

Health Care Fraud Risk Management

Tahir Ekin

Associate Professor of Quantitative Methods
McCoy College of Business
Texas State University

tahirekin@txstate.edu
www.tahirekin.com

SAMSI GDRR Course
Duke Stat 7090 Week 7
Durham, NC

February 25th, 2020

- Coverage of GDRR course so far
- Health care fraud
- Unsupervised data mining
- Supervised data mining
- Sampling and overpayment estimation
- Decisions and games for fraud detection

Health care fraud



Philip Esformes: in 2016, accused of \$1.3 billion 14 year fraudulent network of skilled-nursing and assisted-living facilities by filing false claims for services that were not necessary, convicted in April 2019.

Why health care fraud, waste and abuse

- U.S. health care spending: \$3.65 trillion, or \$11,212 per person in 2018
- Three to ten percent lost to fraud, waste and abuse
- Health care fraud: Intentional deception or misrepresentation made by a person or an entity, with the knowledge that the deception could result in some kinds of unauthorized benefits
- Health care abuse and waste: poor/unnecessary practices that are not consistent with benchmarks
- Global health care overpayments estimated to be around \$ 450 billion
 - Complexity, heterogeneity of the system, unconditional trust on providers, lack of resources for investigations
 - Low probability of detection, relatively low probability of being convicted once detected and the low severity of punishment even if convicted
 - Adaptiveness of fraudsters to anti-fraud schemes

Types of health care fraud

- Identity fraud
- Providing incorrect care or manipulating billing rules
- Managed care fraud
- **Improper coding**: upcoding, unbundling, multiple (double) billing, phantom (ghost) billing
- **Providing unnecessary care**
- **Kickback schemes and self-referrals**

Improper coding and providing unnecessary care

- Home health agencies: Dr. Jacques Roy from Rockwall, Texas billed more home health services through Medicare than any other medical practice in the U.S between 2006 and 2011 for more than \$350 million.
- Florida, Dr. Salomon Melgen: Aggressive Medicine or Malpractice? Incorrect diagnoses and falsified tests, he made \$8.4 million (vs \$6,061) during the six years by treating people with lasers and \$57.3 million (vs \$3 million) by treating patients with Lucentis injection

Kickback schemes and self-referrals

- Texas, 2005- : 160 ongoing Medicaid dental fraud investigations, Texas spending more on braces than other 49 states combined, Nevaeh Hall, a 4-year-old, got overtreated and had complications (brain damage) during a routine appointment at a local dental clinic
- Arizona, Florida, Texas: power wheelchairs
 - Cooper Medical Supply: fraudulent prescriptions and medical documents to submit false claims to Medicare for expensive, high-end power wheelchairs. More than 80 percent of the beneficiaries lived over 100 miles away from Cooper Medical Supply, and most were not even given the wheelchairs.
 - Positive Home Oxygen and Dr. Robert Lyle Cleveland: In exchange of Dr. Cleveland signing Certificates of Medical Necessity for power wheelchairs for patients who did not meet the coverage requirements, patients would be referred to him.

Statistical Health Care Fraud Assessment

- Fraud, waste and abuse (overpayments) result with:
 - Direct cost implications to the government and to the tax-payers
 - Diminished ability of the medical systems to provide quality care to the deserving patients
 - Adverse impact on health
- Audits and investigations by CMS, Office of Inspector General and contractors
- Statistical issues in fraud assessment:
 - Sampling and overpayment estimation
 - Use of data mining for fraud detection
 - Decisions and games for fraud assessment

Overview of Fraud Detection and Prevention Systems

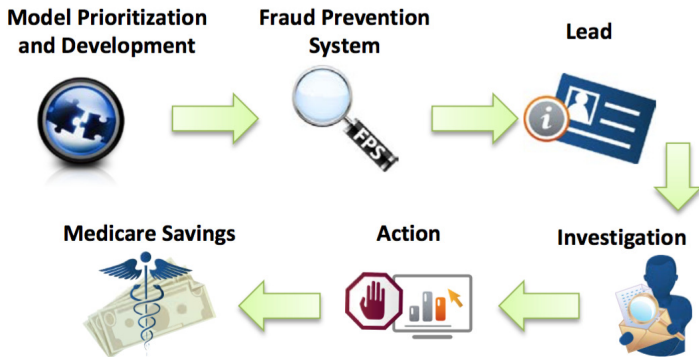


Figure: Overview of the Fraud Prevention System

Health Care Fraud Data

- Unique challenges of health care fraud systems
 - Heterogeneity and complexity: Many programs serving various populations within different payment systems
 - Trade-offs between accuracy and speed: requirement of paying providers in a certain timeframe
 - Lack of incentives to report fraud
- Dynamic patterns due to legislative, policy and population health changes
- Lack of labelled data
- Data quality issues
- Fraud as a rare event: Imbalanced class sizes

Descriptive Statistics

- Reveal billing behaviors and find providers that are different than the standard (expected)
- Initial screening using distribution, mean and variance of a variable, ie payment

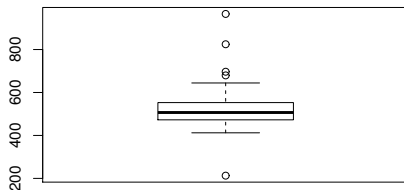


Figure: Boxplot of average Medicare payment for a given provider type (Ambulatory Surgical Center) and a procedure code (removal of excessive skin)

Medical claims data: dynamic, heterogenous, skewed, multi-layered

- Unsupervised methods
- Supervised methods
- Sampling and overpayment estimation
- Decisions and games for fraud

Unsupervised Data Mining Motivation

- Dynamic nature of fraud and expensive labeling: Li et al. (2008)
- Unsupervised methods mainly used for initial screening
 - Anomaly detection: Bauder et al. (2017), Ekin et al. (2017)
 - Clustering: Musal (2010), Macedo et al. (2015)
 - Latent Dirichlet allocation: Ekin et al. (2019)
 - Structural topic models: Zafari and Ekin (2019)
 - Bayesian coclustering: Ekin et al. (2013)
- Identify the hidden patterns among providers (doctors), procedures and patients
- Reveal billing behaviors and find providers that are outliers

