

ORIGINAL RESEARCH Cohort study

Evaluation of editors' abilities to predict the citation potential of research: manuscripts submitted to The BMJ

Sara Schroter,¹ Wim E J Weber,¹ Elizabeth Loder,¹ Jack Wilkinson,² Jamie J Kirkham²

Objective To evaluate the ability of *BMJ* editors to predict the number of times submitted research manuscripts will be cited.

Design Cohort study.

Setting Manuscripts submitted to *The BMJ*, reviewed, and scheduled for discussion at a prepublication meeting.

Participants 10 *BMJ* research editors.

Main outcome measures Reviewed manuscripts were rated independently by attending editors for citation potential in the year of first publication plus the next year: no citations, below average (<10 citations), average (10-17), or above average (>17). Predicted citations were subsequently compared with actual citations extracted from Web of Science (WOS).

Results Of 534 manuscripts reviewed, 505 were published (219 in *The BMJ*) and indexed in WOS, 22 were unpublished, and one was withdrawn. Among the 505 manuscripts, 277 (55%) were cited <10 times, 105 (21%) were cited 10-17 times, and 123 (24%) were cited >17 times. Manuscripts accepted by *The BMJ* were cited more highly (median 12 (interquartile range 7-24) citations) than those rejected (median 7 (3-12) citations). For all 10 editors, predicted ratings tended to increase in line with actual citations, but with considerable variation within categories; nine failed to identify the correct citation category for >50% (range 31-52%) of manuscripts. Editors more often rated papers that achieved high actual citation counts as having low citation potential than the reverse. Collectively, for 160 (32%) manuscripts at least 50% of editors predicted the right category.

Conclusions When it comes to *BMJ* editors, there is no wisdom of the crowd.

Introduction

"Impact factor mania" is a "debilitating and destructive epidemic."¹⁻⁵ One of the criticisms of the impact factor is that it is easy to game,⁶⁻⁸ and editors select papers that will attract numerous citations.⁸ *The BMJ*'s impact factor rose from 5 in 2000 to 96 in 2021, suggesting that our editors are quite brilliant in predicting citability of unpublished research. To prove our own brilliance scientifically, we tested our ability to predict citations of unpublished research papers submitted to *The BMJ*.

Methods

At *The BMJ*, research papers with favourable external peer reviews are discussed at a weekly meeting. Between 27 August 2015 and 29 December 2016, the research editors were invited to independently assess the citation potential of manuscripts for discussion at these meetings. They were asked to indicate how many citations they thought each manuscript would generate in the first year of publication plus the next year: no citations, below average (<10 citations), around average (10-17), or more than average (>17). Our sample size was based on when editors seemed bored with this weekly request.

Participants comprised a self-selected sample of 10 fiercely competitive, paid *BMJ* research editors. We excluded the paid statistical advisors who attended meetings in case they were better than us.

For articles rejected from *The BMJ*, we tried to match submissions with subsequent publications. All searches were completed by 10 May 2022, when we extracted citation data from Web of Science (WOS).

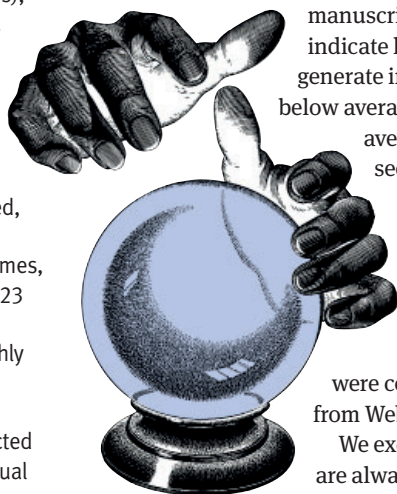
We excluded submissions for *The BMJ*'s Christmas issue as these are always brilliant, those for which we could find no publication, those not indexed in WOS, those not published as full length journal articles, and those published after 2019.

Statistical analysis

For each editor, we calculated the number of manuscripts they assigned to the correct citation category and κ statistics (95% confidence intervals), using Fleiss-Cohen weights,²² as implemented in the R package *vcd*.²³ We calculated how often each editor's classification was "extremely wrong" (a highly cited paper was estimated to have low citation potential or vice versa). We compared the editors' ability to predict citation potential for articles published in *The BMJ* versus elsewhere. To assess the editors as a group, we calculated the mean percentage of editors identifying the correct category per manuscript, and the number and percentage of manuscripts for which at least 50% of the editors were correct.

