

## Editorial Podcast

### In conversation with Dr. John P. Ioannidis: On meta-research, metrics, gaming, and top 2 per cent scientists

Albina Arjuman

*Executive Editor, Indian Journal of Medical Research, Indian Council of Medical Research New Delhi, India*

Received December 19, 2025; Accepted December 19, 2025; Ahead of print December 27, 2025; Published January 31, 2026

*"When a measure becomes a target, it ceases to be a good measure", the Goodhart's Law.*

The massive growth of research literature brings great responsibility towards the accountability of data and their translational value in terms of quality and performance outcomes. The evolving digital environment has bestowed 'quantifiable outcomes' upon the scientific community. However, as everything viable, these too embody both the *Yin* and the *Yang* of all translations. Today there are a number of indicators which can quantify research output and consider various facets of a given piece of research to pin a numerical figure against it denoting its value. In this context, Journal level and Author level metrics, authorship credits, and diverse performance assessment indicators are very popular. While such indicators originated from the scientific fraternity itself to measure and hence assess the need for enhancing accountability in different thematic areas of research, these have now generationally been handed down with distorted convictions due to varying levels of understanding within that same scientific fraternity. Furthermore, these performance indicators may exert immense pressure on researchers to perform in specific ways. This distortion of understanding has unfortunately been leveraged to 'commodify' research and researcher performances themselves, a hefty trade-off for profit making. These trade-offs are most visible today through the gaming of various quantifiable metrics.



Dr. John P. Ioannidis

*Professor of Medicine, of Epidemiology and Population Health, and (by courtesy) of Biomedical Data Science; Co-Director, Meta-Research Innovation Center at Stanford (METRICS); Stanford University, California, USA*

In this context, there are often those who navigate at the precipice of such affairs through denial and avoidance. However, there are those who untiringly strove for decades to generate evidence to support transparency and accountability in research and to educate as well as caution researchers about countering the erosion of integrity. One such individual dedicated his life and decades of service in enriching the domain of meta-research by generating evidence in context of friability and gaming of metrics, while also actively

investing in formulating various guidelines for transparent and accountable conduct, reporting, and assessment of research. This is the same individual who is today popularly known as the brain behind the much discussed ‘Stanford/Elsevier Top 2 per cent Rankings’, Dr. John P.A. Ioannidis.

Dr. Ioannidis is a man with an impressively long list of affiliations and accomplishments (<https://profiles.stanford.edu/john-ioannidis>), but an embodiment of simplicity, focus and clarity. He currently serves as Professor of Medicine at Stanford Prevention Research Center, Professor of Epidemiology and Population Health and (by courtesy) Professor of Biomedical Data Science at Stanford University, California, USA and is the Co-Director of the Meta-Research Innovation Center at Stanford (METRICS). His work has received more than 670,000 citations in Google Scholar and more than 320,000 citations in Scopus. He has conducted pioneering research in biomedicine, data science, meta-science and scientometrics and has headed various relevant research governing societies and initiatives. Having delivered more than 700 lectures on various topics across the globe, most recently, he had a new Centre for Excellence in Medicine and Public Health inaugurated in his honor at the Woxsen University in Hyderabad, India.

#### **SPOTLIGHT: Dr. John P. Ioannidis**

***Over the years through your exhaustive stalwart work in meta-research you have always advocated for a balance in the use of quantitative metrics for research evaluation. What according to you is a ‘soft point’ for considering metric-based evaluations for research output as a whole? Is there a certain ‘right combination of metric indicators’ for evaluating research output/researcher’s performance?***

It is important to capture multiple dimensions. Metrics are useful but usually not enough to tell the whole story. Several initiatives have tried to sensitize people about different dimensions of doing good research including reproducible research practices and some ways to achieve them, for example, sharing data, codes, and protocols and having rigorous methods. There is also a huge literature on scientometrics trying to capture the impact of the work in the scientific literature. None of these is perfect and there are some very thoughtful efforts like the Leiden manifesto<sup>1</sup> that describes how much we can achieve with different types of measurements and how cautiously they should be interpreted. Research is very difficult to

measure because the real impact may take a very long time to be visible. Sometimes it may take decades, and often it happens in unpredictable ways. It is very unfortunate that there are just simple, naïve and overtly spurious gaming approaches to metrics. So, if we can avoid those and have a more thoughtful and balanced approach to understanding what we can measure, what the limitations might be and differentiate between what each metric means and what it doesn’t, I think this would be a step forward.

