has been proposed in observational studies of public health (Eboreime and Abimbola, 2021). This need not limit the range of outcomes or questions to be studied.

Finally, prioritize the construction and deployment of theories of systemic change, making explicit the connections between projects across space and time (Hess and Ostrom, 2006; Sugimoto, 2016). Projects that have the goal of changing scholarly communication should consider their potential contributions — intellectual as well as material — to basic science questions with testable implications. Over time these efforts may cohere into standing bodies or institutes, themselves treated as public goods infrastructure.

In short, scholarly communications needs an open, global data system for monitoring and evaluation, one we believe could be modeled on public health data systems. Such a system would comprise a set of desirable and adverse outcomes, based on comparable measures, with consistent tracking and open data. Under these conditions, each intervention can contribute cumulatively to new knowledge.

Change is needed, and there are real costs to inaction. Scholarly communication in science is too slow (Sarabipour, *et al.*, 2019), too expensive (Wenzler, 2017), and dominated by the concerns of powerful actors — publishers, universities, and governments — in wealthy industrialized nations (Collyer, 2018). In addition, problems of trust undermine public confidence in science (Cohen, 2018). Perfection is not a reasonable goal, and real-world developments can threaten any well-planned intervention, but careful design to capture systemic effects is essential to understanding and improving the impact of our efforts.

About the authors

Authors are listed in alphabetical order.

Micah Altman is a social and information scientist at the Center for Research on Equitable and Open Scholarship at MIT Libraries, Massachusetts Institute of Technology in Cambridge, Mass.
E-mail: escience [at] mit [dot] edu

Philip N. Cohen is Professor of Sociology at the University of Maryland, College Park, and the director of SocArXiv. In addition to open science, he writes about demographic trends, family structure, and health disparities.
E-mail: pnc [at] umd [dot] edu

Jessica Polka is a biochemist and the Executive Director of ASAPbio (Accelerating Science and Publication in biology), a non-profit initiative promoting innovation and transparency via preprints and open peer review.
E-mail: jessica [dot] polka [at] asapbio [dot] org

Acknowledgments

We describe contributions to the paper using a standard taxonomy (see Allen, *et al.* [2014]). All authors take equal responsibility for the article in its current form. All authors collaborated in creating the first draft of the manuscript; All authors contributed to review and revision. All authors contributed to the conception of the article (including core ideas, analytical framework, and statement of research questions). All authors contributed to the project administration and to the writing process through direct writing, critical review, and commentary. We thank Richard Sever, David Weinberger, and members of the CREOS research seminar for comments on earlier drafts.

Competing interests

The authors have no substantial conflicts of interest but wish to thank the Sloan Foundation, Chan-Zuckerberg Initiative, and the U.S. Institute of Museum and Library Services (IMLS) (award LG-250130-OLS-21) for their research support.

Notes

1. Data sharing infrastructure in health and medicine is extensive, including summary indicators databanks (such as provided by the World Bank (https://data.worldbank.org/indicator), OECD (https://stats.oecd.org/), and the World Health Organization's Global Health Observatory (https://www.who.int/data/gho); official statistics (such as the National Vital Statistics System, https://www.cdc.gov/nchs/nvss/index.htm), and open repositories for sharing research data (see, for example, the catalog maintained by the National Library of Medicine, https://www.nlm.nih.gov/NIHbmic/domain_specific_repositories.html).

2. The guiding document on the research ethics, the revised Declaration of Helsinki, https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/, require trial registration, as do most major funders (*e.g.*, registration is one of the 10 EViR guiding principles: https://evir.org/our-principles/), and most medical journals (per the International Committee of Medical Journal E@ditors (ICMJE) requirements https://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html).

3. One such issue is dominance by private interests. For example, in 2020-21 the Gates Foundation contributed more to the World Health Organization than did any other country except Germany — 33 percent more than the U.S. government (World Health Organization, 2022).

4. This brief history should not give the impression that preprints and APCs arrived suddenly to disrupt a previously stable system of scholarly communication. In fact, the practice of journal peer review as commonly recognized today dates only to the 1970s, and thus was only a few decades old when preprints arrived on the scene (Baldwin, 2018).