Grouping medical claims data

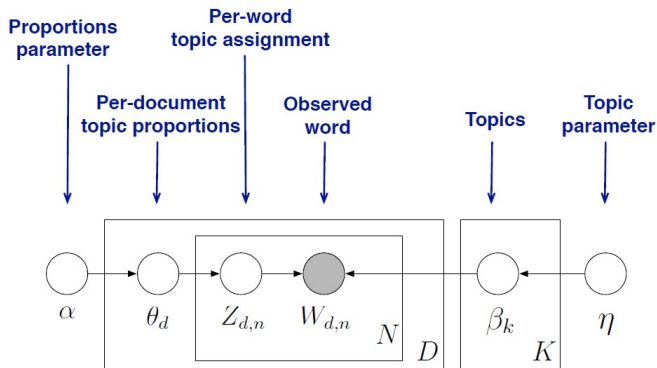
- Group providers with respect to their billing patterns and detect outliers
- Identification of proper peer comparison groups to classify providers as within-the-norm or outliers
- Berenson-Eggers Type of Service (BETOS) system: covers all HCPCS codes (Health Care Procedure Coding System), Macedo et al. (2015); SAS PROC FASTCLUS
- Challenges: The boundaries between some medical specialties are not well defined and a number of physicians may be qualified to practice in more than one medical specialty. Ex: General practitioners (GPs) render services related to the diagnosis and management of heart failure

Topic models: Bayesian hierarchical mixed membership models

- Latent Dirichlet Allocation (LDA): Blei et al. (2003)
 - Topics: mixtures over words where each word within a given document belongs to all topics with varying probabilities
 - Documents: a mixture of latent topics (groups, clusters)
- Correlated Topic Models (CTM): Lafferty and Blei (2006)
 - Extends LDA to consider relationships among topics
- Structural Topic Models (STM): Roberts et al. (2016)
 - Extends CTM to consider document level covariates

LDA for Fraud Detection

- Topics: Collection of Words
- Documents
- Words



$$p(\beta, \theta, \mathbf{z}, \mathbf{w}) = \left(\prod_{i=1}^K p(\beta_i | \eta) \right) \left(\prod_{d=1}^D p(\theta_d | \alpha) \prod_{n=1}^N p(z_{d,n} | \theta_d) p(w_{d,n} | \beta_{1:K}, z_{d,n}) \right)$$

LDA: Generative process

- Observed words, $W_{d,n}$ and a given number of topics, K
- Aim: reveal the hidden topic structure
- Choose distributions for each topic: $(\beta_k \sim Dir(\eta); k=1,\dots,K)$
- For each document d :
 - Draw topic proportions $(\theta_d \sim Dir(\alpha))$
- For each (n^{th}) word in the (d^{th}) document:
 - Draw a topic, k with respect to the proportions; $Z_{d,n} \sim Mult(\theta_d)$
 - For that particular (k^{th}) topic, you already have the topic distribution, $\beta_k \sim Dir(\eta)$
 - Draw a word with respect to the chosen topic's distribution; $W_{d,n} \sim Mult(\beta_{Z_{d,n}})$

$$p(\mathbf{Z}|\mathbf{W}) \propto \int_{\boldsymbol{\theta}} \int_{\boldsymbol{\beta}} \left(\prod_{d=1}^{d=D} \prod_{n=1}^{n=N} p(z_{d,n}|\boldsymbol{\theta}_d) p(W_{d,n}|\boldsymbol{\beta}_{1:K}, z_{d,n}) d\boldsymbol{\beta} d\boldsymbol{\theta} \right)$$

$$p(z_{d,n} = k | \mathbf{Z}_{(-d,n)}, \mathbf{W}) = \frac{(q_{k(-d,n)} + \alpha_0) \frac{(q_{k(-d,n)} + \beta_0)}{\sum_{n=1}^N (q_{k(-d,n)} + \beta_0)}}{\sum_{t=1}^K (q_{t(-d,n)} + \alpha_0) \frac{(q_{t(-d,n)} + \beta_0)}{\sum_{n=1}^N (q_{t(-d,n)} + \beta_0)}}$$

$$E[\theta_{d,k}] = \frac{q_{k(d,\cdot)} + \alpha_0}{\sum_{t=1}^K [q_{t(d,\cdot)} + \alpha_0]}$$

$$E[\beta_{k,n}] = \frac{q_{k(\cdot,n)} + \eta_0}{\sum_{t=1}^K [q_{t(\cdot,n)} + \eta_0]}$$

Steps of Collapsed Gibbs LDA

- The \mathbf{Z} variables are initialized to determine the initial state of the Markov chain.
- The chain is then run for a number of iterations, each time finding a new state by sampling each $z_{d,n}$ from the specified distribution
- After a fixed number of iterations, convergence of the Markov chain is checked.
- When convergence is attained, the respective counts of \mathbf{Z} are recorded and the posterior mean values of θ and β are computed.

LDA Collapsed Gibbs notation

- $\mathbf{Z}_{(-d,n)}$: vector of all other topic assignments other than the one for n^{th} medical procedure for the d^{th} doctor
- $q_{k(d,n)}$: count of assignments to k^{th} topic for the n^{th} medical procedure in the d^{th} document
- $q_{k(-d,n)}$: count of all assignments to k^{th} topic excluding $q_{k(d,n)}$
- $q_{k(d,\cdot)}$: total count of assignments to the k^{th} topic for all N procedures and the d^{th} doctor
- $q_{k(\cdot,n)}$: total count of assignments to the k^{th} topic for all D doctors and the n^{th} procedure

LDA for Fraud Detection, Ekin et al. (2019)

- Topics: Collection of Medical Procedures
- Documents: Doctors (or a set of physicians, hospital)
- Words: Procedures

- Observed words and a given number of topics
- Aim: reveal the hidden topic structure

LDA for Medical Fraud Detection

- Which procedures are frequently billed together? : β_k
- Can we detect doctors that have unusual behavior? : θ_d
- Which doctors are similar? : Similarity assessment via Hellinger distance (Hellinger, 1909)

- Data: CMS Medicare 2012 Part B claims; ≥ 9 M unique providers and 27 attributes for each
- Focus is on Vermont: 1,493,224 separately billed claims, 72 provider types
- 1,055 procedure codes that are billed by 2,268 unique doctors
- Python and R (tm and wordcloud packages) for data pre-processing and analysis

Medical Procedures Cloud

- “g9008”: “Coordinated Care Fee, Physician Coordinated Care Oversight Services”
- “a0425” : “Ground Mileage, Per Statute Mile” .