Patient and public involvement

No patients were specifically involved in this study as we doubt many would lose sleep over citations of research papers.



¹BMJ, BMA House, London

²University of Manchester, Manchester Academic Health Science Centre

Correspondence to: S Schroter sschroter@bmj.com

Cite this as: *BMJ* 2022;379:e073880

Find this at doi: 10.1136/bmj-2022-073880

Results

Of 534 manuscripts discussed, 23 were excluded because we could find no subsequent publication ($n=18$), or only an abstract or preprint ($n=3$), or a substantially different article ($n=1$), or a withdrawn abstract ($n=1$). A further six were excluded as the journal of publication was not indexed in WOS ($n=4$) or publication was after 2019 ($n=2$). Among the 505 eligible manuscript-publication pairs, 219 (43%) were published in *The BMJ*. Accepted manuscripts were cited more highly (median 12 (interquartile range 7-24) citations) than those rejected (7 (3-12) citations).

Ability to estimate actual citations

Of the 505 papers, 277 (55%) generated <10 citations, 105 (21%) generated 10-17, and 123 (24%) generated >17. For all 10 editors, predicted ratings tended to increase in line with actual citations, but with considerable variation within categories (figure). Nine editors were unable to identify the correct citation category for >50% (range 31-54%) of manuscripts. Agreement between estimated and actual categories for all editors ranged from $\kappa=0.01$ to 0.19. It was uncommon for editors to rate papers with low actual citation counts as having high citation potential—more usually it was the reverse.

None of the editors predicted the citation category correctly for $\geq 50\%$ of *The BMJ* articles (range 35-45%). For non-*BMJ* articles, two editors rated $\geq 50\%$ correctly (26-63%).

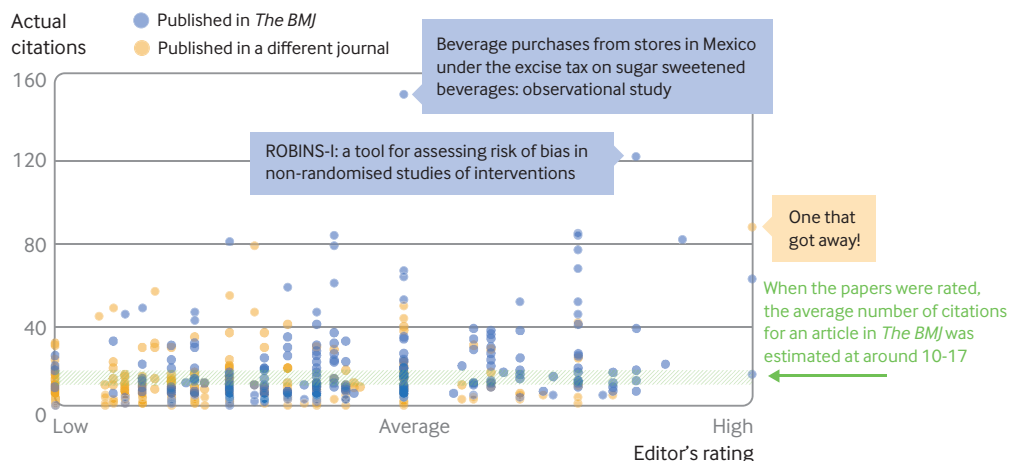
For the collective assessment of editors across all 505 manuscripts, the mean percentage of editors predicting the correct citation category was 43%, and for only 160 (32%) manuscripts at least 50% of editors predicted the right category (table).

WHAT IS ALREADY KNOWN ON THIS TOPIC

- Impact factor mania is a common disorder, and severely affected journal editors might be tempted to accept only highly citable research manuscripts

WHAT THIS STUDY ADDS

- The findings suggest that *The BMJ* editors are not good at predicting the citation potential of accepted or rejected manuscripts
- When it comes to *The BMJ* editors, there is no wisdom of the crowd



BMJ editors' collective assessment of manuscripts

| | Mean percentage of correct ratings per manuscript | No (%) of manuscripts where $\geq 50\%$ of editors predicted correct category |
|--|---|---|
| All manuscripts ($n=505$) | 43 | 160 (32) |
| Citation category: | | |
| Manuscripts with low citations ($n=277$) | 55 | 133 (48) |
| Manuscripts with average citations ($n=105$) | 34 | 17 (16) |
| Manuscripts with high citations ($n=123$) | 23 | 10 (8) |
| Manuscript decision: | | |
| Published in <i>The BMJ</i> ($n=219$) | 36 | 44 (20) |
| Rejected by <i>The BMJ</i> ($n=286$) | 47 | 116 (41) |
| Superstar manuscripts: | | |
| With >30 citations ($n=55$) | 31 | 8 (15) |
| With >50 citations ($n=20$) | 39 | 6 (30) |

