Newton, Sophia
ISYE6501x
Course Project
Due July 18, 2018

Predicting patients who may be at risk of opiate abuse

https://www.sas.com/en_us/customers/competitive-health-analytics.html

> Studying medical and prescription use for patients later diagnosed with opiate abuse. CHA and a partner pharmaceutical company wanted to see if there was a pattern of medical utilization before diagnosis that could inform physicians of at-risk patients. CHA and its partner plan to publish the results of the study.

Longitudinal medical study of patients where they have a patient id as their identifier. Okay so what are we really studying? Doctors visits, prescription use, dates of each, provider identifier number (NPI) that you did the visit with, diagnoses for visits, CPT codes for visits, binary variable for pain pill prescription at all, categorical variable for class of prescribed pain pill, binary variable for past opiate abuse, binary predictor variable from each regression model of whether that patient is a likely opiate abuser based on the other factors.

Data gathering:

- Claim data has nearly all of the data we need: NPI, Date of service of visit, Diagnosis code(s), CPT code(s).
- If you're a partner with the EHR, you can also get their prescription history, their diagnosis history, their procedure code history, RXNORM code for their prescription histories.
- What about prescriptions from other providers? Even if that EHR doesn't have it, if it went through an insurance, it's likely in their prescription database. Dr. First is a major one, Allscripts is another major one, both have a way to report on a patients' prescriptions EVEN from other providers.

Patterns we're after to predict future opiate abuse:

- Pattern to past opiate abuse: need to model it in the first place
- Pattern of escalation of number of providers
- Diagnosis correlation

Analytics Models to determine patterns:

- Regression: does past opiate abuse depend on visit frequency, correlate with a particular diagnosis, do a large number of past opiate abusers originate from a particular combination of providers?, Does history of use of particular class of prescription in the past predict future opiate abuse? This is at least three regressions alone but I'm sure that's not what we were looking for as the whole project.
- Design of experiment: Multi-armed bandit approach first to determine which experiments even to start with.

- Clustering: I think you could actually probably both find a pattern to past opiate abuse and find a predictive model for at-risk patients using a clustering model.
- Exponential Smoothing: I recognize this isn't a model, but it is a technique we've used in this course and actually it seems like it should be relevant - in the same way that relapse risk tends to rely more heavily on recent behavior than prior (off the wagon) behavior, looking at a patient's risk level for opiate abuse should probably be exponentially smoothed, because lots of risky behavior in the past and no indication of abuse, shouldn't necessarily indicate that lots of risky behavior today isn't an indicator that they're at risk today, and your recent behavior should mean more.

So first we'd need to gather the raw data we could from claims data and other information in their EHR - That's most of the first few variables listed out.

Then, we'd need to create variables for past opiate abuse - that's harder, how do we decide if they did? If they self report? That's historically unreliable. Claims data indicating a rehab stint? Most insurers don't actually cover rehab, so it's possible that they wouldn't have bothered sending it to their insurer. Alternately, if they didn't want it to adversely affect their work, they could've intentionally not submitted so that their employers didn't find out. This would honestly be the toughest data to get, because it would be the toughest decisions the researchers would have to make. Really, though, once they've selected a way to create this variable, it's easy enough to gather for each patient.

Other calculated/compiled variables: class of prescriptions - there's definitely a class system already in place for controlled prescription drugs, and most prescription drug monitoring programs already have classification models in place, so let's just assume we inherited that from the PDMP we mined from (it's a lazy assumption but honestly probably a good one). What about the non-controlled prescription pain pills? Those are tracked by PDMP's too, as it turns out, so we can still stick with that assumption of heritability. Binary for prescribed pain pills/opiates - that's also going to be heritable from the PDMP, so we're good there too.

Finally, the predictive models! Once you've gathered the data, it should be simple enough to run the models, decide the riskiness of patients (and providers!) and have that data to report on as often as you want or need it. How often should that be? That's actually not federally mandated right now, but it is certainly determined on a state-by-state basis, and in order to keep their electronic prescribing privileges, most providers need to check the PDMP for patients they're prescribing controlled drugs each time they prescribe a controlled substance. How often do the providers need to turn in that they're checking into it, and what about pharmacies, you might ask? Well I'll tell you: most states require that pharmacies run it AT LEAST every 48 hours, but because that's such a tight turn around, and because some states are stricter, most that I've come across run these types of reports at least once every 24 hours. Yes, that's every single day. Because this is America, and we're really concerned about making sure they're not popping out pills to people who don't need them. What about the providers? Normally they've

got to check into it every 72 hours or so, but most providers I've worked with literally have worked it in so that every prescription they write that may be even a little questionable, they check the PDMP.

So, that's a summary, sorry it wasn't shorter, but I think it was thorough, on how if I worked for Competitive Health Analytics, a Humana company, who partnered with SAS, I would've carried out this study. I think because they could've just created the models and handed them off to the PDMP's, the EHRs, etc, I would have gathered data maybe twice a month for a year, just to check if there were seasonal effects (spoiler alert: there definitely are) but if they needed a super quick turnaround and didn't have a year's worth of EHR data, they could work around that.