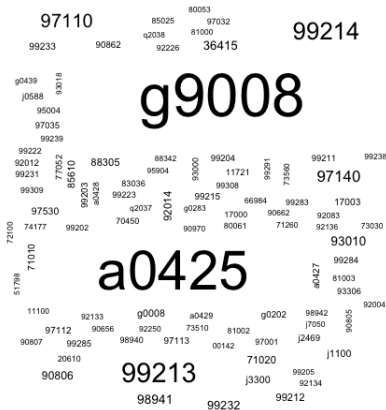


Figure: Medical Procedure cloud

LDA Results: Analysis of Topics

HCPCs Code	Description	β_4
92014	Eye exam and treatment	0.351
92012	Eye exam established patient	0.136
92083	Visual field examination(s)	0.075
66984	Cataract surgery w/iol 1 stage	0.061
92133	Computerized ophthalmic imaging optic nerve	0.060

Table: List of procedures in Topic 4, a portion of β_4 with terms sorted descendingly

LDA Results: Analysis of Topics

HCPCs Code	Description	β_4
92014	Eye exam and treatment	0.351
92012	Eye exam established patient	0.136
92083	Visual field examination(s)	0.075
66984	Cataract surgery w/iol 1 stage	0.061
92133	Computerized ophthalmic imaging optic nerve	0.060

Table: List of procedures in Topic 4, a portion of β_4 with terms sorted descendingly

- Procedures related to “eye exam and treatment”
- Which specialty of doctors bill for this topic?

LDA Results: Analysis of Topics

HCPCs Code	Description	β_4
92014	Eye exam and treatment	0.351
92012	Eye exam established patient	0.136
92083	Visual field examination(s)	0.075
66984	Cataract surgery w/iol 1 stage	0.061
92133	Computerized ophthalmic imaging optic nerve	0.060

Table: List of procedures in Topic 4, a portion of β_4 with terms sorted descendingly

- Procedures related to “eye exam and treatment”
- Which specialty of doctors bill for this topic?
- Ophthalmologists: eye exams, diagnose and treat disease, prescribe medications and perform eye surgery
- Optometrists: certain eye problems and diseases, and may participate in your pre- and post-operative care

LDA Results: Analysis of Topics

- 105 providers of which 100 are “eye” related providers
- 32 out of the 36 ophthalmologists and 68 out of the 78 optometrists.
- Remaining five: a physician assistant, an interventional radiologist, an ambulatory surgical center and diagnostic radiologists: Potential red flags

- What about the rest of the optometrists and ophthalmologists who have not billed for Topic 4 the most?
- 4 ophthalmologists billed for Topic 1 the most.
- 2, 1, 7 optometrists have billed the most for topics 1, 8 and 10

LDA Results: Analysis of Topics

HCPCs Code	Description	β_1
J3300	Triamcinolone A inj PRS-free	0.443
92226	Special eye exam subsequent	0.127
92134	Cptr ophth dx img post segmt	0.108
90805	Psytch off 20-30 min with E &M	0.103
67028	Injection eye drug	0.067

Table: List of procedures in Topic 1, a portion of β_1 with terms sorted descendingly

- Procedures 92226 and 67028 within Topic 1 are related to eye exams and injections.
- Topic 10 includes code “142” that corresponds to “lens surgery”: 7 optometrists

LDA Results: Analysis of Topics

HCPCs Code	Description	β_8 ,
99204	Office/outpatient visit new	0.304
92557	Comprehensive hearing test	0.070
99203	Office/outpatient visit new	0.067
45380	Colonoscopy and biopsy	0.060
43239	Upper gi endoscopy biopsy	0.052

Table: List of procedures in Topic 8, a portion of β_8 with terms sorted descendingly

- How to explain the 1 optometrist that does not behave like peers?

LDA Results: Analysis of Topics

- Can reveal characteristics that are different than the overall pattern

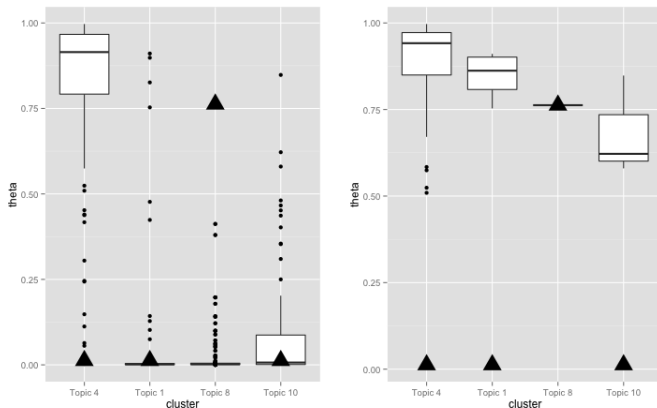


Figure: Select topic proportions, θ_d for the peer group (left) and maximum topic proportions θ_d (right)

LDA Results: Similarity Assessment

- Hellinger Distance between the topic proportions of two doctors

- $$d_{ij} = \frac{1}{\sqrt{2}} \sqrt{\sum_{k=1}^{k=K} [(\sqrt{\theta_{i,k}} - \sqrt{\theta_{j,k}})^2]}$$

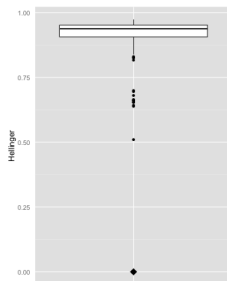


Figure: Hellinger Distance between the outlier and peers

- Average: 0.899, maximum: 0.973 and minimum: 0.510.
- Reveals further insights: outliers, teamwork

LDA: Limitations and Extensions

- How to sample from the posterior distribution:
 - Sampling based algorithms (MCMC): MC, dependent sequence of RV, with a limiting distribution as posterior: Steyvers and Griffiths, 2006
 - Variational algorithms which find the distribution that closely mimics the posterior distribution via optimization: Blei et al. (2003), Hoffman et al. (2010)
- Model evaluation: Perplexity, lack of labeled data
- **Correlation among topics**
- **Co-variates: topic proportions by provider type**
- **One type of cluster, does not address dyadic data:**
Co-clustering

“The level of urgency is greater than ever to develop creative solutions based on exploiting modern data mining and communication proficiencies.”
President’s Commission on Combating Drug Addiction & Opioid Crisis

“The level of urgency is greater than ever to develop creative solutions based on exploiting modern data mining and communication proficiencies.”

