Understanding Medical Patients with Machine Learning

September 18, 2016

Agenda

- **Context and Overview**
- Patient Profiles
- J Codes

A health insurance claim is a bill, submitted to a health plan by a provider, for health care services rendered

Claim Definition

- A claim is a bill submitted by a provider to a patient's insurance company for reimbursement for delivered services (e.g. diagnoses, medical procedures)
  - A professional claim is a claim submitted by a doctor or other health care professional
  - An institutional claim is filed by a hospital or institution for facility and equipment use
- The amount of reimbursement for the same procedure or diagnosis can vary depending on the patient's benefit plan, co-pay, deductible and the provider's contract with the health plan

Perspectives on Claims

Claims processing has varying implications for members of the healthcare system:

- Health plan: A claim represents the analysis, processing and payment of bills from contracted providers
- Provider: A claim is a bill or invoice, delivered from a provider with an explanation of services rendered
- Member: A claim is a financial statement that outlines services rendered and payment status

Unsupervised learning methodologies could produce actionable insights when applied to medical claims

What questions can we explore with claims data?

Patient Profiles

- Do one or more customer segments drive a disproportionate share of medical cost?
- How are “expensive” patients similar or dissimilar?
- Which elements of a claim best explain member groups / clusters? E.g., diagnoses, DRGs, procedures, revenue codes?
- Are there logical next steps to develop differing clinical pathways based on patient clusters?

Methodology Used:
K-Means Clustering

Injectable “J Code” Procedures

- Which are the most expensive and highest volume J Codes?
- What places of service tend to deliver injections?
- Is there a statistically significant difference in cost amongst places of service delivering the same injection?
- What claim elements do we tend to see with a higher than expected frequency prior to receiving J Code claims?
- Are there ways to drive patients to lower cost settings?

