



**EJW**

ECON JOURNAL WATCH  
Scholarly Comments on  
Academic Economics

ECON JOURNAL WATCH 20(2)  
September 2023: 460–464

# *Journal of Accounting Research's* **Report on Its Own Research- Misconduct Investigation of an Article It Published**

Daniel B. Klein<sup>1</sup>

[LINK TO ABSTRACT](#)

I comment here on a document by *Journal of Accounting Research* (JAR). The document presents itself as a report on its own research-misconduct investigation of a problem in an article it published in 2020. Here I refer to the document as the JAR Report. JAR's investigation was undertaken in response to a call by Stephen Walker, a call which he made publicly in the September 2022 issue of *Econ Journal Watch* (EJW) (see [here](#)).

On June 5, 2023, JAR emailed the JAR Report to Walker. Walker forwarded it to me. I refrain from quoting the JAR Report, which is 1,077 words, because doing so is not necessary to clarify its nature and the judgment it conveys, because it is not published, and because the first one to publish it ought to be JAR itself.

The JAR Report is a response to a request submitted by Walker to JAR to investigate a problem that Walker identified in the 2020 JAR article.

The present comment is not to allege misconduct on the part the author group of the 2020 JAR article. The issue here is JAR's conduct. Walker called for a research-misconduct investigation and JAR took up the call and sent Walker a report on the investigation. The question is: Did JAR conduct and report on a research-misconduct investigation in a way that is at all satisfactory? I will argue that it did not. I go further and say the JAR Report is a disgrace.

The issue is of cardinal importance to science, because if a problem in research raises the question of research misconduct and that question is not meaning-

---

1. George Mason University, Fairfax, VA 22030.

fully and faithfully addressed by the guardians of science, the scientific project becomes a sham.

In the present comment I speak of “the author group,” because the focus here is on the JAR Report, not that author group.

Walker made his call for a research-misconduct investigation, by email, to JAR on August 11, 2022. Eight weeks later, JAR emailed Walker, clarifying how to submit such a request. Walker promptly replied to JAR’s satisfaction on October 5, 2023. Eight months later, on June 5, 2023, JAR sent the JAR Report to Walker.

Walker has published three articles on this matter, all three in EJW. The author group has published one article in EJW. Here is a list of the six articles that preceded Walker’s request for a research-misconduct investigation:

1. The 2020 JAR article ([link](#)).
2. Walker’s March 2021 critique in EJW of the 2020 JAR article ([link](#)).
3. The author group’s March 2021 reply in EJW to Walker’s March 2021 critique ([link](#)).
4. Walker’s September 2021 rejoinder in EJW to the author group ([link](#)). (Before the rejoinder was published, the author group was invited to reply to it, either for concurrent publication or anytime afterward, but they declined.)
5. The author group’s September 2022 “Erratum” in JAR ([link](#)).
6. Walker’s September 2022 critique in EJW of the “Erratum” ([link](#)). (Immediately after Walker’s critique of the “Erratum” was published, the author group was invited to reply to it for subsequent publication, but they have not done so; the invitation remains open.)

In the author group’s “Erratum” in JAR, the opening words of the first paragraph (p. 1635) are as follows: “Walker (2021a and 2021b) identified an error **in the program codes** of Bao et al. (2020) posted on Github that led to an overstatement of model performance metrics. This erratum corrects this error...” (boldface added).

The paragraph goes on to say: “Bao et al.’s (2020) **intent was to recode** all spanning serial fraud observations in the training period as nonfraud. However, the **codes** posted on Github **contained an error** that resulted in approximately 10% of these spanning serial fraud cases in the training period not being recoded as zero, which in turn helped model performance” (boldface added).

Thus, in the “Erratum” the author group confess that there was a problem—of some kind—in the 2020 JAR article.

Later in the “Erratum” (p. 1636) they speak of the problem as a “coding error.” Also, the JAR Report itself employs the expression “coding error” to refer

to the problem.

Whether the problem is aptly denominated a “coding error” is one thing that the JAR investigation should have investigated and reported on. But the JAR Report offers nothing on the problem that is at the heart of the entire matter. Again, the JAR Report is 1,077 words. Of that, 492 words are given to the title and a recounting of the materials in question, a description of what JAR has done in their putative investigation, including a listing of eleven documents distributed to Reviewers. Next come the presentations of two quotations from Reviewers of the materials, taking 197 words; both Reviewer quotations express an overall judgment that research misconduct has not been definitively established but not giving a single word to investigation of the “coding error.” Next come 57 words stating that the JAR senior editors agree and that the case is closed. Lastly, there are 331 words given to the matter that, in the “Erratum,” once the author group corrected the core problem, the preferred test sample period changes from 2003–2008 (as in the 2020 article) to 2003–2005. This matter is downstream of correcting the core problem and does nothing toward investigation of the core problem itself. The segments of 492, 197, 57, and 331 words sum to 1,077. There are zero words given to investigating the core problem.

