

A TAO for Data Wrangling: A Practical Routine for Getting Past the ‘Janitor Work’

Memo to Editor and Reviewers

We thank the editors for the opportunity to revise this manuscript and thank the reviewers for their careful attention and helpful comments. This feedback has strengthened the paper. Below, we summarize how we have addressed each of the key suggestions.

Clarifications and Terminology Both reviewers and the editors commented on terms and phrases that could benefit from clearer definitions. In this revised version, we have refined the framing and added explanations where needed.

Following Reviewer 1’s (R1) suggestion, we now clarify the distinction between survey and aggregate data and explain their respective harmonization process in greater detail (see p.2). We also follow Reviewer 2’s (R2) suggestion by adding the term “*retrospective data harmonization*” alongside “*ex-post data harmonization*” to ensure that readers familiar with either expression can follow our discussion (see p.8).

In the comments, R1 also expressed interest in understanding how DCP0tools works. We appreciate this attention to our software. However, word limits constrain us from detailing DCP0tools in the manuscript. To provide guidance, we will publish a software vignette. As noted on page 4, we direct readers who are interested in the operation to the appropriate resources for further information.

Expanded Literature Review We thank R2 for highlighting additional relevant scholarship on data harmonization. We have incorporated these references into the revised manuscript to better engage with the broader literature (see p.1).

Further Validation of Step 2 We appreciate R2’s observation that while automation in Step 2 of our routine reduces errors associated with manual data entry, it cannot fully eliminate all possible mistakes. In the revised version, we now explicitly acknowledge this limitation and discuss the validations for automatic data entry (see p.6 and p.8).

We also thank the editor for noting the connection between our discussion and other contributions in this symposium regarding software usage. We have integrated this point into the manuscript (see p. 6).

Finally, **Polishing** We are grateful for the editors’ and reviewers’ suggestions to further improve the manuscript’s clarity and framing. The revised version has been thoroughly edited and proofread. The revised parts have been italicized and highlighted in the revision. If any parts remain unclear, we would be pleased to make further revisions.

Thank you again for your careful reading and valuable feedback.