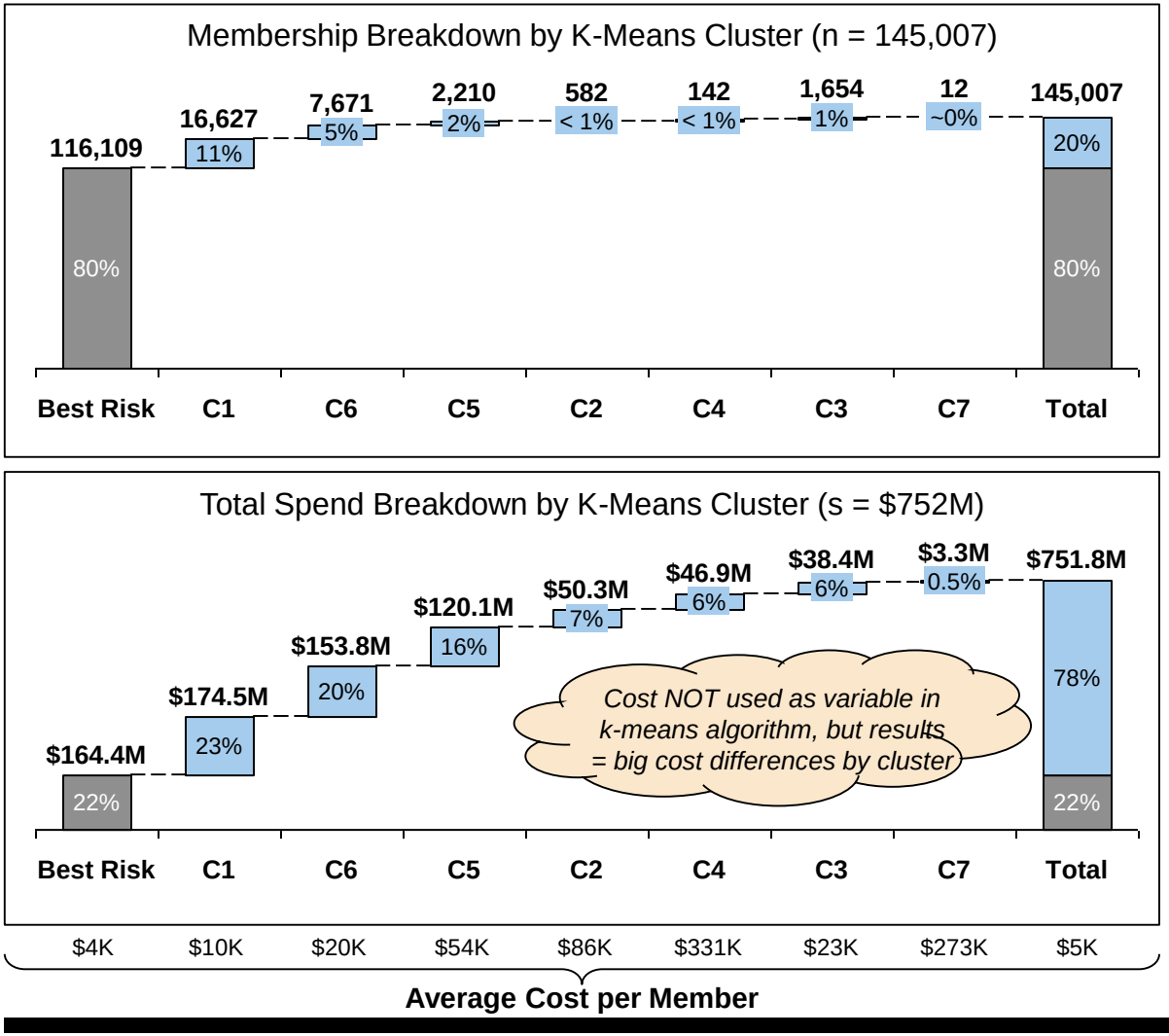
Methodology Used:
Association Rules

Agenda

- Context and Overview
- **Patient Profiles**
- J Codes

20% of members drive 78% of health care spend: clustering this expensive risk pool can help us understand segment characteristics