Number of actual citations are those in the year of publication and the following calendar year. Overall editors' ratings for each paper represent the average score from all editors who rated the papers. Higher ratings denote that most editors thought a paper would be more highly cited

Discussion

The accuracy of our editors' predictions for citation potential was generally poor for papers rejected and accepted for publication in *The BMJ*. Editors did predict the citation potential of manuscripts they went on to accept as higher than those they rejected. While this could indicate that predicted citation potential influences editorial decisions, it could also reflect the methodological quality and importance of the accepted articles.²⁵ Editors were not good at predicting citation potential either alone or as a committee (no wisdom of the crowd when it comes to *BMJ* editors).

Strengths and limitations of this study

Our study has several potential limitations. Identification of published articles can be difficult as article titles often change substantially from submission, but we made use of all the additional submission data authors needlessly have to upload to our database on first submission to help with the searches. *The BMJ* has a large international audience, a high impact factor, and a low acceptance rate, which may also have influenced citations received and the generalisability of our results.

Policy implications

Our new editor in chief should consider hiring different editors if he wants to publish more highly cited papers, but there is no evidence that these people exist. Editorial decision making is a complex process, yet editors, like peer reviewers, rarely bother to seek training.³⁰ Core competencies for biomedical editors—a useful framework for assessing editors, were only developed in 2017, so we can usually get away with having inadequate training.³¹

Conclusions

Most editors tended to be on the cautious side; they more often rated papers that achieved high actual citation counts as having low citation potential than the reverse. We see this as a good thing as editors do try to focus on the quality of manuscripts and the importance of the content for their readership rather than be swayed by impact factor mania. That said, we probably can't resist the temptation to unmask the data and seek advice from the single editor who could predict correct citation categories >50% of the time.

On the 12th day of Christmas a statistician sent to me...

The BMJ's statistical editors relish a quiet festive season, so make their wish come true and avoid the common faux pas presented here by **Richard Riley and colleagues**

In a December 2019 meeting, statistical editors at *The BMJ* agreed that an article showcasing common statistical issues—and offering advice that authors can open again and again—might be helpful. After some procrastination and sharing of an initial list, a dozen items were selected to represent each of the 12 days of Christmas.

Richard D Riley, professor of biostatistics, University of Birmingham
r.d.riley@bham.ac.uk

Tim J Cole, professor of medical statistics, UCL Great Ormond Street Institute of Child Health, London

Jon Deeks, professor of biostatistics, University of Birmingham

Jamie J Kirkham, professor of biostatistics, University of Manchester

Julie Morris, honorary reader in medical statistics, University of Manchester

Rafael Perera, professor of medical statistics, University of Oxford

Angie Wade, emeritus professor of medical statistics, UCL Great Ormond Street Institute of Child Health, London

Gary S Collins, professor of medical statistics, University of Oxford

Cite this as: *BMJ* 2022;379:e072883



1 Clarify the research question

As with Christmas, statistical reviews provide a time for reflection—in this instance on research questions and objectives. Aim to be clear about the focus of your research (eg, descriptive or causal)

2 Focus on estimates, confidence intervals, and clinical relevance

An under-cooked Christmas turkey will be sent back to the kitchen, and articles focusing on “statistical significance” will be sent back to authors. Consider estimates, 95% confidence intervals, and potential clinical relevance of findings

3 Carefully account for missing data

Quantify the amount of missing data and explain how such data were handled. It is spooky how many submissions fail to do this—the ghost of Christmas articles past, present, and future

4 Do not dichotomise continuous variables

Santa likes dichotomisation (naughty or nice), but do not do the same for continuous variables by splitting into groups defined by being above and below an arbitrary cut point. This method wastes information, reduces statistical power to detect associations, and attenuates the performance of prediction models

5 Consider non-linear relationships

Some statistical relationships are simple (eg, a linear association). However, as with family members at Christmas dinner, some relationships are more complex (eg, non-linear) and this should be examined to avoid them being missed

6 Quantify differences in subgroup results

Many submitted articles include results for subgroups, such as those who do and do not eat Brussels sprouts. However, it is also important to quantify differences in results between subgroups

