Establishing a workflow for recording and analysing bioarchaeological data

Christianne Fernee *based on peer reviews by 2 anonymous reviewers*

The paper by Xanthopoulos and colleagues [1] presents an approach to establish a pipeline for the analysis of osteological and biochemical data. This approach integrates novel data collection, FAIR principles for data longevity and accessibility, utilises R markdown and cloud webware. Following the changes recommended by the reviewers this paper presents a welcome contribution to osteoarcheology and bioarchaeology.

Osteoarchaeology and bioarchaeology often involves the collection of vast amounts of data both in the field and from consequential analysis in the lab. From this data we can reconstruct many aspects of past human experiences. However, issues often arise when bringing together these diverse types of data. In this regard, this paper proposes are useful methodology in which osteoarchaeological researchers can bring their data together as part of a streamlined process, from data collection to analyses.

References:

[1] Xanthopoulos, K., Georgiadou, A. and Papageorgopoulou, C. (2024). An approach to establishing a workflow pipeline for synergistic analysis of osteological and biochemical data. The case study of Amvrakia in the context of Corinthian colonisation between 625-189 BC in Epirus, Greece. Zenodo, 11156506, ver. 3 peer-reviewed and recommended by Peer Community in Archaeology. https://doi.org/10.5281/zenodo.8298579

Reviews

Evaluation round #1

Authors' reply, 09 May 2024

Dear Christianne Fernée

I am writing to inform you that I have resubmitted my paper after taking into consideration all the suggestions and corrections the reviewers had noted. I used the zenodo platform to upload the new file as version 2 and a new doi has been generated (10.5281/zenodo.11156506). I have edited my article data accordingly.

I remain at your disposal.

Best regards

Kiriakos Xanthopoulos

Decision by **Christianne Fernee**, posted 18 December 2023, validated 19 December 2023

This paper broaches a subject that is very needed in archaeology/osteoarchaeology. The paper shows a lot of promise, however some revisions are required prior to recommendation. In particular, the paper would benefit from:

- Abstract: a more concise extract that reflects the content that is found in the paper

- Introduction: An acknowledgement of the attempts made in in osteoarchaeology to aid data collection and standardisation

- Methods: an explanation of the pipeline in more detail

Reviewed by anonymous reviewer 2, 11 November 2023

I would like to express my sincerest gratitude to the authors for broaching the subject. The problem of data collection and the exploration of datasets in bioarchaeological research is of paramount importance for many osteo-archaeological projects. The huge amount of data that we collect in the field and in the laboratory are difficult to interpret in their entirety and an R tool which facilitates the process of exploring the patterns of both osteological and isotopic data in one go would be a welcome addition to the bioarchaeological toolbox. However, the promise made in the abstract and in the introduction has not been delivered convincingly within the text itself.

A considerably shortened abstract, that is more to the point, would be better. The key sentence being 'we propose a tool that enables the researcher to automatically find any correlations between dental pathologies and isotope values'. The keywords are misleading as no paleodemography, paleopathology, isotope, and database architecture are discussed within the text. I would recommend using bioarchaeological keywords instead.

[line 48] The introduction would clearly benefit from more references concerning the previous studies using both 'stable isotopic and dental health data'.

[lines 51-52] It would be great to see the tool and the outcome it generates and the rationale behind the analysis mentioned in the passage 'we generated a tool that enables the researcher to automatically find if there are any correlations between dental pathologies and isotope values'.

[line 60] It would be great to see a North pointing arrow on the picture.

[lines 66-68] These sentences are quite confusing. First you mention 'integrating different types of recorded data into one common database' and then you write about coping with this using: web storage 'Google Drive, Dropbox', a group collaboration suite 'Box', and a code versioning solution 'GitHub'. Please, clarify.

[line 77] What do you mean by 'most undivided'. Please, elaborate.

[line 79] What do you mean by: tomb, burial, and assemblage? It would help if you could explain the differences. Can assemblages and burials be found within the tomb? Do you identify all commingled burials as assemblages?

[line 80] Could you clarify what you mean when writing about an individual? Do the lines 79 and 80 together mean that you have been able to identify the MNI of 100 when investigating 200 tombs? Did you in addition to these 200 tombs investigate a number of burials and assemblages? Please clarify.

[lines 81-83] The citation in line 83 should be moved to line 81 and come immediately after 'individual burial'.

[line 84] The picture would benefit from the area 1 label being more visible. It also would be helpful to see the North pointing arrow on the picture.

[lines 87-93] I would suggest this paragraph be reworked into a longer description of how you comply with all the FAIR principles. While doing this, when possible, please, keep the division between data and data analysis clear for the reader.

[lines 97-98] Needs rephrasing. I think I know what you mean but your point is not stated clearly. The second sentence is especially confusing. Could you please elaborate a little more about the data you collect and the way you do it?

[lines 99-103] The whole paragraph is very confusing. It could be rewritten in a few paragraphs explaining your 'pipeline'. You should also describe exactly how, and why, you use the software listed in the paragraph.

[line 106] How do you integrate your data if you allow researchers to use any means available out there when they want to collect data?