Patients with Continuous Enrollment: Waterfall Charts

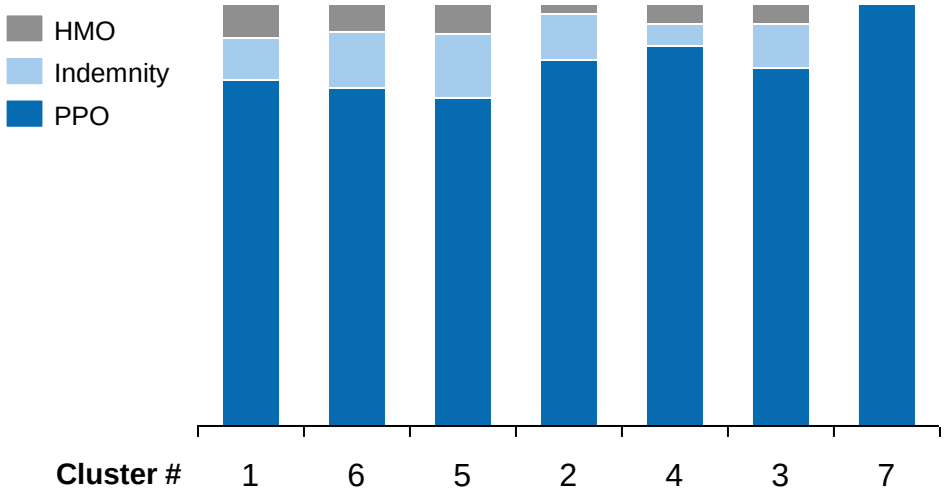


Membership Clusters

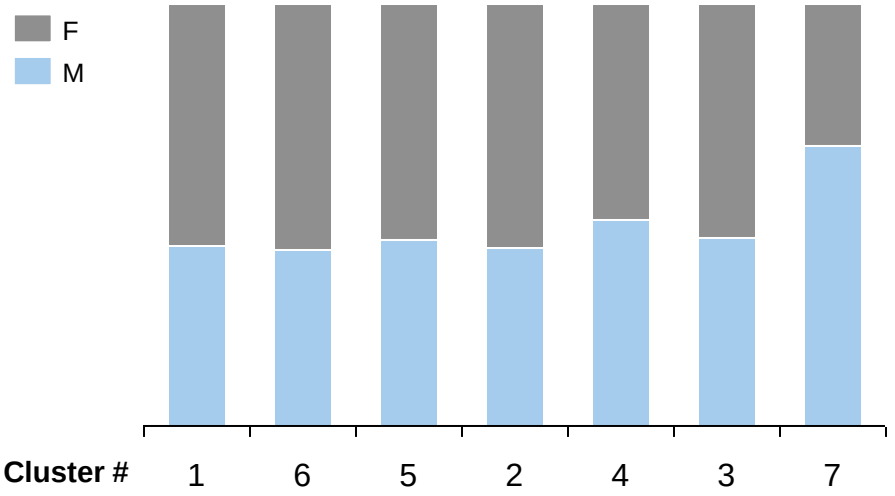
- Massive skew in cost distribution: 20% of patients drive 78% of costs
- K-Means clustering applied to most expensive 20% of members produces 7 key sub-segments
- K = 7 chosen based on scree plot and visual inspection
- Appears to be huge variance even amongst expensive segments:
  - C1 has highest absolute spend ($175M) but is least expensive on per member basis ($10K)
  - C5, C2, and C4 have high absolute and per member spend figures
  - C7 appears to be an outlier with $273K per member but only 12 observations
  - Per member spend ranges from **2 to 83 times** the “best risk” rate

Most products are PPO though Cluster 5 has high indemnity; 60% of expensive members are female, except in Clusters 4 and 7

Cluster Total Spend Composition: Product Type



Cluster Total Spend Composition: Gender



Product Type Takeaways

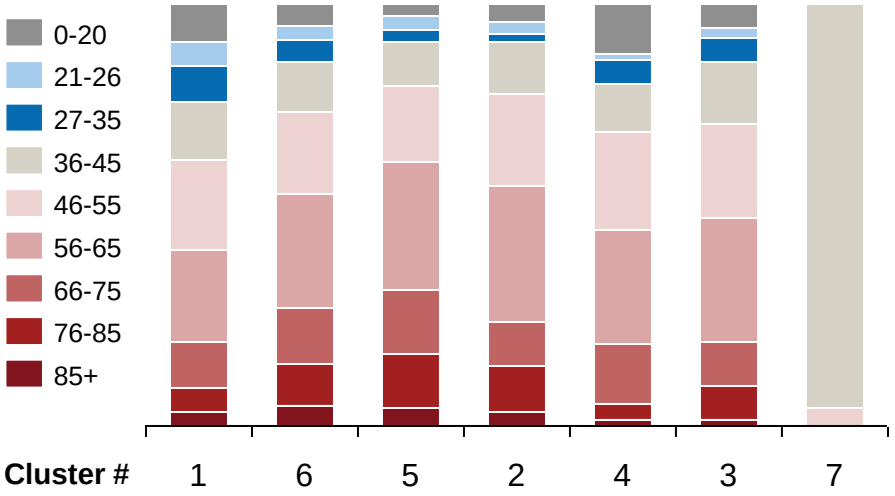
- Clusters 2 and 4 have higher proportion of PPO product
- 5 has more Indemnity than all other clusters
- 2 has minimal HMO

