**Feature Reduction through Imaginary Network Models**

Farrokh Alemi, Ph.D.

George Mason University Department of Health Administration and Policy

[falemi@gmu.edu](mailto:falemi@gmu.edu)

Sankeerthi Mummidisetty [smummidi@gmu.edu](mailto:smummidi@gmu.edu)

**Results for Second Data Set**

In the second data set, we evaluate the effectiveness of an antidepressant on reducing depression symptoms.

**Source of Data:** We used the National Institute of Mental Health’s STAR\*D data, the largest available data set capturing response to antidepressants. These data include genetic markers and phenotypes of 3,678 patients with major depression. The STAR\*D cohort data was split into the training (2931 patients) and research data (747 patients) sets. The training dataset was derived from approximately 80% of STAR\*D data, while the evaluation dataset was derived from the remainder (~20%) of STAR\*D data. The STAR\*D includes up to 4 episodes of treatment for patients who do not respond to the initial treatment. The unit of analysis was the episodes of treatment; thus, we are comparing patients treated with one antidepressant to patients treated with another antidepressant, what in the literature is known as placebo studies.

**Outcome:** The outcome of interest is depression symptom remission, measured in STAR\*D through 50% reduction in Hamilton’s index of depression symptoms.

**Treatment:**We focus on the effectiveness of citalopram, one of the antidepressants provided to patients in STAR\*D trials.

**Covariates:** The effectiveness of the antidepressant depends in part on gender and patient’s history of illness, captured by baseline comorbidities, listed below:

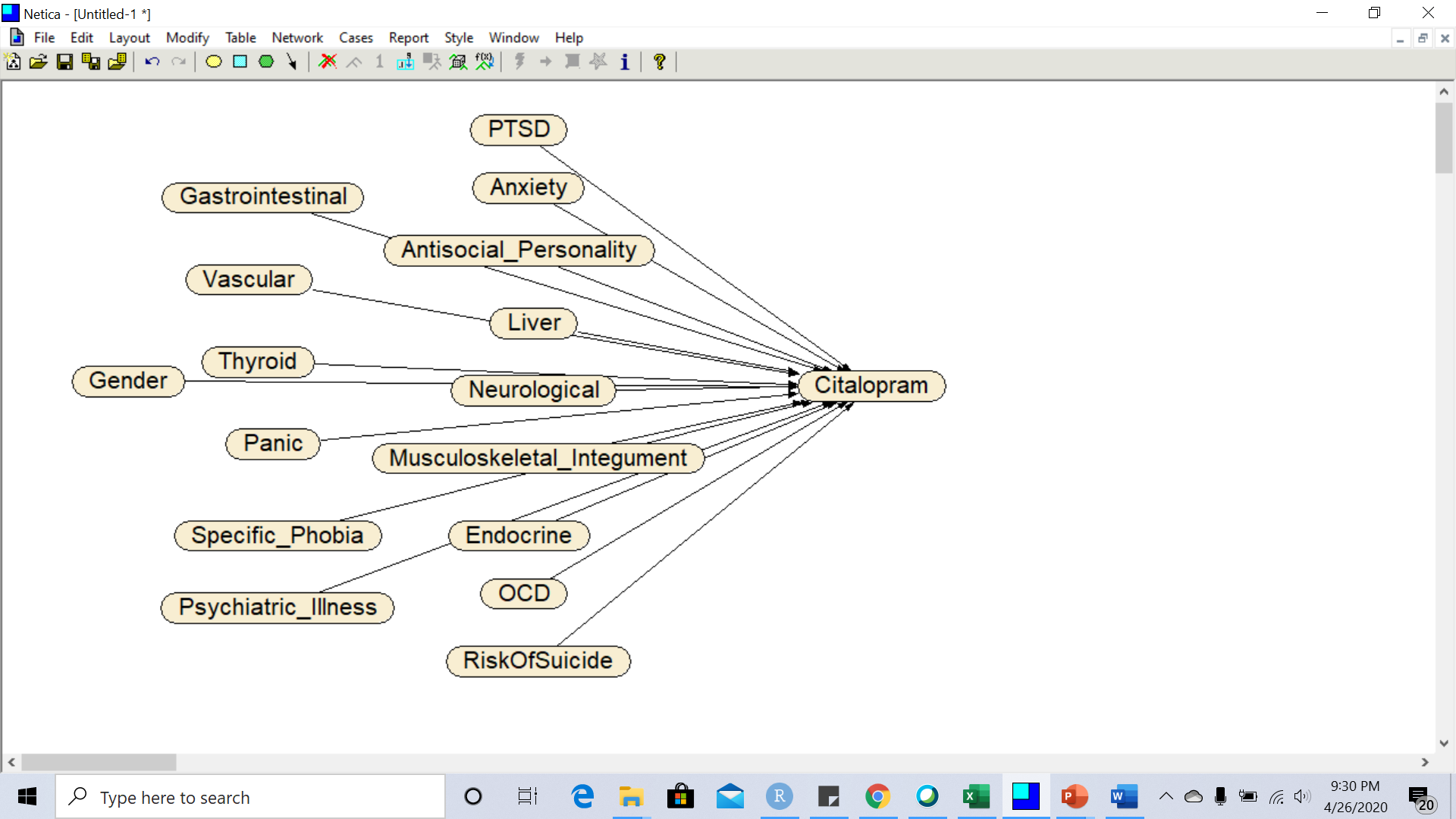
|  |  |
| --- | --- |
| Risk of Suicide | Thyroid |
| Vascular | Heart |
| Eyes Ears Nose Throat Larynx | Hematopoietic |
| Renal | Gastrointestinal |
| Musculoskeletal Integument | Genitourinary |
| Psychiatric Illness | Neurological |
| Liver | Respiratory |
| Alcohol | Endocrine |
| Cannabis | Amphetamine |
| Panic | Opioid |
| Social Phobia | Specific Phobia |
| PTSD | OCD |
| Borderline Personality | Anxiety |
| Antisocial Personality | Dependent Personality |
| Personality Disorder | Paranoid Personality |
| Bulimia | Anorexia |
|  | Cocaine |

