Negative Data Initiative [Letter v0.1]

Trainee collective

Abstract An ongoing community letter for Negative Data Initiative. Current focus : Negative Data in Computational Biophysics.

1 Why this initiative?

The core idea is to help the trainees to showcase the negative results and pitfalls they encountered while having good visibility and feedback from the community. Based on an earlier tweet discussion, a group of trainees came towards to brainstorm on this idea. The group thinks that the following model might work better:

Short-term goals:

1. Form a community consensus on the definition of negative data

2. Once the consensus is made, prepare a checklist for the trainees to follow to make sure what they have is negative data

3. Form an advisory board to assist the trainee group by inviting senior members of the community

4. Initial call for moderators/volunteers to the trainee community to organize a virtual webinar/conference. This virtual gathering allows the trainees to discuss and share their negative data. Applications to the virtual gathering are based on abstract submission. This call for volunteers is to organize the event. The advisory board will be kept in the loop during this entire process

Long term goals:

5. Organizing the event

6. Potential publication via LiveCoMs "lessons learned" article by grouping the negative data

7. Use this model to advocate for more visibility in major conferences

The group decided to run a pilot program of this model with the computational biophysics community since the initial core trainees and potential advisory board members will be coming out of this community.

2 Discussion

The trainee group is requesting feedback from the community on (1) the overall model, (2) the definition of negative data, and (c) a potential (LiveCoMs-like) checklist for trainees to identify the negative data at their hands. Following the LiveCoMs protocol, we will use the Github issue tracker for feedback, suggestions, and further discussions.

2.1 Feedback on the overall model

Regarding the feedback on our proposed model, described in the previous section, the discussions are hosted here at github issue #2

2.2 Feedback on negative data definition

The trainee group discussed on the definition of negative data and suggested the following. Again, we are using this to initiate a discussion and will finalize it after extensive feedback from the community.

Definition 2.1. Negative Data does not support the initial hypothesis or expected outcome after careful and reproducible meticulous scientific research. However, the authors believe that this can drive science forward or are lessons learned that help future trainees while designing their experiments.

We hope to reach a clear, concise definition of negative data through the discussions hosted here at github issue #3. We ask the community to share their definition of negative data here.

2.3 Feedback on the checklist

From the brief discussion during the brainstorming meeting, we propose a general outline for a checklist to identify negative data and invite community suggestions. Please find the checklist on the next page. The discussion for the checklist feedback is hosted here at github issue #4

INITIAL HYPOTHESIS

☐ An initial testable hypothesis backed up by strong rationale is formed (based on literature review, prior knowledge, research gap identified, etc.)

☐ A detailed, careful experiment design is formed to test the hypothesis with statistical rigor. (Included independent replicates, following best practices in the community, ...)

☐ Benchmarks for testing the validity of results is set up a priori.

ANALYSING THE DATA

☐ Sanity checks have been done to the best of your abilities. Code/method review, debugging, testing..etc. Independent replicates

☐ Identified or tried different analysis schemes when the initial scheme failed the anticipated results.

☐ Additional testing scheme was added to check for a potential flaw in the design

☐ Considered alternative hypotheses and tested them

INTERPRETING THE DATA

☐ You have not done any "cherry-picking" of data. Analysis schemes are done on data without any bias

☐ Your analysis indicates the rejection of your initial hypothesis with statistical rigor. Or, you see an absence of anticipated features or characteristics that helped you formulate your hypothesis. Including simple, semiquantitative checks that describe them.

REPORTING THE DATA

☐ Honest and complete reporting of all the attempted tests and rationale

☐ Report alternative hypothesis tested if any. Discuss such possibilities.

☐ Discuss the potential flaws, pitfalls, and other features of the method that was used.