President’s Commission on Combating Drug Addiction & Opioid Crisis

- 1 Finding drug associations
- 2 Finding drug prescription patterns across medical specialties
- 3 Finding the providers with different drug prescription distributions versus their specialty peers
 - Documents: D providers
 - Words: N_d drugs, V unique drugs (vocabulary)
 - Topics: K collections of drug prescriptions
 - An example document (provider): $\{A, A, A, B, B\}$ with that doctor prescribing 3 A’s and 2 B’s

Proposed model

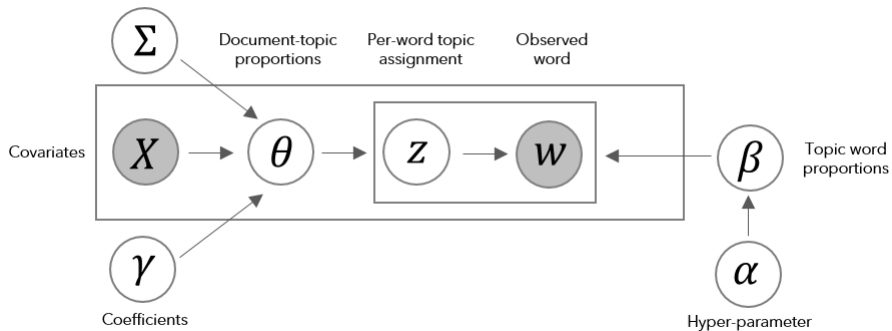


Figure: Graphical plate of the proposed model

Proposed model

Logistic Normal model for topic prevalence, θ_d : mixture over K topics

$$\begin{aligned}\theta_d &\sim \text{LogisticNormal}_{K-1}(\Gamma' \mathbf{X}_d', \Sigma), \\ \gamma_k &\sim \text{Normal}(0, \sigma_k^2), \text{ for } k = 1, \dots, K-1,\end{aligned}$$

\mathbf{X}_d : Covariate matrix to be used to model θ_d , provider specialty

Multivariate normal linear model with a single shared variance-covariance matrix with parameters that have half-Cauchy(1,1) hyper-priors

$$\beta_k \sim \text{Dirichlet}_K(\alpha),$$

The topical content β_k is assumed to follow Dirichlet distribution where we use a corpus-level conjugate Dirichlet prior with flat hyperparameters of $\alpha = 1$: all drugs have the same probability of being assigned to a group

$$\begin{aligned}z_{d,n} &\sim \text{Multinomial}_k(\theta_d), \text{ for } n = 1, \dots, N_d, \\ w_{d,n} &\sim \text{Multinomial}_v(\beta_{z_{d,n}}), \text{ for } n = 1, \dots, N_d.\end{aligned}$$

Model selection

Semantic coherence: co-occurrence of the frequent words of a given topic

Exclusivity: frequency of the top words of a given topic versus other topics

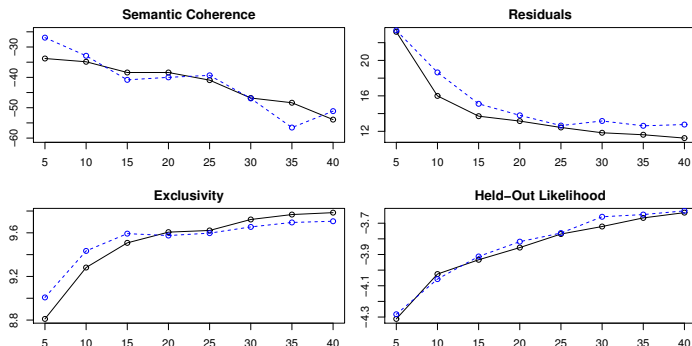
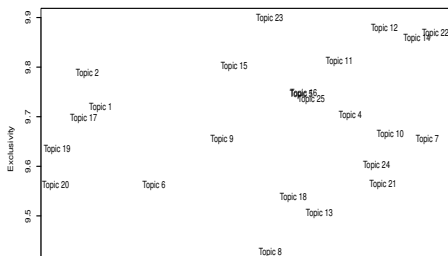
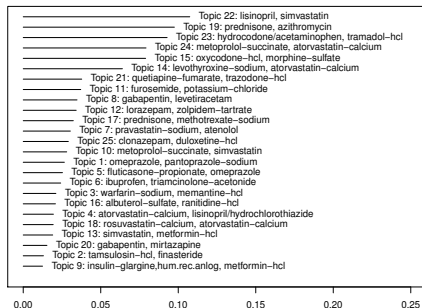


Figure: Diagnostic by the number of topics and initialization methods (— LDA vs. - - - Spectral)