**Methods of Identifying Markov Blanket**: We used the glmnet R software to conduct the LASSO regression. Demographic variables were assumed to have occurred prior to baseline comorbidities and baseline comorbidities were assumed to occur prior to treatment with citalopram. The initial LASSO regression was treatment on variables that occur prior to it. In the second LASSO regression, the response variable was the outcome, i.e. symptom remission. The independent variables were the variables that were significant in the initial LASSO regression plus the treatment variable.

**Methods of Measuring Accuracy:** We have evaluated the performance of the model using coefficient of determination (*r*2 ) and Root Mean Squared Error (RMSE). R squared is the proportion of the variance in the dependent variable that is predictable from the independent variables Ideally, lower RMSE indicates a better model. In our study we have observed that the RMSE on the training data are 0.465. The results on the test data is 0.467 when considering citalopram as target variable and RMSE on the training data are 0.483. The results on the test data is 0.486 when considering remission as target variable

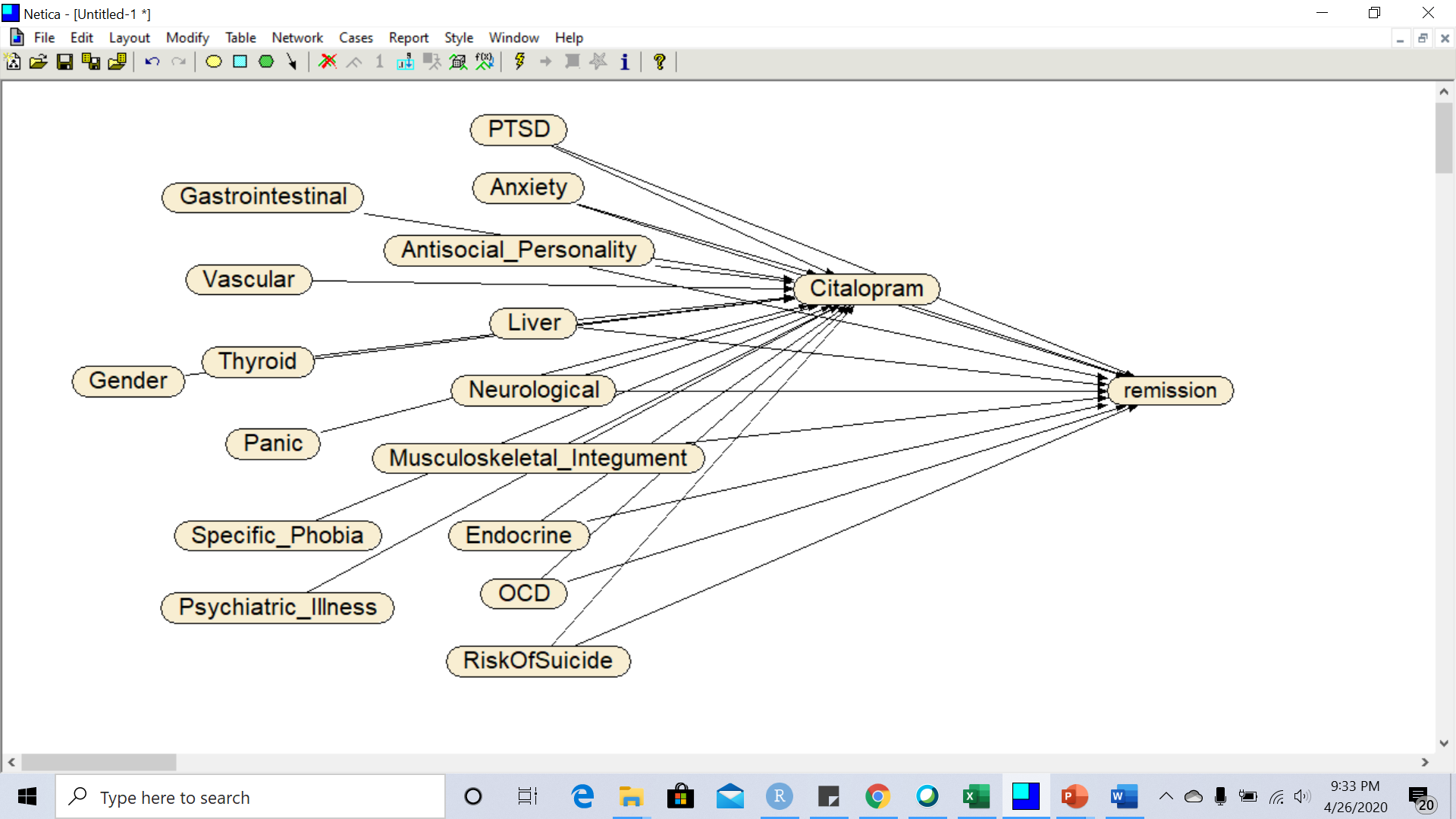