Gender Takeaways

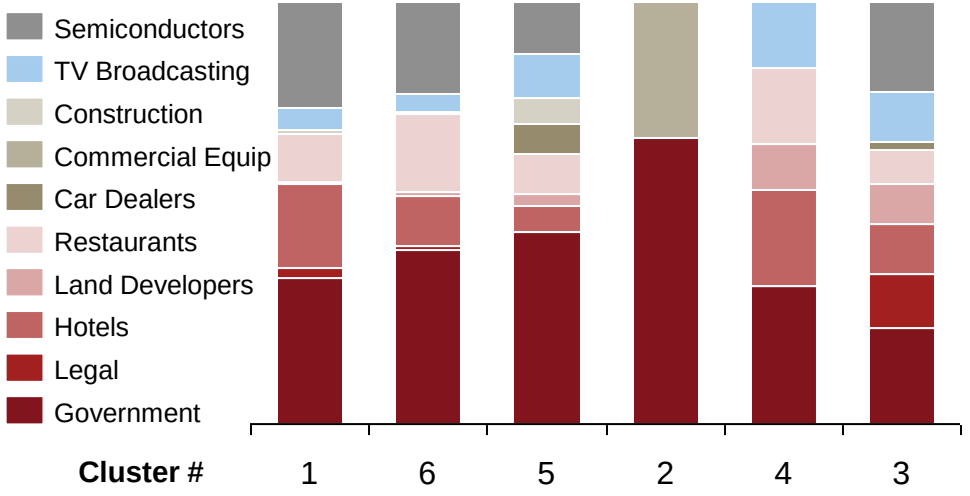
- Overall, more females than males in expensive segment
- 4 and 7 seem to have higher proportion of males than other clusters

There appears to be very different age and industry compositions across clusters

Cluster Total Spend Composition: Age Group



Cluster Total Spend Composition: Top Industries



Age Group Takeaways

- Cluster 4 has a lot of 0 – 20 year olds, but is lower in the 76 – 85 year old range
- Clusters 5 and 2 seem to be skewed towards older patients

Industry Takeaways

- Selected top codes identifying which industries the members belong to (most are “other” → excluded)
- Semiconductors group appears to be expensive overall
- High concentration of government workers in Cluster 2

Disease state (diagnoses) and treatment (procedures / revenue codes) explain key differences between clusters

Cluster Analysis: Top Disease and Treatment Codes

Top Features	C1	C6	C5	C2	C4	C3	C7
Procedure	• Office visit • ER visit	• Office visit • Therapeutic exercises	• Sub hospital care • Chemo	• Chemo • Trastuzumab injection	• Dialysis • Sub hospital care • ALS1 emergency	• Therapeutic exercises • Chiropractic manipulation • Manual therapy	• Alcohol / drug services • Group psy-therapy • Drug screening
Diagnosis	• Newborns / childbirth	• End stage renal disease • Rehab	• Antineoplast chemo • Prostate cancer	• Breast cancer • Colorectal cancer	• End stage renal disease	• Bone spurs • Back pain • Spinal / ortho issues	• Opioid dependence • Drug / alcohol dependence
Revenue Code	• Emergency room • Room and board	• OR services • Ambulatory surgery	• Radiation therapy	• Chemo • Organ transplant	• Intensive care • Hemodialysis	• Room and board • Ambulatory surgery	• Clinical lab • Psychiatric treatment
Summary:	Expensive “healthy” patients	Unclear	Cancer treated by radiation	Cancer treated by chemo	Kidney failure	Elective surgery (knee replacement)	Drug treatment

