# Identifying Over-prescription in Healthcare Claims Using Bayesian Network

HUANYE LIU

# Research Question

How to use Bayesian network model to detect a doctor's over-prescription of a target drug?

References:

Sakshi Babber and Sanjay Chawla's *On Bayesian Network and Outlier Detection*, 2010

Jing Li etc.'s *A Survey on statistical methods for health care fraud detection*, 2007

# Motivation

Why meaningful?

1) Over-prescription is rampant and incurs a large amount of cost/waste in the healthcare system.

2) The illegal deals between the drug company and the healthcare provider still exist in many countries.

Therefore, we need effective computer aided methods to measure and detect the over-prescription behavior.

# Data

Data used in this project can be downloaded from the U.S. Centers for Medicare and Medicaid Services (CMS.com) website:

https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Part-D-Prescriber.html

In particular, they published a dataset of prescriptions under Medicare Part D in 2013 data.cms.gov

Part D Pre | Bayesian net... | high dimensi... | persp-resear... | Your Reposit... | web.engr.ore... | Zika Virus Ep... | Part D Opioi... | 2013 Part D... | Research, St... | All-Payer Cla... | How to take...

# Data.CMS.gov

| Home | Sign In to Data.CMS.Gov

## Medicare Provider Utilization and Payment Data: 2013 Part D Prescriber

Please visit https://data.cms.gov/view/36cg-9reu to view this data. This dataset is from http://download.cms.gov/Research-Statistics-Data-and-

Find in this Dataset

Manage | More Views | Filter | Visualize | Export | Discuss | Embed | About

| | Specialty Description | Description Flag | Drug Name | Generic Name | Beneficiary Count |
|---|---|---|---|---|---|
| 332 | Family Practice | S | PROAIR HFA | ALBUTEROL SULFATE | |
| 333 | Family Practice | S | FENOFIBRATE | FENOFIBRATE | |
| 334 | Family Practice | S | PAROXETINE HCL | PAROXETINE HCL | |
| 335 | Family Practice | S | METFORMIN HCL | METFORMIN HCL | |
| 336 | Family Practice | S | VENTOLIN HFA | ALBUTEROL SULFATE | |
| 337 | Family Practice | S | PRAVASTATIN SODIUM | PRAVASTATIN SODIUM | |
| 338 | Family Practice | S | HYDROCHLOROTHIAZIDE | HYDROCHLOROTHIAZIDE | |
| 339 | Family Practice | S | SPIRIVA | TIOTROPIUM BROMIDE | |
| 340 | Family Practice | S | ONGLYZA | SAXAGLIPTIN HCL | |
| 341 | Family Practice | S | CLONAZEPAM | CLONAZEPAM | |
| 342 | Family Practice | S | AMITRIPTYLINE HCL | AMITRIPTYLINE HCL | |
| 343 | Family Practice | S | BYSTOLIC | NEBIVOLOL HCL | |
| 344 | Family Practice | S | LISINOPRIL | LISINOPRIL | |
| 345 | Family Practice | S | COMBIVENT RESPIMAT | IPRATROPIUM/ALBUTEROL SULFATE | |
| 346 | Family Practice | S | TRAMADOL HCL | TRAMADOL HCL | |
| 347 | Family Practice | S | LANTUS | INSULIN GLARGINE,HUM.REC.ANLOG | |
| 348 | Family Practice | S | ATENOLOL | ATENOLOL | |

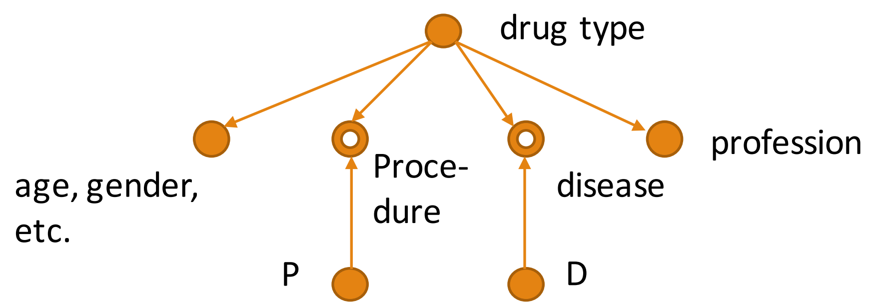
# Theory to interpret the data

We use Bayesian statistics to capture the probabilistic dependency among different variables, and detect outliers from the data based on their small posterior probabilities.

# Analysis and Computational tools

1)Bayesian Network

Among many outliers/anomaly detection tools, Bayesian network model can directly map the relationship between multiple variables to a graph. And there are ready-to-use algorithms for model learning and inference.



# Analysis and Computational tools

2) Dynamic Programming

We use dynamic programming to accumulate each individual posterior probability of using a target drug $r$ prescribed by one doctor to detect if the doctor's over-prescription of this particular drug.

We need to calculate the probability that the expected number of prescriptions for the target drug $r$ based on the model inference is greater than the number of observed ones:

$\Pr(e_r \geq o_r) = \sum_{k=o_r}^{T} \Pr(e_r = k)$=$\sum_{k=o_r}^{T} Pr_{k,i=T}$

Where $Pr_{ki} = \Pr(drug_T = r|y_T)\, Pr_{k-1,i-1} + \Pr(drug_T \neq r|y_T)\, Pr_{k,i-1}$