**Results:** In the LASSO regression of citalopram, the following independent variables were statistically significant: thyroid, PTDS, gender, neurological, OCD, risk of suicide, anxiety, psychiatric-Illness, vascular, musculoskeletal Integument, gastrointestinal, specific phobia, antisocial personality, panic, liver and endocrine disorders. Figure 2 shows the network model of variables that affect selection of citalopram.

**Figure 2: Variables that Affect Selection of Citalopram**



In the next LASSO regression, the variables that were statistically significant were risk of suicide, PTSD, anxiety, antisocial personality, liver, neurological, musculoskeletal integument, endocrine, citalopram, and OCD. In addition, citalopram was also a significant predictor of symptom remission. The combined LASSO regressions describe the network in Figure 3.

**Figure 3: Combined Network Model for Predicting Effect of Citalopram on Remission**



The net accuracy of predictions can be assessed from the percentage of variation explained by each of the regressions, as seen in Table 3:

**Table 3: Impact of Citalopram**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Independent Variables in the Model** | **Number of Covariates (Percent Reduced)** | **Impact of Citalopram on Remission** | **Percent of Variation Explained** | **10-fold Cross-Validated AROC** | **AROC divided by Degrees of Freedom** |
| Comprehensive: Gender, 32 Baseline Diagnoses, & Citalopram | 34 (100%) | 16(47.05%) | 0.532 | 61.00% | 1.79 |
| Only Parents of Citalopram & Citalopram | 16(100%) | 10(62.5%) | 0.514 | 57.00% | 3.56 |