Topic distributions



Topic Correlations

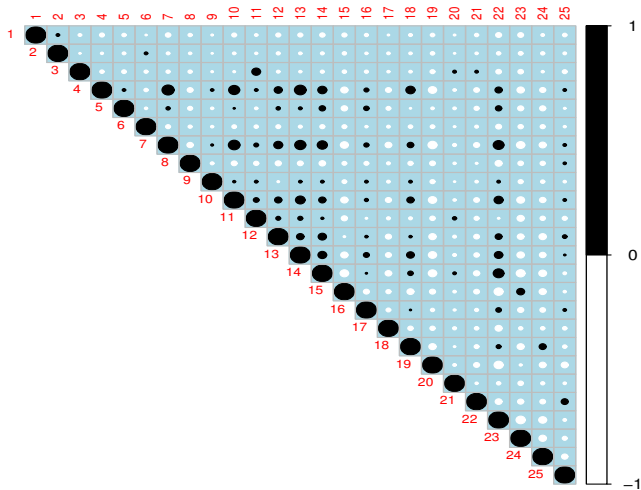


Figure: Topic Correlations

Drug Categories

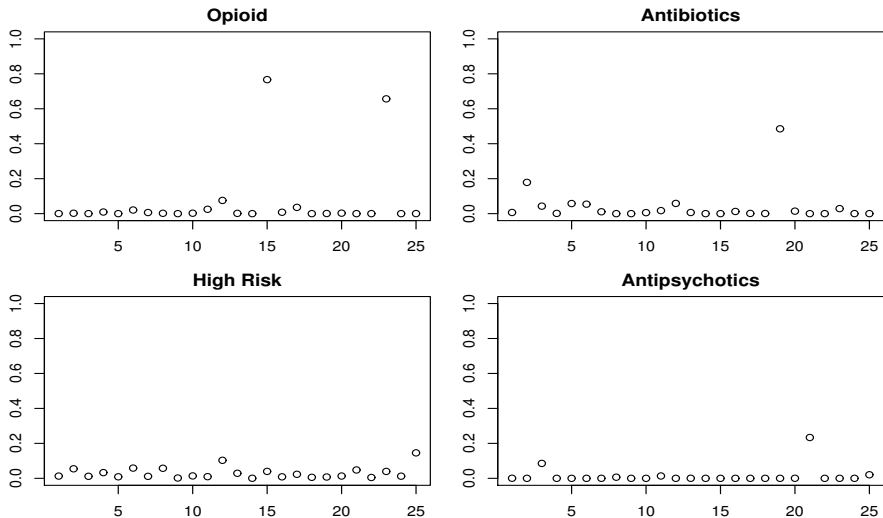
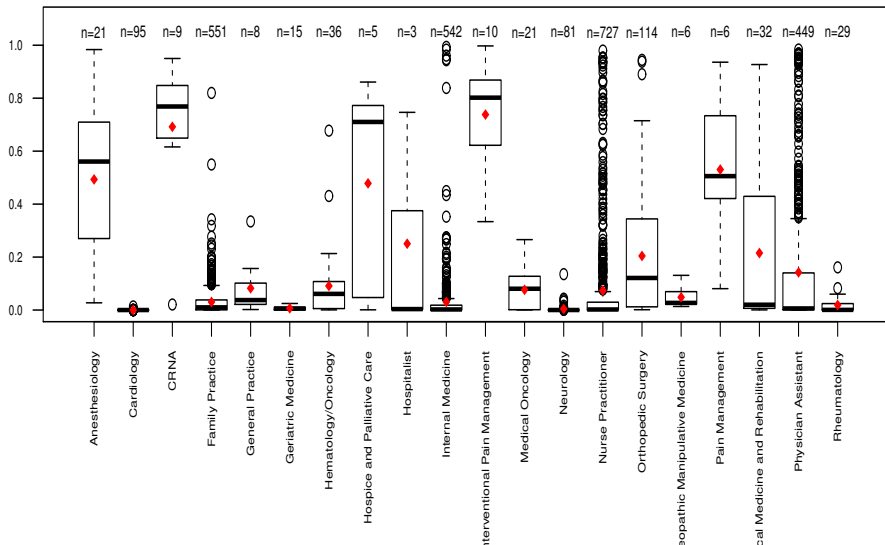


Figure: Expected Proportion of Main Drug Categories Across Topics

Distribution of posterior means of all providers ($\theta_{d,15}$) Across Specialties



Outlier measures for medical prescriptions

- Outlier detection methods that are based on comparisons with predetermined peer groups
- Outlier detection based on clustering output and similarity measures: Ekin et al. (2019) SAM
- How to use associations of providers and drug prescriptions within an outlier detection framework:
 - Lorenz curve (Marshall et al. (1979)): comparison of the income distribution with a uniformly distributed income
 - Concentration function (Cifarelli and Regazzini (1987)): generalization of Lorenz curves to compare any pair of distributions
 - Gini coefficient (Gini (1914)), Pietra's index (Pietra (1915))
 - Ekin et al. (2017) AmStat: first work to use within medical fraud domain
- This paper extends their adoption to use with topic modeling output, and utilizes it to identify prescription discrepancies among medical providers.

Concentration function

- We want to compare distributions of interest and benchmark (discrete uniform income)
- For a given population, we order the income levels $x_{(k)}$ for each $k = 1, \dots, K$ in an ascending order from the poorest to the richest individual.
- We define $S_0 = 0$ and $S_k = \sum_{i=1}^k x_{(i)}$. Therefore, S_K is the total income of the population and S_k/S_K is the fraction of wealth owned by the k poorest individuals.
- The concentration function is drawn by connecting these respective cumulative distribution points $(k/n, S_k/S_K)$, $k = 0, \dots, n$. For a given k , the plot displays the fraction S_k/S_K of the total income owned by the $k/K \cdot 100\%$ of the poorest part of the population.
- We obtain a convex, increasing function connecting the points $(0, 0)$ and $(1, 1)$.

Outlier detection

- The increasing curve connecting these points is called the concentration function of probability measure of interest.
- The distance between the concentration function and the straight line quantifies the difference between the distributions of providers of interest.
- We compute Gini's area of concentration as

$$1/2 - 1/2 \sum_{i=1}^K (Q_i - Q_{i-1})(\vartheta_i + \vartheta_{i-1}) \quad (1)$$

and Pietra's index as

$$\sup_{1 \leq i \leq K-1} (Q_i - \vartheta_i) \quad (2)$$

Outlier detection based on STM output

Aim: To measure the distance between the probability measures of the topic membership distribution of provider d , $\underline{\theta}_d = (\theta_{d,1}, \dots, \theta_{d,k}, \dots, \theta_{d,K})$ and of the topic distribution of its specialty $\underline{q}_m = (q_{m,1}, \dots, q_{m,k}, \dots, q_{m,K})$.

- For a given prescriber d , we order the topics $k = 1, \dots, K$ in an ascending order based on their likelihood ratios of $r_k = \theta_{d,k}/q_{m,k}$.
- The concentration function is drawn by connecting these respective cumulative topic distribution points (Q_ℓ, ϑ_ℓ) where $Q_0 = \vartheta_0 = 0$,
 $Q_\ell = \sum_{i=1}^{\ell} q_{m,(i)}$, and $\vartheta_\ell = \sum_{i=1}^{\ell} \theta_{d,(i)}$ for the ordered topics $(i) \in \{1, \dots, K\}$.
- The increasing curve connecting these points is called the concentration function of probability measure $\underline{\theta}_d$ with regards to \underline{q}_m .
- Next, we compute Gini's area of concentration and Pietra score.

