Peer Community In Paleontology

Questioning isotopic data from the end-Cretaceous

Christina Belanger *based on peer reviews by* ***Thomas Cullen*** *and 1 anonymous reviewer*

Melanie A. D. During, Dennis F. A. E. Voeten, Per E. Ahlberg (2024) Calibrations without raw data - a response to "Seasonal calibration of the end-cretaceous Chicxulub impact event". OSF Preprints, ver. 5, peer-reviewed and recommended by Peer Community in Paleontology. https://doi.org/10.31219/osf.io/fu7rp

Submitted: 29 June 2023, Recommended: 26 March 2024

Cite this recommendation as:

Belanger, C. (2024) Questioning isotopic data from the end-Cretaceous. *Peer Community in Paleontology*, 100221. 10.24072/pci.paleo.100221

Published: 26 March 2024

Being able to follow the evidence and verify results is critical if we are to be confident in the findings of a scientific study. Here, During et al. (2024) comment on DePalma et al. (2021) and provide a detailed critique of the figures and methods presented that caused them to question the veracity of the isotopic data used to support a spring-time Chicxulub impact at the end-Cretaceous. Given DePalma et al. (2021) did not include a supplemental file containing the original isotopic data, the suspicions rose to accusations of data fabrication (Price, 2022). Subsequent investigations led by DePalma's current academic institution, The University of Manchester, concluded that the study contained instances of poor research practice that constitute research misconduct, but did not find evidence of fabrication (Price, 2023). Importantly, the overall conclusions of DePalma et al. (2021) are not questioned and both the DePalma et al. (2021) study and a study by During et al. (2022) found that the end-Cretaceous impact occurred in spring.

During et al. (2024) also propose some best practices for reporting isotopic data that can help future authors make sure the evidence underlying their conclusions are well documented. Some of these suggestions are commonly reflected in the methods sections of papers working with similar data, but they are not universally required of authors to report. Authors, research mentors, reviewers, and editors, may find this a useful set of guidelines that will help instill confidence in the science that is published.

References:

DePalma, R. A., Oleinik, A. A., Gurche, L. P., Burnham, D. A., Klingler, J. J., McKinney, C. J., Cichocki, F. P., Larson, P. L., Egerton, V. M., Wogelius, R. A., Edwards, N. P., Bergmann, U., and Manning, P. L. (2021). Seasonal calibration of the end-cretaceous Chicxulub impact event. Scientific Reports, 11(1), 23704. https://doi.org/10.1038/s41598-021-03232-9

During, M. A. D., Smit, J., Voeten, D. F. A. E., Berruyer, C., Tafforeau, P., Sanchez, S., Stein, K. H. W., Verdegaal-Warmerdam, S. J. A., and Van Der Lubbe, J. H. J. L. (2022). The Mesozoic terminated in boreal spring. Nature, 603(7899), 91–94. https://doi.org/10.1038/s41586-022-04446-1

During, M. A. D., Voeten, D. F. A. E., and Ahlberg, P. E. (2024). Calibrations without raw data—A response to "Seasonal calibration of the end-cretaceous Chicxulub impact event." OSF Preprints, fu7rp, ver. 5, peer-reviewed by PCI Paleo. https://doi.org/10.31219/osf.io/fu7rp

Price, M. (2022). Paleontologist accused of fraud in paper on dino-killing asteroid. Science, 378(6625), 1155–1157. https://doi.org/10.1126/science.adg2855

Price, M. (2023). Dinosaur extinction researcher guilty of research misconduct. Science, 382(6676), 1225–1225. https://doi.org/10.1126/science.adn4967

Reviews

Evaluation round #1

DOI or URL of the preprint: https://osf.io/fu7rp/
Version of the preprint: 1

Authors' reply, 31 January 2024

Download author's reply
Download tracked changes file

Decision by Christina Belanger, posted 03 September 2023, validated 04 September 2023

Invitation to revise preprint submitted to PCI Paleo

Greetings -

Your pre-print "Calibrations without raw data – A response to "Seasonal calibration of the end-Cretaceous Chicxulub impact event" has received two reviewers. Both reviewers were generally supportive of the pre-print, but also raised some important points that should be addressed before recommendation, thus my decision is to ask for revision.

1. The reviewers see many of the issues raised with the DePalma et al. paper as reasonable criticisms, however feel that alternative interpretations to intentional fabrication of the data should be acknowledged in the pre-print. Both Reviewers also express that the degree of offence (mistake, sloppy handling, manipulation, fabrication) cannot be definitively established with the present evidence. I agree with this point, and this uncertainty should be clearly expressed in the pre-print.

1a. Reviewer 1 (who provides their comments as a .pdf) provides some alternative interpretations as well as benign explanations for why some observations made by the present manuscript. Reviewer 2 (who provides comments as plain text) points out an alternative explanation of the d13C values of the sturgeon. These explanations should be incorporated into the pre-print.

1b. Some criticisms raised by the pre-print point out what may be called "best practices" – for example, similar analytical details are missing in many papers (see R1) but are expected by others (see R2). I think an impact of this pre-print could be to recommend community standards for reporting methods and data that will help future authors and reviewers.