It is now agreed by all that there was a core problem in the 2020 JAR paper. Walker has described at length the nature of that problem. Neither the author group nor the JAR Report, however, so much as identify what the core problem was, why it occurred, nor explain why it should be called a “coding error.” The author group simply denominate the core problem a “coding error” and then proceed to correct it.

At the outset of his 2022 critique of the “Erratum,” Walker quotes the “Erratum” where it says that he (Walker) “identified an error **in the program codes.**” Walker then says:

The part in boldface is false. I discovered no errors in their program codes. Rather, the ‘error’ was in their dataset—specifically, how they identified fraud cases in their sample. A coding error may be accidental, but the method used to define their fraud cases was deliberate. To this date, the authors have offered no explanation as to why they did what they did, nor offer any explanation as to what prompted their misidentifying of fraud cases. (Walker, p. 190 [here](#))

Later in the piece, Walker writes:

In fact, ‘error’ is not the correct word to describe the issue. The code worked fine. Rather, it was the manipulation of the dataset itself where 17 unique fraud cases in the dataset received two identifiers, a manipulation of the underlying data which made no sense. (Walker, p. 193 [here](#))

Walker is pointing out some of the questions that a real research-misconduct investigation would ask:

1. What is the nature of the problem? That should be clearly laid out in any Report on a research-misconduct investigation.
2. When the actions giving rise to the problem were taken, what was intended, as opposed to what was actuated? As an analogy, if someone mistypes a word, what was the word he intended to type? If someone misquotes a source, what is the quotation he intended to reproduce? If the problem arises from an error in judgment, what was the state of mind when the person erred? What was he thinking? In other words, why did the problem arise?
3. Walker in 2022 notes: “If an erratum were in order, why didn’t they say that in their March 2021 reply to me? In that reply, they act as though nothing in their 2020 JAR article is amiss. Furthermore, when invited by *Econ Journal Watch* to supply a second reply, why didn’t they take that opportunity to say that an erratum was in order?” JAR should have put these very questions to the author group and reported on their replies.

When a problem that arouses suspicion of misconduct occurs, investigators must ask and report on the questions posed above. If the parties in question say, “Sorry, we made an innocent mistake,” or a “coding error,” and the investigators do not probe the credibility of those excuses, they simply are not investigating. If all research-misconduct investigations are handled in such a way that misconduct is never to be found, misconduct is not handled justly. There must be a line somewhere, and the responsibility of investigators is to test matters against that line. In doing so, they both establish and evaluate evidence in the case at hand, and they help to clarify and establish the line that we all need to preserve integrity in our common project of scholarship and science.

My present remarks echo some of the insights and concerns expressed by Ian Gow in his 2022 SSRN paper “Should Bao et al. (2020) Be Retracted?” ([link](#)) and in his heavily downloaded 2023 SSRN paper “The Elephant in Room: P-Hacking in Accounting Research” ([link](#)). The latter includes the following block quotation. The quotation refers to things that I have not referenced but makes several of the points made here:

That the research community is relatively understanding about coding errors perhaps explains efforts by researchers to attribute issues in their papers to “coding errors.” For example, the code repository for Boissel and Matray (2022; retraction [here](#)) included the line *replave*  $B = B/1.8$  if  $t > -3$  &  $t < 0$ . This code modifies two coefficients in a way [that] enhances a plot used to support

a claim of “parallel trends.” While the responding author attributes this to a “coding error” it is difficult to imagine what the correct version of this line of code would be. The “coding error” in Bao et al. (2020) differs from that in Boissel and Matray (2022) in a number of respects. First, no code containing the claimed error was provided. Second, producing code with this issue accidentally seems even less plausible than with the line above from Boissel and Matray (2022; see Gow, 2022 for an attempt to replicate the “coding error”). Third, the authors’ efforts to attribute the issue to a “coding error” is belied by earlier efforts to suggest it was an appropriate research design choice (Bao et al., 2021). (Gow 2023 [here](#), p. 12 n.31)

Again, I encourage JAR to make its Report public. Also, EJW invites JAR to reply to my comment here.

In conducting and reporting the research-misconduct investigation, JAR is, in a significant way, acting as judge in its own case, for if misconduct occurred, it reflects on the editors and reviewers at JAR, not only of the 2020 article but also the 2022 “Erratum.”

In England in 1775, Parliamentarians and other magistrates had to act as judge in their own case. These words from Edmund Burke ([link](#)) seem apropos:

We are indeed, in all disputes with the Colonies, by the necessity of things, the judge. It is true, Sir. But I confess, that the character of judge in my own cause is a thing that frightens me. Instead of filling me with pride, I am exceedingly humbled by it. I cannot proceed with a stern, assured, judicial confidence, until I find myself in something more like a judicial character.

## About the Author



**Daniel Klein** is the chief editor of *Econ Journal Watch* and professor of economics and JIN Chair at the Mercatus Center at George Mason University, where he and Erik Matson lead a program in Adam Smith. His email address is [dklein@gmu.edu](mailto:dklein@gmu.edu).

[Go to archive of Watchpad section](#)  
[Go to September 2023 issue](#)