Detection of outlier providers

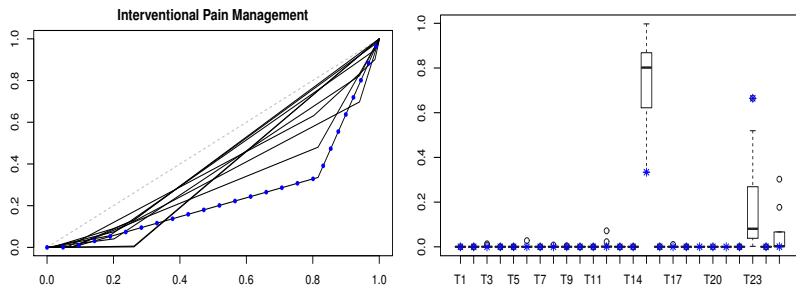


Figure: Concentration functions (left) and topic distribution of the outlier provider (right)

Gini and Pietra Scores

	1	2	3	4	5	6	7	8	9	10
Gini Score	0.18	0.11	0.13	0.12	0.15	0.07	0.25	0.07	0.12	0.08
Pietra's Index	0.33	0.17	0.25	0.16	0.24	0.13	0.48	0.13	0.25	0.13
Opioid Score	0.22	0.06	0.19	0.03	0.00	0.10	0.31	0.09	0.19	0.06

Table: Outlier measures for interventional pain management providers

Opioid score

Major drug types: opioids, antibiotics, high risk meds, antipsychotics
How to capture excessive opioid prescriptions:

The 'opioid score' for a given prescriber with known specialty m is introduced as:

$$\sum_{k=1}^K E_k(\text{opioid}) \times \max(0, \theta_{d,k} - q_{m,k})$$

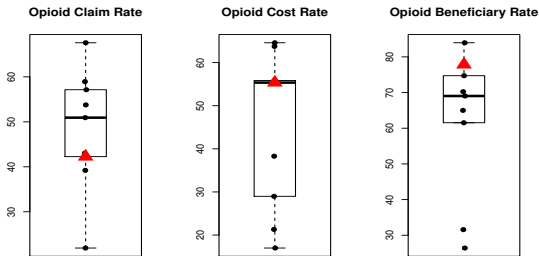
where, for a given topic k :

$$E_k(\text{opioid}) = \sum_{v=1}^V \mathbb{1}_v \times \beta_{k,v}$$

$\mathbb{1}_v$ is a binary indicator to show whether the v^{th} drug is opioid or not.

Practical evidence

Comparison of the the provider that has the highest opioid score (triangle in the boxplot) with his/her peers with respect to the ratio of the claims billed with an opioid drug, percentage of the total opioid drug costs, and the ratio of beneficiaries that were prescribed at least one opioid drug.



Practical evidence for the effectiveness of the proposed method in flagging providers for audits without considering the monetary measures and by only taking their billing pattern into account.

Bayesian co-clustering

- Objective is to reveal common billing patterns
- Bayesian co-clustering allows mixed membership for clusters of providers and of procedures
- Soft clustering
- Conspiracy fraud
- Financial investigations
- U.S. Medicare Part B (outpatient) claims data
 - Anesthesiologists in Texas that provide services in a facility
 - Providers that have billed for at least 10 unique procedures and the procedures that are billed by at least 20 unique providers
 - Binary billing matrix that lists whether each of $J=94$ procedures are billed by $I=376$ providers or not

Bayesian co-clustering

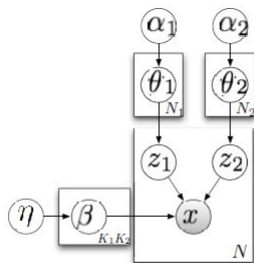


Fig. 2. Latent Dirichlet Bayesian Co-clustering Model

- Wang et al. (2010), Ekin et al. (2013)
- Procedures-Providers, Providers-Patients (Beneficiaries)
- Consumer-Consumer, Consumer-Merchant

Bayesian co-clustering: Model

$X_{ij} = 1$ if provider i bills for procedure j (customer i shops from merchant j)

$$X_{ij} | Z_{1i} = k, Z_{2j} = l, \beta_{kl} \sim \text{Ber}(\beta_{kl})$$

Priors:

$$\theta_1 \sim \text{Dir}(\alpha_{1k}; k = 1, \dots, K), \theta_2 \sim \text{Dir}(\alpha_{2l}; l = 1, \dots, L)$$

$$\beta_{kl} \sim \text{Beta}(a_{kl}, b_{kl}), k = 1, \dots, K, l = 1, \dots, L.$$

Posteriors:

$$\beta_{kl} | \mathbf{Z}_1, \mathbf{Z}_2, \mathbf{X} \sim \text{Beta} \left(a_{kl} + \sum_{i,j} X_{ij} \mathbf{I}(Z_{1i} = k, Z_{2j} = l), \right. \\ \left. b_{kl} + \sum_{i,j} (1 - X_{ij}) \mathbf{I}(Z_{1i} = k, Z_{2j} = l) \right)$$

Bayesian co-clustering: Model

$$\theta_1 | \mathbf{Z}_1 \sim \text{Dir}\left(\alpha_{1k} + \sum_{i,j} \mathbf{I}(Z_{1i} = k); k = 1, \dots, K\right),$$

$$\theta_2 | \mathbf{Z}_2 \sim \text{Dir}\left(\alpha_{2l} + \sum_{i,j} \mathbf{I}(Z_{2j} = l); l = 1, \dots, L\right).$$

The full conditionals of (Z_{1i}, Z_{2j}) can be obtained as

$$p(Z_{1i} = k, Z_{2j} = l | \theta_1, \theta_2, \beta, X_{ij}) = \frac{\beta_{kl}^{X_{ij}} (1 - \beta_{kl})^{1 - X_{ij}} \theta_{1k} \theta_{2l}}{\sum_{r=1}^K \sum_{c=1}^L \beta_{rc}^{X_{ij}} (1 - \beta_{rc})^{1 - X_{ij}} \theta_{1r} \theta_{2c}}.$$

Bayesian co-clustering Results

$$\beta = \begin{bmatrix} 0.145 & 0.143 \\ 0.140 & 0.142 \\ 0.149 & 0.146 \end{bmatrix}$$

- The most frequent occurrences between provider-procedure pairs are in co-cluster (3, 1).
- The provider-procedure pairs that are in co-cluster (3, 1) are 6.4% more likely to bill compared to the pairs in co-cluster (2, 1).