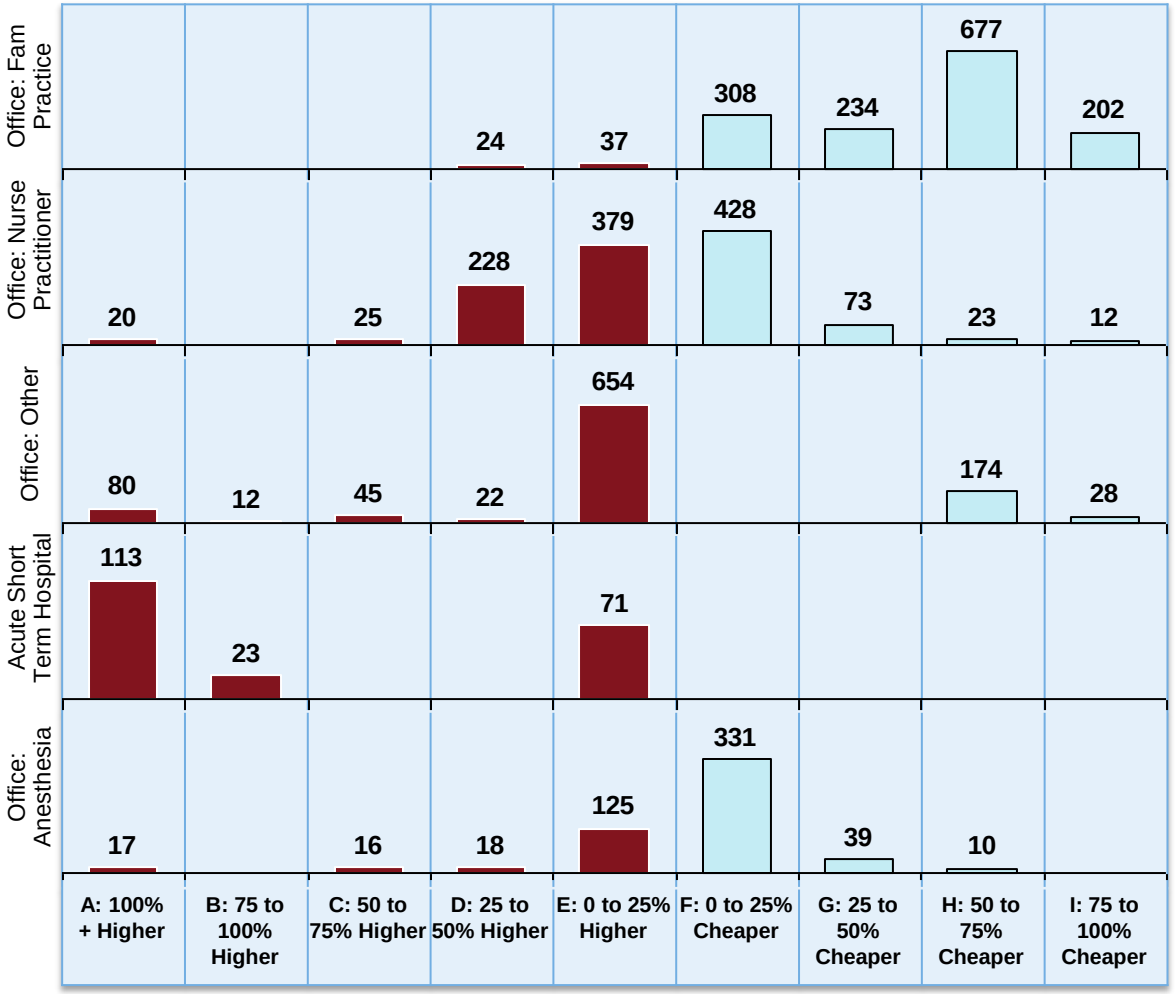
Agenda

- Context and Overview
- Patient Profiles
- **J Codes**

Top 25 injectable code cost varies significantly by place of service, even when holding disease state constant

Cost Comparison for Select Places of Service

(Relative to Avg. Cost per Injection for Same Diagnosis; count = # claim lines)



Place of Service Discussion

- Considered only top 25 injectable codes based on total spend
- Summarized avg. cost per line by grouping of procedure / diagnosis / revenue codes and place of service
- Eliminated records with fewer than 10 lines per group (239 groups remained)
- Anova indicated which places of service had significant cost variance
  - Compared avg. cost per line for each place of service within group
  - Significance threshold = 0.0002 (0.05 / 239) to correct for multiple testing error
- Some places of services are more consistently expensive / cheaper than others, however it seems to vary
- Therefore: must examine by individual J code to understand behavior

Drilling down by J Code and diagnosis gives a sense for which places of service to focus on

