PREDICTIVE MODELING: BASICS & BEYOND

Ian Duncan FSA FIA FCIA MAAA

November 2014

Agenda

1. Lecture 1: Health Risk.
2. Lecture 2: Condition and Risk Identification.
3. Lecture 3: Grouper Models.
4. Lecture 4: Model Construction.

If time permits:

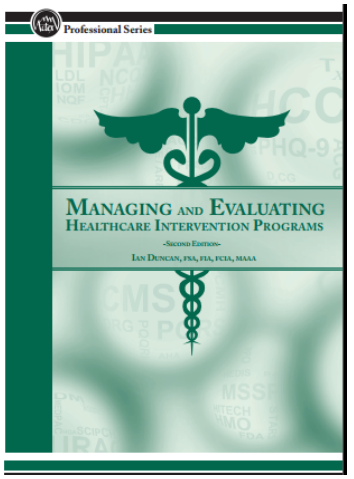
Applications – case studies:

- Predictive Modeling for ACOs
- Bundled Payments
- Massachusetts Health Reform Risk Adjustment
- Case finding/opportunity analysis

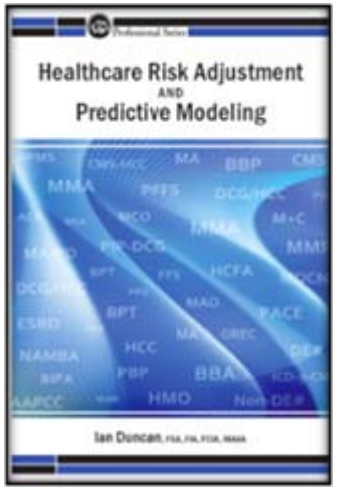
Introductions

Ian Duncan, FSA, FIA, FCIA, MAAA

- Professor, Actuarial Statistics, University of California at Santa Barbara.
- Founder and former president, Solucia Consulting (now SCIO Health Analytics).
- Author of several books and a number of peer-reviewed studies on healthcare management and predictive modeling
- Principal investigator of a multi-year actuarial analysis of Massachusetts Reform



2nd Edition 2014



What we will cover in lecture 1

- What is health risk?
- Typical Claims cost distributions.
- Member transitions over time.
- Traditional (Actuarial) methods of risk prediction:
  - Age/sex
  - Prior Cost
  - Prediction using Clinical Conditions.

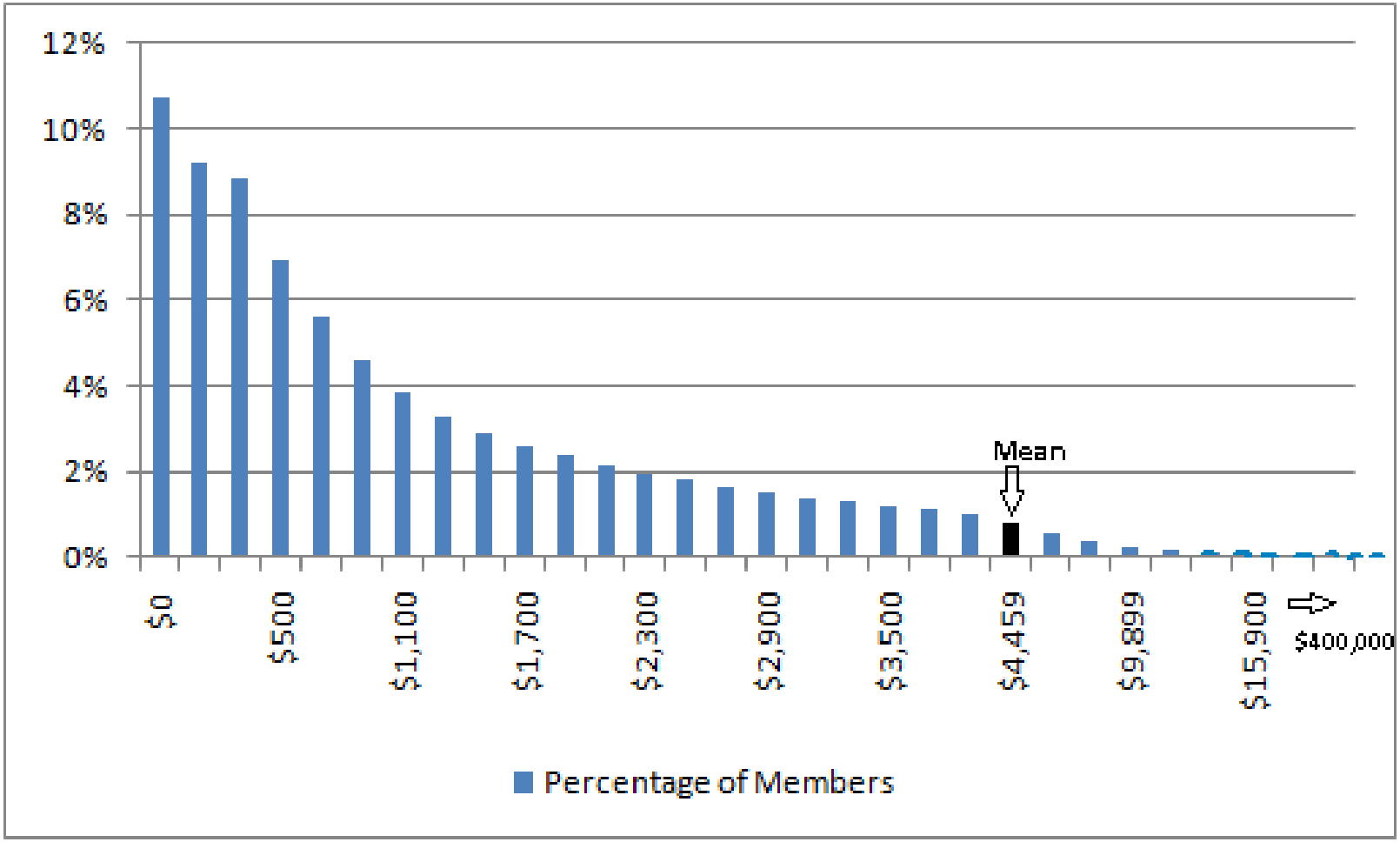
Definition of Risk

RISK = F (Loss amount; Probability)

•Another way of saying this is that Risk is a function of Frequency (of occurrences) and Severity of the occurrence.

•In healthcare, we are interested in many different states. Most frequently actuaries are interested in Financial Loss, which occurs because an event imposes a cost on an individual (or employer or other interested party). To a clinician, however, a **loss** could have a different meaning: it could be a loss of function, such as an inability to perform at a previous level or deterioration in an organ.

Typical Distribution of Population health cost



* Distribution of allowed charges within the Solucia Consulting database (multi-million member national database).

Key Concept: Member Transition

- A key concept in understanding health risk is that, while the *shape* of the distribution remains stable over time, the *composition* of the distribution changes constantly.
- Said another way: yesterday's high-cost individual isn't going to be tomorrow's.

We often hear statements about "x% of members responsible for y% of cost" in healthcare. Two problems with this statement:

- So what?
- Yesterday's high cost individuals may not be tomorrow's.

Let's take a look at some real data.

Key Concept: Member Transition

	Baseline Year	Sequent Year		
Baseline Year Cost Group	Baseline Percentage Membership	LOW <$2,000	MODERATE $2,000-$24,999	HIGH $25,000+
LOW <$2,000	69.5%	57.4%		
			11.7%	
				0.4%
MODERATE $2,000-$24,999	28.7%	9.9%		
			17.7%	
				1.1%
HIGH $25,000+	1.8%	0.2%		
			0.9%	
				0.6%
TOTAL	100.0%	67.6%	30.3%	2.2%

Key Concept: Member Transition

	Baseline Year	Sequent Year		
Baseline Year Cost Group	Baseline Percentage Membership	LOW <$2,000	MODERATE $2,000-$24,999	HIGH $25,000+
LOW <$2,000	69.5%	57.4%		
			11.7%	
				0.4%
MODERATE $2,000-$24,999	28.7%	9.9%		
			17.7%	
				1.1%
HIGH $25,000+	1.8%	0.2%		
			0.9%	
				0.6%
TOTAL	100.0%	67.6%	30.3%	2.2%

Key Concept: Member Transition

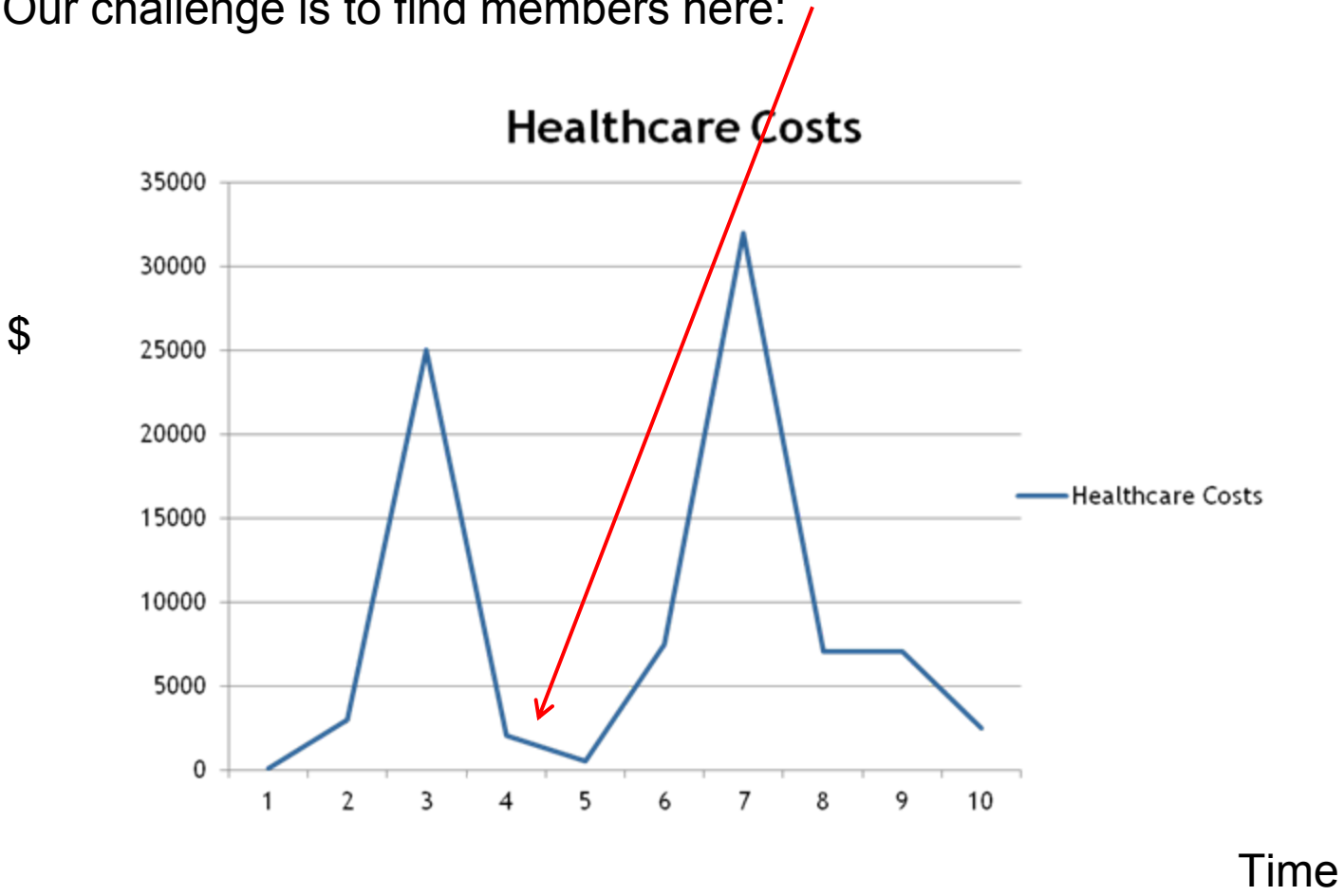
	Baseline Year	Sequent Year		
Baseline Year Cost Group	Baseline Percentage Membership	LOW <$2,000	MODERATE $2,000-$24,999	HIGH $25,000+
LOW <$2,000	69.5%	57.4%		
			11.7%	
				0.4%
MODERATE $2,000-$24,999	28.7%	9.9%		
			17.7%	
				1.1%
HIGH $25,000+	1.8%	0.2%		
			0.9%	
				0.6%
TOTAL	100.0%	67.6%	30.3%	2.2%

Key Concept: Member Transition

	Baseline Year	Sequent Year PMPY CLAIMS		
Baseline Year Cost Group	Mean Per Capita Cost	LOW <$2,000	MODERATE $2,000-$24,999	HIGH $25,000+
LOW <$2,000	$510.37	$453.24		
			$5,282.58	
				$56,166.54
MODERATE $2,000-$24,999	$6,157.06	$888.30		
			$6,803.91	
				$49,701.87
HIGH $25,000+	$55,197.12	$907.47		
			$10,435.51	
				$73,164.49
TOTAL		$518.72	$6,325.46	$57,754.19
		67.6%	30.3%	2.2%
AVERAGE	$3,090.36			$3,520.09
TREND				13.9%

Key Concept: Member Costs Fluctuate

Our challenge is to find members here:



Aside regarding member transitions

The above is a very simple application of a multi-state (Markov) model. Markov models are actually not well-suited to healthcare analysis because the Markov property – that transition depends only on the current state and not on any previous states – is violated by healthcare experience.

I have been interested for a while in the problem of construction of mathematical models of chronic disease member transition; we should be able to construct something like "actuarial tables" with expected values for the likelihood that a member aged (x) with a condition (y) will transition to a more severe form of (y) or to conditions (y) + (z).

Currently looking at this as an application of a semi-Markov process (unlike Markov processes, which are memoryless, semi-Markov models take account of history and length of time in the current state. Applying to UK data (which has the advantage of being available for long periods).

Traditional (actuarial) Risk Prediction

Age/Sex: although *individuals* of the same age and sex represent a range of risk profiles and costs, *groups* of individuals of the same age and sex categories follow more predictable patterns of cost. In a gross way, these age/sex rates are predictable (just not very accurately).

Relative Cost PMPY by Age/Sex			
	Male	Female	Total
< 19	$1,429	$1,351	$1,390
20-29	$1,311	$2,734	$2,017
30-39	$1,737	$3,367	$2,566
40-49	$2,547	$3,641	$3,116
50-59	$4,368	$4,842	$4,609
60-64	$6,415	$6,346	$6,381
Total	$2,754	$3,420	$3,090

Accuracy of Traditional Risk Prediction

Traditional (Age/Sex) risk prediction is somewhat accurate at the population level. Larger group costs are more predictable than smaller groups.

Demographic Factors as Predictors of Future Health Costs								
		Age/Sex Factors		Factor Ratio			Difference** (Predicted-Actual)	
Employer	Number of lives	Baseline	Subsequent Year	Subsequent/ Average	Predicted Cost*	Actual Cost	$	%
1	73	1.37	1.42	138%	$4,853	$23,902	**($19,049)**	-392.5%
2	478	0.74	0.76	74%	$2,590	$2,693	**($102)**	-3.9%
3	37	0.86	0.87	84%	$2,965	$1,339	$1,626	54.8%
4	371	0.95	0.97	95%	$3,331	$3,325	$6	0.2%
5	186	1.00	1.03	100%	$3,516	$3,345	$170	4.8%
6	19	1.80	1.85	180%	$6,328	$10,711	**($4,383)**	-69.3%
7	359	0.95	0.97	94%	$3,315	$3,401	**($87)**	-2.6%
8	543	0.94	0.96	93%	$3,269	$3,667	**($398)**	-12.2%
9	26	1.60	1.64	159%	$5,595	$5,181	$414	7.4%
Average		1.00	1.03	1.00	$3,520	$3,520	$ -	0.0%
Sum of absolute Differences (9 sample groups only)							$26,235	

Prior Experience adds to accuracy

To account for the variance observed in small populations, actuaries typically incorporate prior cost into the prediction, which adds to the predictive accuracy. A "credibility weighting" is used. Here is a typical formula:

$$\text{Expected Cost} = \text{Prior Year Cost} \times \text{Trend} \times Z + \text{Book of Business Cost} \times (1 - Z)$$

where $Z = \left(\frac{N}{2000}\right)^{0.5}$ and N is the number of members in the group.

Combination of Age, Sex, and Prior Cost as Predictor of Future Experience

			Cost PMPY			Difference vs. Actual	
Employer	No. of lives	Credibility Factor	Baseline	Subsequent Year Pre-dicted	Subsequent Year Actual	Difference	Difference (% of Actual)
1	73	0.19	$27,488	$9,908	$23,902	**($13,994)**	-141.2%
2	478	0.49	$1,027	$2,792	$2,693	$100	3.6%
3	37	0.14	$1,050	$2,724	$1,339	$1,385	50.9%
4	371	0.43	$2,493	$3,119	$3,325	**($205)**	-6.6%
5	186	0.30	$3,377	$3,617	$3,345	$271	7.5%
6	19	0.10	$11,352	$6,971	$10,711	**($3,739)**	-63.6%
7	359	0.42	$2,008	$2,880	$3,401	**($522)**	-18.1%
8	543	0.52	$2,598	$3,108	$3,667	**($559)**	-18.0%
9	26	0.11	$3,022	$5,350	$5,181	$169	3.2%
….	….	….	….	….	….	….	….
Average			$3,090	$3,520	$3,520	$ -	0%
Sum of absolute Differences (9 sample groups only)						$20,944	

What does Clinical information tell us about risk?

Having information about a patient's condition, particularly chronic condition(s) is potentially useful for predicting risk.

How well do standardized costs predict individual member costs?

Relative Cost PMPY by Age/Sex			
	Male	Female	Total
< 19	$1,429	$1,351	$1,390
20-29	$1,311	$2,734	$2,017
30-39	$1,737	$3,367	$2,566
40-49	$2,547	$3,641	$3,116
50-59	$4,368	$4,842	$4,609
60-64	$6,415	$6,346	$6,381
Total	$2,754	$3,420	$3,090

What does Clinical information tell us about risk?

Having information about a patient's condition, particularly chronic condition(s) is potentially useful for predicting risk.

Condition-Based Vs. Standardized Costs						
Member	Age	Sex	Condition	Actual Cost (Annual)	Standardized Cost (age/sex)	Condition-Based Cost/ Standardized Cost (%)
1	25	M	None	$863	$1,311	66%
2	55	F	None	$2,864	$4,842	59%
3	45	M	Diabetes	$5,024	$2,547	197%
4	55	F	Diabetes	$6,991	$4,842	144%
5	40	M	Diabetes and Heart conditions	$23,479	$2,547	922%
6	40	M	Heart condition	$18,185	$2,547	714%
7	40	F	Breast Cancer and other conditions	$28,904	$3,641	794%
8	60	F	Breast Cancer and other conditions	$15,935	$6,346	251%
9	50	M	Lung Cancer and other conditions	$41,709	$4,368	955%

Risk Groupers predict relative risk

Commercial Risk Groupers are available that predict relative risk based on diagnoses. Particularly helpful for small groups.

Application of Condition Based Relative Risk						
			Cost PMPY		Difference (Predicted-Actual)	
Employer	Number of lives	Relative Risk Score	Predicted	Actual	$	%
1	73	8.02	$28,214	$23,902	$4,312	15.3%
2	478	0.93	$3,260	$2,693	$568	17.4%
3	37	0.47	$1,665	$1,339	$326	19.6%
4	371	0.94	$3,300	$3,325	**($25)**	-0.8%
5	186	1.01	$3,567	$3,345	$222	6.2%
6	19	4.14	$14,560	$10,711	$3,850	26.4%
7	359	0.84	$2,970	$3,401	**($432)**	-14.5%
8	543	0.80	$2,833	$3,667	**($834)**	-29.4%
9	26	1.03	$3,631	$5,181	**($1,550)**	-42.7%
Average			$ -	0.0%	$ -	0.0%
Sum of absolute Differences (9 sample groups only)					$12,118	

Lecture 2: Condition and Risk Identification

- At the heart of predictive modeling! How do we introduce members' diagnostic information into the identification of risk?
  - Who? What common characteristics?
  - How do we define a group of patients with those characteristics?
  - What are the implications of those characteristics?
- There are many different algorithms for identifying member conditions. THERE IS NO SINGLE AGREED FORMULA. (Grouper models are useful but not essential.)
- Condition identification often requires careful balancing of sensitivity and specificity.

Identification – example (Diabetes)

Diabetics can be identified in different ways:

Diagnosis type	Reliability	Practicality
Physician Referral/ Medical Records/EMRs	High	Low
Lab tests	High	Low
Claims	Medium	High
Prescription Drugs	Medium	High
Self-reported	Low/medium	Low

Medical and Drug Claims are often the most practical method of identifying candidates for predictive modeling.