[lines 121-123] For this paragraph. Could you please provide the rationale behind your proposition to use cloud server environments; 'cloud server' means nothing specific in terms of technology or the services it provides. Do you recommend it for providing spreadsheet functionality to allow online data collection by all collaborators?

[line 126, 127] By 'parametrization' do you mean standardization and data cleaning?

[lines 129-138] The bibliography given there, although great from a bioarchaeological point of view, is not needed for the subject presented. These bibliographical references would be invaluable if you were to rewrite this article to be about your solution from the point of view of the bioarchaeological analysis it performs.

[line 146] In this line you write about the tool created. Is this available for download anywhere on the internet? If so, please, provide a reference to it.

There are many important problems mentioned in the text: data collection, cleaning, analysis, and publication but none of them is described in depth. Moreover, the 'pipeline' stated in the subject is not satisfactorily elaborated upon in the text. I would recommend that the authors select a subject

and concentrate on it. They could follow up on what they promise in the introduction and describe the tool they devise as a focal point in the 'pipeline'. I can see at least two potential approaches to this. They could describe the tool and the 'pipeline' from the point of view of their technical development and use. They could also elaborate on the rationale behind the 'pipeline' they propose and the bioarchaeological solution behind the R code they have created – or behind the R code they intend to create.

I strongly recommend that the text be corrected by an English native speaker with good writing skills. There are many sentences in the text that needs clarification.

The article has great potential but I strongly recommend that the authors refrain from publishing it in its current form. The whole text gives the impression of being a draft document and that the actual subject matter has not yet been fully clarified.

Download the review

Reviewed by anonymous reviewer 1, 01 November 2023

Summary

Overall, this is a welcome contribution to the field of bioarchaeology, both in regards to material from the Aegean and beyond. The study's technical merit and scientific significance is unquestionable. The field could greatly benefit from such a remarkable idea and I will be looking forward to the final publication. However, I would recommend some minor revisions to the manuscript in order for it to be more precise and complete. Thus, please accept some suggestions for improvements that I hope would greatly improve the paper.

Title and abstract

In more detail, the title and the abstract suggest an innovative and promising idea that is greatly lacking in the field of bioarchaeology. They both provide an adequate synopsis of the manuscript as submitted and a comprehensive outline of the content and the aims of the paper. No results are provided, but this does not pose a problem for the present study and its scope.

Introduction

The introduction refers to the limitations the current state of research faces regarding data collection and standardization. However, it does not contain sufficient literature review on the matter and suggest any publications that have been trying to overcome such limitations.

Following there are some publications towards this direction:

Standardized Osteological Database (SOD), Annual Review Anthropology 1996 article
`https://www.jstor.org/stable/2155819?seq=17`

Wellcome Osteological Research Database
`https://www.museumoflondon.org.uk/collections/other-collection-databases-and-librari es/centre-human-bioarchaeology/about-osteological-database`

Caruso, A., Karligkioti, A., Selempa, G. and Nikita, E., STARC OSTEOARCH: An open access resource for recording and sharing human osteoarchaeological data. International Journal of Osteoarchaeology. `https: //airtable.com/shr4mDZga3uMFN35n`

Paleopathology Association efforts for standardization
Rose JC, Anton S, Aufderheide A, Buikstra JE, Eisenberg L, et al. 1991. Paleopathology Association Skeletal Database Committee Recommendations. Detroit: Paleopathology Association.

White, W. 2008. Databases. In Advances in Human Paleopathogy, ed. R Pinhasi and S. Mays, 177-188. John Wiley and Sons. `https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470724187.ch9`

Methods, Analyses and Publication

The procedures regarding decision-making and the approach implemented are stated clear enough. However, the methodologies implemented need to be clarified in more precision. Furthermore, even though the authors are using the colonization of Ambrakia as a case study, they do not offer specific examples coming from the application of their methodologies in the case study. E.g. the way they correlate oral pathologies with isotopic

values, as they proposed on line 59 of the introduction. Therefore, the suggested methodology for a pipeline of workflow analysis based on an Ambrakia, as the abstract and title suggest, is merely given at a theoretical level, while the practical part is critically missing from the study. Such an addition would greatly help the scientific community understand and implement the authors' methods in order to facilitate data sharing and correlation.

The sections concerning Recording Data, Importing and Tidying Datasets, Analyses and Publication are not very clear and could greatly benefit from some additional details.

Thus, I would like to suggest the following additions:

a) The names of the parameters the authors utilize and the analyses they are related to should be given in detail.

b) The specific descriptive statistics and/or specific statistical tests that are utilized should be mentioned.

c) The specific R packages the authors utilize should be mention in more precision, e.g. the authors refer to tidying datasets. Is this a reference to the TidyR R package?

d) In the beginning of the paper the authors mention R Scripts. For the task the authors are describing R markdown notebooks seem appropriate. Have the authors considered them?

e) Figure 3 could be replaced with a more detailed workflow diagram.

Finally, the authors provide no discussion regarding tools for data querying tables (such as SQL) in order to discover correlations in relation to age, sex, archaeological data and osteological biomarkers that are necessary for researchers to identify, interpret, and report. Even though this does not suggest a mandatory

Download the review