Biostatistical Methods for Clinical Research III

Written Project

This will be a written report of a data analysis, similar to a brief, focused research manuscript but with more statistical detail than usual. Ideally, the project should be an original analysis of data relevant to your own research interests, potentially leading to a submitted manuscript. Address a single substantive issue or a few closely related issues; this should not be a comprehensive report of all findings from a study.

The report should:

- carefully define the key issues, questions, or hypotheses to be investigated
- describe the study design and the data that were collected
- explain the statistical methods used and why you chose them, including problems or difficulties and how you handled them
- provide valid and informative summaries of relevant results
- interpret the results appropriately, noting caveats
- discuss substantive conclusions and implications; this will often include noting what is different from the most relevant previous studies and explaining possible reasons, along with implications for further research

Text length should be **less than 2000 words**, with a total of **no more than four figures and/or tables**. Include a title and references, but no abstract. Details of non-statistical methods, such as laboratory or surgical techniques, should be omitted if possible, along with lengthy details of inclusion/exclusion criteria that are not essential for interpreting the results. Limit the background to the minimum necessary for understanding the key issues, avoiding extensive summaries of previous literature.

You will be assigned a mentor from the Biostatistics Division faculty to provide guidance for your project. Please make use of this resource.

Due Date: **May 22, 2009**

Turn in hard copy to Olivia De Leon.

Project sessions will be scheduled with your advisor.

Expectations for Written Project

Major plusses

- Appropriate use of difficult methods, such as multiple random effects, difficult time-varying covariates, bootstrapping, cross-validation, propensity scores, etc.
- Insightful interpretation of the statistical results, especially if sensibly synthesizing biological knowledge or results of prior studies.
- Devising non-obvious ways to focus on issues that aren't optimally addressed by typical approaches.
- Recognition of subtle but important possible limitations or biases.
- Exploration of alternative approaches, with good assessment of what differences or agreements show.
- Correct use of multiple imputation, particularly if it makes some noticeable difference versus case-wise deletion.

Minor plusses

- Interpretation of estimates and confidence limits.
- Appropriate use of intermediate methods, such as mixed effects models, ordinal logistic regression, simple time-varying covariates, etc.
- Checking major assumptions of analyses.
- Appropriate handling of assumption violations.
- Clear, informative graphs or figures.

Tie-breakers

- Interesting question or findings.
- Clear writing.
- Spurious precision.

Minor problems

- Failure to check assumptions.
- Indirect analyses or summaries when more direct ones are possible. e.g., giving separate summaries of before and after values without summaries of changes.
- Confidence intervals not provided for some key results.
- Poor summaries, such as mean +/- SD when data are severely skewed.
- Definitions of important variables not clear enough.
- Inadequate details about follow-up and/or event ascertainment and definition.
- Poor writing.

Major problems

- A data set and question that does not permit very interesting analysis.
- Exclusive focus only on p-values when estimates and CI's could be obtained and examined.
- Failure to address obvious violations of model assumptions. e.g., using PH regression or logrank test when K-M curves show severe non-proportionality.
- Ignoring dependence in the data, e.g., clustering.
- Interpretations that exaggerate how conclusive the results are.

Severe problems

- Inappropriate use of only cursory statistical analyses. e.g., only correlations or t-tests when multivariate modeling is both possible and needed.
- The main questions or issues don't make sense or can't be discerned.
- Conclusions that contradict the results of the analyses.

Guidelines for Oral Presentation

Prepare a 25 minutes talk. Each room will have a laptop available for presentation. It is ideal for you arrive with your talk on a USB drive or make an arrangement to e-mail it to your project advisor or Olivia DeLeon in advance.

Apportion the time approximately as follows

5 minutes on background
(the scientific question, available data)

5 minutes on statistical methods
(how you analyzed the data and why)

10 minutes on major results
(need not be comprehensive as in the written report)

5 minutes for questions and reflections

The talk should have a methodological bent. You'll want to highlight any subtle of difficult choices you made in the analysis and explain why you made this. Nearly every project will present some issues which we haven't covered in the class. Take the opportunity to explain to fellow students any novel issues you encountered and the techniques you used to address them.