Identification – example (Diabetes)

Inpatient Hospital Claims – ICD-9 Claims Codes

CODE	CODE TYPE	DESCRIPTION
250.xx	ICD-9 CM	Diabetes Mellitus
357.2	ICD-9 CM	Poyneuropathy in diabetes
362.0; 362.0x	ICD-9 CM	Diabetic Retinopathy
366.41	ICD-9 CM	Diabetic Cataract
648.0 - 648.04	ICD-9 CM	Diabetes Mellitus (as other current condition in mother) classifiable elsewhere but complicating pregnancy or childbirth

Codes in the 250 series ending in 1 or 3 (e.g. 250.x3) denote Type 1 diabetes

Diabetes – drug codes

Insulin or Oral Hypoglycemic Agents are often used to identify members. A simple example follows; for more detail, see the HEDIS code-set.

This approach is probably fine for Diabetes, but may not work for other conditions where off-label use is prevalent.

Insulin	
2710*	Insulin**

OralAntiDiabetics	
2720*	Sulfonylureas**
2723*	Antidiabetic - Amino Acid Derivatives**
2725*	Biguanides**
2728*	Meglitinide Analogues**
2730*	Diabetic Other**
2740*	ReductaseInhibitors**
2750*	Alpha-Glucosidase Inhibitors**
2760*	Insulin Sensitizing Agents**
2799*	Antiadiabetic Combinations**

Algorithm Development: Diabetes Example

Not all diabetics represent the same level of risk. Different diagnosis codes help identify levels of severity.

Codes for Identification of Diabetes Severity	
Diagnosis Code (ICD-9-CM)	**Code Description**
250.0	Diabetes mellitus without mention of complication
250.1	Diabetes with ketoacidosis (complication resulting from severe insulin deficiency)
250.2	Diabetes with hyperosmolarity (hyperglycemia (high blood sugar levels) and dehydration)
250.3	Diabetes with other coma
250.4	Diabetes with renal manifestations (kidney disease and kidney function impairment)
250.5	Diabetes with ophthalmic manifestations
250.6	Diabetes with neurological manifestations (nerve damage as a result of hyperglycemia)
250.7	Diabetes with peripheral circulatory disorders
250.8	Diabetes with other specified manifestations
250.9	Diabetes with unspecified complication

Algorithm Development: Diabetes Example

Relative Costs of Members with Different Diabetes Diagnoses			
Diagnosis Code ICD-9-CM	**Description**	**Average cost PMPY**	**Relative cost**
250	A diabetes diagnosis without a fourth digit (i.e., 250 only).	$13,258	105%
250.0	Diabetes mellitus without mention of complication	$10,641	85%
250.1	Diabetes with ketoacidosis (complication resulting from severe insulin deficiency)	$16,823	134%
250.2	Diabetes with hyperosmolarity (hyperglycemia (high blood sugar levels) and dehydration)	$26,225	208%
250.3	Diabetes with other coma	$19,447	154%
250.4	Diabetes with renal manifestations (kidney disease and kidney function impairment)	$24,494	195%
250.5	Diabetes with ophthalmic manifestations	$11,834	94%
250.6	Diabetes with neurological manifestations (nerve damage as a result of hyperglycemia)	$17,511	139%
250.7	Diabetes with peripheral circulatory disorders	$19,376	154%
250.8	Diabetes with other specified manifestations	$31,323	249%
250.9	Diabetes with unspecified complication	$13,495	107%
357.2	Polyneuropathy in Diabetes	$19,799	157%
362	Other retinal disorders	$13,412	107%
366.41	Diabetic Cataract	$13,755	109%
648	Diabetes mellitus of mother complicating pregnancy childbirth or the puerperium unspecified as to episode of care	$12,099	96%
TOTAL		$12,589	100%

Algorithm Development: Diabetes Example

Which leads to a possible relative risk severity structure for diabetes:

A Possible Code Grouping System for Diabetes			
Severity Level	**Diagnosis Codes Included**	**Average Cost**	**Relative Cost**
1	250; 250.0	$10,664	85%
2	250.5; 250.9; 362; 366.41; 648	$12,492	99%
3	250.1; 250.3; 250.6; 250.7; 357.2	$18,267	145%
4	250.2; 250.4	$24,621	196%
5	250.8	$31,323	249%
	TOTAL (All diabetes codes)	$12,589	100%

Risk (Condition) Groupers

Codes define important variables like Diagnosis (ICD-9 or 10; HCPS; V and G codes); Procedure (CPT); Diagnosis Group (DRG – Hospital); Drug type/dose/manufacturer (NDC; J codes); lab test (LOINC); Place of service, type of provider, etc. etc.

As we have seen with the simple diabetes example, the identification of codes can be time-consuming.

Identification Algorithms and pre-defined "Grouper" models sort-through the raw material and consolidate it into manageable like categories.

Risk Groupers (examples: DCGs; HCCs; ACGs; CRGs; CDPS) do some of this work for you. We will look at them more closely in Lecture 3.

Risk (Condition) Groupers

Risk Groupers are associated with a close relative of Predictive Modeling, Risk Adjustment. Risk Adjustment has been practiced for many years in the Medicare Advantage and many managed Medicaid programs, but not Commercial Insurance. This will change in 2014 when State-based exchanges offer small group (< 50 lives) and individual purchasers the opportunity to compare and purchase insurance from a single source.

Collectivizing insurance purchasing like this gives government the opportunity to control all premiums and to re-distribute it to participating insurers, to better match the risks that they write.

If there is time, we will look at the operation of risk adjustment in the Massachusetts Exchange (Connector) later in this seminar. For now, let me try to differentiate between Risk Adjustment and Predictive Modeling.

Risk Adjustment vs. Predictive Modeling

The same techniques and tools are used for both RA and PM: both use underlying risk factors and diagnoses to predict the future utilization trajectory of at-risk members (remember Lecture 1?)

Predictive Modeling identifies the stratum into which a member may fall in the future (and the member's likely cost). Health plans may use this information to design and deliver a program aimed at changing the member's behavior and reducing the predicted utilization.

Risk Adjustment comes in two flavors: Concurrent and Prospective. Both provide a *normalized* cost, to compare with the *actual* cost of a population, providing a basis on which to move premiums between plans. Prospective Risk Adjustment is similar to Predictive Modeling: based on the member's historic risk factors a projected cost is estimated for the coming year. Concurrent Risk Adjustment looks back at the end of the year just completed and calculates a normalized cost for the member, based on the member's actual risk factors.

Risk Adjustment vs. Predictive Modeling

Risk adjustment (basic concept) revenue transfer

1. Calculate individual member risk scores.
2. Calculate weighted average risk scores for the plan (assume 1.10)
3. Calculate weighted average risk scores for the State (assume this is normalized to 1.0 for simplicity).
4. Baseline premium (average for the State): $1,000.
5. Plan net claims cost: $1,250.
6. Plan adjustment (Plan Relative Risk Score – 1)(Baseline Premium) = (0.10)($1,000) = $100.
7. Premium adjustment: + $100 = $1,100. In this case, although premium is transferred, it is insufficient to offset all additional claims.

- Under the ACA, transfers will be revenue neutral.

Risk Adjustment vs. Predictive Modeling

To understand Risk Adjustment, let's use the previous example:

Applying Risk Scores						
Mem	Age	Sex	Condition(s)	Actual Cost	Standardized Cost	Risk Score
1	45	F	None	$2,864	$4,842	0.59
2	45	M	Diabetes	$5,024	$2,547	1.97
3	40	M	Diabetes + CAD	$23,479	$2,547	9.22
4	40	M	CAD	$18,185	$2,547	7.14
5	60	F	Breast Ca. + other	$15,935	$6,346	2.51
6-25			All other members	$2,500	$3,500	0.39
			TOTAL	$115,487	$88,829	**1.17**

This group is expected to have costs about 1.17 times the average for a group with the same age/sex distribution. Assuming that this plan collected only 100% of the required premium (standardized cost), the Exchange authority will transfer additional premium equal to 17% of the basic premium ($15,100). Not nearly sufficient to offset actual costs.

Risk Adjustment vs. Predictive Modeling

The prior example points up the need for predictive modeling, even when Risk Adjustment is applied to a population. In our example, Risk Adjustment accounted for $15,100 of the difference in claims between the standardized and actual claims. The balance (about $14,558) is due to our high-risk members experiencing claims in excess of their predicted level.

If we knew which members were likely to have high claims next year we could try to moderate that utilization, reducing the excess. We could use the risk scores generated by the Risk Adjuster model to risk-rank our members; we would then try to manage Members 3 and 4 (CAD+Diabetes and CAD), and possibly other members, if we have the resources.

The use of Risk Adjuster model is not the only way to identify high risk members, as we shall see later.

All people are not equally identifiable

An important issue with any claims-based identification algorithm is that you are *imputing,* rather than *observing* a diagnosis. Thus you are always at risk of including false positives, or excluding false negatives, from the analysis.

One consequence of using a grouper model is that you are at the mercy of the modeler's definition of diagnoses, and thus cannot control for false positives or negatives. **An important draw-back of typical Grouper models is that they assign a diagnosis based on a *single instance of the diagnosis.* They are therefore more sensitive than specific.** How does this work in practice? Let's look at some data.

Prevalence of Chronic Conditions Identified Using Different Claims Algorithms				
	Number of Claiming Events in the Year			
Condition	4 or more	3 or more	2 or more	1 or more
Asthma	2.4%	2.9%	3.9%	6.1%
Cardiovascular disease	0.8%	1.2%	1.7%	2.8%
Heart Failure	0.2%	0.2%	0.3%	0.6%
Pulmonary Disease	0.2%	0.3%	0.5%	1.0%
Diabetes	3.3%	3.7%	4.1%	4.9%
All	6.3%	7.4%	9.2%	13.1%

All people are not equally identifiable (2)

A less-rigorous algorithm will identify more people with the condition (more than twice as many in the example above). But it runs the risk of sweeping in false positives. This table shows the likelihood re-qualifying with the condition in the following year (remember that these are members with Chronic conditions that (theoretically) are permanent):

Probability that a Member Identified with Chronic Condition in Year 1 will be Identified with that Condition in Year 2				
All Chronic Conditions				
No. Claiming Events in Year 2	Number of Claiming Events in Year 1			
	4 or more	3 or more	2 or more	1 or more
4 or more	59.7%	26.3%	15.7%	7.2%
3 or more	65.8%	35.9%	22.9%	10.6%
2 or more	72.0%	47.9%	34.3%	17.2%
1 or more	78.0%	62.3%	49.9%	30.9%
Do not re-qualify	22.0%	37.7%	50.1%	69.1%

All people are not equally identifiable (3)

Why is the issue of sensitivity and specificity important?

For Predictive Modeling, we may not mind much if we use a sensitive algorithm and identify members who may not have the condition; if members are followed up by a program, their true condition can be determined.

For Risk Adjustment, however, specificity matters. Medicare's Risk Adjustment process has given rise to an industry that finds additional diagnoses (that drive the risk score). Increases in the Medicare Advantage plan's average risk score will increase its revenue. Interestingly, unlike Medicare Advantage plans, the proposed State Exchange Risk Adjustment is a zero-sum game: if another plan increases its average risk score, your plan will lose revenue. With the typical risk adjuster assigning a diagnosis based on a single instance of a diagnosis, it may be possible for aggressive plans to "game" the risk adjustment process.

Algorithm Development: Diabetes Example

Example of an identification algorithm:

Example of a Definitional Algorithm			
Disease	**Type**	**Frequency**	**Codes**
Diabetes Mellitus	Hospital Admission or ER visit with diagnosis of diabetes in any position	At least one event in a 12-month period	ICD-9 codes 250, 357.2, 362.0, 366.41, 648.0
	Professional visits with a primary or secondary diagnosis of diabetes	At least 2 visits in a twelve month period	CPT Codes in range of 99200-99499 series E & M codes or 92 series for eye visits
	Outpatient Drugs: dispensed insulin, hypoglycemic, or anti-hyperglycemic prescription drug	Three or more prescriptions in a twelve month period	Diabetes drugs (see HEDIS or similar list of drug codes).
EXCLUDE gestational diabetes.	Any (as above)	As above	648.8x

Sources of Algorithms

- NCQA – HEDIS.
- DMAA (Now CCA; Chronic definitions).
- Grouper Models.

Additionally, there has been an explosion of rules-based quality metric reporting in recent years. Just a few examples:

- PQRS (208 measures in 22 categories);
- STAR Measures (36 Medicare Advantage; 15 Medicare Part D);
- ACO quality reporting (33 measures);
- HEDIS Measures (75 measures in 8 domains).

All require risk-adjustment for their application.

Grouper Construction

Society of Actuaries Studies:

Note 1: the SOA tests both *Concurrent* (retrospective) and *Prospective* models. Concurrent model correlations tend to be higher.

Note 2: there are some issues with models that you should be aware of:

- They tend to be less accurate at the "extremes" (members with high or low risk scores);
- We have observed an inverse correlation between risk-score and $'s across a wide range of members.
- As we have discussed previously, "sensitive" models are open to inclusion of more false-positives.
- A well-managed patient who fails to develop a more serious form of a condition will have a lower risk score.

Commercial Groupers: SOA studies

The Society of Actuaries studies show:

1. Risk grouper modeling tools use *different algorithms* to group the source data. For example, the Symmetry models are built on episodes of care, DRGs are built on hospital episodes, while other models are built on diagnoses.

2. Similar performance among all leading risk groupers.

3. Accuracy of prediction has increased since the publication of the original study. In part, this is due to more accurate coding and the inclusion of more claims codes.

4. Risk groupers use *relatively limited data* sources (e.g. DCG and Rx Groups use ICD-9 and NDC codes but not lab results or HRA information).

5. Accuracy of **retrospective (concurrent)** models is now in the 30%-40% R^2 range. **Prospective** model accuracy is in the range of 20% to 25%.

A note about Prospective and Concurrent Models

Both have their place. Neither is perfect.

1.Concurrent models are also called Retrospective.

The concurrent model is used to reproduce **actual historical costs.** This type of model is used for assessing relative resource use and for determining compensation to providers for services rendered because it normalizes costs across different members. Normatively, the concurrent model provides an assessment of what costs *should* have been for members, given the health conditions with which they presented in the past year. It is also used in program evaluation, which is performed once all known conditions may be identified.

2.The Prospective model predicts what costs w*ill* be for a group of members in the future. The Prospective model is predicting the unknown, because the period over which the prediction is made lies in the future. The Concurrent model, by contrast, provides an estimate of normalized costs for services that have already occurred. For prospective prediction, members with no claims receive a relative risk score component based on age/sex alone.

A note about Prospective and Concurrent Models

Concurrent models have the advantage that they represent all the known information about the member in the completed year. However, when they are used to compensate providers (for example) for managing a group of members, there is a risk to the provider that if the provider does a good job and prevents the exacerbation of the member's condition, the member risk score (and therefore the provider's compensation) will be lower than it would be if the provider does not prevent the exacerbation.

Prospective models are often used to allocate revenue to different managed care plans. The drawback to this approach is that members' prospective risk scores are based on historical data, and do not take account of developing (incident) conditions that emerge during the year.

Commercially-available Risk Groupers

Commercially Available Grouper Models		
Developer	**Risk Grouper**	**Data Source**
CMS	Diagnostic Risk Groups (DRG) (There are a number of subsequent "refinements" to the original DRG model.	Hospital claims only
CMS	HCCs	Age/Sex, ICD -9
3M	Clinical Risk Groups (CRG)	All Claims (inpatient, ambulatory and drug)
IHCIS/Ingenix	Impact Pro	Age/Sex, ICD-9 NDC, Lab
UC San Diego	Chronic disability payment system Medicaid Rx	Age/Sex, ICD -9 NDC
Verisk Sightlines™	DCG RxGroup	Age/Sex, ICD -9 Age/Sex, NDC
Symmetry/Ingenix	Episode Risk Groups (ERG) Pharmacy Risk Groups (PRG)	ICD – 9, NDC NDC
Symmetry/Ingenix	Episode Treatment Groups (ETG)	ICD – 9, NDC
Johns Hopkins	Adjusted Clinical Groups (ACG)	Age/Sex, ICD – 9

Grouper Algorithms

As an alternative to commercially-available risk groupers, analysts can develop their own models using common data mining techniques. Each method has its pros and cons:

There is a considerable amount of work involved in building algorithms from scratch, particularly when this has to be done for the entire spectrum of diseases. Adding drug or laboratory sources to the available data increases the complexity of development.

While the *development* of a model may be within the scope and resources of the analyst who is performing research, use of models for production purposes (for risk adjustment of payments to a health plan or provider groups for example) requires that a model be maintained to accommodate new codes. New medical codes are not published frequently, but new drug codes are released monthly, so a model that relies on drug codes will soon be out of date unless updated regularly.

Commercially-available clinical grouper models are used extensively for risk adjustment when a consistent model, accessible to many users, is required. Providers and plans, whose financial stability relies on payments from a payer, often require that payments be made according to a model that is available for review and validation.

Grouper Algorithms

An analyst that builds his own algorithm for risk prediction has control over several factors that are not controllable with commercial models:

1.Which codes, out of the large number of available codes to recognize. The numbers of codes and their redundancy (the same code will often be repeated numerous times in a member record) makes it essential to develop an aggregation or summarization scheme.

2.The level at which to recognize the condition. How many different levels of severity should be recognized? . The analyst will also need to determine how to handle different levels of severity of the same diagnosis: should each be included, or should the higher-severity code “trump” the lower-severity code?

3.The impact of co-morbidities. Some conditions are often found together (for example heart disease with diabetes). The analyst will need to decide whether to maintain separate conditions and then combine where appropriate, or to create combinations of conditions.

Grouper Algorithms

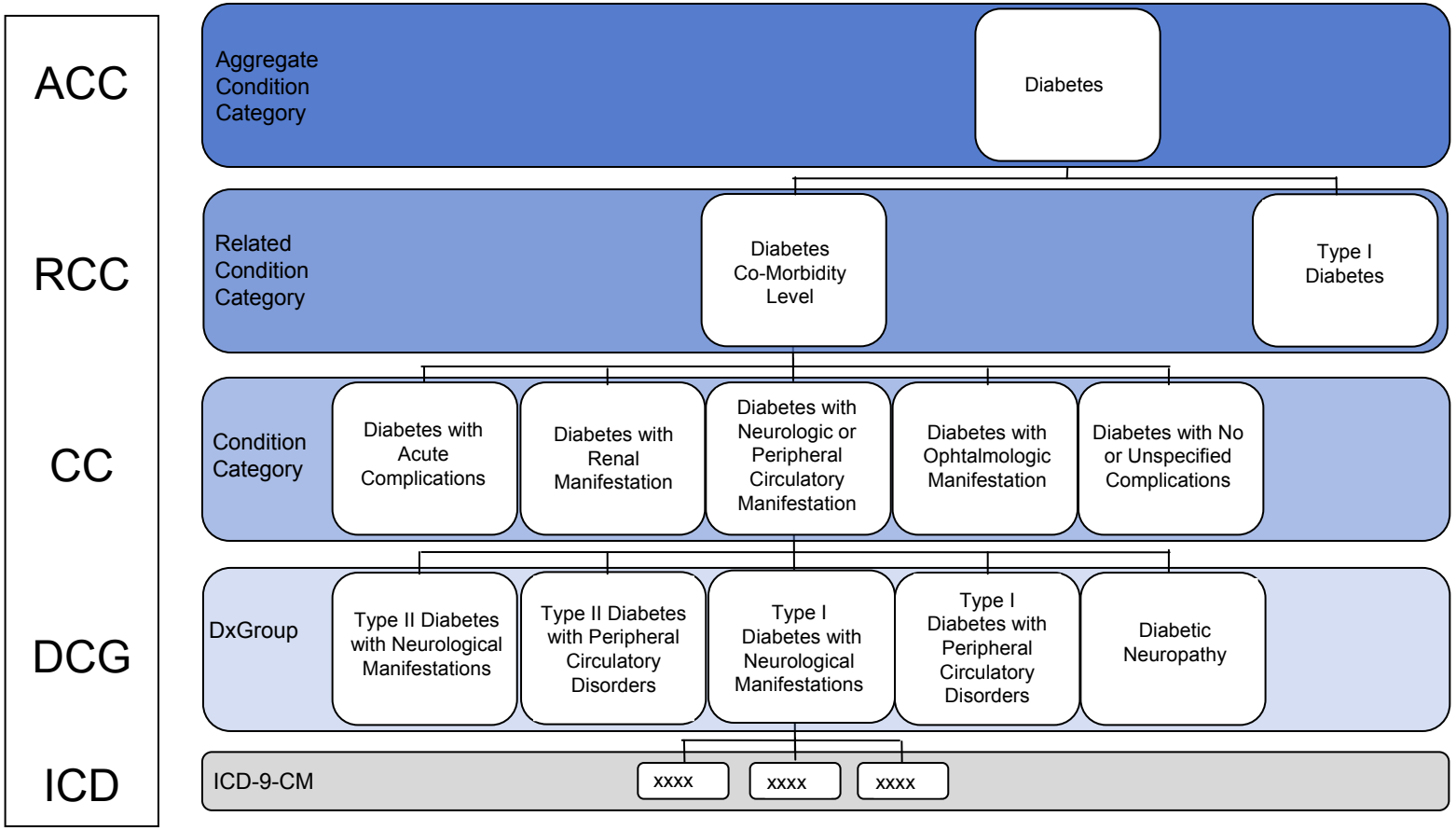
An analyst that builds his own algorithm for risk prediction has control over several factors that are not controllable with commercial models:

4.The degree of certainty with which the diagnosis has been identified (confirmatory information). (We have already seen the implication of this issue when looking at the relationship between sensitivity/specificity and the number of discrete instances of a diagnosis in the member record.) The accuracy of a diagnosis may differ based on who codes the diagnosis, for what purpose and how frequently a diagnosis code appears in the member record. The more frequently a diagnosis code appears, the more reliable the interpretation of the diagnosis. Similarly, the source of the code (hospital, physician, laboratory) will also affect the reliability of the diagnostic interpretation.