5. For these counts we add arXiv submissions from https://arxiv.org/stats/monthly_submissions and biorxiv.org (search results) to other preprints listed by Europe PMC at https://europepmc.org/Preprints. The *Web of Science* count reflects all articles listed for each year.

6. The history of the disclaimer is available on the Internet Archive at https://web.archive.org/web/20220000000000*/https://connect.biorxiv.org/relate/content/181.

References

R.J. Abdill and R. Blekhman, 2019. "Tracking the popularity and outcomes of all bioRxiv preprints," *eLife*, volume 8, e45133 (24 April).
doi: https://doi.org/10.7554/eLife.45133, accessed 8 August 2023.

C. AbouZahr and T. Boerma, 2005. "Health information systems: The foundations of public health," *Bulletin of the World Health Organization*, volume 83, number 8, pp. 578–583, and at https://apps.who.int/iris/handle/10665/269465, accessed 8 August 2023.

C. Alexander, 1979. *The timeless way of building*. New York: Oxford University Press.

L. Allen, J. Scott, A. Brand, M. Hlava, and M. Altman, 2014. "Publishing: Credit where credit is due," *Nature*, volume 508, number 7496 (16 April), pp. 312–313.
doi: https://doi.org/10.1038/508312a, accessed 8 August 2023.

M. Alsan and M. Wanamaker, 2018. "Tuskegee and the health of Black men," *Quarterly Journal of Economics*, volume 133, number 1, pp. 407–455.
doi: https://doi.org/10.1093/qje/qjx029, accessed 8 August 2023.

M. Altman, 2022. "Designing community tracking indicators for open and inclusive scholarship," *Proceedings of the Association for Information Science and Technology*, volume 59, number 1, pp. 393–397.
doi: https://doi.org/10.1002/pra2.640, accessed 8 August 2023.

M. Altman and C. Bourg, 2018. "A grand challenges-based research agenda for scholarly communication and information science" (17 December), at https://grandchallenges.pubpub.org/pub/final, accessed 19 December 2019.
doi: https://doi.org/10.21428/62b3421f, accessed 8 August 2023.

ASAPbio, 2021. "Funder policies," at https://asapbio.org/funder-policies, accessed 15 January 2021.

M. Baldwin, 2018. "Scientific autonomy, public accountability, and the rise of 'peer review' in the cold war United States," *Isis*, volume 109, number 3, pp. 538–558.
doi: https://doi.org/10.1086/700070, accessed 8 August 2023.

A.V. Banerjee and E. Duflo (editors), 2017. *Handbook of economic field experiments*. Amsterdam: North-Holland.

J. Beard, 2006. "DDT and human health," *Science of the Total Environment*, volume 355, numbers 1–3, pp. 78–89.
doi: https://doi.org/10.1016/j.scitotenv.2005.02.022, accessed 8 August 2023.

G.S. Birkhead, C.B. Morrow, and S. Pirani, 2022. *Turnock's public health: What it is and how it works*. Seventh edition. Burlington, Mass.: Jones & Bartlett Learning.

J. Brainard, 2020. "For €9500, *Nature* journals will now make your paper free to read," *Science* (24 November), at https://www.science.org/content/article/9500-nature-journals-will-now-make-your-paper-free-read, accessed 25 April 2022.
doi: https://doi.org/10.1126/science.abf8491, accessed 8 August 2023.

L. Brierley, F. Nanni, J.K. Polka, G. Dey, M. Pálfy, N. Fraser, and J.A. Coates, 2022. "Tracking changes between preprint posting and journal publication during a pandemic," *PLoS Biology*, volume 20, number 2, e3001285 (1 February).
doi: https://doi.org/10.1371/journal.pbio.3001285, accessed 8 August 2023.

Budapest Open Access Initiative, 2002. "Read the Declaration," at https://www.budapestopenaccessinitiative.org/read/, accessed 5 April 2022.

E. Callaway, 2020. "Will the pandemic permanently alter scientific publishing?" *Nature*, volume 582, number 7811 (11 June), pp. 167–168.
doi: https://doi.org/10.1038/d41586-020-01520-4, accessed 8 August 2023.