2. Given these valid alternative explanations, I agree with Reviewer 1 that “softening” the tone of the pre-print is needed. Reviewer 1 provides examples such as changing definitive terms, like “demonstrates” to terms that allow for alterative interpretations, such as “suggests.” R2 uses the phrases “gives the impression” and “suggests” in their review as well.

3. Both reviewers express that responsibility for the quality of the DePalma et al. article also lies with Scientific Reports and that many of the issues raised by the present manuscript should have probably been raised in the original review process. I agree with this point and shared responsibility should be more clearly expressed in pre-print.

In addition to the three themes summarized above, the reviewers also provide some detailed comments and suggestions that will improve this pre-print.

I invite you to respond to the reviewers’ comments and revise the pre-print for the further consideration in revised form.

Thank you for your submission,

Christina Belanger

Reviewed by Thomas Cullen, 17 August 2023

In this manuscript by During et al., the authors document various irregularities, omissions, and outright unusual aspects of the ‘data’ presented in a prior study by DePalma et al. which covered a very similar subject (and the same site). I think they present the evidence supporting their arguments very well, and nicely outline the issues in the DePalma et al. study. It is unfortunate that their submission was apparently not considered sufficient by Scientific Reports to be featured as an official reply/rebuttal to DePalma et al.’s study (or as sufficient grounds to have DePalma’s study investigated and potentially retracted based on these issues) [or perhaps that is still ongoing?], but at least in the current form the manuscript will be available for others wishing to examine the issues in detail. I think this manuscript has merit, and have only minor comments/changes suggested.

Overall Review / Comments:

I agree with the authors in their criticisms and concerns with the primary data itself, as well as the lack of analytical outputs. I also agree with their concerns regarding the notably incomplete methods section, which lacks information on sample weights, pretreatment and dissolution procedures, or analytical standards (all of which would be considered a basic reporting requirement in a stable isotope geochemistry study). The amount of samples they obtained via microdrilling is unusually high given the small size of the specimens (as pointed out here by During et al.), and I agree that it too requires some corroboration, as such density of sampling would be very difficult (to somewhat implausible) if done using ‘typical’ microsampling approaches for d13C and d18O analysis of bioapatite structural carbonate on specimens of the size indicated.

The combination of an absence of primary data and the substandard methodological reporting alone should have been sufficient to call the results of DePalma et al.’s study into question and functionally invalidate their paper as a result of incompleteness, regardless of whatever post-hoc excuses may or may not have been since offered by DePalma and colleagues (e.g. collaborator who ran the analyses is now deceased, no one kept analysis records, a dog ate their homework, etc). Indeed, one wonders how these issues escaped the notice of the original reviewers of the DePalma et al. study in Scientific Reports, as these problems should have been more than sufficient to reject their paper (or at least delay it until the missing information was provided to support their results/claims), going off of the editorial and data availability standards purported to be required for publication in that journal.

That above is before one even begins to evaluate the arguably more serious allegations made here that DePalma et al. manipulated or outright fabricated their data. Considering those more serious allegations, It is hard to disagree with the conclusions presented here by During et al., as there are far too many irregularities to dismiss as being the product of chance or some sort of image artefact from uploading/publishing. As the

authors note, there are myriad examples of datapoints for d13C and d18O which are mismatched/misaligned in the plots of sampling location, of outright missing datapoints (despite line dips/angles suggesting the presence of a data point), of datapoints and error bars which are only partially present on the figure or off-centre, and of places where the same specimen analysis data are presented in the main figures and the supplement but somehow have different numbers of datapoints and isotopic composition patterns.

At best, it seems fair to say that the images presented as results in DePalma et al. are manipulated in some way which has led to an inaccurate/inconsistent reporting of the original data (which of course cannot be confirmed due to the primary data not being provided by DePalma et al.), and which does suggest that this manipulation may go beyond image/figure issues and represent direct manipulation of the data itself. To that end, I largely agree that it gives the impression that the plots were 'hand-made' rather than being representative of some original data output being plotted up by a program. I think the authors' points concerning the issues with the data overlays between individuals (and their remarkably implausible consistency), as well as the similarly unlikely 'coincidence' that curves from allegedly distinct individuals happen to align perfectly when stretched or compressed, are very well-explained and raise serious questions of the validity of the data plots presented in DePalma et al.'s study. The distinction between manipulation vs. outright fabrication of data cannot be firmly established of course since the primary data are not provided by DePalma et al. (which as noted above, should itself be enough of a reason for the study results to be considered questionable/invalid/irreproducible), but it is any case fairly damning.

Minor / more specific comments:

- a few places in the manuscript the authors write "carbon and oxygen isotopes" without including "stable", which I think should be corrected given the context of these discussions does not concern radiogenic isotopes such as 14C.

