Statistics 215A, Fall 2023 Final Project
Learning and Evaluating Clinical Decision Rules

Due: Friday December 8, 11:59 PM

Like Lab 4, you will be put into several groups to work on this project. Each group only needs to submit one version of the project. Whoever is submitting for your group must push a project_final/ folder to their Github Repo with the following files by the deadline:

- **project_final.Rmd**: the raw report and code required to create the writeup PDF.
- **project_final.pdf**: the output of project_final.Rmd. This output should be no more than 12 pages and should not contain any code output.
- **R/** : a folder containing any . *R* scripts (e.g. load_data.R and clean_data.R) that will be sourced in project_final.Rmd and any other pieces of code you use for your analysis and report.

You may create additional folders if you wish (figures etc.), but please keep your project directory organized. Push everything necessary to reproduce your report, but nothing else (e.g., do not push your data/ or documents/ folders). Note also that you do not have to create blinded files.

1 Introduction

The Pediatric Emergency Care Applied Research Network (PECARN) performs research into acute injuries and illnesses among children in a wide range of demographics and institutions. Practitioners in clinical medicine must often prescribe treatment based on imperfect information and often weight tradeoffs when the treatment has the possibility of adverse side effects. For example, the benefit of a CT scan must be weighed against the risk that ionizing radiation from the scan will cause further health problems in the patient. In order to aid clinical practitioners, it is of interest to develop algorithmic risk scores in order to screen patients based on their clinical characteristics and identify patients of high risk. In this project, you will be able to choose among three PECARN datasets and will develop an interpretable clinical decision rule for your dataset and evaluate the rule in a range of settings. Each of the three datasets explores a different injury common among pediatric patients, one of which you have explored in Lab 1: traumatic brain injury (TBI).

2 Data

Your group must select one of the three datasets detailed below in order to carry out the project. Note that you will need to clean the data in each case, but if the TBI data is selected, and you agree with the cleaning steps done in your Lab 1, you can use that method of cleaning. **As an incentive for the groups that decide to use a new dataset, if your group uses one of the new datasets, you will recieve 3 points of extra credit added to your project score ($\sim$ 5%) of the total points**. The data available comes from three different injuries:

- **(CSI)** Predicting Cervical Spine Injury. Data can be downloaded here: `https://pecarn.org/datasets/` "Predicting Cervical Spine Injury (CSI) in Children: A Multi-Centered Case-Control Analysis". Original papers:
  - Factors associated with cervical spine injury in children after blunt trauma
  - Utility of plain radiographs in detecting traumatic injuries of the cervical spine in children
- **(IAI):** Identifying children at very low risk of clinically important blunt abdominal injuries . Data can be downloaded here: `https://pecarn.org/datasets/` "Identifying children at very low risk of clinically important blunt abdominal injuries". Original paper:
  - Identifying children at very low risk of clinically important blunt abdominal injuries
- **(TBI):** Identification of children at very low risk of clinically-important brain injuries after head trauma. Data can be downloaded here: `https://pecarn.org/datasets/` "Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study". Original paper:
  - Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study

Prior to starting your project, your group should decide which of the above datasets you will use.

3 Tasks

How can we best vet and/or improve the clinical decision rule for your given problem? Most importantly, the clinical decision rule should be highly predictive and minimize the amount of missed diagnoses (i.e. have a very high sensitivity). It should also be easy-to-use, using variables that clinicians can readily have access to when making their decisions. Finally, the interpretability of the rule helps to check whether its predictions will make sense for new patients and makes it easier to apply in new settings.

Specifically, for this project, your group will have to deliver the following:

- **Introduction:** Introduce the domain problem for the dataset you have selected. Motivate why this is a relevant problem to solve and give some background of previous work in the area.
- **Data:** Download the raw data. Detail your data cleaning process and document all judgement calls made. Describe the features in the processed data and the outcome features. Detail how you will split the data for the development and testing of your new clinical decision rule. Select a hold out test set that will reflect how the model will be used in practice.
- **Modeling:** Implement the baseline clinical decision rule from the original paper for the dataset you selected. Develop and implement a newly derived model for the dataset. Explain in detail how you developed your model, and describe any intermediate models you did not end up using that led to your final model.
- **Interpretation:** Examine the interpretability of your model. This should help clinical practitioners use your model in practice. Is your model a simple interpretable form? If not, how do you recommend interpreting how it obtains the predictions it does?
- **Stability under Model Perturbation:** Introduce a perturbation to your final model, and summarize the effects of this perturbation on the predictions of your model.
- **Stability under Data Perturbation:** Study your final model under three perturbations of the data:
  - What happens if the covariate distribution of the test set changes from what the model was trained on? Simulate this and present the implications for your model.

  - What happens if your model is used on only a subgroup of the patients the model was trained on? Simulate this and present the implications for your model.
  - Create an additional stability check and show how it affects your model.

- **Evaluation:** On your held out test set, evaluate and present the final performance of your model. Is the accuracy similar to that in the training/validation sets? If not is there any pattern in the errors? Discuss the implications this evaluation has for any possible use of your model in real life.

As you work on this project, keep the following ideas about the data from Lab 1 in mind:

3.1 Data Collection

What are the most relevant data to collect to answer your domain question?

Ideas from experimental design (a subfield of statistics) and active learning (a subfield of machine learning) are useful here. The above question is good to ask even if the data has already been collected because understanding the ideal data collection process might reveal shortcomings of the actual data collection process and shed light on analysis steps to follow.

The questions below are useful to ask: How were the data collected? At what locations? Over what time period? Who collected them? What instruments were used? Have the operators and instruments changed over the period? Try to imagine yourself at the data collection site physically.

3.2 Meaning

What does each variable mean in the data? What does it measure? Does it measure what it is supposed to measure? How could things go wrong? What statistical assumptions is one making by assuming things didnât go wrong? (Knowing the data collection process helps here.)

Meaning of each variable – ask students to imagine being there at the ER and giving a Glasgow coma score, for example, and also a couple of variables – ask students what could cause different values written down.

How were the data cleaned? By whom?

3.3 Relevance

Can the data collected answer the substantive question(s) in whole or in part? If not, what other data should one collect? The points made in (3.2) are pertinent here.

3.4 Comparability

Are the data units comparable or normalized so that they can be treated as if they were exchangeable? Or are apples and oranges being combined? Are the data units independent? Are any two columns of data duplicates of the same variable?

4 Note on Grading

As the final project is a group project, being a good collaborator on the project will be taken into account for each individual's grade on the final project. After the project is submitted, we will send a Google form for each group member to evaluate the collaboration received from other members of their group. The final score for each student will be a combination of the student's collaboration score from their group mates, and the overall project score.