5.Data may come from different sources with a range of reliability and acquisition cost. A diagnosis in a medical record, assigned by a physician, will generally be highly reliable. Other types of data are not always available or as reliable.

Example of Grouper Construction

Grouper models are constructed in a similar fashion to that illustrated above. Below we show the hierarchical structure of the DxCG model for Diabetes:



Example of Grouper Construction

Grouper models are constructed in a similar fashion to that illustrated above. Below we show how the risk score is developed for a patient with diagnoses of Diabetes, HTN, CHF and Drug Dependence, illustrating the hierarchical and additive structure of the DxCG model:

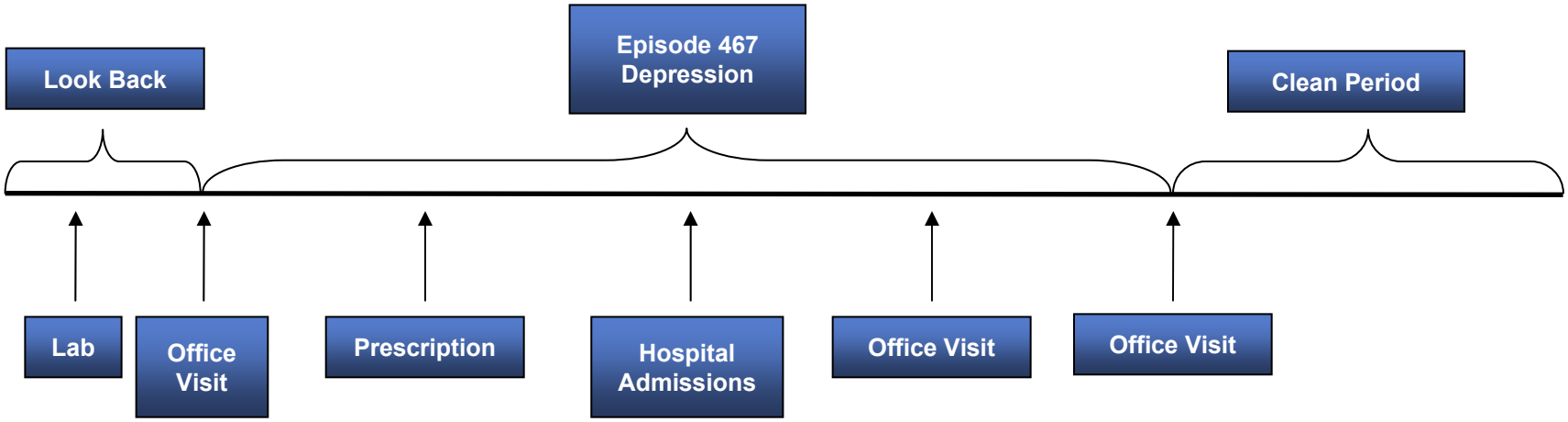
Example of Construction of a Relative Risk Score		
Condition Category	**Risk Score Contribution**	**Notes**
Diabetes with No or Unspecified Complications	0.0	Trumped by Diabetes with Renal Manifestation
Diabetes with Renal Manifestation	2.1	
Hypertension	0.0	Trumped by CHF
Congestive Heart Failure (CHF)	1.5	
Drug Dependence	0.6	
Age-Sex	0.4	
Total Risk Score	**4.6**	

Different Approach to Model Construction:

- Episode Groupers represent a different approach to model construction. Instead of a model being constructed around diagnoses, it is constructed around episodes of care. These episodes are often acute (surgery, for example) but may also be chronic or preventive.
- In the case of an acute episode, the episode grouper can be thought of as a normative tool, in that it specifies all the different services that make up the episode. For example, it may start with an office visit, followed by diagnostic testing, then inpatient admission for surgery, rehabilitation, and additional follow up visits for check-ups.
- The episode approach clearly had the advantage that it specifies the services typically ordered in the case of a specific treatment, thus enabling us to compare the utilization patterns of physicians treating the same illness. The utilization of services associated with different episodes can also be translated into a relative risk score.
- While Episode Groupers handle acute episodes well, for chronic illnesses (where there is no "episode," or the episode could be the year of treatment) the approach is very similar to condition-based models.

Episode-based Groupers

An example of an episode Group: the Symmetry Grouper.



Episode-based Groupers

Application of the Symmetry Grouper. Risk Scores are developed similarly to DxCG.

Construction of Relative Risk Scores Using ETGs						
Example: Male Aged 58						
ETG	Severity Level	Description	ERG	Description	Retrospective Risk Weights	Prospective Risk Weights
163000	2	Diabetes	2.022	Diabetes w/significant complication/co-morbidity I	0.9874	1.2810
386800	1	Congestive Heart Failure	8.043	Ischemic heart disease, heart failure, cardiomyopathy III	2.2870	2.0065
238800	3	Mood Disorder, Depression	4.033	Mood disorder, depression w/ significant cc/cb	0.8200	0.7913
473800	3	Ulcer	11.022	Other moderate cost gastroenterology II	2.3972	0.6474
666700	1	Acne	17.011	Lower cost dermatology I	0.1409	0.1023
666700	1	Acne	17.011	Lower cost dermatology I		
Demographic risk: Male 55-64						0.7331
					6.6325	5.5616

One more very useful grouper…

Drug groupers group 100,000s NDC codes into manageable therapeutic classes

Example of Therapeutic Classes Within the GPI Structure					
Group	**Class**	**Sub Class**	**Group**	**Class**	**Sub Class**
GROUPS 1- 16 ANTI-INFECTIVE AGENTS					
01	00	00	***PENICILLINS***		
01	10	00		Penicillin G	
01	30	00		PENICILLINASE -RESISTANT PENICILLINS	
01	50	00		AMINO PENICILLINS/BROAD SPECTRUM PENICILLINS	
01	20	00		Ampicillins	
01	40	00		EXTENDED SPECTRUM PENICILLINS	
01	99	00		*Penicillin Combinations**	
01	99	50			*Penicillin-Arninoglycoside Combinations***
01	99	40			*Penicillin-NSAIA Combinations***
02	00	00	***CEPHALOSPORINS***		
02	10	00		*Cephalosporins -1st Generation**	
02	20	00		*Cephalosporins -2nd Generation**	
02	30	00		*Cephalosporins -3rd Generation**	
02	40	00		*Cephalosporins -4th Generation**	
02	99	00		*Cephalosporin Combinations**	
03	00	00	***MACROLIDE ANTIBIOTICS***		
03	10	00		*Erythromycins**	
03	10	99			*Erythromycin Combinations***
03	20	00		*Troleandomycin**	
03	30	00		*Lincomycins**	
03	40	00		*Azithromycin**	
03	50	00		*Clarithromycin**	
03	52	00		*Dirithromycin**	
Etc.	Etc.	Etc.			

Rules-based vs. Statistical Models

What about rules-based models?

1. First, all models ultimately have to be converted to rules to apply in an operational setting.
2. What most people mean by "rules-based models" is actually a "Delphi*" approach. For example, application of "Gaps-in-care" or clinical rules (e.g. ActiveHealth).
3. Rules-based models have their place in Medical Management. One challenge, however, is risk-ranking identified targets, particularly when combined with statistical models.

* Meaning that experts, rather than statistics, determine the risk factors.

On the predictive ability of experts vs. the machine, see next slide!

Providers are not good at predicting risk*

- Researchers from the VA System assessed the predictions made by
  - Physicians
  - Case managers
  - Nurses
- *"...none of the AUC values were statistically different from chance"*

* Although they believe that they are.

Allaudeen N, Schnipper JL, Orav EJ, Wachter RM, Vidyarthi AR. Inability of providers to predict unplanned readmissions. J Gen Intern Med. 2011;26(7):771-6

Inability of Providers to Predict Unplanned Readmissions

Nazima Allaudeen, MD[1,2], Jeffrey L. Schnipper, MD, MPH[3], E. John Orav, PhD[4], Robert M. Wachter, MD[2], and Arpana R. Vidyarthi, MD[2]

[1]Department of Medicine, VA Palo Alto Healthcare System, Palo Alto, CA, USA; [2]Division of Hospital Medicine, Department of Medicine, University of California, San Francisco, CA, USA; [3]BWH Academic Hospitalist Service and Division of General Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; [4]Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA.

BACKGROUND: Readmissions cause significant distress to patients and considerable financial costs. Identifying hospitalized patients at high risk for readmission is an important strategy in reducing readmissions. We aimed to evaluate how well physicians, case managers, and nurses can predict whether their older patients will be readmitted and to compare their predictions to a standardized risk tool (Probability of Repeat Admission, or P_{ra}).

METHODS: Patients aged ≥65 discharged from the general medical service at University of California, San Francisco Medical Center, a 550-bed tertiary care academic medical center, were eligible for enrollment over a 5-week period. At the time of discharge, the inpatient team members caring for each patient estimated the chance of unscheduled readmission within 30 days and predicted the reason for potential readmission. We also calculated the P_{ra} for each patient. We identified readmissions through electronic medical record (EMR) review and phone calls with patients/caregivers. Discrimination was determined by creating ROC curves for each provider group and the P_{ra}.

RESULTS: One hundred sixty-four patients were eligible for enrollment. Of these patients, five died during the 30-day period post-discharge. Of the remaining 159 patients, 52 patients (32.7%) were readmitted. Mean readmission predictions for the physician providers were closest to the actual readmission rate, while case managers, nurses, and the P_{ra} all overestimated readmissions. The ability to discriminate between readmissions and non-readmissions was poor for all provider groups and the P_{ra} (AUC from 0.50 for case managers to 0.59 for interns, 0.56 for P_{ra}). None of the provider groups predicted the reason for readmission with accuracy.

CONCLUSIONS: This study found (1) overall readmission rates were higher than previously reported, possibly because we employed a more thorough follow-up methodology, and (2) neither providers nor a published algorithm were able to accurately predict which patients were at highest risk of readmission. Amid increasing pressure to reduce readmission rates, hospitals do not have accurate predictive tools to guide their efforts.

Received August 27, 2010
Revised January 27, 2011
Accepted January 28, 2011
Published online March 12, 2011

KEY WORDS: readmission; unplanned; prediction.
J Gen Intern Med (26)7:771-6
DOI: 10.1007/s11606-011-1663-3

BACKGROUND

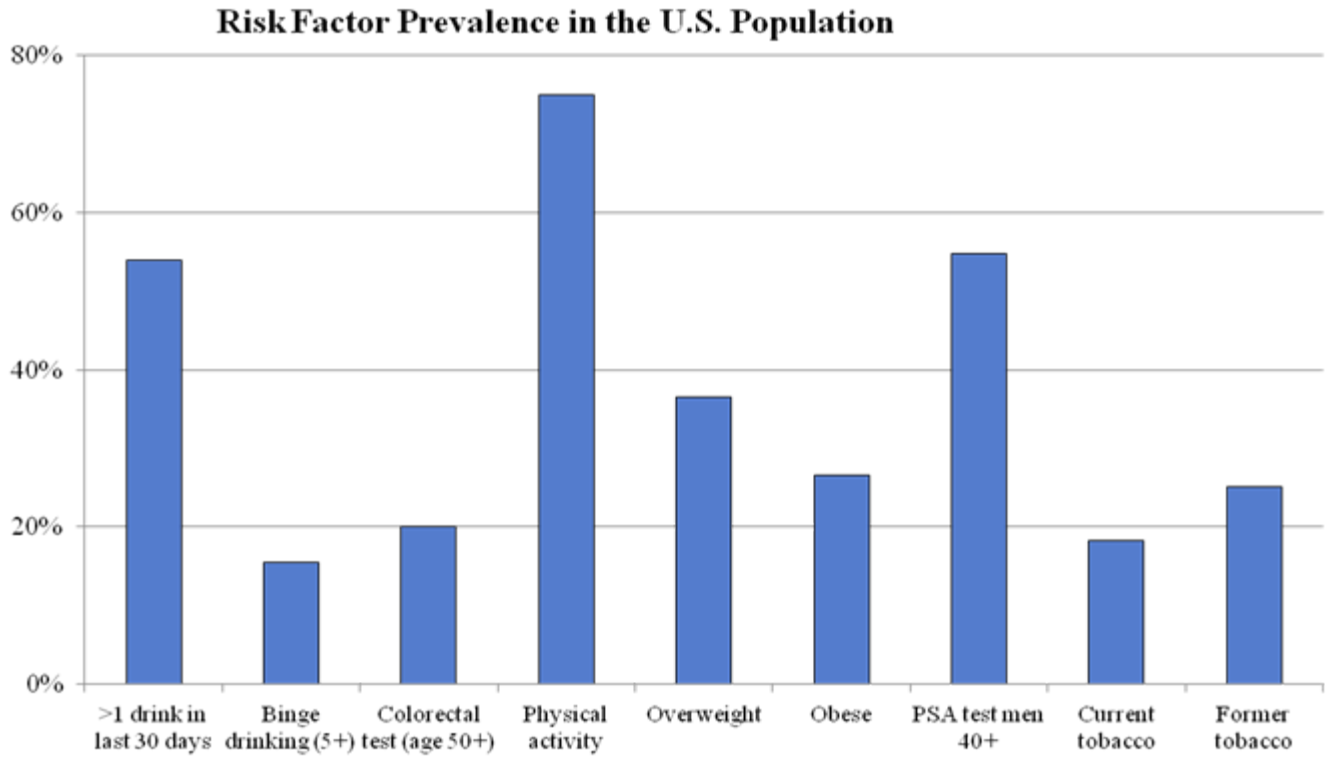
Against the background of rising concerns about both the cost and quality of American medical care, hospital readmissions have come under increasing scrutiny from both outside and within the government.[1–3] Hospital readmissions may be a marker for poor quality care, are disunifying for patients and families, and increase health care costs. Medicare estimates that $15 billion is spent on the 17.6% of patients who are readmitted within 30 days.[4]

Although it would be ideal to develop interventions that improve the hospital-to-home transition for all patients, given limited resources, some have argued for targeting intense efforts—such as comprehensive discharge planning, post-discharge phone calls or home visits, and early clinic visits—towards high risk patients. However, such strategies require that we have accurate methods to identify patients at highest risk.

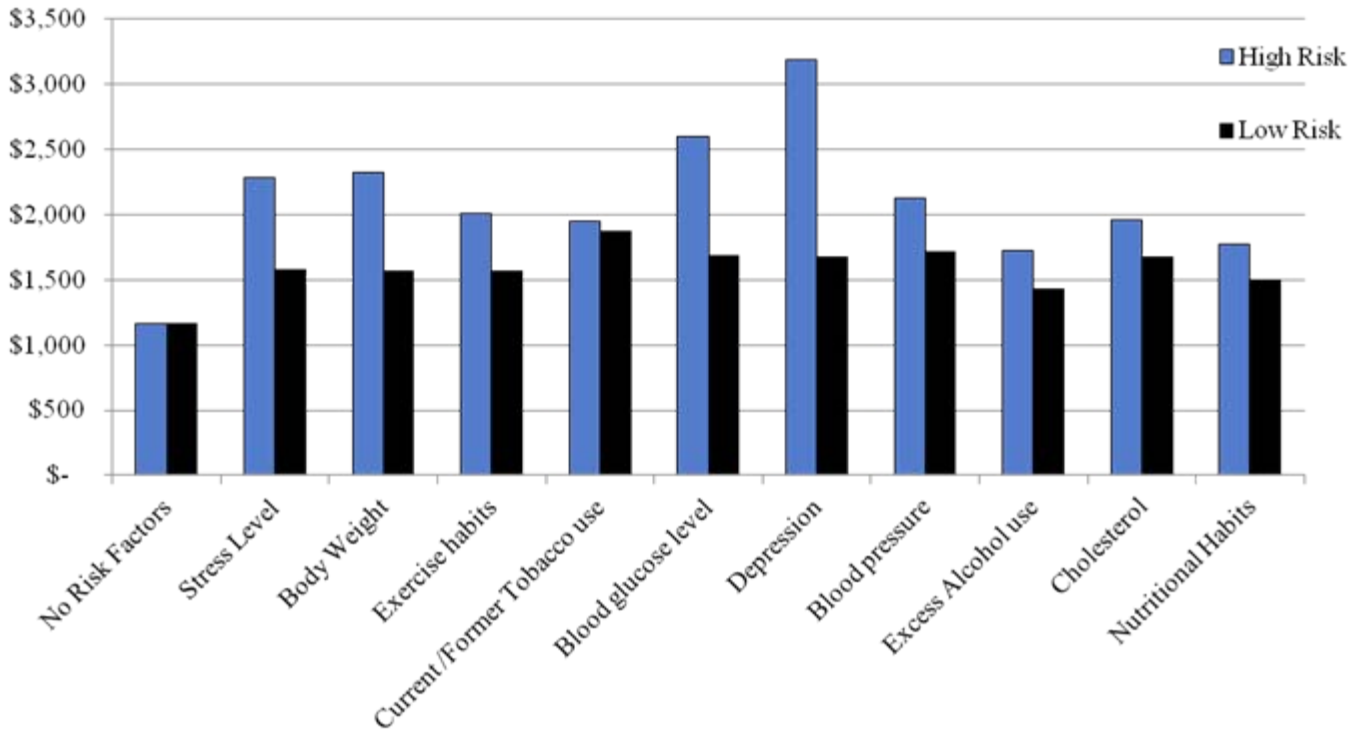
Anecdotal evidence suggests that inpatient providers (physicians, nurses, discharge planners) currently make informal predictions of readmission that affect discharge planning. Such predictions are not new: providers have tried to predict other outcomes, such as mortality and length of stay, in several settings (e.g., intensive care unit, emergency department), with varying success.[5–10] However, the accuracy of informal predictions of hospital readmission is unknown. Several algorithms have also been developed in recent years to predict hospital readmissions, but their use has been limited, because they require information not typically gathered during clinical care, their models are complex and difficult to use, and/or because they are not accurate. A few studies have compared providers with algorithm-based tools to predict readmission and mortality in other settings,[9] but it remains unknown how well providers' predictions of readmission for general medicine patients compare with published algorithms or how the predictions of multiple disciplines compare with one another.