R. Carson, 1962. *Silent spring*. Greenwich, Conn.: Fawcett Publications.

M.M. Cassell, D.T. Halperin, J.D. Shelton, and D. Stanton, 2006. "Risk compensation: The Achilles' heel of innovations in HIV prevention?" *BMJ*, volume 332, number 7541 (9 March), pp. 605–607.
doi: https://doi.org/10.1136/bmj.332.7541.605, accessed 8 August 2023.

Cell Press, n.d. "Cell Press sneak peek," at https://www.ssrn.com/index.cfm/en/cell-press-sneak-peek/, accessed 15 July 2021.

Centers for Disease Control and Prevention, 2022. "National public health performance standards," at https://www.cdc.gov/publichealthgateway/nphps/, accessed 10 October 2022.

Centers for Disease Control and Prevention, 1999. "Framework for program evaluation in public health, 1999," *Morbidity and Mortality Weekly Report: Recommendations and Reports*, volume 48, number RR-11 (17 September), pp. 1–40, and at https://www.cdc.gov/evaluation/framework/index.htm, accessed 8 August 2023.

B.C.K. Choi, 2012. "The past, present, and future of public health surveillance," *Scientifica*, volume 2012, e875253.
doi: https://doi.org/10.6064/2012/875253, accessed 8 August 2023.

P.N. Cohen, 2018. "The widening political divide over science" (14 December).
doi: https://doi.org/10.31235/osf.io/u95aw, accessed 8 August 2023.

F.M. Collyer, 2018. "Global patterns in the publishing of academic knowledge: Global North, global South," *Current Sociology*, volume 66, number 1, pp. 56–73.
doi: https://doi.org/10.1177/0011392116680020, accessed 8 August 2023.

Committee on Toward an Open Science Enterprise, Board on Research Data and Information, Policy and Global Affairs, and National Academies of Sciences, Engineering, and Medicine, 2018. *Open science by design: Realizing a vision for 21st century research*. Washington, D.C.: National Academies Press, and at https://www.nap.edu/catalog/25116, accessed 19 December 2019.

Demographic and Health Surveys, 2022. "Survey process," at https://dhsprogram.com/Methodology/survey-process.cfm, accessed 10 October 2022.

E.A. Eboreime and S. Abimbola, 2021. "A priori registration of global health research — Necessity or absurdity?" *BMJ Global Health*, volume 6, number 5, e006199.
doi: http://dx.doi.org/10.1136/bmjgh-2021-006199, accessed 8 August 2023.

eLife, 2022. "eLife's new model: Changing the way you share your research" (20 October), at https://elifesciences.org/inside-elife/54d63486/elife-s-new-model-changing-the-way-you-share-your-research, accessed 3 April 2023.

Elsevier, 2022. "Elsevier journal and article ecosystem," at https://www.elsevier.com/__data/assets/pdf_file/0010/616474/Journal-and-article-ecosystem.pdf, accessed 8 August 2023.

Z. Faulkes, 2016. "A pre-print experiment: Will anyone notice?" *NeuroDojo* (17 March), at http://neurodojo.blogspot.com/2016/03/a-pre-print-experiment-will-anyone.html, accessed 17 August 2021.

N. Fraser, L. Brierley, G. Dey, J.K. Polka, M. Pálfy, F. Nanni, and J.A. Coates, 2021. "The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape," *PLoS Biology*, volume 19, number 4 (2 April), e30000959.
doi: https://doi.org/10.1371/journal.pbio.3000959, accessed 8 August 2023.

J.A. García, R. Rodriguez-Sánchez, and J. Fdez-Valdivia, 2021. "Quality censoring in peer review," *Scientometrics*, volume 126, number 1, pp. 825–830.
doi: https://doi.org/10.1007/s11192-020-03693-1, accessed 8 August 2023.

P. Ginsparg, 2021. "Lessons from arXiv's 30 years of information sharing," *Nature Reviews Physics*, volume 3, number 9 (4 August), pp. 602–603.
doi: https://doi.org/10.1038/s42254-021-00360-z, accessed 8 August 2023.

