Goals:

1. Replicate, as far as possible, the work of Prayle, Hurley and Smyth in the paper titled “Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study”, BMJ 2011; 344:d7373, January 2012. Do it twice as follows:

   a) Use the same set of 1465 trials as in the paper (data will be provided). This will allow you to compare your results with theirs.

   b) do it with a comparable set of trials where the Completion Date is given and is December 31, 2015.

2. Conduct analysis with the clinical trials data on a topic of your choice. Some options I can suggest are:

   a) a topic that relates to the paper on Neglected Tropical Diseases by Verrest and Dorlo.

   b) analyze a small sample of trials that should have but do not report results and have at least 3 publications after the Completion Date. Checking PubMed to see if they have results, if these are positive or negative results. Hypothesis: these are largely negative results.

3. Conduct analysis with the clinical trials data on a topic of your choice that relates to the paper on Learning Disease Relationships from Clinical Drugs by Haslam and Perez-Breva. (Data available)

Each group will explore their own interests in goals 2 and 3. You should let me know what these are as soon as possible but definitely by Feb 10. I will guide you on scope.

What you will learn at a minimum:

Challenges of working with a real (and important) dataset (clinicaltrials.gov)
Defining small scale data analytics problems that you can solve.
Parsing XML
Basic statistical analysis
Analysis and presentation of your results

What to submit:

A paper (written in formal ACM style; double column etc.), that has a brief overview of your goals, your methods, results, analysis and a description of limitations in your work. Each goal will be described in a separate section. Please do not cut and paste any sentences from the papers you read. Instead summarize and paraphrase.

ACM template files are available at: https://www.acm.org/publications/proceedings-template
Older versions are at: https://www.acm.org/publications/proceedings-template-16dec2016
Work Distribution:

A cover sheet must be included that describes the contributions of each member of the group. The cover sheet needs to be signed by all members.