To reach the ultimate goal of preventing readmissions, identifying the highest risk patients is the first of a multistep process. Providers would next need to speculate on the reason for readmission before then targeting an effective

Don’t overlook non-condition-based Risk

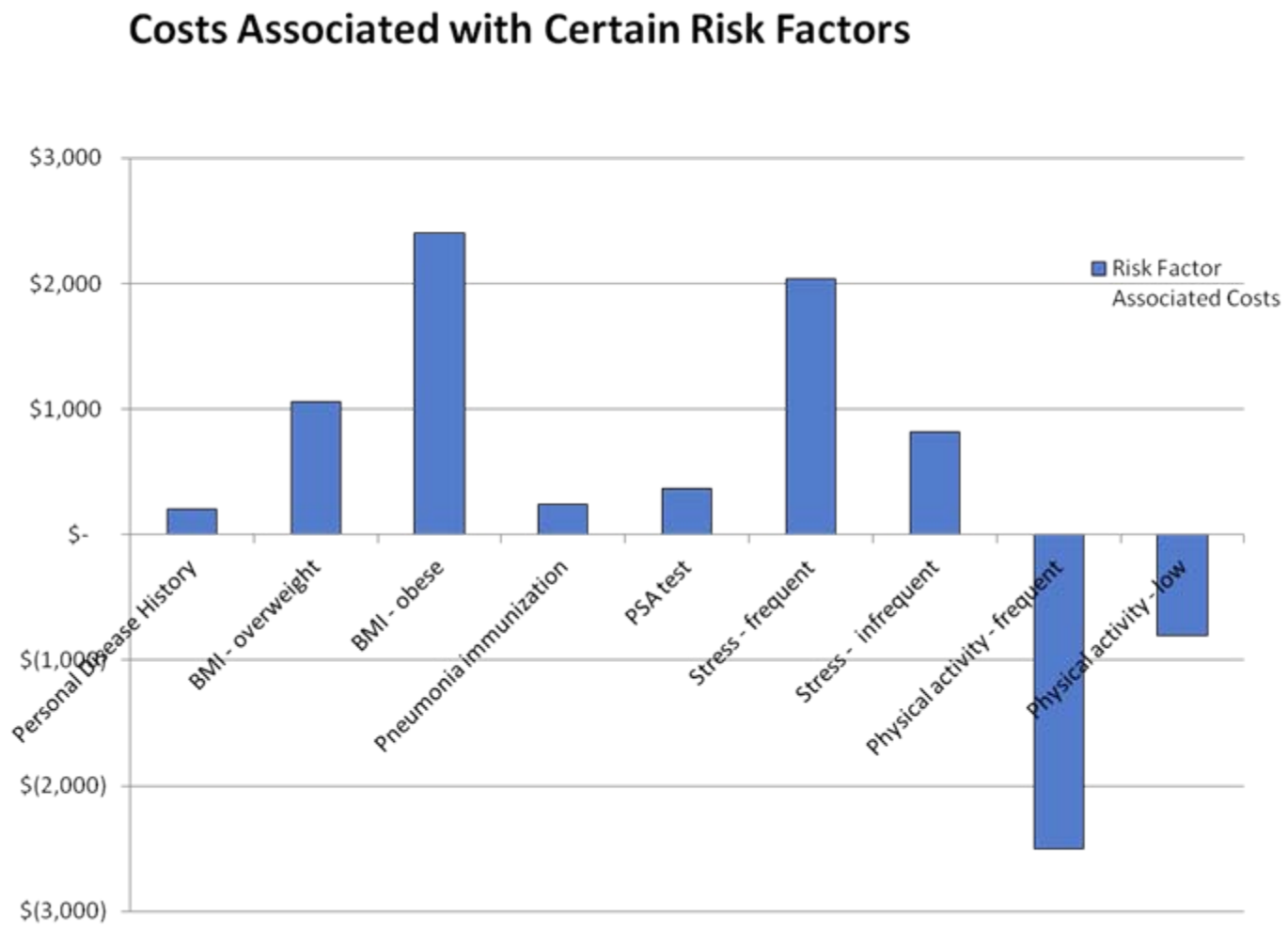


Non-condition-based Risk



Relationship Between Modifiable Health Risk Factors and Annual Health Insurance Claims

Non-condition-based Risk



On the non-linearity of risk adjusters

1. Risk adjustment (e.g. in the ACA context) depends on a linearity assumption for the underlying risk adjuster. *Meaning that:*
2. There is a linear relationship between relative risk score and cost.
3. If this hypothesis is true, then a 1% increase in relative risk score should be matched with a 1% increase in income transfer. So for a plan with a relative risk score of 1.05, a 1% increase would result in a next-year relative risk score of 1.0605. This increase in turn translates into a 1% income transfer.
4. But what if the relationship is non-linear? If, for example, higher risk members cost proportionally more than average (or lower) risk members?

On the non-linearity of risk adjusters

5. Consider the following simple model:

Members	Risk Score	Cost/ member	Risk Score	Cost/ member
900	1.05	1,050.00	1.040	1,040.00
100	1.05	1,050.00	1.245	1,245.00
1,000	1.05	1,050.00	1.0605	1,060.50

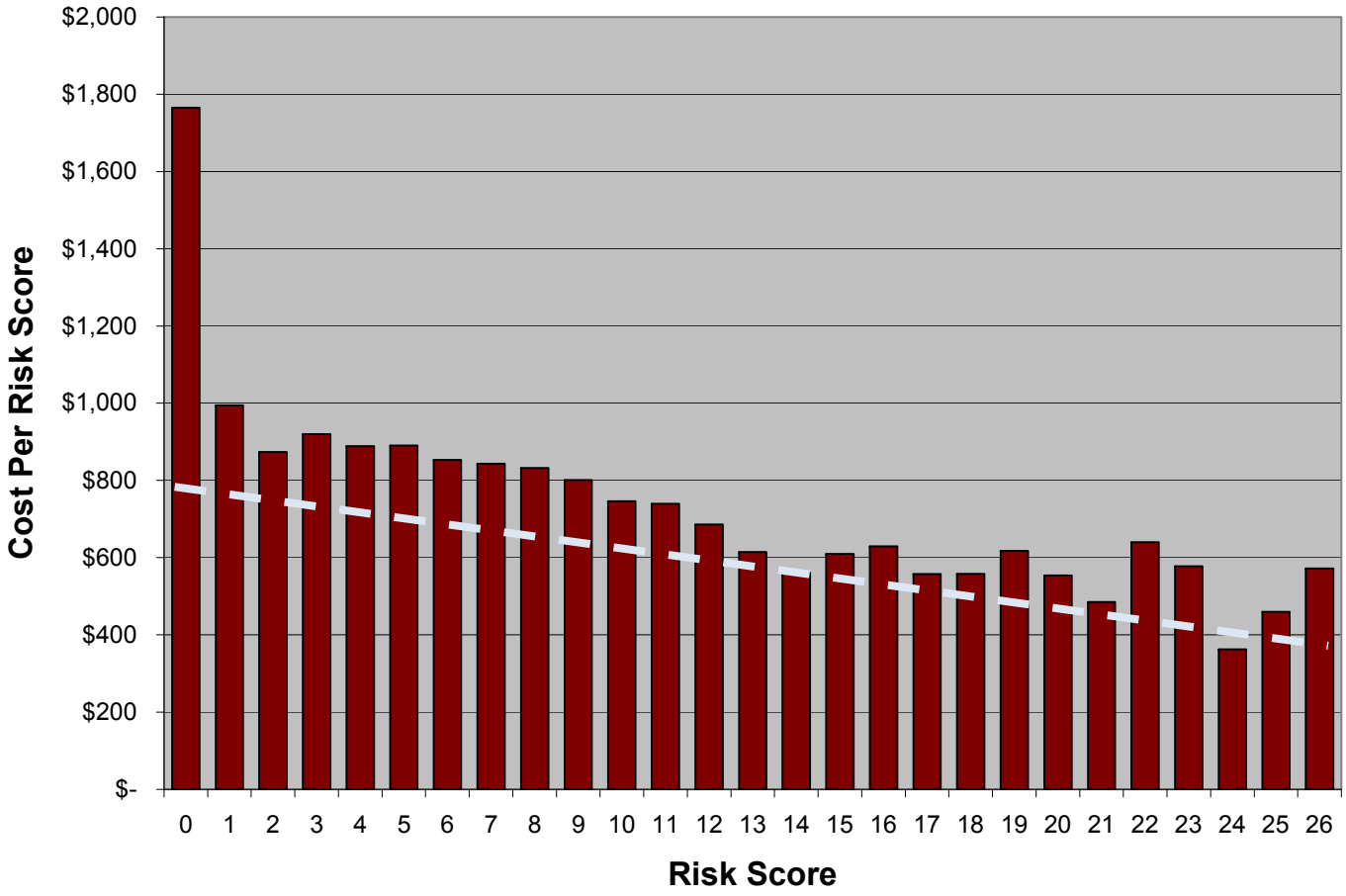
In this case cost and risk is linearly related at all points in the distribution.

If, however, high cost members cost proportionally more, say $1,500 per member, then average risk score in year 2 would be 1.0605, reimbursement would be $1,060.50 but cost would be $1,086.00.

Reimbursement at the average rate (based on risk score 1.0605) would under-reimburse the plan for its high-risk members.

On the non-linearity of risk adjusters

Some real-world Risk/Cost data



Lecture 4: Model Construction

"All models are wrong, but some are useful." George E.P. Box.

(Professor of statistics at the University of Wisconsin –Madison)

In this lecture we will look (briefly) at model construction. This is a very significant topic, worthy of its own course. We will just skim the surface.

We will look at some common tools that are used for model building and then follow a practical example. An example that fits Box's epigram is linear regression.

Lecture 4: Model Construction

Following section is taken from our paper (under review)

Predictive Modeling of Healthcare Costs using Hierarchical Regression

I. Duncan M. Loginov M. Ludkovski

September 20, 2014

Abstract

Predictive models of healthcare costs have become mainstream in much healthcare actuarial work. The Affordable Care Act requires the use of predictive modeling-based risk-adjuster models to transfer revenue between different health exchange participants. While the predictive accuracy of these models has been investigated in a number of studies, the accuracy and use of models for applications other than risk adjustment has not been the subject of much investigation. We investigate predictive modeling of future healthcare costs using a number of different statistical techniques. Our analysis was performed based on a dataset of 30,000 insureds containing claims information from two contiguous years. The dataset contains over a hundred covariates for each insured, including detailed breakdown of past costs and causes encoded via coexisting condition (CC) flags. We discuss statistical models for the relationship between next-year costs and medical and cost information to predict the mean and quantiles of future cost, ranking risks and identifying most predictive covariates. A comparison of multiple models is presented, including (in addition to the traditional linear regression model underlying risk adjusters) Lasso GLM, multivariate adaptive regression splines, random forests, decision trees, and boosted trees. A detailed performance analysis shows that the traditional regression approach does not perform well and that more accurate models are possible.

JEL Classification: I13, C53

Developing your own model

As we have seen, there are pros and cons to using a commercially-available model. There is no reason not to develop your own predictive or risk adjustment model. Chapters 7-12 of my book address different statistical models that are encountered in this work. Examples of statistical models used frequently:

- Linear Regression. Advantage: everyone understands this.
- Generalized linear model and Logistic regression: more sophisticated models often used for healthcare data.
- Tree models: more difficult to apply operationally than regression models.

In addition, we will briefly touch on some other methods that Mike Ludkovski, Michael Loginov and I (UCSB) include in our paper* reporting on results of applying different models to a claims dataset.

* *Predictive Modeling of Healthcare Costs using Hierarchical Regression"* Under review by NAAJ.

Regression

Basic Model:

Response $Y = A + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$

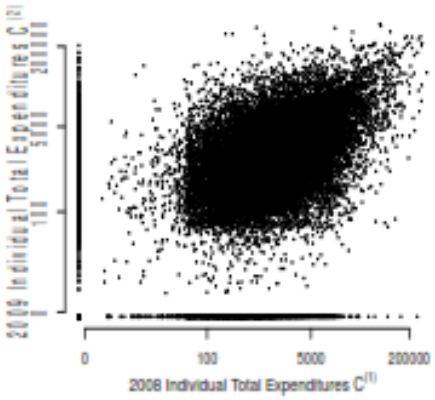
- Simple model, well-understood.
- Parameters supplied by analyst.
- Various variable transformations can be made to reduce model to a linear form.

<u>Requirements for the linear model</u>

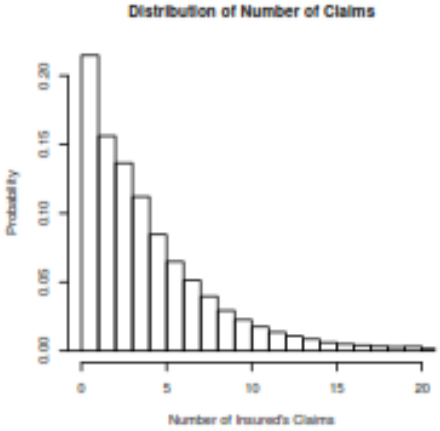
1. Valid model design and variable selection.
2. Additivity and linearity.
3. Independence of errors.
4. Equal variance of errors (homoscedasticity).
5. Normality of errors.

Regression – data evaluation

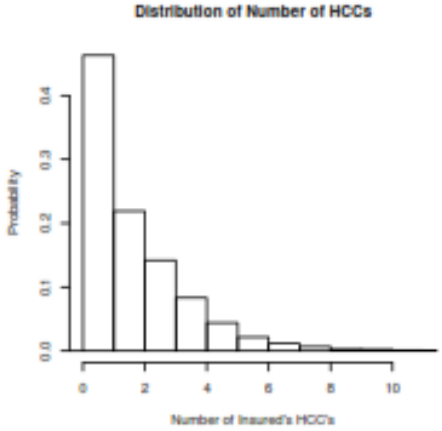
Data Features



COST



NUMCL



NUMHCC

Regression – Model Objectives

Basic Model:

- Predict Individual aggregate (sum of all services) next year expenditures.
- Classify insureds into risk groups.
- Identify predictive co-variates. (e.g. how useful is the diagnostic (HCC) information?)

Challenges:

- Many predictor variables (e.g. 83 HCCs). Many are highly correlated.
- Non-linear relationships (e.g. age).
- Impact is state dependent (e.g. having diabetes at 30 is not the same as having diabetes at 70).
- Parameterizing relationships is difficult.
- Errors are non-normally distributed.

Tested Linear model with some (manual) interaction terms; $R^2 = 0.17$.

Alternative Models

We considered two classes of solutions to the fit problem:

1.Regularized Regression (LASSO; MARS);

2.Tree models (Random Forest; Decision Trees; Boosted Trees).

We also considered (but did not pursue) Logistic Regression models because these models did not show improvement over OLS. See for example work of Jed Frees and Margie Rosenberg ("Predicting frequency and amount of health care expenditures;" NAAJ 15 (3) 2011. 377-392.)

Regularized Regression

Regularized Regression aims to improve the predictive accuracy of a model by jointly optimizing the Mean-squared Error and Model Complexity. It is possible (for example) to increase the "fit" of a model by adding more variables (at the risk of over-fitting). Efficiency requires the minimum number of predictor variables. The problem becomes a trade-off between Bias and Variance of the model.

We applied the LASSO model (**L**east **A**bsolute **S**hrinkage and **S**election **O**perator.)

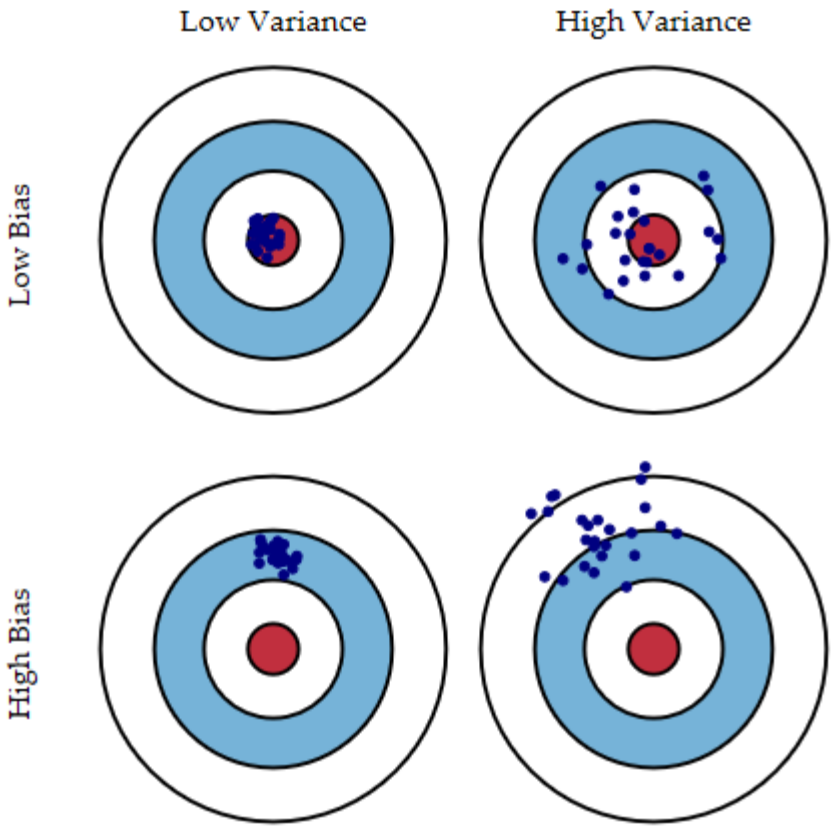


Fig. 1 Graphical illustration of bias and variance.

LASSO

LASSO successively removes variables from the regression while optimizing MSE. The surviving coefficients are then a proxy for the importance of each variable in the model.

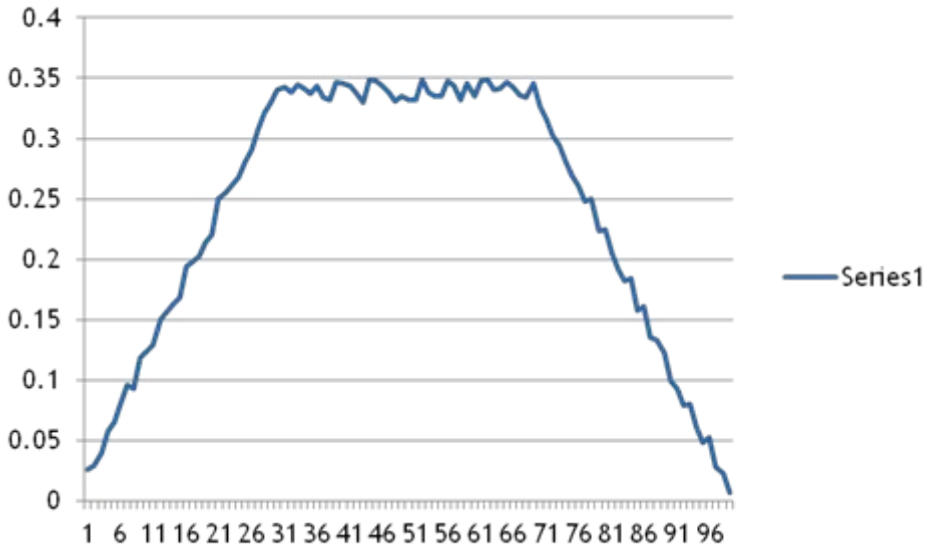
This method is helpful when (as in this case) there is high correlation between the predictor variables.

Applying the LASSO model to the data improves fit slightly; $R^2 = 0.197$.

MARS

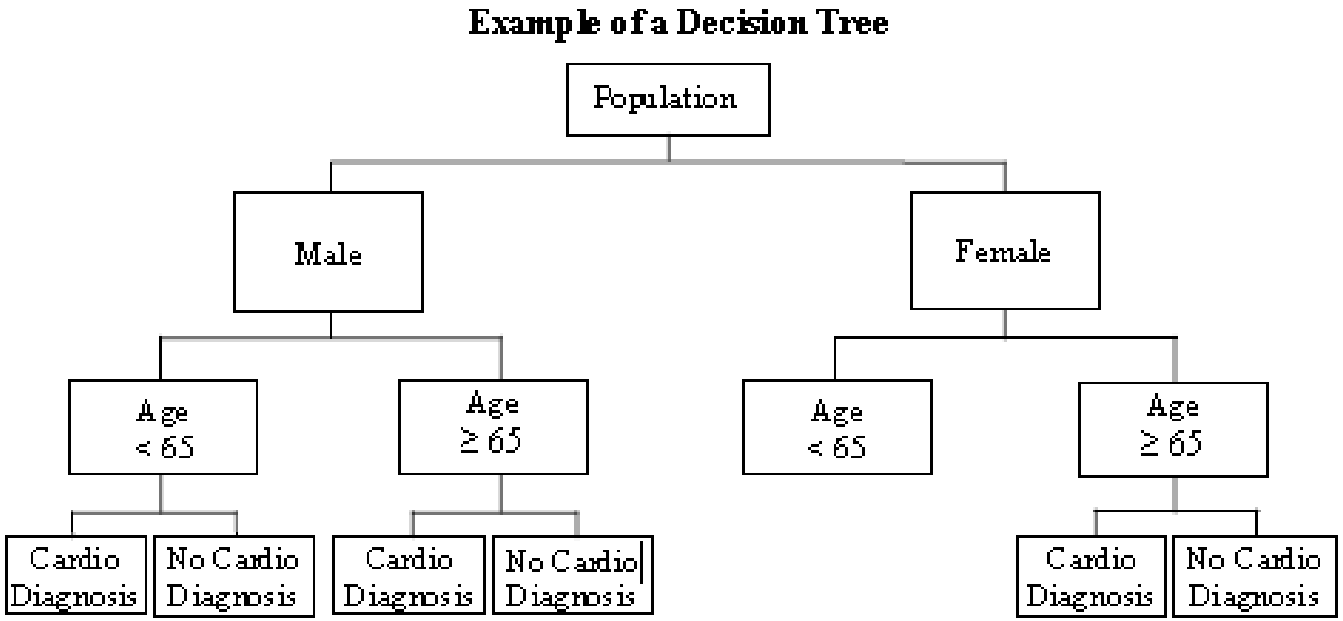
MARS stands for **M**ulti-variate **A**daptive **R**egression **S**plines. The method holds a lot of promise; it functions by dividing the model into a series of linear components, joined at points called “knots” or “hinges.”

An example of a distribution that would be difficult to model using OLS; however, MARS identifies the knots and then fits piece-wise functions between them. Despite its intuitive appeal, MARS was less-successful at fitting the healthcare data.