J. Glanz and C. Robertson, 2020. "Lockdown delays cost at least 36,000 lives, data show," *New York Times* (20 May), at https://www.nytimes.com/2020/05/20/us/coronavirus-distancing-deaths.html, accessed 21 September 2021.

K. Glanz and D.B. Bishop, 2010. "The role of behavioral science theory in development and implementation of public health interventions," *Annual Review of Public Health*, volume 31, pp. 399–418.
doi: https://doi.org/10.1146/annurev.publhealth.012809.103604, accessed 8 August 2023.

J.M. Green, 2018. "The rise and future of glyphosate and glyphosate-resistant crops," *Pest Management Science*, volume 74, number 5, pp. 1,035–1,039.
doi: https://doi.org/10.1002/ps.4462, accessed 8 August 2023.

R. Hammonds, J. Hanefeld, and G. Ooms, 2019. "Accountability as a driver of health equity," *World Health Organization Regional Office for Europe*, at https://apps.who.int/iris/bitstream/handle/10665/312282/9789289054096-eng.pdf, accessed 26 July 2022.

J. Hedlund, 2000. "Risky business: Safety regulations, risk compensation, and individual behavior," *Injury Prevention*, volume 6, number 2, pp. 82–89.
doi: https://dx.doi.org/10.1136/ip.6.2.82, accessed 8 August 2023.

C. Hess and E. Ostrom (editors), 2006. *Understanding knowledge as a commons: From theory to practice*. Cambridge, Mass.: MIT Press.
doi: https://doi.org/10.7551/mitpress/6980.001.0001, accessed 8 August 2023.

K. Ho and M. Neidell, 2009. "Equilibrium effects of public goods: The impact of community water fluoridation on dentists," *National Bureau of Economic Research, Working Paper*, number 15056.
doi: https://dx.doi.org/10.3386/w15056, accessed 8 August 2023.

L.A. Jorgenson, C.D. Wolinetz, and F.S. Collins, 2021. "Incentivizing a new culture of data stewardship: The NIH policy for data management and sharing," *Journal of the American Medical Association*, volume 326, number 22 (4 November), pp. 2259–2,260.
doi: https://dx.doi.org/10.1001/jama.2021.20489, accessed 8 August 2023.

R. Kiley, T. Peatfield, J. Hansen, and F. Reddington, 2017. "Data sharing from clinical trials — A research funder's perspective," *New England Journal of Medicine*, volume 377, number 20 (16 November), pp. 1,990–1,992.
doi: https://dx.doi.org/10.1056/NEJMsb1708278, accessed 8 August 2023.

S.-J. Kim and K.S. Park, 2021. "Open access status of journals and articles in Journal Citation Reports," *Science Editing*, volume 8, number 1, pp. 26–31.
doi: https://doi.org/10.6087/kcse.226, accessed 8 August 2023.

J.J. Kirkham, N. Penfold, F. Murphy, I. Boutron, J.P. Ioannidis, J. Polka, and D. Moher, 2020. "Systematic examination of preprint platforms for use in the medical and biomedical sciences setting," *BMJ Open*, volume 10, number 12, e041849.

doi: http://dx.doi.org/10.1136/bmjopen-2020-041849, accessed 8 August 2023.

I. Kodvanj, J. Homolak, D. Virag, and and V. Trkulja, 2022. "Publishing of COVID-19 preprints in peer-reviewed journals, preprinting trends, public discussion and quality issues," *Scientometrics*, volume 127, number 3, pp. 1,339–1,352.
doi: https://dx.doi.org/10.1007/s11192-021-04249-7, accessed 8 August 2023.

D. Koo and S.F. Wetterhall, 1996. "History and current status of the national notifiable diseases surveillance system," *Journal of Public Health Management & Practice*, volume 2, number 4, pp. 4–10.
doi: https://dx.doi.org/10.1097/00124784-199623000-00004, accessed 8 August 2023.

K. Kupferschmidt, 2016. "After 40 years, the most important weapon against mosquitoes may be failing," *Science* (13 October), at https://www.sciencemag.org/news/2016/10/after-40-years-most-important-weapon-against-mosquitoes-may-be-failing, accessed 5 August 2021.