#### ***What are your reflections on metrics gaming?***

We see that there is a lot of gaming of metrics around the globe. Any metric that acquires some importance will be susceptible to gaming. Some metrics are easy to game compared to others. I believe that it is like virus and anti-virus, developing metrics that can capture the gaming of metrics. To try to understand how and why it happens, what it means, not just for a single scientist but also for the micro-environments (their teams or departments) and macro-environments (their institutions, universities or countries) in terms of the incentives that they provide and how they push people to behave in a given way or misbehave in a given way – how they incentivize good behaviors or bad behaviors. Unfortunately, there are a lot of misaligned incentives in scientific research and the publication system at the moment. Lots of publish or perish pressure, with reward systems linked to very superficial and naïve type of indicators.

***Do you think that these misaligned incentives you mention are key to greater instances of gaming particularly given the greater access to Artificial Intelligence (AI)?***

Artificial Intelligence is a tool like so many other technological tools that have become available. It may be more revolutionary and maybe it has more capabilities than other tools we have had before. When it comes to doing research AI does offer some nice possibilities. We have seen some great success stories like predicting protein structure or creating some new types of computational capacity that we have not envisioned before. However, we also see a lot of misuse and even fraudulent use. AI has facilitated the growth of a market for fake papers. Paper mills are thriving practically selling fake papers to people who wish to get a name as an author in these. We also see not only fraudulent work but also low quality automated work that people may be using these tools to generate, a least publishable unit and very weak type of analysis and

publications. While AI tools may help people to catch up with native speakers in terms of the quality of the language, they may also create literature that is replete with mediocrity if not with fraud as well. I think this is a major risk. I don't think that we know whether the benefits will outweigh the risks, we just have to be very watchful.

***Pulling on to the same thread, do you think that the thriving of paper mills through misuse of AI in addition to sparing instances of something like plagiarism given easy AI access are dissuasive to well performing researchers in the field? Has the churn rate of such resources utilizing AI gone down? Do you see the number of researchers wanting to exclude AI from their research starting to go up?***

Well, I think that the answer is different in different micro-environments. I believe that a very large share of people use different AI tools currently in one way or another and I think we don't have full disclosure of these uses. Perhaps they also use these tools to try to masquerade that they haven't used these tools! This is also a possibility. As I said, I don't want to demonize AI. It may have some nice research applications, much like any other tool. Then one may describe the methods of how some research results were achieved in the laboratory or *in silico*: 'this is what I used and this is what I got from using it', I think this is quite legitimate, provided one knows what the limitations of the method used are, what safeguards are in place, how one secures the validity of the work, and so forth. I think the problem is when AI gets misused and manipulated and used for the wrong reasons, either to subvert review or subvert literature or spread misinformation or to do other weird stuff. Unfortunately, there is plenty of room for this type of misbehavior.

***Being associated with a Journal, I find that it's a lot about trust. We as Editors are heavily reliant on declarations of various kinds. Even major Editorial bodies largely recommend the same. There is no mandatory guideline to follow which can determine the code of conduct of a researcher. In practice we often witness that the ethics surrounding participants (more so around clinical trials and some select study designs) requires stringency. However, ethics as a concept is far more challenging now in scientific publishing particularly given the evolving AI enabled environment. How far do you feel can we go with just editorial gut feelings and trust? Should there be some kind of a binding towards researcher code of conduct?***

Science was always dependent on trust since its very beginning. But at the same time, we see that trust is violated too often. I think that we should be using tools to detect and probe and there are many such tools being developed both for use and misuse of AI and for other research practices and problems. These include tools that look at the statistical machinery or tools that look at various aspects of potential misconduct. Some of these have been very effective. For example, if you take plagiarism, it used to be a substantial problem in the past, but it almost disappeared after the advent of iThenticate or other tools which are quite good. It's no longer such a major problem like other problems. I think in the same way we should be using tools to try to detect new forms of misconduct and of questionable and detrimental practices. In principle asking for more transparency which means sharing the raw data, sharing the code, the algorithms, methods materials, protocols – if a protocol does exist, while if it doesn't then one must say that this is an exploratory type of research. These investments in transparency probably strengthen trust and they allow for better review, better post publication review, better utilization of research, better synthesis with other types of studies where evidence can be merged at a more granular level. We have seen improvement in all of these fronts. Unfortunately, most scientific papers being published still do not share data or code or register protocols. However, we see a sizeable proportion that do. It doesn't mean that these are always the best papers but I think these have some advantage compared to the rest that are less transparent.