Some Other Tools: Decision Trees

Here is a simple example:



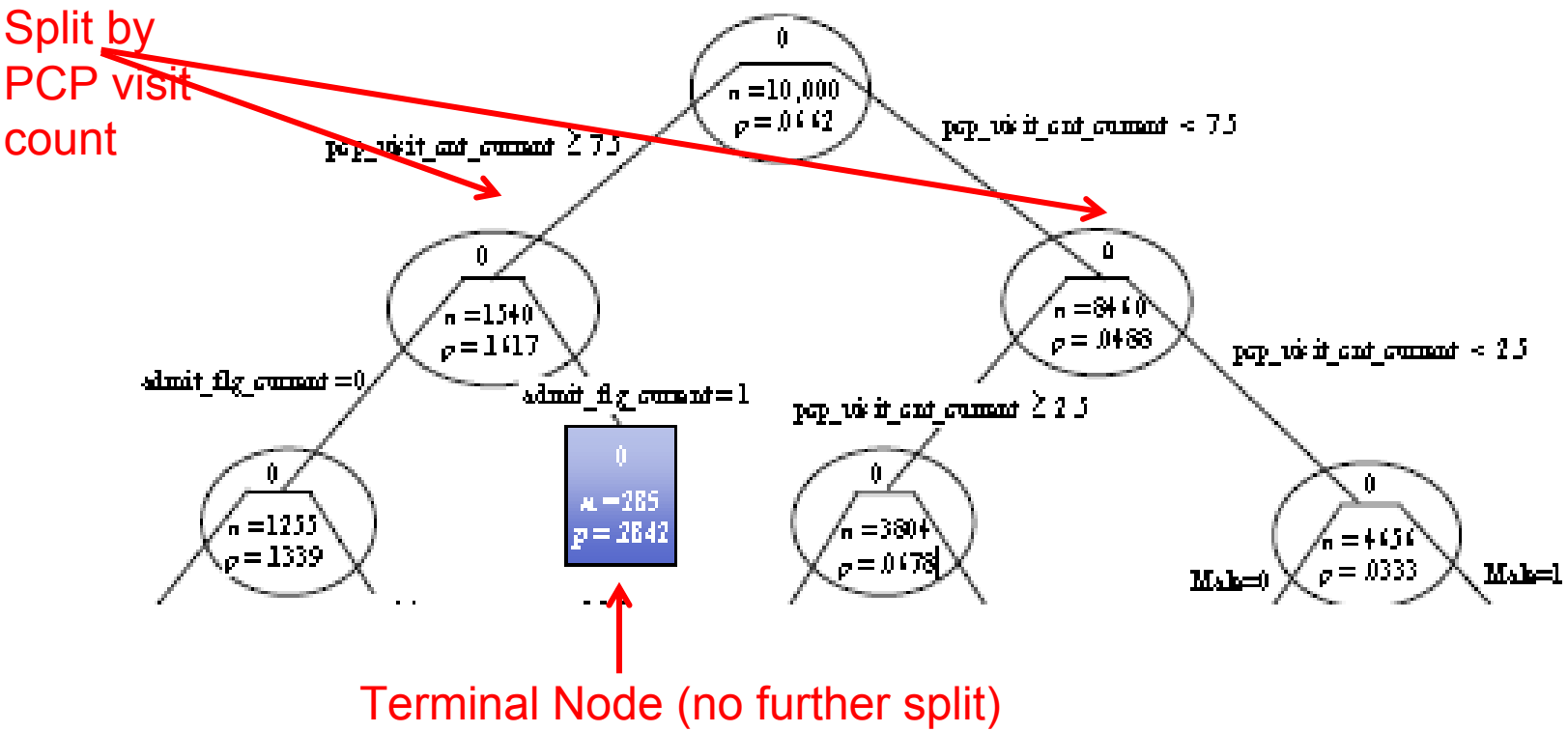
Some Other Tools: Decision Trees

Decision Trees have some advantages over other methods:

1. Tree methods are data-adaptive by giving data more freedom to choose the suitable bases that best approximate the true regression function.
2. Easy and meaningful interpretations may be extracted by tracing the splitting rules down the path to each terminal node. These combined rules help determine what has led to the event or nonevent.
3. The tree method provides a more efficient way to optimize the usage of categorical predictors, including the automatic merging of redundant levels, than regression models.
4. Trees are invariant to monotone transformations on continuous predictors.
5. The tree structure provides a natural and optimal way of grouping data, making them attractive to applications such as medical diagnosis or prognosis and credit scoring.
6. Trees excel in dealing with interactions of high order and complexity. Interactions are automatically and *implicitly* handled by the hierarchical tree architecture.
7. The tree method is *unstable* as a modeling tool, in that a small perturbation in the data could lead to dramatic changes in the final tree model. Although a single tree often does not attain satisfactory accuracy in a prediction task, multiple trees can be built and combined in model ensembles to achieve high accuracy of prediction.

Some Other Tools: Decision Trees

The model is hard to read, but here is a snapshot. Remember that in the logistic regression model, current year cost, admissions, emergency room visits, PCP visits and age were significant.



Decision Trees

The Random Forest (A tree model) was the best-fitting model to our dataset (R^2 = 0.215). The random forest model works by bootstrapping the learning dataset, creating multiple models and (ultimately) aiming for model stability. As in the picture above, Random Forest can be interpreted through the nodes. This is not true of the next tree model, GBT.

Generalized Boosted Trees (another tree model) was the second-best model.

Interpreting the Results

Once we move beyond OLS, it is difficult to interpret the results of a particular model.

- Highly-skewed data makes R^2 unreliable.
- A number of different measures were calculated, including several that cap the error term or outlier observations. Refer to the paper for more detail.
- Most influential co-variates: different models selected different co-variates as predictors, although there was considerable overlap. Interestingly, there is conflicting evidence regarding the usefulness of diagnostic data in the final model. This is largely due to the inclusion of base year cost as a predictor variable (and its correlation with diagnostic information).

Interpreting the Results (2)

A useful test of predictive accuracy is the ability to classify risks.

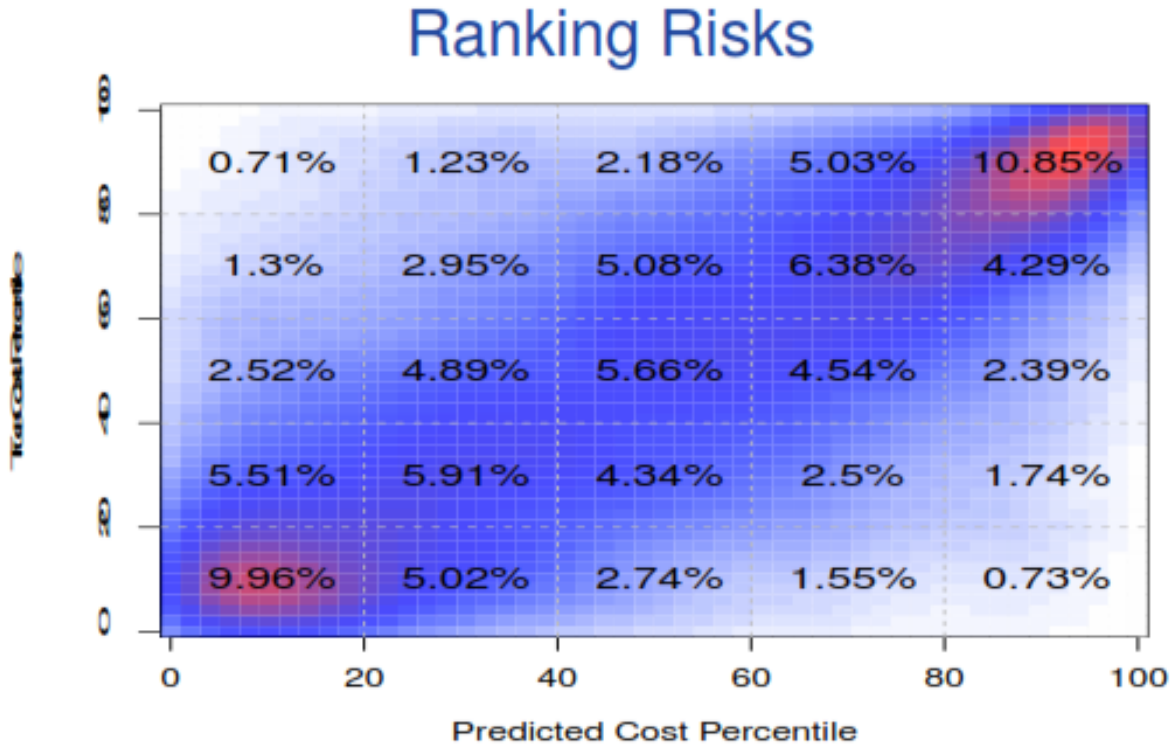


Figure : Classifying risks into quintiles. The numbers indicate percentage of test population ($M = 10000$ individuals) that fall into the particular cell (grouped by predicted quintile versus the actual quintile) so that the total sum is 100% and each row/column adds up to 20%. Colors provide a more granular description of the same information, with dark red indicating highest density.

Some Quick Comments

1. Linear Regression remains popular because it is simple, effective and practitioners understand it. Remember the George Box quote: even if Linear Models are biased and inaccurate, they may remain sufficient for some applications. Linear models have a difficult time with heteroscedasticity and may produce inappropriate predictions.
2. Generalized Linear Models (GLM): in these models, the linear relationship between the dependent and independent variables (basis of Linear Models) is relaxed, so the relationship can be non-linear. However, these models also have poor performance for low-and high-risks. "Local" or interpolating models may fit better (e.g. MARS).
3. Some more recent advances in regular regression (such as LASSO and MARS) could be helpful to improving the model.
4. Decision Trees are a means of classifying a population using a series of structured, successive steps. Our results point to an ensemble of tree models as being most effective for fitting health claims data.

Case Studies

Case Studies

In this section we will look at 4 different case studies, depending on time.

1. Predictive Modeling in ACOs.
2. Predictive Modeling for Bundled Payments.
3. Risk Adjustment in Healthcare Reform: the example of Massachusetts.
4. Case management case identification and the application of Opportunity Analysis.

Example 1: Predictive Modeling for ACOs

Example 1: Predictive Modeling for ACOs

Statement of the Problem:

Financially, ACOs are provider-based organizations that are charged with managing the care of their (attributed) members. Medicare Shared Savings (MSSP) ACOs share gains relative to a projected baseline with CMS.

Although ACOs have a strong quality orientation, to save money and share in gains, the ACO needs to reduce utilization of high-utilizing members. Which members are future high utilizers? Where should the ACO apply its limited care management resources?

This is a classic Predictive Modeling/Opportunity Analysis application.

Overview - The Role of Analytics

Analytics supports the goals of the ACO through the following processes:

Conducting opportunity analysis to identify (and then quantify) potential clinical programs;

1. Aggregating and warehousing data from multiple sources;
2. Predictive modeling/risk stratifying at the patient level for implementation of clinical programs;
3. Identifying gaps in care at the patient level;
4. Developing baseline quality measures for outcomes reporting (33 quality measures);
5. Providing ongoing reporting for program management and outcomes.

Overview – Predictive Modeling Defined

Predictive models stratify the patient population according to their likelihood of experiencing the target event. The process includes:

1. Using a similar dataset, identify all potentially correlated independent variables that predict the dependent (outcome) variable.
2. Derive scores for each patient (i.e. likelihood of experiencing the event) under numerous combinations of variables.
3. Compare the actual outcomes to the scores, to determine the scenario with the best positive predictive value. (PPV)
4. Operationalize the method for application to actual ACO data.

Develop a program to manage the targeted members.

Example 1: ACOs – Step 1

First Step: Analysis of current situation

- Where is the money spent today?
- Which high-cost patients represent an opportunity to change behavior or outcome?

For Medicare, analyze the Medicare 5% file.

Many analyses focus on *conditions.* We have found it useful to focus instead on groups of conditions (e.g. chronic) and co-morbidities.

Overall membership distribution by condition

- Members were classified into 8 hierarchical categories:
  - Nearly 25% of members do not have an identified acute, chronic or mental health condition.
  - 35.3% of the <65 segment do not have an identified condition

Membership Distribution

Condition	<65	65+	Overall
1) Acute MH Chronic	11.0%	11.9%	11.7%
2) Acute MH	5.8%	3.6%	4.1%
3) Acute Chronic	7.3%	17.6%	15.2%
4) Acute Only	6.5%	9.1%	8.5%
5) MH Chronic	7.8%	4.3%	5.1%
6) MH Only	10.3%	3.9%	5.4%
7) Chronic Only	8.8%	15.7%	14.1%
8) EHC	7.2%	13.8%	12.2%
No Condition	35.3%	19.9%	23.6%

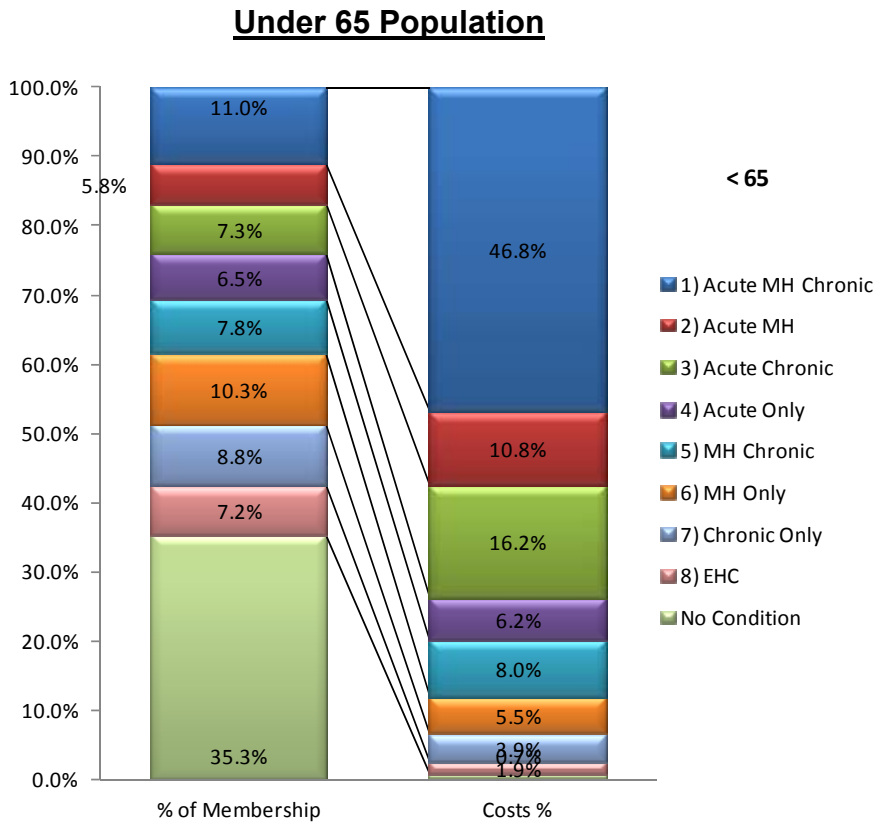
Overall cost distribution by condition

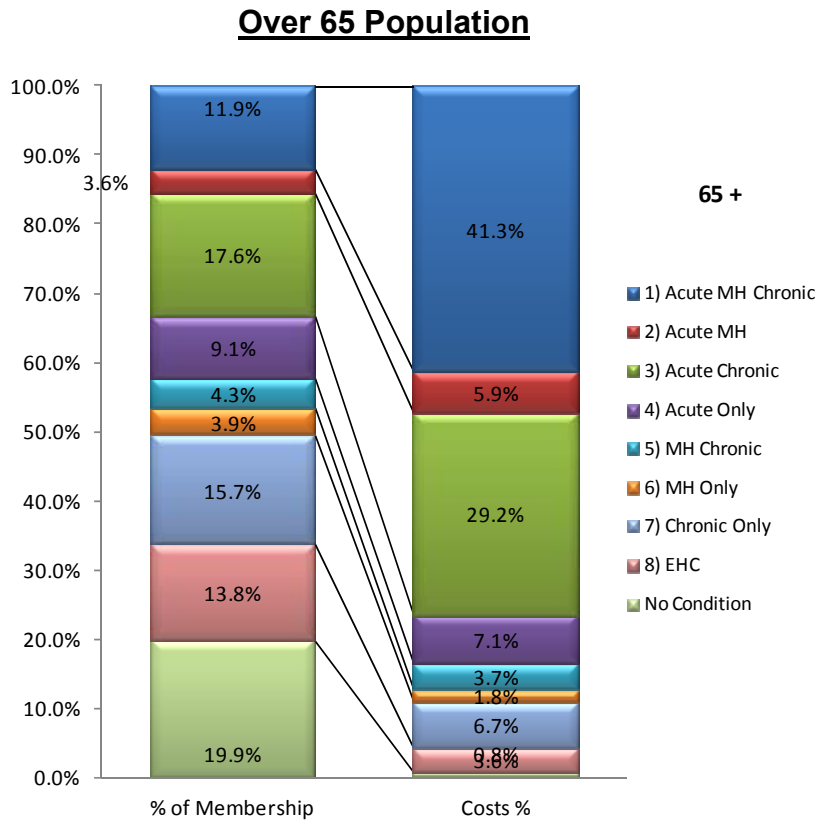
- When focusing on costs…
  - More than $4 of every $10 dollars is spent on most complex members.
  - Acute members with chronic and mental health comorbidities account for nearly 76% of all spend – care management is critical for these members

Condition	<65	65+	Overall
1) Acute MH Chronic	46.8%	41.3%	42.3%
2) Acute MH	10.8%	5.9%	6.8%
3) Acute Chronic	16.2%	29.2%	26.8%
4) Acute Only	6.2%	7.1%	6.9%
5) MH Chronic	8.0%	3.7%	4.4%
6) MH Only	5.5%	1.8%	2.4%
7) Chronic Only	3.9%	6.7%	6.2%
8) EHC	1.9%	3.6%	3.3%
No Condition	0.7%	0.8%	0.8%

Distribution for Aged and Disabled Population

Comparison between <65 members and >65 population.





Preventing over-medicalized End-Of-Life care

End-of-Life

A small percentage of all members represent those at end-of-life (≈ 4%). However, these patients account for 23% of all costs.

	END OF LIFE - HOSPICE			END OF LIFE - NON-HOSPICE		
	< 65	65 +	Total	< 65	65 +	Total
Average Lives	1,035	15,091	16,126	2,752	18,139	20,892
% of Overall	0.1%	0.5%	0.4%	0.3%	0.6%	0.6%
Total Allowed	$ 7,322	$ 5,826	$ 5,922	$ 5,072	$ 5,550	$ 5,487
% of Overall	1.5%	3.9%	3.4%	2.8%	4.4%	4.1%

•Note: these lives represent 6 months of deaths; to derive the annual total double the prevalence.

•These numbers represent 6 months of claims. To derive the last 12 months of claims, multiply by 3.0.

End-of-Life

The most complex members are a significant portion of the end-of-life population, and total cost.

	Members		Costs	
	< 65	65 +	< 65	65 +
ALL	39.5%	46.9%	62.2%	60.1%
HOSPICE	43.8%	50.4%	56.2%	59.8%
NON-HOSPICE	56.2%	58.9%	65.5%	61.2%

Overall, close to 60% of end-of-life costs are generated by the most complex patients; end-of-life accounts for 23% of all costs, with the complex patients generating about 14% of all costs.

End-of-Life (all)

The most complex members: Comparative Utilization.

Complex end of life patients have a high frequency of hospital admissions (2500 per 1000). Most of these are for medical DRGs. Also very high specialist visit frequency.

Under 65 patients are an even high-utilizing group.