M. Laakso and B.-C. Björk, 2012. "Anatomy of open access publishing: A study of longitudinal development and internal structure," *BMC Medicine*, volume 10, article number 124.
doi: https://dx.doi.org/10.1186/1741-7015-10-124, accessed 8 August 2023.

M. Levenson, T. Parker-Pope, and J. Gorman, 2020. "What we know about your chances of catching the virus outdoors," *New York Times* (15 May), at https://www.nytimes.com/2020/05/15/us/coronavirus-what-to-do-outside.html, accessed 9 February 2021.

M.L.H. Mabaso, B. Sharp, and C. Lengeler, 2004. "Historical review of malarial control in southern African with emphasis on the use of indoor residual house-spraying," *Tropical Medicine & International Health*, volume 9, number 8, pp. 846–856.
doi: https://doi.org/10.1111/j.1365-3156.2004.01263.x, accessed 8 August 2023.

M.S. Majumder and K.D. Mandl, 2020. "Early in the epidemic: Impact of preprints on global discourse about COVID-19 transmissibility," *Lancet Global Health*, volume 8, number 5, pp. e627–e630.
doi: https://dx.doi.org/10.1016/S2214-109X(20)30113-3, accessed 8 August 2023.

L. McKenzie, 2021. "Big deal for open access," *Inside Higher Ed* (16 March), at https://www.insidehighered.com/news/2021/03/17/university-california-reaches-new-open-access-agreement-elsevier, accessed 15 March 2022.

D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, and the PRISMA Group, 2009. "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *Annals of Internal Medicine*, volume 151, number 4 (18 August), p. 264.
doi: https://doi.org/10.7326/0003-4819-151-4-200908180-00135, accessed 8 August 2023.

F. Momeni, S. Dietze, P. Mayr, K. Biesenbender, and I. Peters, 2022. "Which factors drive open access publishing? A Springer Nature case study," *arXiv*:2208.08221at (17 August).
doi: https://doi.org/10.48550/arXiv.2208.08221, accessed 8 August 2023.

F. Momeni, N. Fraser, I. Peters, and P. Mayr, 2019. "From closed to open access: A case study of flipped journals," *arXiv*:1903.11682 (27 March).
doi: https://doi.org/10.48550/arXiv.1903.11682, accessed 8 August 2023.

Nature, 2018. "Challenges in irreproducible research" (18 October), at https://www.nature.com/collections/prbfkwmwvz, accessed 29 September 2020.

L. Nelson, H. Ye, A. Schwenn, S. Lee, S. Arabi, and B.I. Hutchins, 2022. "Robustness of evidence reported in preprints during peer review," *Lancet Global Health*, volume 10, number 11, pp. e1684–e1687.
doi: https://doi.org/10.1016/S2214-109X(22)00368-0, accessed 8 August 2023.

A. Nishikawa-Pacher, T. Heck, and K. Schoch, 2022. "Open editors: A dataset of scholarly journals' editorial board positions," *Research Evaluation*, rvac037 (4 October).
doi: https://doi.org/10.1093/reseval/rvac037, accessed 8 August 2023.

B.A. Nosek, D.T. Mellor, T.M. Errington, C.K. Soderberg, T. Krall, F.L. Grasty, R.E. Brooks Iii, R. Davis, S. Kousta, S. Swaminathan, V. Kiermer, J. hompson, W.B. Mendes, D. Heim, D.B. Samuel, N. Bansal, J. Wolfe, R.E. Lucas, C. Chambers, D. Rapp, M. Back, M. Arterberry, I. Nikolaou, C.C. Morey, K. Arceneaux, and J.W.B. Lang, 2020. "NSF Science of Science proposal: Randomized trial of registered reports" (10 February), at https://osf.io/preprints/metaarxiv/jvxab/, accessed 2 May 2022.
doi: https://doi.org/10.31222/osf.io/jvxab, accessed 8 August 2023.

B.A. Nosek, G. Alter, G.C. Banks, D. Borsboom, S.D. Bowman, S.J. Breckler, S. Buck, C.D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D.P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B.A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E.J. Wagenmakers, R. Wilson, and T. Yarkoni, 2015. "Promoting an open research culture," *Science*, volume 348, number 6242 (26 June), pp. 1,422–1,425.
doi: https://doi.org/10.1126/science.aab2374, accessed 8 August 2023.