- concerning the d13C compositions of the sturgeon not showing a strong shift representative of feeding in marine vs freshwater settings (and particularly when the d18O does show a stronger cyclicity), I would mention that it is not impossible for a marine vertebrate to have d13C compositions in the –4 to –1 per mil range. For example, mosasaurs have been reported in the –12 to +2 per mil range, depending on location/time. That doesn't necessarily contradict the overarching concerns, just noting that I don't think that lack of cyclicity is necessarily something that supports charges of data manipulation per se.

- on lines 94-95 you note that one of the re-used samples (FAU.DGS.ND.755.57.T) has 43 sampling spots in Fig 2 but has 29 sampling spots in the Supp Materials. When I look at the graph for that specimen in SUP MAT 9 of DePalma et al.'s paper I see 35 sampling spots for d13C and 35-36 for d18O (the line angle suggests a point but one is not marked). I do agree with you, however, that the plots provided for the same specimen in Fig 2 and SUP MAT 9 do not match and show different numbers of samples, which is very irregular and requires some sort of explanation (particularly when it is identical for portions of the interval, but widely different in other select locations).

Reviewed by anonymous reviewer 1, 29 August 2023

Thank you for the opportunity to review this manuscript, which is a response to a recent paper published in Scientific Reports (DePalma et al 2022). The manuscript calls attention to a number of issues regarding the Scientific Reports publication, especially regarding the lack of data availability, methods details, and possible inconsistencies in the presentation of results. The manuscripts argues that these issues represent evidence that the results may be fabricated, a very serious allegation indeed. I approach this as someone with 10+ years of experience conducting stable isotope analysis on biological and paleontological samples, including the collection, handling, and (micro)sampling of samples, running (and customizing/repairing) isotope ratio mass spectrometers and associated peripheral devices (including those mentioned in this manuscript), and handling datasets ranging in size from tens to thousands of stable isotope measurements.

My overall impression of this manuscript (perhaps shared by the authors), is that all of these issues ought

to have been raised during the peer review and editorial process at Scientific reports prior to the publication of DePalma et al 2022, and thus are reasonable to raise in some form. However, a number of these issues are minor (not naming the analytical facility, not providing sample weights, not naming specific standards used) and do not either individually, or in combination, provide evidence one way or the other regarding the possibility of data fabrication. Some issues raised in this manuscript regarding the graphs in DePalma 2022 are potentially more serious, and are indeed worth raising, but I don't see a smoking gun. As such, I would ask the authors of this manuscript consider revising their manuscript such that it clearly acknowledges alternative interpretations of the issues raised, such as unintentional mistakes, database (copy/paste) errors, or graphing software misuse cannot be discounted.

Specific comments:

Lines 67-69: I agree the lack of data availability is unfortunate, and that the authors of the Scientific Reports publication should have included results with their paper. Some fault here also lies on the editor and reviewers of that paper, and as such this issue does not itself constitute evidence of fault solely on the part of the authors of the Scientific Reports publication.

Lines 71-74: I agree that it is good practice to include this information, but many papers do not and this does not constitute a major anomaly as long as there is some clear indication where the analyses were conducted and by whom, which the original Scientific Reports publication does clearly provide.

Lines 77-78: I agree that it is good practice to include such information, but this does not constitute a major anomaly but rather a minor omission that is often caught in the course of the peer review/editorial process. The authors of the PCI preprint might specify what other information they would wish to know regarding the techniques. For instance, one might wish to see a statement explaining that phosphoric acid was used to analyze carbonate component of fossil samples, and the reaction temperature.

Lines 83-89: This is a reasonable question to raise, and I agree here that additional information should have been provided by the authors of the Scientific Reports publication regarding their sampling strategy, especially regarding the typical area over which powder was collected for each analysis.

Lines 110-113: How do the authors define failure of either measurement? Do they mean the software does or does not provide a value? Regardless, it is not correct to say that situations where either carbon or oxygen analyses fail (however defined), the other cannot still be used. Rather, it depends on why how failure is defined. For instance, if high inter-peak variation is observed in d18O for an individual sample, the d13C value could still be used if the its inter-peak variation is 'normal'.

Lines 113-117: I wonder if this could also be explained by repeated micro sampling of the same areas, measured multiple times, or potentially by errors in spreadsheet management and/or data use in graphing software.

Lines 117-118: Could this also be the result of 'sloppy' use of graphing software?

Lines 119-127: I agree the difference in error bar length is an issue worth raising, but as with the other issues raised here more innocent explanations such as simple sloppy graphing software use cannot be discounted. The parenthetical statement is not relevant here and should be removed.

Lines 128-133: The conclusions here are one possibility, but their case is very far from conclusive. I do not mean to suggest that such errors are unimportant, but sloppy handling of data and graphic software (perhaps by a student) could very easily result in such issues, which indeed should be corrected but are nonetheless not equivalent to intentional forgery. Thus, the authors should soften their language a bit, especially by changing "demonstrate" to "suggests the possibility" and also by acknowledging other possible explanations.

Lines 137-138: This sentence is too vague, please provide more information.

Lines 141-154: These are interesting points raised here, which are perhaps the most (really, only) compelling evidence to even raise the possibility of data fabrication.

Figures: Could the authors please define "misaligned" data points?

Download the review