***Coming to your brainchild, the Stanford/Elsevier Top 2 per cent rankings, these are highly revered the globe over, but have also garnered a lot of criticism related to methodological limitations/biases, for devaluing true academic scholarship, perceived misrepresentation of research quality, inaccuracies, etc. What are your reflections on these comments?***

These are datasets that we generate and we release in an updated version annually. There are some strengths and some weaknesses. We have published a large number of papers describing the methodology, the validation processes, the error rates, the accuracies, the validity issues and how one can actually use the datasets to detect and map questionable and fraudulent research practices<sup>2-10</sup>. Certainly, nothing is perfect. Everything has shortcomings. These data were never intended to be worn as medals or as signs of recognition or to offer some sort of reward. We are very explicit about that and people should read very carefully at a minimum

the Frequently Asked Questions document that we have on the website where we release these datasets (<https://elsevier.digitalcommonsdata.com/datasets/btchxktzyw/8>). These datasets are basically meant to capture a very large number of citation indicators that adjust for scientific field and take into account not just citations but also co-authorship-adjusted citation impact, and relative contributions based on authorship position. Moreover, they include information on self-citations, retractions, and impact of retractions. You can see if people are outliers based on their profile characteristics. They could be outliers for good reasons or for bad reasons. I think that unfortunately, these lists are seen in some circles as signs of recognition. These are not medals! We very specifically say that not being included in these does not mean that one is not a good scientist. Similarly, being included in the list does not prove one is certainly a great scientist. In fact, some horrible scientists are also included in the lists and the presented data make this totally transparent, if one wants to read them carefully. We present data that can be used to understand diverse aspects of impact in the scientific literature and put things into context. I want to stress that the databases can also offer evidence for gaming of scientific literature and impact. Unfortunately, I do not see this side of the evidence being used that often. It's also unfortunate that there are several sites that have popularized these metrics ignoring that multi-dimensionality. These sites have nothing to do with me or my colleagues, at the analytics Research Intelligence unit at Elsevier or my team at Stanford. Some of these popular sites just repost our metrics with very flashy titles. They may even sell certificates of recognition, which is something that I would never have wished to see because that is not the purpose of creating and updating these databases. If used prudently, they may offer some information that I believe is better to have procured in a standardized, centralized way with some common criteria and with some standard field adjustment and with some cleaning of the data sets, rather than have each scientist or each team or each institute produce their own metrics, their own preferences and manipulations and make claims that are completely unfounded.

***Springing back to the debate on 'quality vs. quantity' of research output, what are your thoughts on the actual translation of research? Particularly translation of research into policy because that is our ultimate goal, one which could take several years! How best do you think we can evaluate this translation of research into policy and other useful deliverables?***

It depends on what research one is doing. If someone is occupied with very early discovery basic science, it's unlikely that translation to policy will be their immediate goal. It will be wonderful if that happens at some point and this is why we do basic science: because out of these discoveries, some of these might be translated into useful deliverables, sometimes into policy, interventions, medications, or other things that work. However, I wouldn't want to judge someone who is working on early discovery research with whether they had their work translated into policy and interventions. Once we are in the late translational spectrum with clinical trials and similar types of work, then, yes, you do want to see that you have some type of impact eventually that you do move the needle. If it comes to clinical interventions, you want to see that you save lives and you improve community outcomes, you improve life expectancy or diminish burden of disease. There are hardcore outcomes that one would wish to measure. Do we measure those in relation to scientific work? Sometimes we do, but it is very erratic. But we know that some important scientific work may just take too long to get there. So, we cannot just say that unless we achieve such major impact one should not be rewarded with being able to continue their work. We need to have some level of trust, not completely blind trust but really allowing researchers to do good work and to give them incentives also to translate that work into something that eventually will have a useful application. I am not sure that we do that in most countries and institutions around the world. Researchers are mostly told that you need to prove in your grant applications that you will save the world. This is asking us to lie in a way. Cumulatively the scientific community does help the world in major ways, does save lives and probably may even save the world from major crises that are threatening humanity. However, for each researcher to promise that they will do that alone, I think this is completely misconstrued and it is unlikely to happen to a vast majority of us who are well intentioned and try to do our best but perhaps 99 of us will fail and one would succeed. It is the cumulative success that matters and kind of mirrors the effort of everyone on board that ship that we call science.