P.T. O'Shaughnessy, 2008. "Parachuting cats and crushed eggs; The controversy over the use of DDT to control malaria," *American Journal of Public Health*, volume 98, number 11, pp. 1,940–1,948.
doi: https://doi.org/10.2105/AJPH.2007.122523, accessed 8 August 2023.

D. Parsons, D. Willis, and J. Holland, 2011. "Benefits to the private sector of open access to higher education and scholarly research," *UK Open Access Implementation Group*.

PeerJ Staff, 2013. "'PeerJ PrePrints' — a new Preprint Server for the Biological and Medical Sciences," *PeerJ blog* (3 April), at https://peerj.com/blog/post/47030855181/peerj-preprints-a-new-preprint-server-for-the-biological-and-medical-sciences/, accessed 4 June 2020.

S. Pei, S. Kandula, and J. Shaman, 2020a. "Differential effects of intervention timing on COVID-19 spread in the United States," *medRxiv* (29 May).
doi: https://doi.org/10.1101/2020.05.15.20103655, accessed 8 August 2023.

S. Pei, S. Kandula, and J. Shaman, 2020b. "Differential effects of intervention timing on COVID-19 spread in the United States," *Science Advances*, volume 6, number 49 (4 December).
doi: https://doi.org/10.1126/sciadv.abd6370, accessed 8 August 2023.

H. Piwowar, J. Priem, V. Larivière, J.P. Alperin, L. Matthias, B. Norlander, A. Farley, J. West, and S. Haustein, 2018. "The state of OA: A large-scale analysis of the prevalence and impact of open access articles," *PeerJ*, volume 6, e4375.
doi: https://doi.org/10.7717/peerj.4375, accessed 8 August 2023.

J.K. Polka, G. Dey, M. Pálfy, F. Nanni, L. Brierley, N. Fraser, and J.A. Coates, 2021. “Preprints in motion: Tracking changes between posting and journal publication,” *bioRxiv* (20 February).
doi: https://doi.org/10.1101/2021.02.20.432090, accessed 8 August 2023.

R. Poynder, 2019. &dquo;Open access publishing — Beyond Plan S for the Global South,” *University World News* (27 April), at https://www.universityworldnews.com/post.php?story=20190422104333956, accessed 8 August 2023.

P. Pradhan, A.K. Pandey, A. Mishra, P. Gupta, P.K. Tripathi, M.B. Menon, J. Gomes, P. Vivekanandan, and B. Kundu, 2020. “Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag,” *bioRxiv* (31 January).
doi: https://doi.org/10.1101/2020.01.30.927871, accessed 8 August 2023.

W. Pree, 1995. *Design patterns for object-oriented software development*. Reading, Mass: Addison-Wesley.

G.D. Price and S. Chaytor, 2015. “Open access: Is a national licence the answer?” *Higher Education Policy Institute (HEPI), Occasional Paper*, number 10, at https://www.hepi.ac.uk/2015/03/30/open-access-national-licence-answer/, accessed 8 August 2023.

J. Priem, H. Piwowar, and R. Orr, 2022. “OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts,” *arXiv*:2205.01833 (4 May).
doi: https://doi.org/10.48550/arXiv.2205.01833, accessed 8 August 2023.

Public Library of Science, 2003. “PLoS launches *PLoS Biology*: New open-access journal will increase access to scientific research and speed scientific discovery” (13 October), at https://web.archive.org/web/20031016063325/https://plos.org/news/announce_pbiolaunch.html, accessed 23 November 2021.

I. Puebla, J. Polka, and O.Y. Rieger, 2022. *Preprints: Their evolving role in science communication*. Charleston, S.C.: Against the Grain Press.
doi: https://doi.org/10.3998/mpub.12412508, accessed 8 August 2023.

H. Qian, T. Miao, L. Liu, X. Zheng, D. Luo, and Y, Li, 2021. “Indoor transmission of SARS-CoV-2,” *Indoor Air*, volume 31, number 3, pp. 639–645.
doi: https://doi.org/10.1111/ina.12766, accessed 8 August 2023.

