Data sharing, a four week overview with a critical review of existing practices

### Long term goal

Our long term goal is to enhance the scientific rigor, reproducibility, and responsible conduct of biomedical data science research through a four week curriculum on data sharing that you can use in a stand-alone one credit hour class or incorporate into an existing course on research ethics or methodology. The work will culminate in a team project where students conduct a critical review of representative sample of open data sets in a particular scientific or medical specialty.

### Aim 1. Development and review of training materials.

The research team will develop Powerpoint slides and a script included in the speaker notes covering three major areas: the mandate for data sharing, privacy and ethical concerns, and data documentation standards. We will also develop exercises for each area along with grading rubrics. We will seek external reviewers for this material and revise the materials as needed.

The team project is the heart and soul of the training module. We will recruit small teams of students in three disciplines: Biology, Medicine, and Psychology. These teams will attempt to identify twelve data sets associated with peer-reviewed publications and attempt to reproduce at least one descriptive statistic from the associated paper. The students will document the time spent and will offer feedback on how difficult they found this exercise.

### Aim 2. Deploy materials in Clinical Research Methodology class and conduct evaluation.

The sharing data module will be deployed as part of an existing class on Clinical Research Methodology. Teaching materials will include close-captioned videos, Powerpoint slides with speaker notes, homework assignments with grading rubrics, and the final team assignment. Students will be tested on their retention of key concepts at the end of the class. Students who provide consent will be contacted six months later for an additional assessment of retention.

### Aim 3. Make all resources available publicly.

All the Powerpoint slides, speaker notes, handouts, exercises, and grading rubrics will be developed using RMarkdown and stored on a github site. The videos produced will be posted on a Youtube channel. All products produced by this grant will be licensed under the [Creative Commons Attribution license](https://creativecommons.org/licenses/by/4.0/) (CC-BY 4.0). This is the most liberal license available and will allow users to create derivative works.

### Where to find this document and related documents.

* Bibliography in [html format](http://www.pmean.com/post/r25-bibliography.html) or in [R Markdown code](https://github.com/pmean/peer-reviewed-data-sets/blob/master/src/r25-bibliography.Rmd).
* R25 grant draft in [html format](http://www.pmean.com/post/r25-grant.html) or in [R Markdown code](https://github.com/pmean/peer-reviewed-data-sets/blob/master/src/r25-grant.Rmd).
* An earlier brief overview of the R25 grant in [html format](http://www.pmean.com/post/training-module-summary.html) or in [R Markdown code](https://github.com/pmean/peer-reviewed-data-sets/blob/master/src/training-module-summary.Rmd).
* Draft publication on figshare in [html format](http://www.pmean.com/post/data-sharing-publication.html) or in [R Markdown code](https://github.com/pmean/peer-reviewed-data-sets/blob/master/src/data-sharing-publication.Rmd).
* R25 specific aims in [html format](http://www.pmean.com/post/r25-specific-aims.html) or [R Markdown code](https://github.com/pmean/peer-reviewed-data-sets/blob/master/src/r25-specific-aims.Rmd).