7 Consider accounting for clustering

At *The BMJ*'s Christmas party its statisticians cluster in a corner and avoid eye contact with authors of rejected papers. To avoid biased results or misleading confidence intervals, account for clustering of data

8 Interpret I^2 and meta-regression appropriately

Interpret the I^2 statistic correctly in systematic reviews and meta-analyses; these popular submissions to *The BMJ* often give statisticians a recurring nightmare before (and after) Christmas. Also avoid using meta-regression of study level information to make inferences about patient level effect modifiers

9 Assess calibration of model predictions

To expose inaccurate predictions, fully evaluate model performance when developing or validating clinical prediction models, including an examination of calibration. Focusing only on model discrimination creates an incomplete picture (remember that unfinished 1000 piece jigsaw from last Christmas?)

10 Carefully consider the variable selection approach

Justify why a variable selection method has been chosen, or avoid it completely—as you would that last turkey sandwich on New Year's Day

11 Assess the impact of any assumptions

It's A Wonderful Life is considered a Christmas movie, but not everyone agrees the same about *Die Hard*. Similarly, die-hard analysis assumptions might be debated and reconsidered (eg, in a sensitivity analysis)

12 Use reporting guidelines and avoid overinterpretation

Make use of reporting guidelines (Santa suggests checking these twice). These represent the minimum required details for a better understanding of research and its findings. As when making new year's resolutions, avoid both overinterpretation of findings and spin

Last Christmas statistical editors tweeted this list, but the very next day they got poor submissions anyway. This year, to save them from tears, they've tailored it for someone special — you, *The BMJ* reader

EDITORIAL

A call to action

Must be more than just “slacktivism”

While it is always time to act,¹ it can be hard to know where or how to start in a world facing unprecedented crises. This can be demoralising and frustrating, but there is one form of action we can all turn to at times like these: the call to action.

A title, abstract, and keyword search (with no date restrictions applied) of Scopus for “call to action” in October yielded more than 6000 results, as did a full text search of *The BMJ* archive. The call to action is clearly a durable and important facet of medicine that has preoccupied doctors and researchers for decades. In fact, one of the first calls to action recorded in *The BMJ* was in 1913.²

Looking to more recent history, calls to action have been made for many important reasons, including to prioritise patient experience in nursing practice,³ to reform academic global health partnerships,⁴ and to expand universal health coverage.⁵ Close relatives of the call to action also pepper the literature, including the “call for an end,”⁶ and the arguably more humble “urge”⁷ and “demand” for action.⁸

What is unknown of course, is what follows the call to action. While we don’t have anything approaching a randomised controlled trial, one influential study published in a previous *BMJ* Christmas issue found no conclusions could be drawn on whether the frequent appeals to “act now” resulted in action.¹ The call to action, however, has several obvious advantages over actually acting. Making that call allows you to salve your conscience, to “do something” without the hard work of actually doing something.

The call to action can be bold and decisive without requiring extensive or time consuming collaboration with others (open letters

Care must be taken not to divert well meaning people away from acting



with multiple high profile authors⁹ being the exception here). No compromise is needed, and you do not have to worry about nuances such as how to achieve or implement the action you are calling for, only about how things ought to be: we should all eat better, help refugees, and take the bus rather than driving, for example.

Beyond words

It naturally follows that the call to action is also less onerous and comes with fewer costs than action. This is particularly important when many

of us are already busy acting. Best estimates (by the author) suggest that an average person could make several if not dozens of calls to action in any one day. Some are more time consuming than others, of course, and come with far greater risks. For example, scaling a bridge to demand government action on the climate

crisis¹⁰ is both time consuming and risky. It’s also an example of action calling for action. We should therefore be careful not to dismiss all calls to action as mere “slacktivism.”¹¹

The complex relationships between calls to action and action (along with their relative advantages and disadvantages) deserve far

greater scrutiny from researchers and bioethicists. But calls to action are arguably most problematic when they divert well meaning people away from acting, and care should be taken to mitigate this risk.

Fortunately, numerous options exist for readers who want to do something instead of calling for something to be done. They could donate to *The BMJ*’s appeal for the International Federation of the Red Cross to support people facing humanitarian crises (p 463). Or join others to campaign against the ongoing neglect of the NHS (keepournhspublic.com) or any number of critical issues from racism to militarisation (www.medact.org/get-involved). Over the next few months, action may simply involve standing in solidarity with colleagues working through what will undoubtedly be an extremely challenging winter for healthcare in the UK.