	OVERALL			1) Acute MH Chronic		
	< 65	65 +	Total	< 65	65 +	Total
Average Lives	3,788	33,230	37,017	1,509	15,627	17,137
IP Admits - Overall	2,548	1,929	1,977	3,136	2,499	2,555
Average IP Length of Stay - Overall	7.8	7.2	7.7	8.5	7.6	7.7
IP 30 Days Re-Admits - Overall	800	421	426	998	566	604
Readmit % - Overall	31.4%	21.8%	21.5%	31.8%	22.6%	23.6%
IP Admits - Medical	2,243	1,682	1,680	2,691	2,154	2,201
Average IP Length of Stay - Medical	7.1	6.7	6.9	7.5	6.9	7.0
IP 30 Days Re-Admits - Medical	744	379	378	901	503	538
Readmit % - Medical	33.2%	22.5%	22.5%	33.5%	23.3%	24.4%
IP Admits - Surgical	305	247	297	445	345	354
Average IP Length of Stay - Surgical	13.1	10.6	12.1	14.5	12.1	12.4
IP 30 Days Re-Admits - Surgical	56	43	48	97	63	66
Readmit % - Surgical	18.4%	17.2%	16.1%	21.7%	18.2%	18.6%
LTC Admits	-	0.1	0.3	-	0.4	0.4
ER Visits	311	72	123	356	138	157
OP Services	17,112	15,883	12,816	15,444	15,546	15,537
PCP Visits	2,448	2,605	2,744	3,457	2,927	2,973
Specialist Visits	6,818	4,989	4,850	6,539	4,718	4,878
CT Services	3,750	2,726	2,626	4,003	3,261	3,326
MRI Services	844	482	433	718	489	509
X-Ray Services	10,352	7,916	9,365	15,796	12,251	12,563

Preventing over-medicalized **End-Of-Life** care

Clinical Program:

- Education for physicians and their staff on how to instigate end-of-life conversations.[3]
- Program to encourage patients to complete advance directives, consisting of materials, a helpline, and a registry.[4]
- Patient access to hospice and palliative care.
- Symptom-focused case management for very high-risk patients.[3]

↑ Population health	↓ Administrative burden	↓ Per capita cost	↑ Patient experience
Reduction in inappropriate life-sustaining treatments within 6 months of death, including a reduction in ER visits.[2]	Dedicated case managers to support physicians in caring for complex patients that are at very high risk of over-medicalized end-of-life care as defined by Barnato et al.[2]	Home-hospice care associated with significantly lower average costs ($12,434 versus $4,761 per year in 2007 dollars).[5]	Patients receiving in-home palliative care report significantly higher satisfaction and quality of life.[6]

[1] Zhang B, Wright AA, Huskamp HA, et al. Health care costs in the last week of life: associations with end-of-life conversations. *Archives of Internal Medicine.* 2009;169(5):480
[2] Barnato AE, Farrell MH, Chang CC, Lave JR, Roberts MS, Angus DC. Development and validation of hospital "end-of-life" treatment intensity measures. Medical Care. 2009;47(10):1098-1105
[3] Wright AA, Zhang B, Ray A, Mack JW, Trice E, Balboni T, et al. Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. JAMA 2008; 300(14):1665-73
[4] Nicholas L, Langa KM, Iwashyna TJ, Weir DR. Regional variation in the association between advance directives and end-of-life Medicare expenditures. JAMA 2011; 306(13):1447-53
[5] Shnoor Y, Szlaifer M, Aoberman AS, Bentur N. The cost of home hospice care for terminal patients in Israel. Am J Hosp Palliat Care. 2007 Aug-Sep;24(4):284-90
[6] Brumley R, Enguidanos S, Jamison P, Seitz R, Morgenstern N, Saito S, McIlwane J, Hillary K, Gonzalez J. Increased satisfaction with care and lower costs: results of a randomized trial of in-home palliative care. J Am Geriatr Soc. 2007 Jul;55(7):993-1000.

End of Life Predictive Model - Definition

Over-medicalized death is defined as:

- Chemotherapy for cancer patients within 14 days of death;
- Unplanned hospitalization within 30 days of death;
- More than one emergency department (ED) visit within 30 days of death
- ICU admission within 30 days of death; or
- Life-sustaining treatment within 30 days of death.

- Ho, T. H., Barbera, L., Saskin, R., Lu, H., Neville, B. A., & Earle, C. C. (2011). Trends in the aggressiveness of end-of-life cancer care in the universal health care system of Ontario, Canada. *J Clin Oncol, 29*(12), 1587-1591. doi:10.1200/JCO.2010.31.9897. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3082976/pdf/zlj1587.pdf
- Earle, C. C., Park, E. R., Lai, B., Weeks, J. C., Ayanian, J. Z., & Block, S. (2003). Identifying potential indicators of the quality of end-of-life cancer care from administrative data. Journal of Clinical Oncology, 21(6), 1133-1138. doi: 10.1200/jco.2003.03.059 Retrieved from http://jco.ascopubs.org/content/21/6/1133.long

End of Life Predictive Model - Scoring

- An EOL risk score is calculated for each member.
- Risk scores range in value from 0.0-1.0.
- Model is based on the following member attributes (121 in all):
  - Age and gender;
  - Race;
  - Region
  - Clinical Grouper Flags (65 HCCs);
  - Baseline admission count(s)
  - Baseline readmission count(s)
  - Baseline ER visit count(s)
  - Baseline admission via ER indicator
  - Baseline dollars spent for healthcare resources

Conditions and Attributes that Add Most to Scores

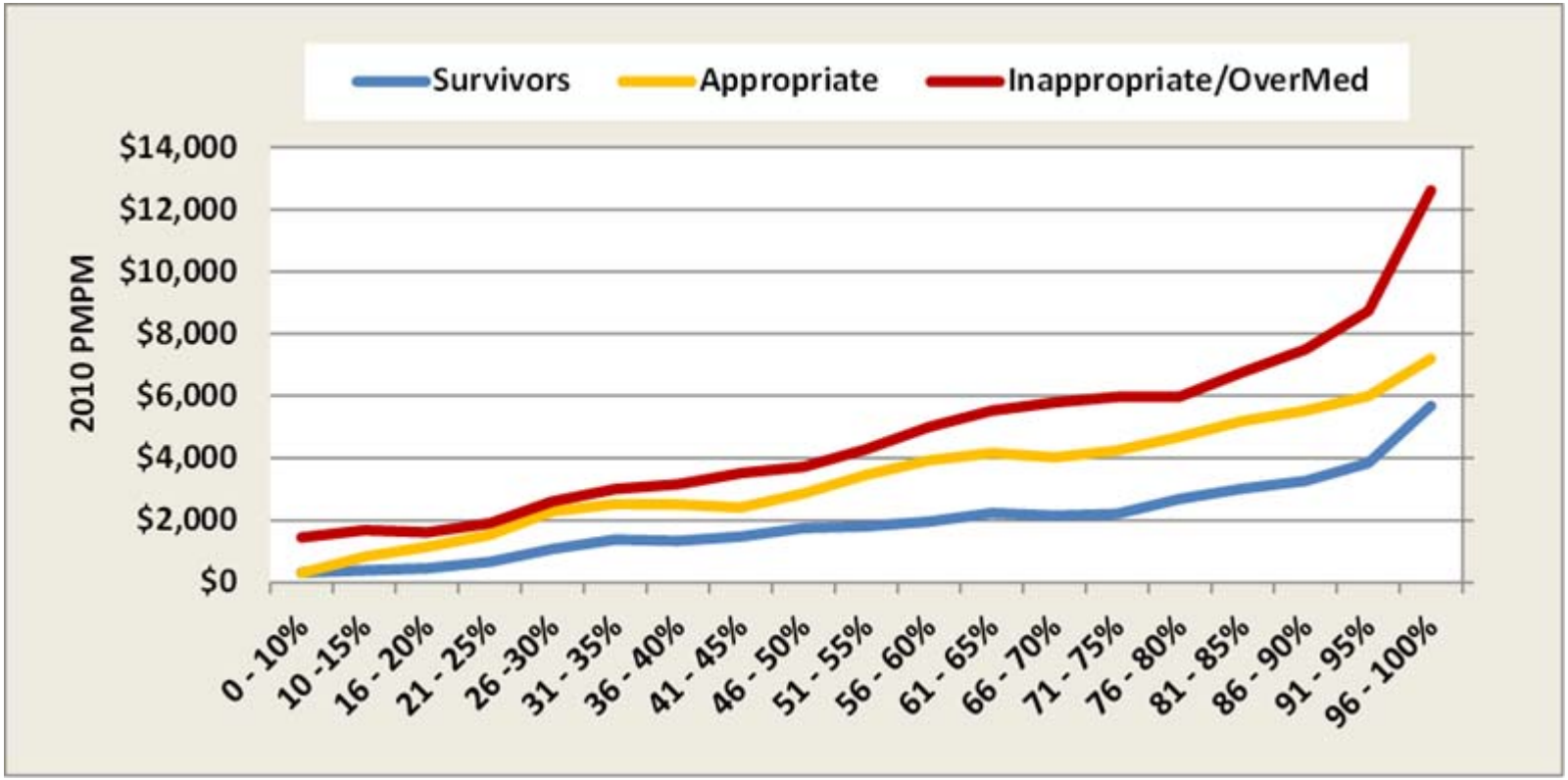
1. Acute Myocardial Infarction
2. Acute Leukemia
3. Craniotomy with major device implant
4. Cardio-Respiratory Failure & Shock
5. Metastatic Cancer & Acute Leukemia
6. Lung, Upper Digestive Tract and Other Severe Cancers
7. Septicemia or Severe Sepsis
8. Number of Admissions

Opportunity as seen from Medicare 5% Database

Medicare Patients and Deaths (based on 50% of the 5% file)			
Categories	**Members**	**% of Total Population**	**PMPM**
Survivors	819,189	92.0%	$684.80
Deceased	71,059	8.0%	$4,323.73
Appropriate	22,989	2.6%	$2,249.62
Inappropriate	**9,832**	**1.1%**	**$3,433.30**
OverMedicalized	**38,238**	**4.3%**	**$5,797.08**
Total	890,248	100.0%	$975.26

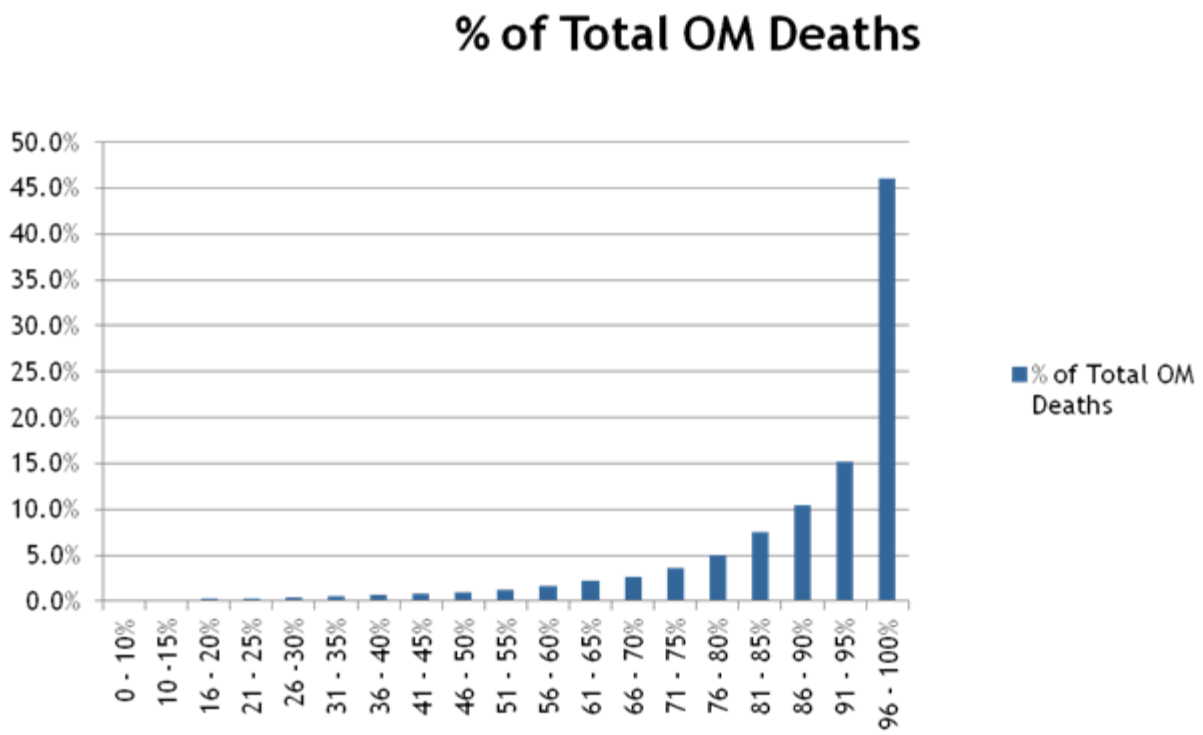
The difference between over-medicalized and appropriate death represents a financial and clinical opportunity. (Inappropriate death also represents an opportunity, although a smaller one).

Member costs by category and risk score



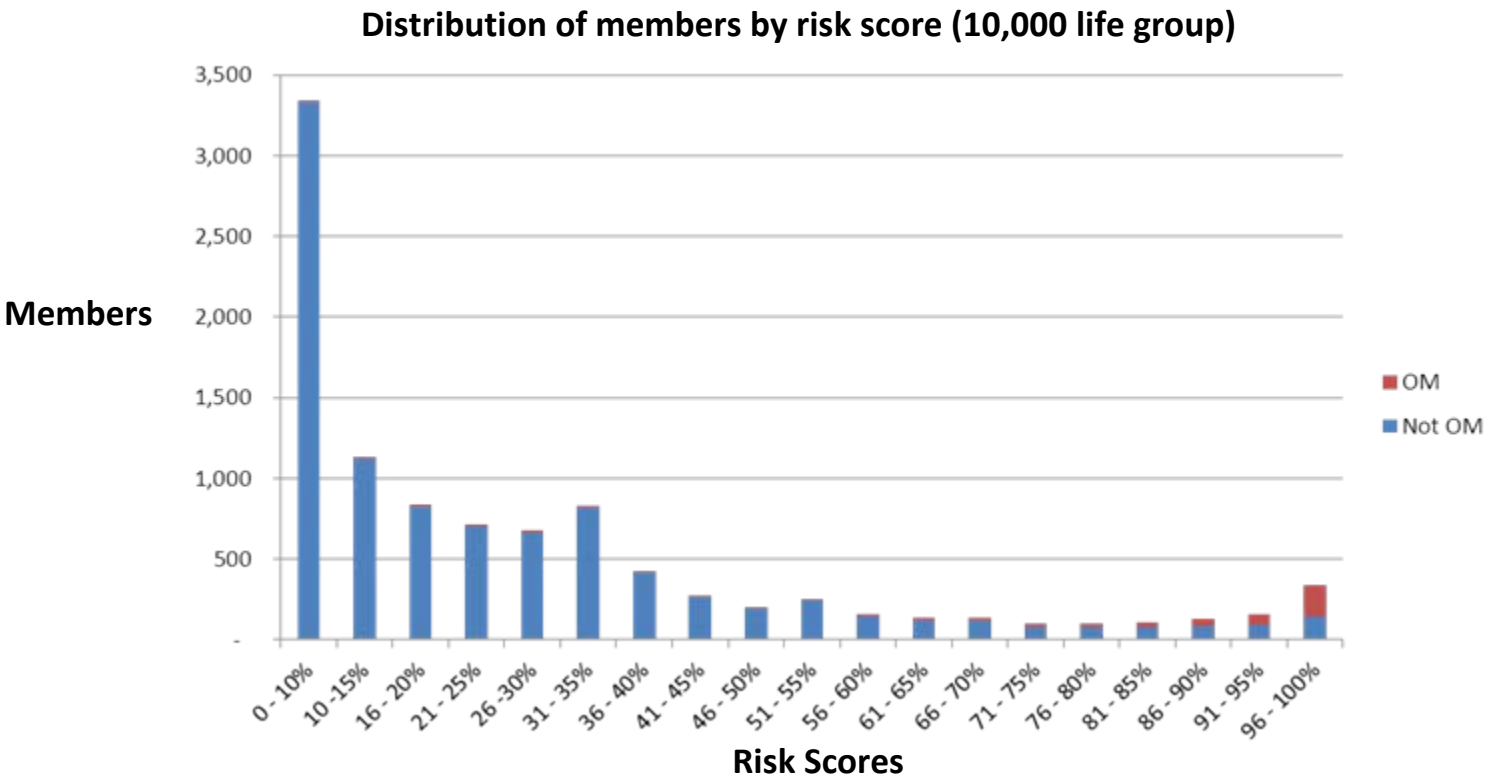
The PMPMs for members in each category vary across the bands of risk scores. The difference in the costs between those that experience overmedicalized deaths versus those that experience appropriate deaths is greatest in members with risk scores >.95.

Performance of Model on Medicare 5% Database



Out of a 10,000 attributed life group, we would expect 430 overmedicalized deaths (4.3%). Based on our model, approximately 46% of these members will have risk scores >.95.

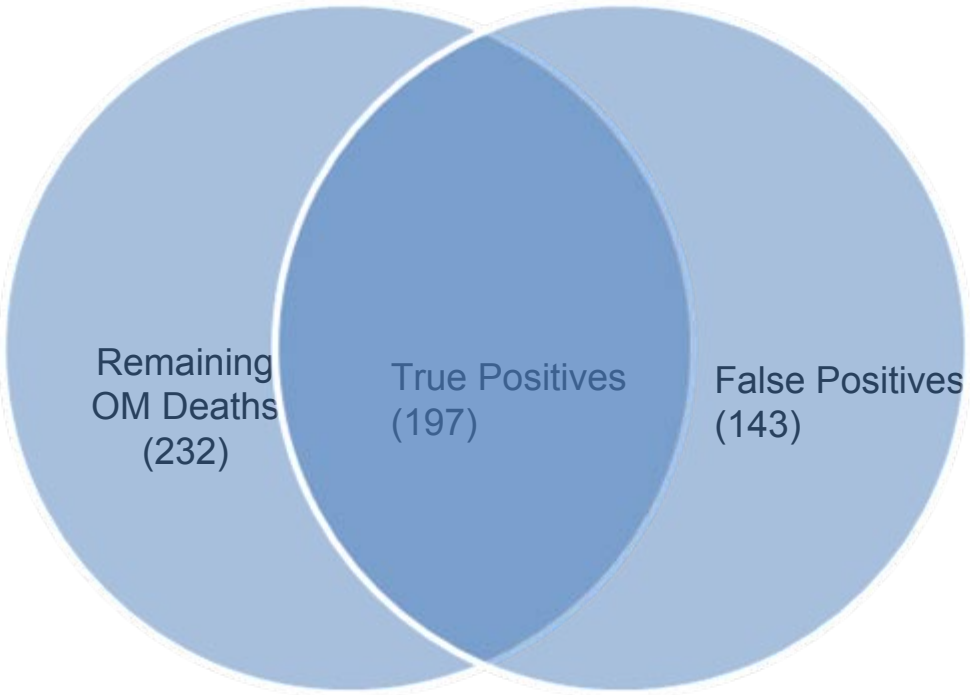
Performance of Model on Medicare 5% Database



Out of a 10,000 attributed life group, we would expect 341 members to have risk scores >.95. Of these members, we expect 197 (57.9%) to be “true positives”; that is, these are the members that represent an opportunity to avoid an overmedicalized death.

End of Life Predictive Model - Targeting

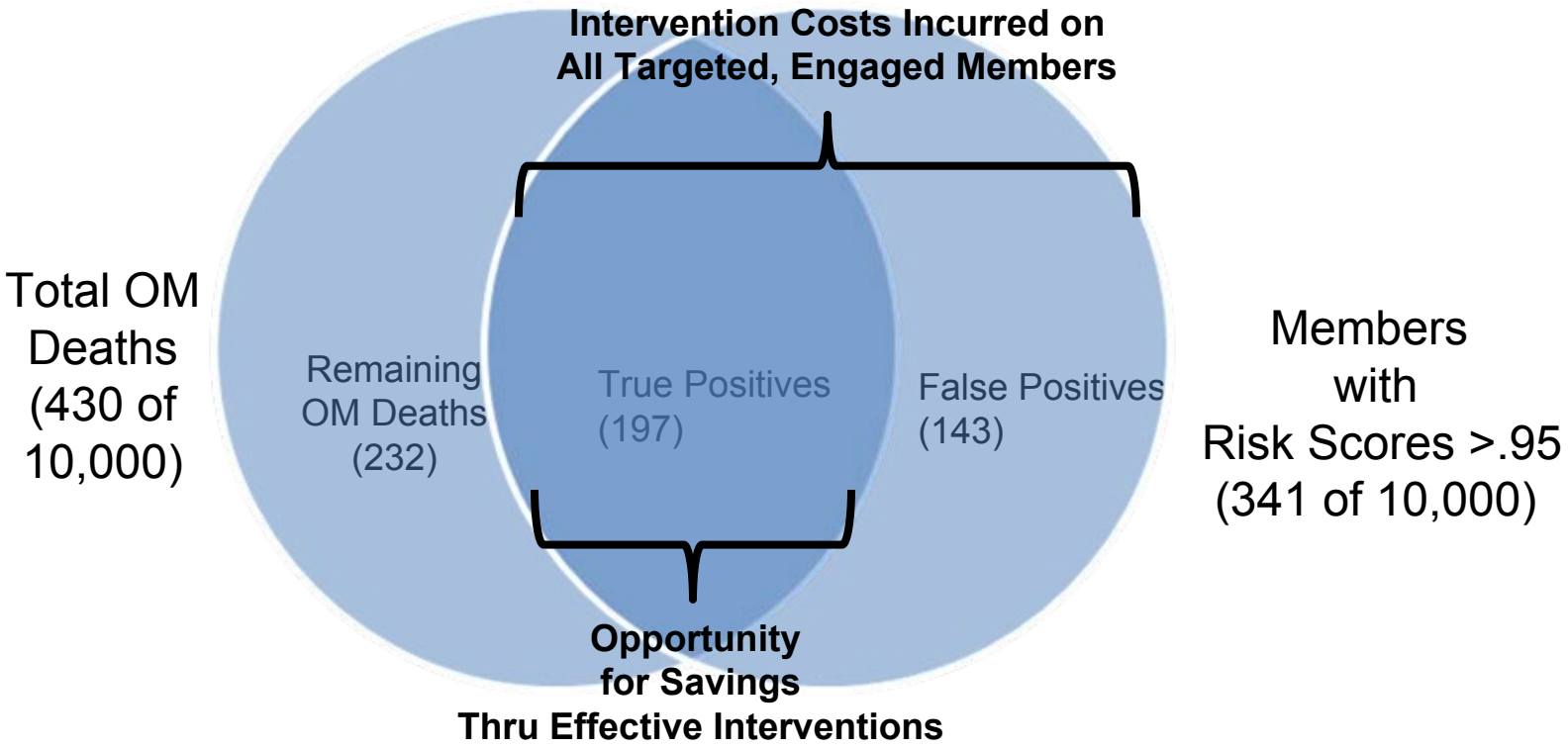
Total OM Deaths (430 of 10,000)



Members with Risk Scores >.95 (341 of 10,000)

Focusing on members with risk scores >.95 allows us to target our resources on only 3.4% of the population in order to "find" nearly half of the members that represent our opportunity.