Bayesian co-clustering Results

- Billing of the Procedure 64941 that corresponds to facet joint injection by Provider with ID 100.
- The posterior modes are $Z_1 = 3$ and $Z_2 = 1$ in line with $\beta_{3,1}$

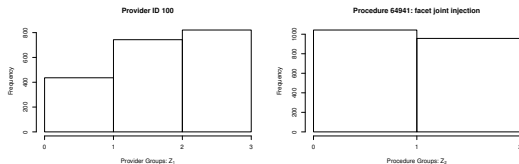


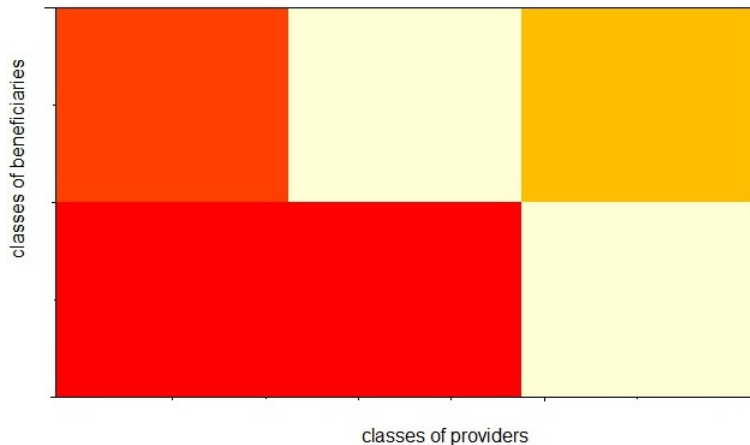
Figure: Posterior distributions of the memberships for Provider 100 and Procedure 64941

Bayesian co-clustering: Potential insights

- Can we detect providers and procedure pairs that have unusual behavior?
 - β_{kl} provides insights for associations
 - The higher the β_{kl} , the more likely is the probability that a member of provider cluster k bills for the members of procedure cluster l .
 - Discrepancies between the expected behavior and the actual behavior of a given provider can provide investigative leads.
 - For a given billing; if the provider does not behave similar to his co-cluster; this may reveal a potential fraudulent behavior.
- Identification of associations among providers and patients
 - Potential flags for unusual memberships in provider and procedure clusters
 - θ_1 for providers, θ_2 for patients
 - Can be useful for conspiracy fraud and kickback schemes

Providers-Patients co-clustering

$X_{ij} = 1$ if provider i serves patient j



Providers-Patients co-clustering

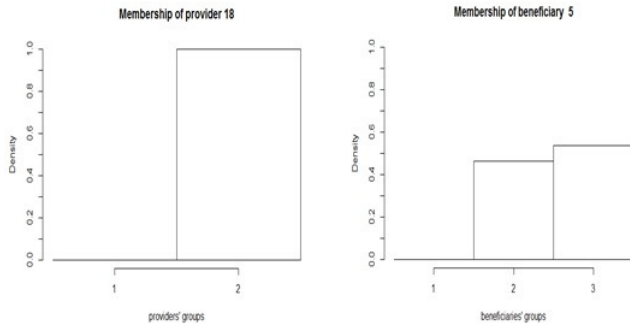


Figure: Posterior distributions of memberships of provider 18 and beneficiary 5

Providers-Patients co-clustering

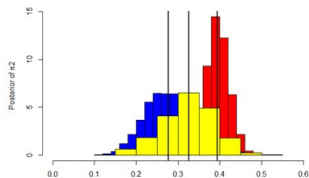
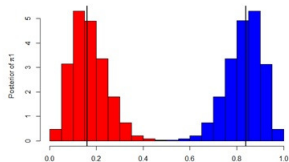


Figure: Posterior distribution of membership probabilities

Supervised Data Mining

- Predicting a fraud score using a number of variables such as provider, patient or claims characteristics
- Classification of the category of a new claim on the basis of labeled data with known category memberships
- Prediction of overpayment amount
- Quantile regression for analyzing billing behavior and aggressiveness

Dealing with Imbalanced data

- Few frauds in large data sets; skewed distributions
- Ex: 400 frauds out of 200,000 transactions: highly imbalanced, with positive class(fraud) accounting for 0.2% of all transactions.
- Scale numerical variables
- Oversampling: Artificially creating more observations from unbalanced class
 - SMOTE: **S**ynthetic **M**inority **O**versampling **T**echnique
 - Finding k-nearest neighbors for minority class observations
 - Randomly choosing one of the k-nearest neighbors and using it to create similar, new observations
- Undersampling: Artificially creating less observations from overrepresented class
 - RandomUnderSampler: performing k-means clustering from overrepresented class and removing data points from high density centroids.
- AUROC, precision, accuracy, false positives, false negatives

Supervised Data Mining: Classification

- Identification of the category of a new claim/application on the basis of labeled data with known category memberships.

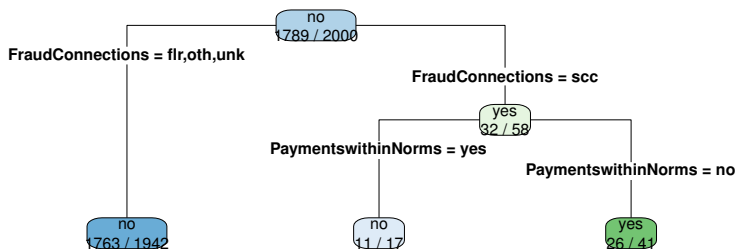


Figure: A decision tree output for classification of health care claims

Quantile Regression, Ekin and Damien (2019)

- “(Billing) aggressiveness (of the provider)” is defined as the ratio of the average submitted charged amount and average payment amount.
- The factors that impact the lower quantiles of the aggressiveness distribution may be different than those at the upper quantiles, as well as their marginal impacts.
- A change in the “payment amount” may have little impact if billing aggressiveness is low, but it could result in a great impact in billing aggressiveness in case it is already high.
- An additional percent increase in mean aggressiveness of a particular provider type may have greater impact on individual provider aggressiveness among conservative providers, but could have a low impact among very aggressive providers.
- Use of BIC for variable selection

Quantile Regression

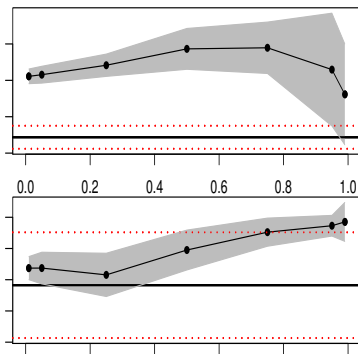


Figure: OLS and quantile regression estimates for “average Medicare standardized payment amount”(top) and “standard deviation of aggressiveness for provider type (sd aggPT)(bottom)”.

