Response to reviewers

Submission ID 217815520

We thank the Editor and the three reviewers for their comments, which have helped substantially to improve the manuscript. In response, we have made considerable changes to the main manuscript as outlined below.

Minor changes:

- We have corrected all minor grammatical errors pointed out by the reviewers.

Major changes:

- Reviewers suggested reframing the problem space so there is more emphasis on the case study.

Reviewer 1: The authors prominently cite (Chatfield 1985) for the definition of the term "IDA". However, Chatfield fairly explicitly defines IDA to be a data summarizing and scrutinizing process; not a cleaning, scrubbing, or munging process. It seems curious to me that the whole paper is framed as being "IDA" based on the author's unsupported statement that data cleaning should be considered party of IDA.

- Supporting information for the lack of data cleaning in papers.

Reviewer 1: The authors also emphasize that "There are few research papers that document the data cleaning". While I agree that data cleaning is under-emphasized and under-documented, it's hardly true that there are not papers on the topic. A quick Google Scholar search for "data cleaning" unearths many scholarly papers and books on the principles of this process, none of which appear to be cited in this work.

- Outlier detection

Reviewer 1: Section 3.2.1 goes into detail on the authors' approach to characterizing erroneous outliers in the data. Their approach is reasonable, but it is not (as far as I can tell) based on any established or tested procedure. This section once again presents a tension between the paper as a case study versus a topical commentary: if it is a case study only, then the "common sense" justification for the approach is appropriate, but if it is meant to establish future norms, these modeling choices must be more formally supported. The reference to a "reasonable degree of fluctuation" particularly struck me as a subjective or "ad hoc" claim. I am particularly a bit concerned by the statement on page 13: "The robust mixed model could be the best model to be employed in this case. However, this method is too computationally and memory expensive, especially for a large data set, like the NLSY79 data." Surely, fitting a mixed-effect model to a dataset with only one predictor and 1188 individuals ought to be very feasible?

- We thank the reviewer for pointing out to us that the data contains information on the work experience – this additional data is added in the extraction of the data in Figure 1. This information was compared with Singer & Willet's original data. Comparison shown in Figure X shows that the information does not match up, however, there is a high correlation between the variables.