End of Life Predictive Model - Targeting



The risk score "cut off" point is determined by evaluating the number of total members above a given risk score with the number of "true positives" found in that group. We will incur intervention costs on all members with risk scores above the cut-off, but only have the opportunity to generate savings on the "true positives" within that group.

Financial Scenario at 95% Risk Score Threshold

Based on Members with Risk Scores >.95	
# of Members (out of 10,000)	341
% of Members (out of 10,000)	3%
Over Medicalized Sensitivity	46.0%
PPV (OM Deaths)	57.9%
# of True Positives (out of 10,000)	197
# of False Positives (out of 10,000)	143
Estimated Gross Savings	
# of True Positives (a)	197
Engagement Rate (b)	40%
Effectiveness Rate (c)	50%
Potential Savings per True Positive (d), (1)	$ 15,981
Estimated Gross Savings (a x b x c x d)	**$630,853**
Estimated Net Savings	
# of Members with p>.95 (e)	341
Engagement Rate (b)	0%
Cost of Case Management (f)	$ 940.67
Total Cost (e x b x f)	$128,234
Net Savings/(Costs)	**$502,619**

(1) Difference in costs between OM death and appropriate death, over 6.5 months (PMPM*6.5).

Example 2: Bundled Payments

- CMS (CMMI) has offered hospitals and other providers the opportunity to receive bundled payments for episodes of care, rather than current fee-for-service reimbursement. Other payers are also interested.
- Under the CMMI Bundled Payments initiative, CMS would link payments for multiple services patients receive during an episode of care. For example, instead of a surgical procedure generating multiple claims from multiple providers, the entire team is compensated with a "bundled" payment that provides incentives to deliver health care services more efficiently while maintaining or improving quality of care. Proposers will have flexibility to determine reimbursement of different providers from the bundled payment.
- The expectation is that bundled payments will align incentives for providers – hospitals, post acute care providers, doctors, and other practitioners– encouraging them to partner closely across all specialties and settings that a patient may encounter to improve the patient's experience of care during a hospital stay in an acute care hospital, and during post-discharge recovery.

Example 2: Bundled Payments

- From the perspective of predictive modeling, the Bundled Payments initiative is no different than other applications: it requires modeling to predict frequency and severity in a population and methods for identifying and mitigating risk.
- The CMS bundled payment initiative allowed proposers to select their own episodes and define those procedures that they wished to exclude.
- Providers have focused on inclusiveness (finding as many DRGs as possible to include in the bundle) to maximize revenue. However, this runs the risk of heterogeneity: including all sorts of different patients, at different risk levels and with potentially variable outcomes in the bundle.
- Providers address the variability/unpredictability problem by excluding many down-stream procedures, re-admissions for complications etc. While this is *a* solution to the problem of variability in outcome, it is not ideal, reducing the provider's risk to the primary admission only and destroying the incentive to coordinate care and reduce risk.

Example 2: Bundled Payments

Optimally, how would we organize a bundled payment initiative?

•Choose the condition or procedure for which bundling is being considered. Generally these tend to be acute episodes, although some bundlers have also undertaken bundling of chronic conditions.
•Ideally the episode or condition should be one for which clear treatment protocols and evidence-based guidelines exist.
•Once the conditions or procedures are identified, investigate what the data tell us:

- Frequency of different procedures
- Severity
- Outliers
- If the episode is admission-based, what is the ideal pre- and post-admission period for the inclusion of services?
- If protocols/guidelines are available, to what extent does actual practice mirror best-practice?

Example 2: Bundled Payments

Optimally, how would we organize a bundled payment initiative?

- Develop a risk adjusted tiered bundle payment that ensures provider is appropriately compensated in caring for sicker patients. (Ideally, the Insurer should assume population risk (frequency and severity) not the provider.
- The number of tiers, and the identification of risk factors for tiering patients will be an empirical exercise.
- Additionally, determine claims distributions in the "tail" to inform the outlier/reinsurance discussion.
- Develop a predictive model for prospectively assigning patients at inception/diagnosis by risk tier.
- Determine the price per bundle to be charged to the payer.

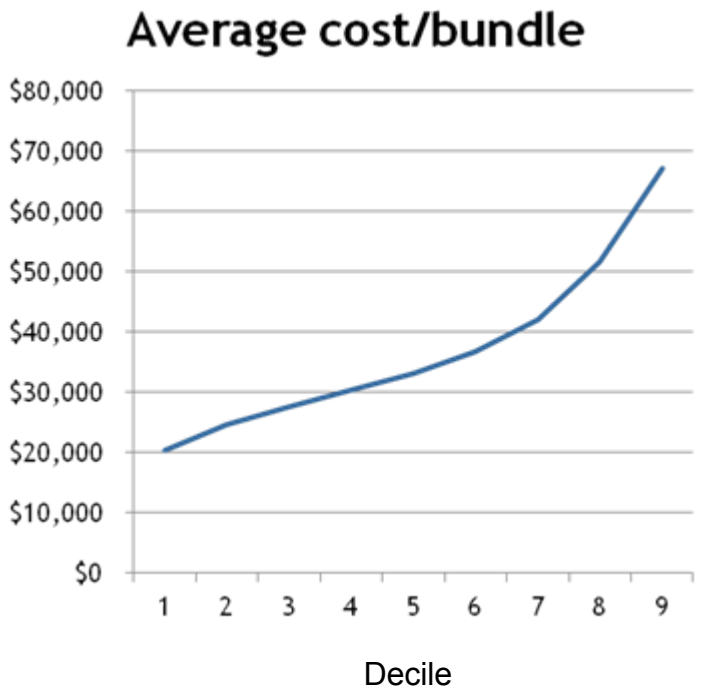
The bundler's reimbursement will be determined by the number of patients accepted at each tier, the price per bundle per tier, and reimbursement (if any) for outliers.

Example 2: Bundled Payments

Payment data indicate that for this state, the reimbursements for the bundle of activities varied from a low of $20,000 to a high of more than $600,000.

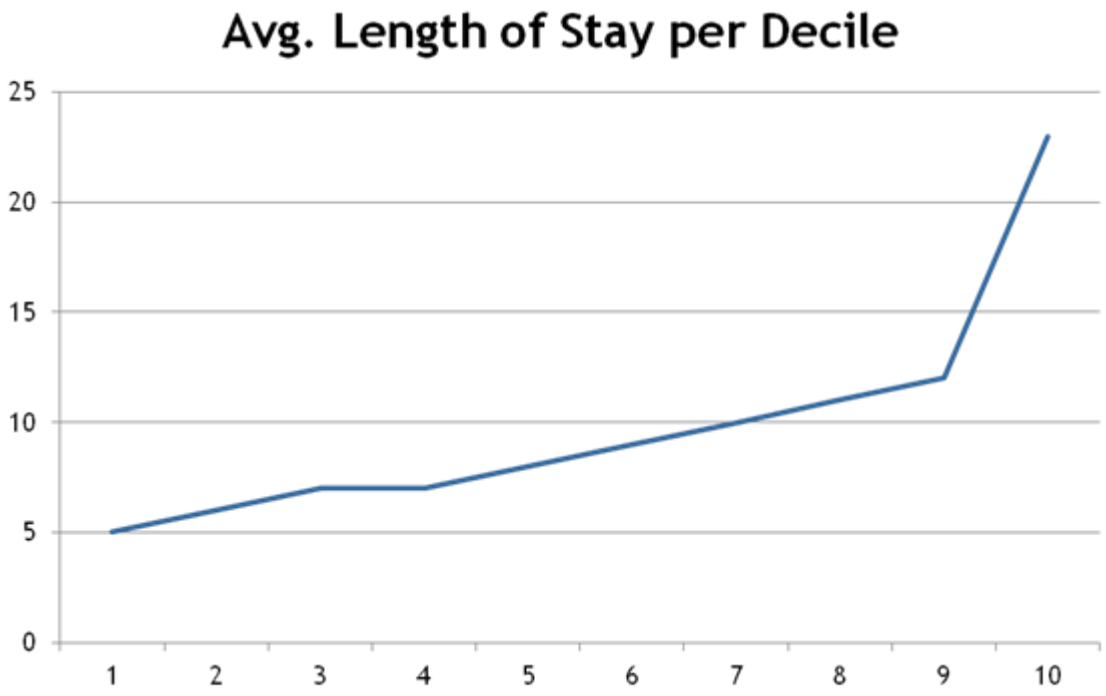
This range underscores the importance of identifying a patient's risk prospectively and of assigning the patient to the appropriate stratum for reimbursement.

The risk adjustment process assigns patients prospectively to the appropriate stratum; providers are then reimbursed for the bundle of services at the rate appropriate for the patient's stratum.



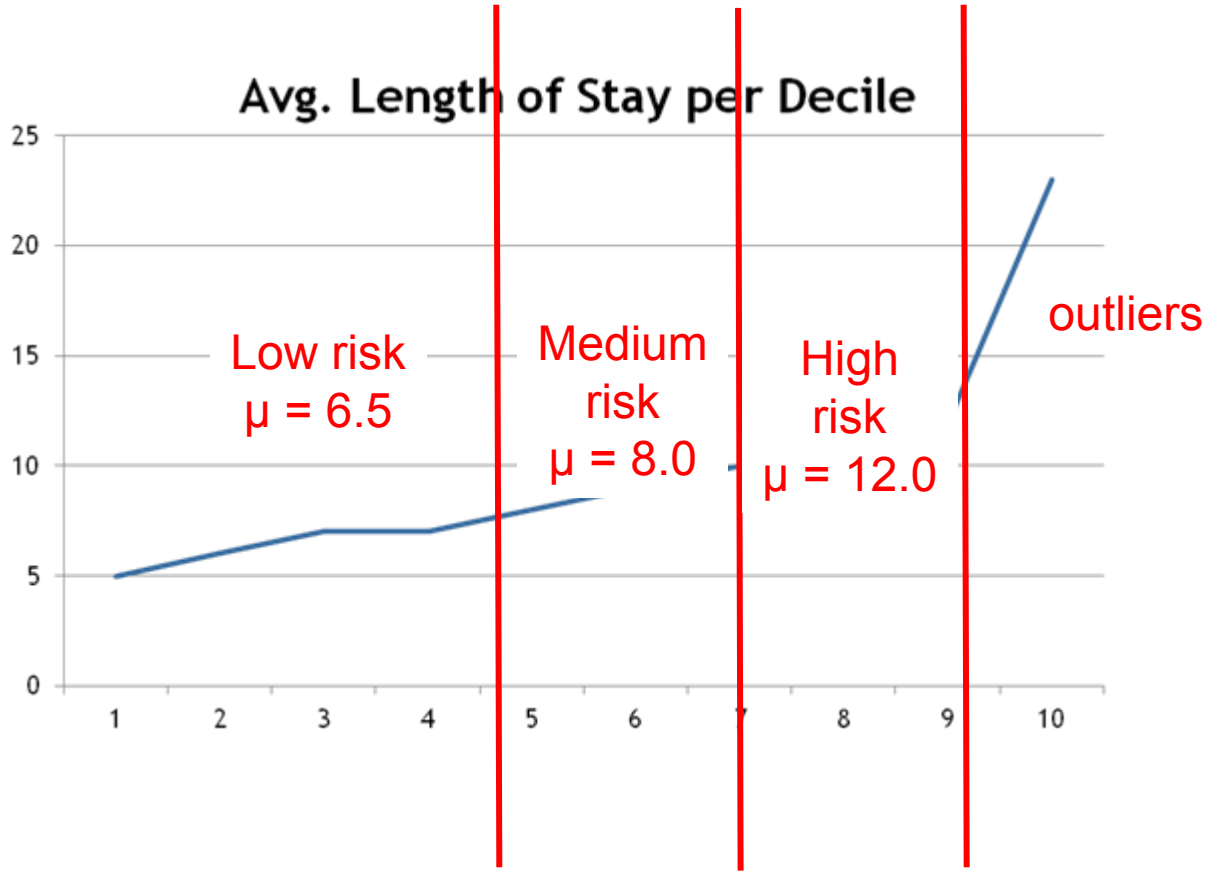
Example 2: Bundled Payments

Looking at average length of stay by decile helps inform the stratification and outlier strategy.



Example 2: Bundled Payments

One possible grouping structure:



Example 2: Bundled Payments

Developing a Model for Bundled Payments

The bundled payment model assigns patients to a risk stratum based on commonly-available risk factors. RISK FACTORS MUST BE COMMONLY-AVAILABLE PRIOR TO ADMISSION.

The ideal risk-adjusted bundled payment model will minimize intra-stratum variation, while maximizing variation among strata. Said another way, the analysis must identify the minimum number of strata that recognize legitimate variation in payments, without introducing an overly complex number of strata.

A key to developing the model was to identify risk factors that were correlated with higher payments. A patient presenting for surgery will then be placed into one of several strata based on his/her risk factors. The assignment to a particular stratum must be based on the types of data available when the patient is scheduled for surgery. Thus, while it is possible to construct sophisticated bundle algorithms, a practical model must be tailored to available payer (or analytics provider) data.

Example 2: Bundled Payments

To develop the model, we used as a measure of relative patient risk the ratio of average payment per episode to the aggregate average payment. (See Distribution of Relative Risk Scores). This relative score for each episode was used as an explanatory variable to predict future payment. A wide range of other independent variables was tested, including member condition, demographic, socio-economic status, and lab and blood screenings.

Variables selected for the model were:

- Age
- Sex
- Socioeconomic status
- Diabetes diagnosis
- Metabolic syndrome diagnosis
- Cancer diagnosis
- Mental Health/Psychiatric diagnoses
- Pulmonary disease diagnosis
- Renal disease diagnosis
- CNS diagnoses diagnosis
- Substance abuse treatment/diagnosis
- Presence of other procedures
- Previous inpatient claim for osteoarthritis

Decile	Risk Score
10	0.490
20	0.593
30	0.662
40	0.728
50	0.793
60	0.880
70	1.007
80	1.240
90	1.613
100	14.726

Example 2: Bundled Payments

The key output of the model is the predicted average payment, which was generated by applying the model weights to each patient's explanatory variable values. The predicted and actual average costs were reasonably close for risk score bands 1 and 2, indicating that the model is likely to perform well as a predictor of cost for patients in those bands. The model somewhat under-predicted costs for members in risk band 3.

Risk Score Bands, Actual and Predicted Payments

Risk Score Band	Number of Members	Risk Score Mean	Actual Average Payment, Truncated	Predicted Average Payment
1 ($\leq$ 0.9)	965	0.80	$29,291	$29,814
2 (0.9-1.2)	1,842	1.00	$36,689	$37,660
3 (> 1.2)	650	1.40	$57,053	$53,524

Example 2: Bundled Payments

It is important to understand how this type of bundling could affect a participating hospital. This table illustrates how a specific hospital's current reimbursement would be affected if it participated in bundling using the predicted risk score and reimbursement by stratum in the prior slide.

Distribution of Episodes and Reimbursement at the Sample Hospital

Risk Score Band	Number of Members	Risk Score Mean	Actual Average Payment Truncated	Predicted Average Payment	Variance
1 (≤ 0.9)	130	0.78	$33,981	$29,814	-$4,167
2 (0.9-1.2)	249	1.00	$40,736	$37,660	-$3,076
3 (>1.2)	123	1.53	$62,187	$53,524	-$8,663

Example 2: Bundled Payments

Overall the model indicates that our sample hospital will receive less reimbursement under the bundling model than under the current fee-for-service model. This provides an incentive for the hospital to increase its efficiency and ensure evidence-based practice.

Actual and Predicted Reimbursement for the Sample Hospital

	Sample Hospital		Ratio
Risk Score Band	**Actual Total Payment**	**Predicted Total Payment**	**Predicted/ Actual**
1 ($\leq$ 0.9)	$4,417,485	$3,875,855	88%
2 (0.9-1.2)	$10,143,202	$9,377,372	92%
3 (>1.2)	$7,649,011	$6,583,457	86%
ALL	$22,209,698	$19,836,684	89%

Example 2: Bundled Payments

Risk Stratification Model Application

The bundle in the Sample Hospital appears to be composed of higher-cost providers. Although providers will determine the price they bid for the entire bundle, this number must fall below the payer's actual average payment to be attractive. In this example there is an opportunity for the hospital to negotiate more cost-effect pre- and post-surgery arrangements.

Descriptions	ALL	Sample Hospital
Number of Episodes	**3,457**	**502**
Payment per Episode	$ 38,453	$ 44,242
Avg. Pre Surgery Payment	$ 6,203	$ 8,352
Avg. Peri-Surgery Payment	$ 28,338	$ 31,165
Avg. Post-Surgery Payment	$ 3,912	$ 4,726

Example 3: Healthcare Reform and Exchanges

3 Rs under national reform

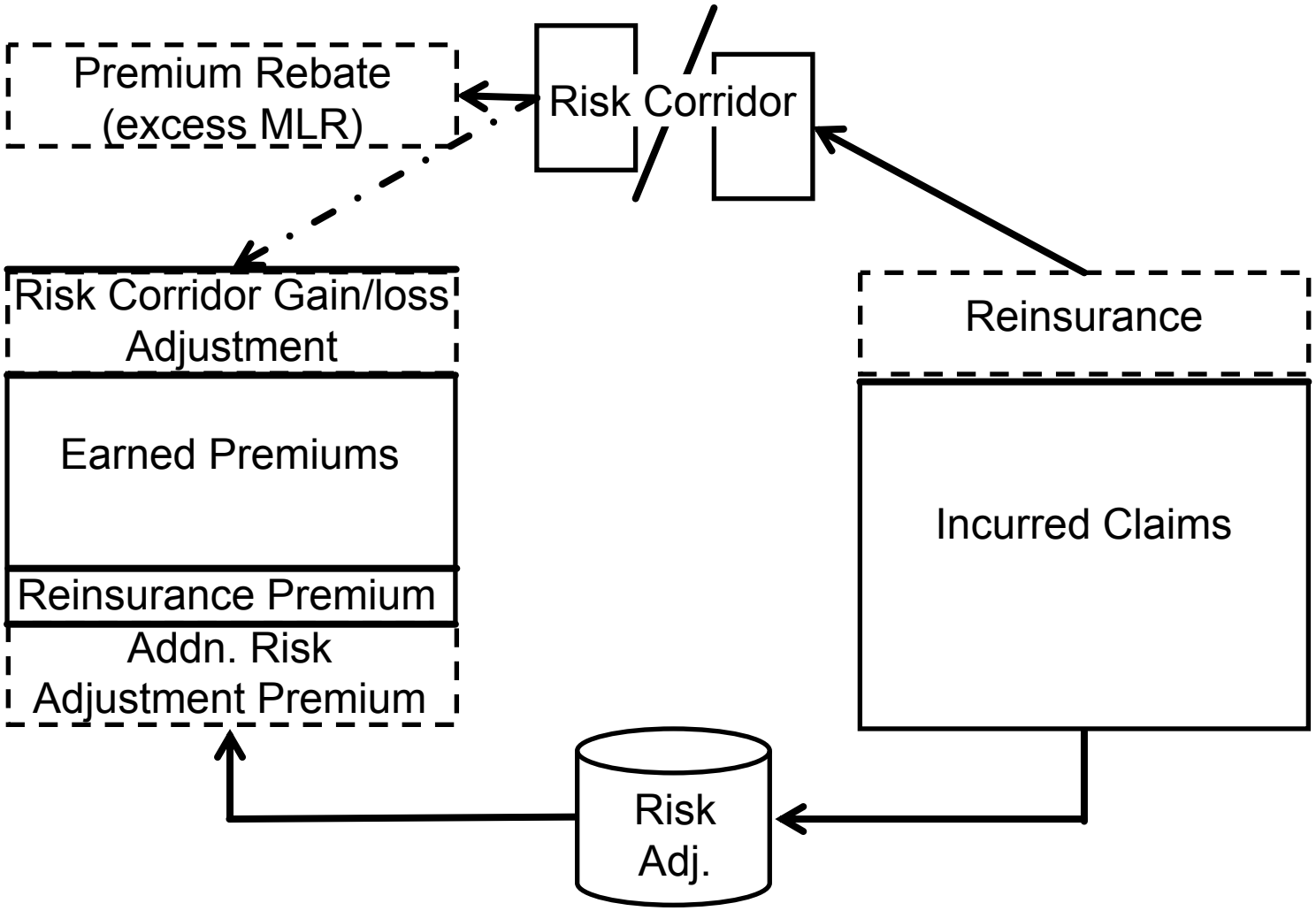
- Three programs intended to reduce the potential risk and encourage participation of a wide range of insurers in exchanges.
- Reduced risk = lower premiums for consumers.
- Three programs:
  - Risk Adjustment
  - Reinsurance
  - Risk corridors (gain sharing)
- Operation will require decisions relative to governance, structure, data, methodology, timing, funding and administration.