Retraction Watch, n.d. “The Center For Scientific Integrity,” at https://retractionwatch.com/the-center-for-scientific-integrity/, accessed 8 August 2023.

K.J. Rothman, T.L. Lash, T.J. VanderWeele, and S. Haneuse, 2021. *Modern epidemiology* Fourth edition. Philadelphia: Wolters Kluwer.

S. Rumsey, 2021. “cOAlition S funders’ repository requirements: Unity in diversity • cOAlition S and repositories (part I),” *sOApbox* (1 July), at https://www.coalition-s.org/blog/coalition-s-and-repositories-part1/, accessed 3 April 2023.

Sage, n.d. “advance: Discover research from Advance,” at https://advance.sagepub.com/, accessed 15 July 2021.

S. Sarabipour, H.J. Debat, E. Emmott, S.J. Burgess, B. Schwessinger, and Z. Hensel, 2019. “On the value of preprints: An early career researcher perspective,” *PLoS Biology*, volume 17, number 2, e3000151 (21 february).
doi: https://doi.org/10.1371/journal.pbio.3000151, accessed 8 August 2023.

N. Schönfelder, 2020. “Article processing charges: Mirroring the citation impact or legacy of the subscription-based model?” *Quantitative Science Studies*, volume 1, number 1, pp. 6–27.
doi: https://doi.org/10.1162/qss_a_00015, accessed 8 August 2023.

R. Sever, T. Roeder, S. Hindle, L. Sussman, K.-J. Black, J. Argentine, W. Manos, and J.R. Inglis, 2019. “bioRxiv: The preprint server for biology” (6 November).
doi: https://doi.org/10.1101/833400, accessed 8 August 2023.

C. Shen and B.-C. Björk, 2015. “‘Predatory’ open access: A longitudinal study of article volumes and market characteristics,” *BMC Medicine*, volume 13, article number 230.
doi: https://dx.doi.org/10.1186/s12916-015-0469-2, accessed 8 August 2023.

R. Short, R. Gurung, M. Rowcliffe, N. Hill, and E.J. Milner-Gulland, 2018. “The use of mosquito nets in fisheries: A global perspective,” *PLoS ONE*, volume 13, number 1, e0191519 (31 January).
doi: https://doi.org/10.1371/journal.pone.0191519, accessed 8 August 2023.

K. Siler, 2020. “Demarcating spectrums of predatory publishing: Economic and institutional sources of academic legitimacy,” *Journal of the Association for Information Science and Technology*, volume 71, number 11, pp. 1,386–1,401.
doi: https://doi.org/10.1002/asi.24339, accessed 8 August 2023.

M. Smith, I. Anderson, B.-C. Bjork, M. McCabe, D. Solomon, G. Tananbaum, C. Tenopir, and M. Willmott, 2016. “Pay it forward: Investigating a sustainable model of open access article processing charges for large north american research institutions” (30 June), at https://escholarship.org/uc/item/8326n305#, accessed 15 July 2021.

Springer Nature, n.d. “In Review: Journal-integrated preprint sharing from Springer Nature and Research Square,” at https://www.springernature.com/gp/authors/campaigns/in-review, accessed 15 July 2021.

C.R. Sugimoto (editor), 2016. *Theories of informetrics and scholarly communication*. Berlin: De Gruyter Saur.
doi: https://doi.org/10.1515/9783110308464, accessed 8 August 2023.

L. Teytelman, A. Stoliartchouk, L. Kindler, and B.L. Hurwitz, 2016. “Protocols.io: Virtual communities for protocol development and discussion,” *PLoS Biology*, volume 14, number 8, e1002538 (22 August).
doi: https://doi.org/10.1371/journal.pbio.1002538, accessed 8 August 2023.

S.B. Thacker and D.F. Stroup, 1994. “Future directions for comprehensive public health surveillance and health information systems in the United States,” *American Journal of Epidemiology*, volume 140, number 5, pp. 383–397.
doi: https://doi.org/10.1093/oxfordjournals.aje.a117261, accessed 8 August 2023.

