This project set out to observe the distributions of p-values derived from top medical literatures within a certain time frame.

Of additional interest was the observation of p-value bias and the over inflated importance of p-values.

Data for this project was sourced from abstracts of 3 major epidemiology journals, APE, JAMA and Lancet by means of web scrapping (using R programing language) and manual review. Abstracts retrieved where from the period of 2000 to 2010.

P-values, effect estimates such as hazard and odds ratios with the addition of confidence intervals (CI) and sample size if available were retrieved from the abstracts.

**/upload directory**

This folder contains raw data utilized for this project.

This includes two files, an edited file with our criteria’s/columns of interest and an original unedited file which has all the criteria’s.

**/Code**

Folder contains code used for the scrapping of p-values and other criteria’s of interest. Original code depicting the extraction of p-values from abstracts was sourced from the Jager and Leek paper titled *“An estimate of the science-wise false discovery rate and application to the top medical literature* “and can be found using this link <https://github.com/jtleek/swfdr/blob/master/getPvalues.R> .

**/upload/Figures**

Folder contains figures derived from this project. Figure 1a and 1b illustrates (get picture details from jordan).

**/upload/tables**

Table 1 highlights (get table information from Jordan).

In summary,