Proposal Outline

Alyssa Forber

October 16, 2017

Background/Problem

(Check out most recent manuscript from paper an citations) call it progression to opioid therapy

Opioid information

-explain issue of opioid addiction

-rare outcome

-whether or not to prescribe to patients if likely to become addicted

Intro

The United States is facing an unprecedented opioid epidemic. According to data from the 2015 National Survey of Drug Use and Health, over 2 million people had a prescription opioid use disorder.[1]

This is particularly important in the hospital where opioids are commonly prescribed for pain.[10] Opioid receipt at hospital discharge has been shown to be associated with an increased risk of chronic opioid use.[11]

Predictive tools to identify hospitalized patients at risk for future COT may have clinical utility to improve hospital-based pain management with a focus on limiting opioid prescribing when non-opioid analgesics, or other non-pharmaceutical options, may be effective for pain control.

Data

Describe dataset (comes from Denver health, more info in paper)

-show a table1 of the data

-number of subjects

-outcome

-variables in dataset

The Data

Design: Denver Health retrospective analysis electronic health record (EHR) data from 2008 to 2014. Patients: Hospitalized patients at an urban, safety-net hospital.

The study had a binary outcome of chronic opioid therapy (COT) one year following the index discharge. We defined COT as receipt of $>= 90$-day supply of opioids with < 30-day gap in supply over a 180-day period or receipt of $>= 10$ opioid prescriptions over one year. This is a rare outcome. 27705 in dataset, 1457 with outcome, about 5%

There are 50 variables, which do I mention? And then we narrowed to start with 35 Demographics and potential predictors Data dictionary I could look at?

Aims?

-accurate predicting, better sens and spec for unbalanced outcome -using
and comparing methods of cutpoints and sampling

Methods

describe analysis approach

Model

-temporal split of data to get train set

-using logistic regression for the easier interpretation

-clinical colleagues understand this

-stepwise selection to reduce variables in data

First approach

Approach 1:

-predict and change cutoff with train set- no sampling

-explore cutoff probability to get the "best"

ROC methods

ROC:

-explain ROC curves and youden

-show an example ROC curve

-could also be chosen if a certain sens or spec in mind instead of youden

Second approach (sampling)

Approach 2:

-predict and use 0.5 cutoff with sampled train set that is balanced

-explain up, down, and smote (explain smote)

Preliminary Results

Explain prelim results (sens, spec, accuracy, AUC, npv, ppv):

Table of–

-non-sampled with 0.5 cutoff

-non-sample with youden cutoff

-up sample with youden cutoff

-down sample with youden cutoff

-smote sample with youden cutoff

- down sample with bagging

Discussion of Results

Depening on situation the clinician may like different sens/spec

Some may want to be more conservative, others may not (cancer patients in a lot of pain who need opioids)

Moving Forward

-simulation of different outcome percents, see when you could run model without sampling or changing cutoffs (though sampling does allow a more parsimonious model)

-trying different sampling other than just the defaults for each method

-different ways to do stepwise selection

-bootstrap aggregate the coefficients and get bootstrap CI (loop through getting new sample, saving coefficients, get mean and sd across 1000 boot samples) With this you may get 12 var with one stepwise and 13 with another

- look up bagging and stepwise selection
- haven't seen much on bagging and down sampling— look that up, if not that'll be interesting
- maybe try lasso with cross validation (cv.lasso)
- lasso has been shown to be better at selecting a model than stepwise
- easy to save all the coefficients and bootstrap

Questions

Should I show code in this?

Am I giving the talk as if for an audience of statisticians or clinicians or mixed?

What more should be expanded on?

Any figures or tables that would be helpful show smooth spline of age and probability of COT and show it has a curve to it and why we added the quadratic age smooth.spline with 3 degrees of freedom

Should I go into detail about what logistic regression and predicting is doing behind the scenes? Have one slide showing the equation, don't spend too much time on it Show what is the predicted probability from logistic regression

Is it ok to show all that I've done or will they think I've already done too much? show some of the results

Other ideas of things to be done moving forward?

NOTES FROM 10/31

SCHEDULE!!! schedule between dec 4 and 15 send susan a note (she'll be hardest) ask what days and times are best could send out a doodle poll then send out her availablity

accumulate other papers in endnote