***Coming from a basic science background, I am aware that somewhere down the line these kind of metric evaluations/rankings at times make us basic scientists, the 'discovery researchers' a little uncomfortable, because like you said the late translational research such as trials, epidemiological surveys, translation***

*of basic innovations into diagnostics, etc get cited more as these are quickly translatable into practice and policy. So there is an assumption that probably the research output of researchers involved in such late-translational research is way higher as compared to someone from discovery research or with someone who is involved in pre-clinical studies. What are your comments on the plausibility of a dissuasive environment towards furthering basic research?*

I think that it is important to compare apples with apples and oranges with oranges. So, one needs to adjust for what field or sub-field a researcher is working on. I would never prioritize putting any emphasis on number of publications for example. If you look at the composite indicator that we have in the datasets that we release annually, number of publications is not one of the features that counts. One may be an amazing scientist that publishes a single paper in their entire career. It could be a paper that is Higgs coming up with Higgs boson. Conversely, other scientists maybe need to publish 5000 papers to adequately present their work, because they want to present that massive volume of analysis that they did and data that they collected. In order to be thorough and complete they need to publish 5000 papers. But that alone does not mean anything. One paper or 5000 does not make a difference provided that it is good science and properly reported. If you compare someone against their peers, that is pretty fair and in the datasets we release we try to do that. Could it be done better? Perhaps! You can start looking at not just the 174 subfields within science as we do, but 5000 sub-fields. One can look at half a million sub-fields. We have created maps of science with my team to have half a million very small sub-fields but then you realize that within these very tiny sub-sub-fields there are only a handful of people. Then we lose the context of the larger discipline that perhaps is more relational and more appropriate for comparisons to happen. Importantly, no metric and no automation will remove the need to eventually look at what someone is doing and try to really appraise it not just quantitatively but also in terms of ‘*So what?*’ ‘*What does this mean?*’, and that is a very difficult decision in many cases. In a meritocratic environment where you have scientists who do care about rigorous science and really good work hopefully they would recognize that. In a non-meritocratic environment this would not be recognized and that’s what makes the difference.

*Playing further on that thought, in the existing environment where technology/innovation driven research is being promoted, social media mentions and translation in terms of communicating science to the general masses (especially since the pandemic) has become the preferred channel. You mentioned considering multi-dimensional considerations towards quantitative metrics. In order to make it a more meritocratic environment, do you think some of the non-citation based metric indicators such as altmetrics or technologies generated, newspaper mentions, etc could be considered?*

I think that they offer an interesting counterpart and of course communicating science to a broader audience or having science discussed by a broader audience is something interesting. But it may have its caveats. Such visibility doesn’t mean that the discussion is meaningful, useful, or eventually adds value. It is also highly gameable. I think you can have influencers who can quickly elevate anything to a very high level even if the work may be mediocre and scientifically even completely wrong. We need to ask ‘*what exactly are we measuring?*’ Some of these things are being measured even if we don’t want them to be measured and we need to understand what it means. There was a lot of excitement about alternative metrics in the past. I tend to be not so excited to be honest. I love measurements of all sorts, so it is good to have them available, but we should ask ‘*so what?*’ again. Is that just another popularity context like a beauty pageant? Is it something that is really improving science and the contribution of science to society? I think media and social media may do that occasionally, but there is also lots of waste and lots of trash, low quality and negative contributions of misinformation or just things that don’t really help humanity in any way.

*To sum it up you are saying that one needs to exercise caution in considering the right indicators for choosing the right metric else things can be blown out of proportion and may enhance the gaming prospects?*

When things are blown out of proportion, gaming is possible and detecting gaming says something about people who do it. Some of the nice-looking values or metrics may actually be negative in terms of how these should be evaluated<sup>9-10</sup>. It only shows that there are some very ambitious people who probably don’t care that much about science but maybe primarily care about other sorts of recognition and power. Identifying

more relevant, more translatable, gaming-resistant indicators is important.