Observant readers will have noticed that this article is little more than a call to action calling for action to think carefully about calls to action. Ultimately, though both action and calls to action have their place in medicine and in life, and perhaps they cannot be disentangled. We should embrace the call to action and where we can, act.

Ryan Essex, research fellow, University of Greenwich, London
r.w.essex@gre.ac.uk

Cite this as: *BMJ* 2022;379:e072288

Find this at doi: 10.1136/bmj-2022-072288



Global health nonsense

Discourse that either underinforms or misinforms is widespread in global health, and jeopardises improvement in its governance, argue **Felix Stein, Katerini Tagmatarchi Storeng, and Antoine de Bengy Puyvallée**

One of the most salient features of early 21st century global health discourse is that there is so much nonsense. Spin, hyperbole, meaningless buzzwords, and technocratic jargon have become common fare. Nonsense is not necessarily marked by a will to deceive. Rather, it is characterised by a “lack of a connection to a concern with truth—[an] indifference to how things really are.”

This kind of discourse is marked by its “unclarifiable unclarity” and tends to be “pointless, unnecessary, or pernicious.” Whatever the intention behind nonsense may be, it usually underinforms or misinforms its audience, without thereby relying on lies.

Attempts to govern global health according to the goals, actors, modalities, and concepts of financialised markets are partially to blame for the spread of nonsense. Short term competitive funding rounds, the fetishisation of performance metrics, and a focus on returns on investment increase pressure to constantly project success. As a result, global health’s leading agencies routinely refer to “accelerators,” “catalysts,” and “investment cases,” emulating the hyperbolic self-praise of Silicon Valley.

Several observers have picked up on this trend and made the case for more meaningful and self-aware discourse. They provide sometimes humorous rejections of vacuous global health speak alongside serious reflection on the way language recapitulates and reinforces existing power hierarchies.

Nevertheless nonsense seems to be proliferating, perhaps because so many of us are implicated in producing it. Taking global public-private partnerships in the response to the pandemic as examples, three main forms of global health nonsense are obfuscation, misrepresentation, and the omission of relevant information. We must call out nonsense because it stifles efforts to understand, critically assess, and improve global health governance.

KEY MESSAGES

- Spin, hyperbole, meaningless buzzwords, and technocratic jargon have become increasingly common in global health discourse. They are part of a broader phenomenon labelled “global health nonsense”
- Three main forms of global health nonsense are obfuscation, misrepresentation, and omission of relevant information
- Global health nonsense must be called out, because it stifles collective efforts to understand, critically assess, and improve global health governance

Obfuscation

Global health nonsense obfuscates reality, often by relying on jargon. Take as an example the Access to Covid-19 Tools Accelerator (ACT-A), which brings together leading global health agencies to speed up the development of and equitable access to covid-19 diagnostics, therapeutics, and vaccines. Jargon obscures the structure of this multibillion dollar health effort, describing ACT-A as an “accelerator,” a “framework,” a “collaboration,” a “partnership,” an “initiative,” or “a platform.” Jargon also obscures how ACT-A works, given that it has a “facilitation council,” “executive hub,” “pillars,” “health systems and response connector,” “pillar leads,” “principals,” “partners,” “key delivery partners,” “co-hosts,” “co-conveners,” “co-chairs,” “sponsors,” and “special envoys.” Each “pillar,” in turn, has its own “agency leads,” “principals,” “coordinating committees,” “workstreams,” and “workstream leads,” as well as the occasional “shareholders council,” “engagement group,” “investors group,” and “consensus group.”

Many of these terms are “floating signifiers” that obscure more than they elucidate, papering over the different interests, mandates, degrees of legitimacy, and lines of accountability of ACT-A’s members. Lastly, jargon obscures what will become of ACT-A. While ACT-A promises to continue to “support countries through the transition to long term covid-19 control,” most of its activities are now being “kept warm,” “kept on

standby,” “sunset,” or “transitioned” to individual agencies. The extent to which these are euphemisms for a simple end to ACT-A is unclear.

Misrepresentation

Global health nonsense also misrepresents reality through words, diagrams, or metrics. A good example is the number of vaccine doses delivered by ACT-A’s “vaccine pillar” Covax.