Cost Comparison for Select Injectable Drugs

prcdr_cd	cpt_descr	diagnosis_descr	cost_per_line	place_of_service|specialty	#_lines	total_cost
J1745	**"INJECTION INFLIXIMAB, 10 MG"**	Regional enteritis of unspecified site	**$3,318**	Patient's Home | Home Health Care Agency	14	$46,457
			$4,787	Office | Gastroenterology	28	$134,029
			$8,019	Outpatient Hospital | Acute Short Term Hospital	71	$569,355
			$15,903	Outpatient Hospital | Childrens Hospital	15	$238,540
		Rheumatoid arthritis	**$1,899**	NULL | Ancillary/Hospital-Based	11	$20,893
			$2,395	Office | Internal Medicine	67	$160,462
			$2,963	Office | Rheumatology	99	$293,358
J1745	**Total**				**305**	**$1,463,094**
J9355	**"INJECTION, TRASTUZUMAB, 10 MG"**	Malignant neoplasm of upper-outer quadrant of female breast	**$2,264**	Office | Urology	16	$36,228
			$4,460	Office | Hematology/Oncology	76	$338,998
J9355	**Total**				**92**	**$375,226**
J0129	**"INJECTION, ABATACEPT, 10 MG**	Rheumatoid arthritis	**$2,147**	Office | Rheumatology	91	$195,395
			$6,286	Outpatient Hospital | Acute Short Term Hospital	20	$125,713
J0129	**Total**				**111**	**$321,108**
J1566	**"INJECTION, IMMUNE GLOBULIN, 500 MG"**	"Hypogammaglobulinemia, unspecified"	**$1,487**	Office | Internal Medicine	10	$14,874
			$3,142	Office | Allergy/Immunology	12	$37,703
J0585	**Total**				**22**	**$52,577**
J3489	**"INJECTION, ZOLEDRONIC ACID, 1 MG"**	"Multiple myeloma"	**$ 430**	Office | Not Mapped	98	$42,123
			$3,870	Outpatient Hospital | Acute Short Term Hospital	20	$77,403
J3489	**Total**				**118**	**$119,525**
J2785	**"INJECTION, REGADENOSON, 0.1 MG"**	"Chest pain, unspecified"	**$ 209**	Office | Cardiology	143	$29,908
			$ 238	Office | Internal Medicine	22	$ 5,237
			$ 620	Outpatient Hospital | Acute Short Term Hospital	10	$ 6,204
		Other chest pain	**$ 204**	Office | Cardiology	64	$13,026
			$1,456	Emergency Room - Hospital | Acute Short Term Hospital	12	$17,474
J2785	**Total**				**251**	**$71,848**

Associating J Code claims with prior 30 days of claims reveals some interesting connections and predictive scenarios

J Code Association Rules Transaction Example

Medical Claim with J Code

J15166 Injection, Immune Globulin	Member_ID 123	Date of Service 1/1/2015

Concatenate

Medical Claims for Member 123 for Prior 30 Days

Procedure XYZ	Diagnosis ABC	Place Service DEF	Date of Service 12/31/14
Procedure ABC	Diagnosis 123	Place Service QRF	Date of Service 12/15/14
Procedure 555	Diagnosis 123	Place Service DEF	Date of Service 12/2/14

A-Rules Transaction File

Methodology Discussion

- Retroactively grouped prior 30 days claims for members that had relevant J code claim
- Created “transactions” file to feed into apriori algorithm in Arules R package
- Objective: understand what types of claims occur more frequently than expected preceding a J code claim

Interesting Globulin A-Rules

- Prior 30 days claims tended to contain:
  - Diagnosis: chronic inflammatory polyneuritis
  - Place of Service: Office | Neurology
  - Procedure: Intravenous infusion
- Relevant rules metrics:
  - **Support**: 6% - 31%
  - **Lift**: 1.7 to 2.1
- So when these combinations are present, more likely to see Globulin J code within 30 days

Example visualization: of the 8 association rules for Globulin injection, one seems to be highly predictive based on POS and DX