Quantile Regression

Quantile	λ^*	BIC	Significant Variables
0.01	0.01	2.872	avg. std. payment amount, GPCI Work
0.05	0.02	4.459	line service count, unique bene count, GPCI Work
0.25	0.02	5.808	unique bene count, avg. std. payment amount, GPCI Work
0.5	0.01	6.118	unique bene count, avg. std. payment amount, GPCI Work
0.75	0.01	5.956	line service count, avg. std. payment amount, GPCI Work, sd agg PT
0.95	0.01	5.08	line service count, avg. std. payment amount, GPCI Work, sd agg PT
0.99	0.01	3.932	line service count, GPCI Work, sd agg PT

Table: BIC and Significant Variables for LASSO Quantile Regression

Quantile Regression

Variables/Quantiles	0.01	0.05	0.25	0.5	0.75	0.95	0.99
(Intercept)	2.258	2.109	-0.683	-0.533	-4.159	12.573	40.318
line service count	0	-0.015	0	0	-0.006	-0.01	-0.016
unique bene count	0	0.017	-0.003	-0.001	0	0	0
avg. std. payment amount	0.005	0	0.137	0.157	0.198	0.156	0
GPCI Work	-0.04	-0.053	-0.069	-0.096	-0.141	-0.58	-1.263
sd agg PT	0	0	0	0	0.705	0.69	0.424

Table: MLE's for chosen variables for LASSO Quantile Regression (Note: All the other variables are not selected, and hence have MLEs of 0.)

Medical Audits and Overpayment Estimation

- Simple random sampling, stratified sampling, and two-stage (probe) sampling
- **Recovery amount:** Lower limit of a one sided 90 percent confidence interval for the total overpayments, CMS (2001)
- What if the payments and/or overpayments are not Normally distributed (and possibly multi-modal) with a modest sample size
- Failure of normal approximation and CLT for small sample sizes: Edwards et al. (2003), Minimum sum method for “All or nothing” (Edwards et al. (2003), Ignatova and Edwards (2008)
- What if there is more than one overpayment pattern?
- What about cases with partial overpayments and possibly with “all or nothing”?
- Zero-One Inflated Mixture Model (Ekin et al.,2015 JAS), Bayesian Inflated Mixture Model (Musal and Ekin, 2017 SM), iterative information theoretic multi-stage sampling (Musal and Ekin, 2018 ASMBI)

Decisions for Fraud: Audit sampling decision models, Ekin and Musal (2020)

- Audit resource allocation problem under uncertainty
- Cases where the auditor only has determined the provider of interest with access to the related payment data.
- Trade-offs between audit costs and expected recovery while deciding how to allocate the sampling resources among the initial and potential additional investigations within the budget

$$\begin{aligned} \max_{n_{(init)}, n_{(add)}} & E[rY_{rec} + (n_{(init)} + n_{(add)})\bar{y}] - c_1 n_{(init)} - c_2 n_{(add)} \\ \text{s.t.} & c_1 n_{(init)} + c_2 n_{(add)} \leq B, (n_{(init)}, n_{(add)}) \in A \end{aligned}$$

Decisions for Fraud

Decision problem where the objective is to find the optimal additional sample size for a given initial sample:

$$\max_{n_{(add)}} E[n_{(add)}\bar{y} + rY_{rec}] - c_2 n_{(add)}.$$

Trade-offs of cost and expected recovery involved with choosing the additional sample size.

- Incorporation of advanced analytical methods into audit decision making frameworks within the limits of law: semi-automated systems, cost of false positives/negatives, historical predictive power, audit costs, deviation from average

- Decision games incorporating the fraudsters response: risk adversarial agents, adaptive thresholds that address gaming by fraudsters
- ARA for Fraud Detection
- Adversarial machine learning in risk analysis
- Adversarial Classification
- Unsupervised Adversarial Learning

Ongoing work

- Connections to GDRR themes and WGs: Risk and decisions in societal systems and policy advisory contexts
- Fraud detection as a health care risk management problem
- Better visualization and user interface
- Spatial temporal analysis with topic models
- Network analysis
- Dealing with data quality issues and missing/confidential data
- Quantile regression to estimate risk adjustments per patient
- Connections to patient health (quality outcomes) and related payment models
- Consideration from the patient perspective and involvement

Open challenges

- Explainable AI-need for interpretability, right to know
- Impact of cyber-security breaches on health care fraud
- Adoption of solutions from related areas such as finance, eligibility assessment
- How to increase the value of sampling approaches in a world of court battles and settlements
- Drug pricing by Pharmacy Benefit Managements (PBMs) and price transparency among insurers-hospitals
- Need for adaptive statistical methods
- Impact of potential blockchain solutions of fraud assessment frameworks
- Real time (at least proactive pre-payment) fraud assessment

- Medicare provider utilization & payment data: Physician and Other Suppliers
 - Services and procedures given to Medicare beneficiaries, including utilization information.
 - Payment amounts (allowed amount and Medicare payment).
 - Submitted charges organized by Healthcare Common Procedure Coding System (HCPCS) code.
- 2011-2016 Medicare provider utilization & payment data: Inpatient Hospital Public Use File (PUF)
- Medicare provider utilization & payment data: Outpatient Hospital PUF
- National Health Expenditures

Thank you!

Contact: tahirekin@txstate.edu

Review paper: Ekin, T., Ieva, F., Ruggeri, F., & Soyer, R. (2018).
Statistical medical fraud assessment: exposition to an emerging field.
International Statistical Review, 86(3), 379-402.