***You have made remarkable contributions towards research guidelines including STROBE, CONSORT, PRISMA, STARD, we name it! Do you envision pooling all your masterful experience into any cross-disciplinary meta-research guideline for evaluating translation of research or simply to redefine performance assessments of researchers?***

Guidelines should follow evidence. For many of these topics that we are discussing they need better evidence rather than just experts who come up with proclamations that they know how exactly things are or how things should be. I am in favor of getting more evidence. It could be observational evidence, it could be just documentation. Many of the datasets that we have produced function like observatories. They allow us to observe science and scientific work at large scale along with mentions of its impact. However, we also need evidence of interventions that try to improve our research practices. We should see what happens if we have interventions at the level of peer-review, editorial policies, institutional policies, and so forth. How much do we change things, especially those things we think are important? What is important depends on what research and type of contribution we are talking about. This could range from basic to translational, applied, evidence synthesis, implementation and other types of research. So, I don't think that we should just aim to create the golden rules, because maybe these do not exist. Instead, we should try to understand the complexity of the system and how we can make some of our research practices better. More essence, greater utility, higher efficiency and less gaming, less waste and less fake! This is even more important because we do have a crisis at the moment, with a lot of waste and a lot of haste.

***In terms of monitoring mechanisms for potential gaming instances, do you think that a centralized metric like the H-index, impact factor, Cite score, etc would be better to consider as against a country level metric?***

More accurate information will be helpful and accurate information can be accumulated at the level of single projects, papers, single authors, but can be extended to teams, departments to organizations and institutions. But once we move to very large teams like a university for example, it is problematic because a university comprises of thousands of faculty and tens

of thousands of students and any sort of effort towards amalgamating such heterogeneous population is problematic. You may have a mix of the best and the worst, meaning there may be some amazing scientists but also some fraudulent. Information can really point out to weak as well as strong areas, to problems that are in fact growing or even getting out of control that we should notice and probably audit to see what is going on. These could be individuals, teams, institutions or countries. If institutions and countries have misaligned incentives and push their scientific workforce, then weird stuff will be done by their scientists as they try to survive.

***Anyone who follows meta-research and loves metrics and their translations, is familiar with the Goodhart's law. Do you feel that the Goodhart's Law has largely been undervalued in context of the business models that now dictate the scientific publishing landscape?***

The Goodhart's law still holds true for any metric and some metrics are more susceptible and probably easier to succumb to the law than others. This is why I believe that we need this multi-dimensionality and we need to be cognizant of the strengths and limitations of each metric and how the Goodhart's law may affect what is happening. Some metrics that were proposed early on, like Journal Impact Factor for example, have lost their value due to gaming. Originally journal impact factor was not a bad metric and for Journal-level evaluations, it provided interesting observations. Now, it is completely bogus and it is very unfortunate that we cannot get rid of it. So each metric may have its life or half-life of decay in a sense based on the pressures it receives, but hopefully we can keep abreast, understanding its utility or lack thereof, by measuring what's going on. Journal impact factor currently is to some extent an indicator of gaming and manipulation, both by prestigious journals such as Lancet or Nature, and by petty journals that suddenly boast unbelievable impact. Of course, the tricks that Lancet and Nature use are different from those used by petty journals, but, essentially, they are both tricksters.

***What are your thoughts and recommendations in context to Journal level metrics gaming? Do you think, with the existing metrics-race, whether journals can compromise their attention to metrics and aim purely for the 'quality' of published research?***

It has been recommended for a long time that we should just get rid of Journal Impact Factors. People should not be using them and it is a pity that people

just capitalize on them and present CVs for evaluation where the Journal Impact Factors are being summed. I think that it is very difficult to get rid of metrics that have run their lives and they should be dead but they are vampires.

***On that thought, what would happen in a scenario where we drop all quantitative metrics? Are we looking at a time where ‘yellow journalism’ may soon be upon us and scholarly peer-review based publishing as a concept may become obsolete?***

We see a lot of different publication prototypes that are being tested or promoted. The landscape of publication has changed a lot already. We see lots of journals which publish massively, voluminously, some of them are just mega-journals. Hopefully they try to do some decent peer-review, but one cannot always be sure. Others are predatory journals which publish anything with no review, they just want to get a bit of money through the publication fee. We also have experimental approaches of journals that claim they are here just for peer-review – not to block your way to publication, but try to offer as good peer-review as possible. eLife is one such example. I think we will see more and more variants and possibilities. My belief is that provided that there is transparency about what is happening, I don't have a problem trying different options and see eventually what best serves science and the wider community. I am very open to possibilities. There are too many stakeholders though who try to exploit the system. We talked a lot about the scientists, but scientists as I said are also under pressure so some of the misbehaviors may be the result of misaligned incentives. Then there are publishers that are also trying to serve but also trying to make money with a huge profit margin. As you know, the big publishers have 30-40 per cent profit margins, among the highest compared to any other industry today. We have institutions that are fighting for rankings, although usually these are just bogus ranking schemes. Plus, so many other stakeholders who try to get an upper hand. I don't want to be a pessimist but it's an explosive mix. What we have had traditionally, the peer-review practice, is far from perfect. However, getting entirely rid of it and just having para-scientific types of assessments run the day and tell us what we should believe or not, well, I don't think that's a good idea. I am happy to be convinced otherwise but I just see a lot of weird things and weird stakeholders taking a grip on science. You have political interference, you