K. Tingley, 2020. “Coronavirus is forcing medical research to speed up,” *New York Times* (21 April), at https://www.nytimes.com/2020/04/21/magazine/coronavirus-scientific-journals-research.html, accessed 21 May 2020.

Keita Tsuji, 2018. “Statistics on open access books available through the Directory of Open Access Books,” *International Journal of Academic Library and Information Science*, volume 6, number 4, pp. 86–100.

C.W. Tubbs, 2016. “California condors and DDT: Examining the effects of endocrine disrupting chemicals in a critically endangered species,” *Endocrine Disruptors*, volume 4, number 1, e1173766.
doi: https://doi.org/10.1080/23273747.2016.1173766, accessed 8 August 2023.

Z. Tufekci, 2020. “We need to talk about ventilation,” *Atlantic* (30 July), at https://www.theatlantic.com/health/archive/2020/07/why-arent-we-talking-more-about-airborne-transmission/614737/, accessed 9 February 2021.

U.S. Institute of Medicine, 2015. *Sharing clinical trial data: Maximizing benefits, minimizing risk*. Washington, D.C.: National Academies Press.
doi: https://doi.org/10.17226/18998, accessed 8 August 2023.

T.W. Valente and S.R. Pitts, 2017. “An appraisal of social network theory and analysis as applied to public health: Challenges and opportunities,” *Annual Review of Public Health*, volume 38, pp. 103–118.
doi: https://doi.org/10.1146/annurev-publhealth-031816-044528, accessed 8 August 2023.

H. van den Berg, 2009. “Global status of DDT and its alternatives for use in vector control to prevent disease,” *Environmental Health Perspectives*, volume 117, number 11, pp. 1,656–1,663.
doi: https://doi.org/10.1289/ehp.0900785, accessed 8 August 2023.

H. van den Berg, G. Manuweera, and F. Konradsen, 2017. “Global trends in the production and use of DDT for control of malaria and other vector-borne diseases,” *Malaria Journal*, volume 16, number 1, article number 401.
doi: https://doi.org/10.1186/s12936-017-2050-2, accessed 8 August 2023.

C.G. Victora, J.-P. Habicht, and J. Bryce, 2004. “Evidence-based public health: Moving beyond randomized trials,” *American Journal of Public Health*, volume 94, number 3, pp. 400–405.
doi: https://doi.org/10.2105/ajph.94.3.400, accessed 8 August 2023.

Wellcome, 2020. “Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak” (31 January), at https://wellcome.org/press-release/sharing-research-data-and-findings-relevant-novel-coronavirus-ncov-outbreak, accessed 9 November 2021.

J. Wenzler, 2017. “Scholarly communication and the dilemma of collective action: Why academic journals cost too much,” *College & Research Libraries*, volume 78, number 2.
doi: https://doi.org/10.5860/crl.78.2.183, accessed 8 August 2023.

M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. Bonino da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. ’t Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, 2016. “The FAIR Guiding Principles for scientific data management and stewardship,” *Scientific Data*, volume 3, number 1, article number 160018.
doi: https://doi.org/10.1038/sdata.2016.18, accessed 8 August 2023.

World Health Organization, 2022. “Programme budget Web portal,” at https://open.who.int/2020-21/contributors/contributor, accessed 10 October 2022.

World Health Organization, 2020. “World malaria report 2020” (30 November), at https://www.who.int/publications-detail-redirect/9789240015791, accessed 12 January 2021.

W. Yan, 2020. “Coronavirus tests science’s need for speed limits,” *New York Times* (14 April), at https://www.nytimes.com/2020/04/14/science/coronavirus-disinformation.html, accessed 28 May 2020.

Editorial history

Received 30 January 2023; revised 23 May 2023; accepted 8 August 2023.

This paper is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Interventions in scholarly communication: Design lessons from public health
by Micah Altman, Philip N. Cohen, and Jessica Polka.
First Monday, Volume 28, Number 8 - 7 August 2023
https://firstmonday.org/ojs/index.php/fm/article/download/12941/11295
doi: https://dx.doi.org/10.5210/fm.v28i8.12941