**Estimating the cost of diagnostic delays for endemic mycoses: Next Steps in modeling**

**Previously**: We played around with basic models like truncated GAM, truncated linear, binary, quartiles etc.

**Problems with each**: GAM doesn’t show any particular good fit or relationship, quantiles introduce bias and cause us to lose lots of information. Is there a way to keep the continuity of our exposure and still achieve a good fit?

**Now**: Thinking about our DAG - we now a causal issue- what if the time component (which can’t be measured by descriptive stats or basic regression) is causing more a difference in cost? We realize our models aren’t enough to account for that

**Question**: How do we account for a difference from day to day that doesn’t appear by just looking at the data?

**Answer**: Inverse probability weighting! Instead of having an exposure which compares treatment A to B and using IPW to balance among covariates – we have an exposure that is continuous (# of days) and need to create a weight to balance among covariates. We are comparing all of our days which are grouped into bins to each other.

**Future**: now that we have a few ideas for models and are familiar with the distribution of our data, the literature points to a few options…

**Our distribution**: We have a continuous zero-inflated exposure – the literature says using [*quantile binning*](https://journals.lww.com/epidem/fulltext/2014/03000/constructing_inverse_probability_weights_for.21.aspx)approach to determine propensity scores(PS)/weights is a good option. This involves using a multinomial logistic (polytomous) regression model to achieve propensity scores comparing multiple quantiles to each other (including zero in it’s own bin)

Secondly, Since our data is non-linear BUT linear when the exposure is zero truncated – we can use that linear model as a supplement

1. Create N=(3,4,5,10,15,20) categories (up to 6 categorical vars). make sure zero is its own category
2. Fit a multinomial logistic regression model to these categories to obtain the predicted probabilities of falling into the observed exposure category (this creates our PS variables)
3. Use inverse probability weighting/other weighting (overlap, etc.) to transform PS into weight variable
4. Check balance of different weights
5. Now re-run our ***previous models*** incorporating the weight variable (maybe now GAM or linear works out? Depends on how weights affect our new exposure-outcome relationship)