have journalists who become more influential than the science itself, you have social media influencers who are clueless, but just because they have a lot of followers they dictate policy or what people decide to do with their lives, their health and other decisions that affect society.

***Lastly Sir, what are your thoughts and recommendations for the research community in India?***

India is a fascinating country and I have great respect for Indian colleagues. It's a rapidly growing scientific community. In terms of volume, it is already third in terms of productivity after China and the USA and by 2029, at the current pace it will outpace the USA. I think that volume though is not really what matters. As I said earlier one paper or 5000 papers should not make a difference. I have met so many scientists from India who are amazing, brilliant minds that I do believe and hope quality, rigorous methods, trying to really do good work and work that matters at the end of the day should be the ultimate goal – and one that can be achieved.

India can really capitalize on that amazing human capital and come up with really great research contributions. But, things can also go wrong.

There's a lot of pressure, I realize, to publish lots of papers and get some of these weird metrics to be nice looking. One has to watch out and one needs the right incentives from the institutions and funding agencies to really support excellent, rigorous, reproducible work. I do want to be optimistic and wish you all the best in that regard.

## References

1. Leiden Manifesto for Research Metrics. *10 principles to guide research evaluation with 25 translations, a video and a blog*. Available from: <https://www.leidenmanifesto.org/>, accessed on December 16, 2025.
2. Ioannidis JPA, Pezzullo AM, Cristiano A, Boccia S, Baas J. Linking citation and retraction data reveals the demographics of scientific retractions among highly cited authors. *PLoS Biol* 2025; 23 : e3002999.
3. Ioannidis JPA, Maniatis Z. In defense of quantitative metrics in researcher assessments. *PLoS Biol* 2023; 21 : e3002408.
4. Ioannidis JPA, Boyack KW, Collins TA, Baas J. Gender imbalances among top-cited scientists across scientific disciplines over time through the analysis of nearly 5.8 million authors. *PLoS Biol* 2023; 21 : e3002385.

5. Ioannidis JPA, Maniadis Z. Quantitative research assessment: Using metrics against gamed metrics. *Intern Emerg Med* 2024; *19* : 39-47.
6. Ioannidis JPA, Boyack KW, Baas J. Updated science-wide author databases of standardized citation indicators. *PLoS Biol* 2020; *18* : e3000918.
7. Ioannidis JPA, Baas J, Klavans R, Boyack KW. A standardized citation metrics author database annotated for scientific field. *PLoS Biol* 2019; *17* : e3000384.
8. Ioannidis JP, Klavans R, Boyack KW. Multiple citation indicators and their composite across scientific disciplines. *PLoS Biol* 2016; *14* : e1002501.
9. Ioannidis JPA. Features and signals in precocious citation impact: A meta-research study. *PLoS One* 2025; *20* : e0328531.
10. Ioannidis JPA, Collins TA, Baas J. Evolving patterns of extreme publishing behavior across science. *Scientometrics* 2024; *129* : 5783-96.

#### In a nutshell

- Research impact is best not judged based on the number of publications; it could be 1 vs. 5000 publications having the same impact.
- Metrics gaming is real! Multi-dimensionality of chosen indicators may help choose a good metric with lesser potential for gaming and metrics can also be used to assess gaming.
- Caution should be exercised while considering the right indicators for choosing a quantitative metric.
- Guidelines should follow evidence. Meta-research is in need of more evidence generated out of rigorous research.
- The Stanford/Elsevier dataset rankings were not endorsed to be worn as medals or differentiate between scientists included versus not included in the lists as a surrogate of good versus bad scientists. The datasets are an effort to provide standardized, centralized, subfield-adjusted evidence based on multi-dimensional considerations to examine trends in the global scientific output while also providing comprehensive evidence for gaming in scientific literature.