Covax initially promised to provide “access to at least two billion doses of safe and effective covid-19 vaccines to the most vulnerable [by the end of 2021]” but ended up delivering less than half that (832.5 million). While it was accused of failing in its mission, Covax celebrated “700 million doses delivered,” and “one billion doses delivered” in early 2022 as “historic” successes, complete with videos of people fist pumping in joy over the arrival of vaccine shipments.

Further, Covax and the countries that donated vaccines to it opted to highlight whatever metric best portrayed their impact. They sometimes emphasised “pledged doses” and sometimes “secured doses,” occasionally “ordered doses” and “delivered doses,” but all too rarely “administered doses.” This led *Politico* to conclude that “a dose is not a dose” in the context of Covax’s vaccine rollout.

Another metric that subtly misrepresents reality in favour of global public-private partnerships like Gavi, the Vaccine Alliance and the Global Fund to Fight AIDS,



Short term competitive funding rounds, the fetishisation of performance metrics, and a focus on returns on investment increase pressure to constantly project success



Buzzword Bingo:
To support readers in calling out nonsense, we suggest they play Buzzword Bingo in their next global health meeting. Put a cross on the square when you hear the terms in question. Whoever fills a horizontal, vertical, or full diagonal row first wins!

Conclusion

The examples of nonsense we have identified will be recognisable to many in the global health community. A certain amount of obfuscation, misrepresentation, and omission may be unavoidable, but it is not innocuous. By fostering “strategic ignorance,” nonsense stifles collective efforts to understand, assess, debate, and improve global health governance.

Indeed, our acceptance of nonsense made it possible for global health leaders to at once claim that we “accelerated vaccine equity” while also maintaining “vaccine apartheid.” Crucially, nonsense contributes to the inequity laid bare in the global response to the covid-19 pandemic.

As global health research, publishing, and policy become more reliant on a smaller number of funders, it gets increasingly difficult to conduct and publish independent analyses of policy initiatives. Challenging the status quo can mean

facing ridicule, censorship, or exclusion from the centres of epistemic and economic power. For example, in a 2021 interview, Bill Gates, whose foundation funds all major ACT-A agencies, responded to the proposal of a temporary waiver of intellectual property rights to increase access to covid-19 vaccines by calling it “the stupidest thing [he] ever heard.”

We are all implicated in the nonsense that permeates global health: policy makers, think tanks, consultants, non-governmental organisations, and universities are increasingly compelled to project success to attract funding and garner influence. Stuck in a “success cartel,” we risk reinforcing the power asymmetries that undermine health equity. All of us therefore need to find the courage to avoid, identify, and call out hogwash when we hear it. It’s time to cut the global health nonsense.

Felix Stein, postdoctoral fellow
felix.stein@sum.uio.no

Katerini Tagmatarchi Storeng, associate professor

Antoine de Bengy Puyvallée, doctoral student, Centre for Development and the Environment, University of Oslo

Cite this as: *BMJ* 2022;379:o2932

Tuberculosis, and Malaria is the number of “lives saved.” This emotionally appealing metric is characterised by high uncertainty, is prone to overestimation, and tends to misattribute positive health outcomes to individual programmes, rather than the host of institutions and interventions involved in bringing them about.

It can even lead to double reporting as a person co-infected with AIDS and tuberculosis who receives treatment for both can be counted as though “two lives” had been saved. As with Covax’s vaccine doses, “lives saved” blurs marketing with unbiased reporting of global health results. It embellishes the truth to reinforce vertical programming, distorting national health priorities and budgets in the process.

Omitting relevant information

A final form of global health nonsense is to leave out relevant information, such as frank discussions of political and economic choices, challenges, and shortcomings. Leaders of high income countries and public-private partnerships repeatedly

insisted on the importance of multilateralism, the urgency of global vaccine equity, and the truism that “nobody is safe until everyone is safe.” They often made such generic points instead of discussing concrete matters like vaccine hoarding; soaring prices for covid-19 diagnostics, treatments, and vaccines; the limits of intellectual property in pandemic times; how publicly funded public-private partnerships spend their budgets; or what exactly the public should expect in return for subsidising the pharmaceutical industry in times of crisis.

Similarly, in the autumn of 2022, the head of the World Bank argued that its new financial intermediary fund for pandemic prevention, preparedness, and response (FIF) would “complement” existing global health institutions, “catalyse investments,” and “serve as an integrator” rather than a new silo. But he did not tackle concrete concerns that FIF competes with existing global health funds and institutions, that it should broaden its base of participating countries, or that its claim to provide “catalytic” funding remains to be substantiated.