3 Rs under national reform

- Rating changes under National Reform (ACA)

- Allowable rating factors
  - Geography
  - Age (3:1 compression)
  - Tobacco Use (up to 150%)
  - Family size
  - Plan choice: (Bronze/Silver/Gold/Platinum)

Risk mitigation requires significant funds transfer



Actuarial implications of reform: the 3 Rs

- Large individual losses are managed with stop-loss pool.

The 3 Rs:

- **Risk-adjustment**
- **Risk-transfer**
- **Reinsurance**

- The initial risk imposed by members without prior experience requires a gainsharing mechanism.
- Ongoing risk may be managed with risk adjusted revenue transfers between plans.

Gainsharing (risk corridors)

- The Connector has operated different corridors/gainsharing arrangements in different years.
- Newly-insured lives represented an unknown risk
- The Connector also implemented a stop-loss pool to which all participating plans contributed.

Illustration of the difference between risks addressed by risk corridors and by risk adjustment:

Risk adjustment

Estimated Claims: 100

Experience

Plan A	Plan B	Average
120	80	100

Address this difference with risk adjustment: transfer revenue from Plan B to Plan A.

Risk corridor

Estimated Claims: 100

Experience

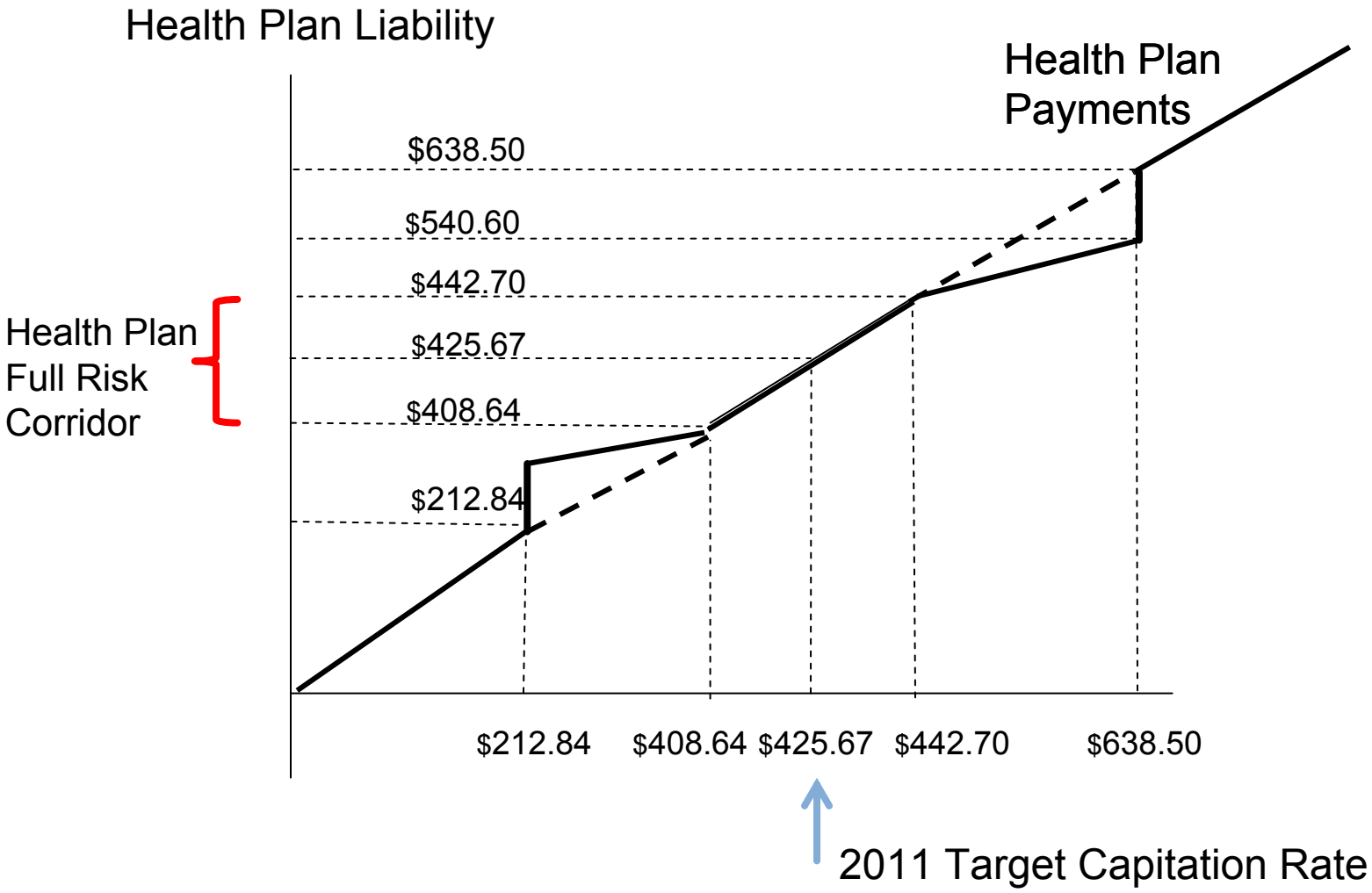
Plan A	Plan B	Average
120	120	120

Address this difference with risk corridor: both plans receive additional reimbursement from the Exchange.

Gainsharing (risk corridors)

- Gainsharing limits and terms have changed since 2006
- Principles remain fundamentally the same.
- Aggregate risk sharing corridors apply to all ***Commonwealth Care*** Health Plans:
  - Aggregate risk sharing corridors of 4% apply above and below the target medical capitation rate ("Health Plan Full Risk Corridor");
  - Connector Authority shares 50% above and below the Health Plan Full Risk Corridor; and
  - Health Plans return to 100% full risk at 50% above and below the medical capitation revenue (closed-end risk sharing).

Gainsharing (risk corridors)



Stop-loss (reinsurance)

- Connector operates a stop-loss pool to which all ***Commonwealth Care*** Health Plans must contribute.
- Health Plans fund the pool at 1.25% of the capitation rate - pool pays 75% of incurred claims above a $150,000 attachment point.
- Experience-rating also occurs if the pool runs a surplus or deficit.

Risk adjustment

- Connector does not fund the *Commonwealth Choice* program.
- Applied based on age and sex of enrolling members.
- Claims-based risk factors were developed, using DxCG Medicaid Model
- New entrants continue to be age-sex adjusted.
- Adjustment calculations performed by the Connector Authority staff using data provided by MassHealth (Medicaid), the enrollment and billing administrator.

Risk adjustment

The PMPM capitation rate to be paid to the Health Plan is equal to:

$$\text{Target PMPM} \times RF_{HP} + \text{Admin.}$$

Where:
Target = Statewide Medical only Target for the Commonwealth Care Program = $394.00
PMPM[1] (before any adjustment for hospital tax, but includes pharmacy costs)

$$RF^{j}_{HP} = \sum_{i=1}^{\text{Total Members}} (Geo_i \times Plan_i \times Disc_i \times Risk_i) \ / \ (\text{Total Members})$$

Where:
$Geo_{i,j}$ = Geographic (region) factor for Health Plan j Member i. [2]
$Plan_{i,j}$ = Plan Type factor for Member i in Health Plan j.
$Risk_{i,j}$ = Risk factor for Member i in Health Plan j.
$Disc_i$ is a factor that represents a discount offered by the Health Plan. It does not apply in FY 2012.

[1] This capitation amount includes prescription drugs, but does not include any adjustment for Health Safety Net (Uncompensated Care) hospital assessment.

[2] Note that in Massachusetts, health plans serve multiple geographic regions and offer different plan types. This calculation develops a plan-wide adjustment factor.

Risk adjustment

Example of calculation of overall adjustment factor

Member	Plan Type	Region	Age	Gender	Rating Factors			
					Plan Type	Geographic	Risk	Total
001	I	North	27	F	1.0619	0.9468	0.8694	0.8741
002 *	I	North	22	F	1.0619	0.9468	0.9970	1.0024
003	II	North	35	M	0.9461	0.9468	0.9108	0.8159
004 *	II	Central	44	F	0.9461	1.1589	1.0350	1.1348
005	III	Central	54	M	0.8909	1.1589	1.2533	1.2941
							Avg	**1.0242**

* Members 002 and 004 had seven or more months of experience during the historic experience period. Therefore, they receive a DxCG risk factor rather than an age/gender risk factor.

Example (member #2): $RF_{i,2}$ = (1.0619)(0.9468)(0.9970) = 1.0024

Payment to this Health Plan in the first quarter of Fiscal Year 2011 is prospectively adjusted for the increased Total Average Rating Factor and becomes: $393.67 x 1.0242 + $32.00 = $435.20

3 Rs: results

Reinsurance (stop-loss) experience:

FY	2007	2008	2009	2010	2011	Cum.
Loss-ratio	80.7%	137.4%	54.7%	67.1%	101.4%	81.8%

Cumulatively the pool has generated a surplus which has been experience-rated back to the plans.

3 Rs: results

Risk-adjustment (Q1 2012):

Plan	1	2	3	4	5
Rel. Risk Score	+5.9%	(8.8%)	15.3%	16.1%	(2.1%)

- Net risk adjustment in Q3 2012 is 3.9%, amounting to a net transfer to the participating plans of $6.8 million.
- The method of risk-adjustment chosen by the Connector does not result in budget neutrality.
- The concurrent, budget-neutral approach of the ACA does not allow for this change in the underlying risk profile of the covered population.

3 Rs: results

- Risk Corridor: Gainsharing has operated according to different parameters in different years.
- *Aggregate* net payments/(receipts) over the last few years have been in the range 1%-2% of premiums.
- Gainsharing results for individual plans have been larger, with several plans being reimbursed by the state between 3% and 5%, and on occasion plans reimbursing the state double-digit percentages.

Example 4: Case Finding and Opportunity Analysis

Example 4: Case Finding and Opportunity Analysis

A recent article about the application of predictive models to the "Triple Aim:"

- Lower cost
- Better Quality Care
- Higher patient satisfaction.

QUALITY & GOVERNANCE

By Gerald Lewis, Heather Kirkham, Ian Duncan, and Rhema Vaithianathan

How Health Systems Could Avert 'Triple Fail' Events That Are Harmful, Are Costly, And Result In Poor Patient Satisfaction

ABSTRACT Health care systems in many countries are using the "Triple Aim"—to improve patients' experience of care, to advance population health, and to lower per capita costs—as a focus for improving quality. Population strategies for addressing the Triple Aim are becoming increasingly prevalent in developed countries, but ultimately success will also require targeting specific subgroups and individuals. Certain events, which we call "Triple Fail" events, constitute a simultaneous failure to meet all three Triple Aim goals. The risk of experiencing different Triple Fail events varies widely across people. We argue that by stratifying populations according to each person's risk and anticipated response to an intervention, health systems could more effectively target different preventive interventions at particular risk strata. In this article we describe how such an approach could be planned and operationalized. Policy makers should consider using this stratified approach to reduce the incidence of Triple Fail events, thereby improving outcomes, enhancing patient experience, and lowering costs.

The "Triple Aim" of health care is to improve individual patients' experiences of care, advance population health, and reduce per capita health care costs.[1] A central tenet of the Triple Aim is to restructure care in ways that lead to improvements across all three of these goals.

The Institute for Healthcare Improvement has worked with organizations in many countries to implement populationwide interventions to foster the Triple Aim.[2] Examples are programs that encourage self-management of chronic conditions,[3] promote e-mail communication between patients and physicians,[4] and encourage greater use of primary care.[5]

Other organizations have adopted a more targeted approach to achieving the Triple Aim. For example, a Commonwealth Fund case study found examples of organizations that were focusing on improving access and care for individual patients who had low incomes, were uninsured, or had complex chronic conditions.[6] Indeed, several authors have argued that success will require both population health and individually focused strategies, such as those employed by Geisinger Health System in Flint, Michigan.[7,8,9] For example, Geisinger increased its primary care capacity (that is, a population approach) and offered health navigators to its high-risk patients (that is, a targeted approach).

The objectives of this article are to propose a third, stratified approach to tackling the Triple Aim and to explore some of the ethical challenges that this new approach presents. The stratified approach to the Triple Aim involves identifying and prioritizing subpopulations according to their risk of experiencing health encounter failures—what we call "Triple Fail" events—and according to their likelihood of benefiting from preventive care.[4,9]

We define a Triple Fail event as a health outcome

Example 4: Case Finding and Opportunity Analysis

Traditionally, predictive modeling for case-finding has aimed at identifying high-risk individuals. This is frequently done by applying a commercial risk-adjuster model and intervening on the highest-risk score population. The following procedures are common:

Model 1: Run a predictive model and stratify members according to their predictive risk score. One potential draw-back of this approach is the high prevalence at the top of that list of members who (although high risk) are minimally intervenable. Even if an intervenability algorithm is applied to the entire population, the resulting list will consist of a mix of members with different conditions, issues and needs.

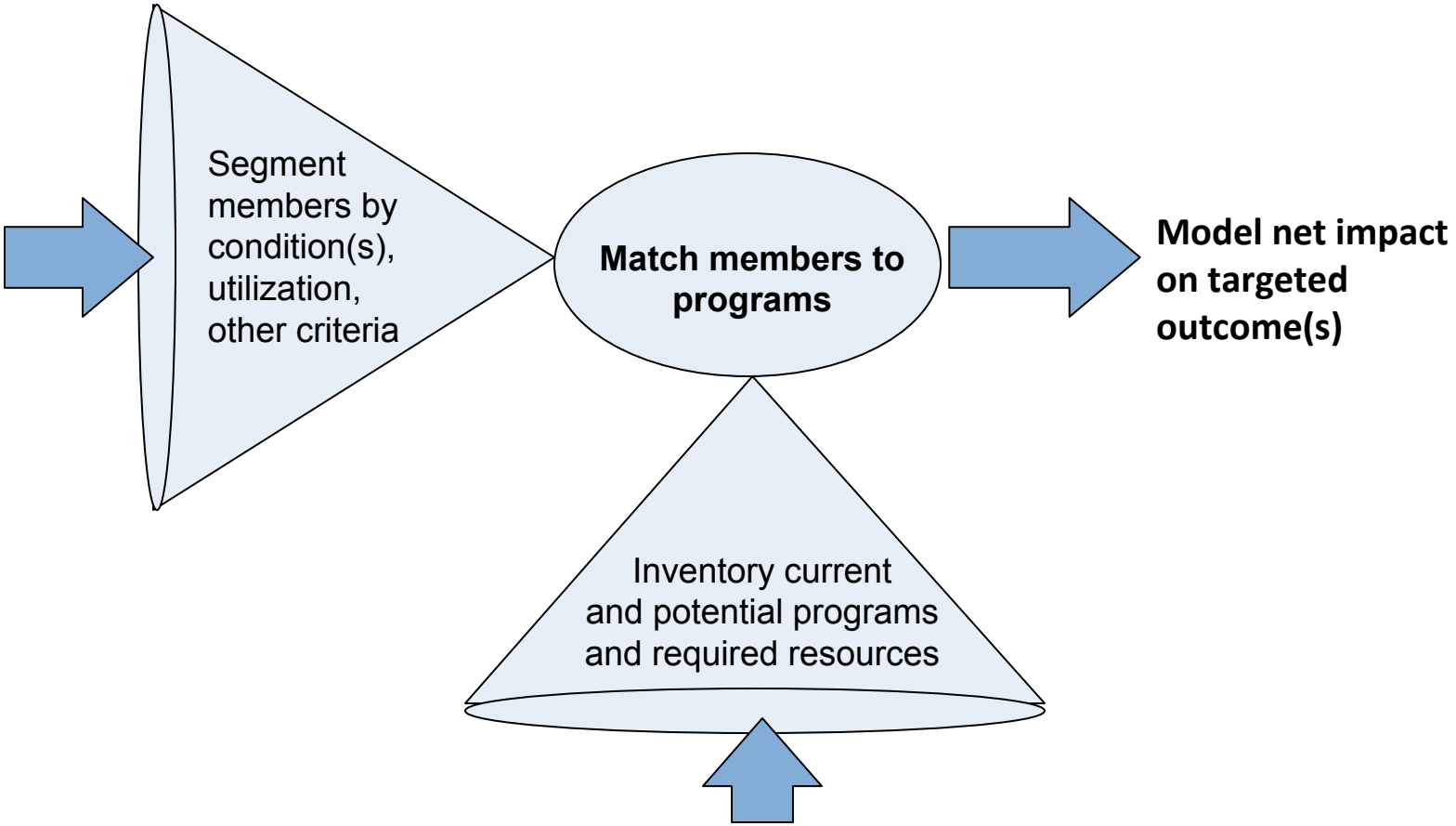
Model 2: Model 2 is a condition-specific model. For simplicity, Program planners frequently want to focus on members with a specific condition, say diabetes. This approach has the benefit of addressing the member heterogeneity inherent in Model 1. However, the high prevalence of co-morbidities in the high-risk population requires that any program targeted at a condition population will ultimately have to be sufficiently broad to address all conditions of the population. Moreover, it is often the interplay between comorbidities that drives the complexity and its associated costs, so a focus on one disease may well miss this greater opportunity.

Model 3: a rules-based approach is often used in case management programs. In this model, program managers determine a set of rules to identify target patients for management. Sometimes the rules are condition-specific; sometimes a financial threshold is used (for example, $50,000). Depending on how the rules are determined, high-opportunity members may or may not be targeted for intervention.

Example 4: Case Finding and Opportunity Analysis

- Opportunity Analysis is designed to address a number of the shortcomings associated with the typical models. It maintains the stratification of Model 1 but adds the element of intervenability by assigning lower priority to those patients with conditions that are less amenable to an intervention program.
- As a general rule, Opportunity Analysis avoids disease-specific programs in favor of programs that target members with common risk profiles (for example, all chronic members or all members at end of life) although there may be a few notable exceptions (for example end-stage kidney disease or some specific preference-sensitive conditions).
- Opportunity Analysis requires research and understanding of the targeting, operation and outcomes of programs that have been implemented in similar populations.
- Finally, Opportunity Analysis takes into consideration the economics of programs: the cost of the intervention vs. the expected reduction in utilization that each individual can be expected to contribute.

Example 4: Case Finding and Opportunity Analysis



Example 4: Case Finding and Opportunity Analysis

A summarization that we have found helpful is to group patients not be risk or condition but by class of condition; this helps with relative intervenability and in the design of programs. Note the disproportionate cost of the "Episodic, Mental Health and Chronic" group, whose costs are about three and one-half times their numbers. Almost half of the total cost of the population is concentrated in this, the most complex segment of members. Unfortunately, this is also the most difficult segment to design programs for or to manage. But any population program that avoids addressing the needs of the most complex patients will be doomed to financial failure.

Condition Category	Population %	Cost %
Episodic, Mental Health and Chronic	12%	42%
Episodic and Mental Health	4%	7%
Episodic and Chronic	15%	27%
Episodic only	9%	7%
Mental Health and Chronic	5%	5%
Chronic only	14%	6%
Mental Health only	5%	2%
Emerging Conditions	12%	3%
None	24%	1%
	100%	100%

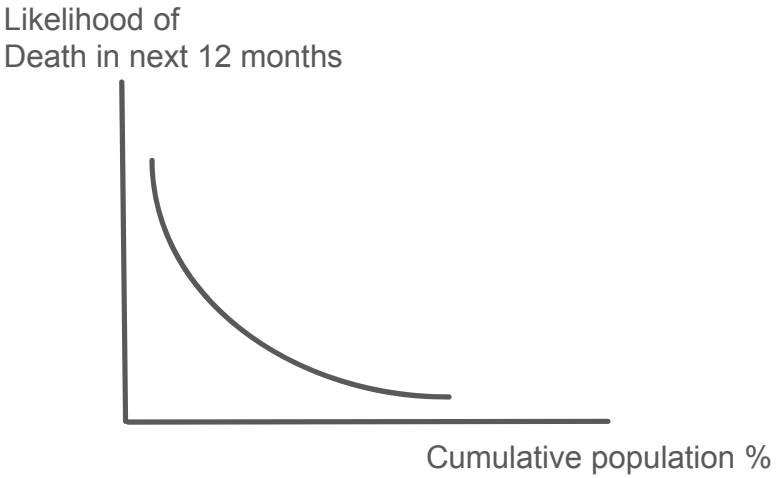
Example 4: Case Finding and Opportunity Analysis

Considerable analytics can be performed to try to identify classes of patients who represent both high cost and an opportunity to change the outcome. One such class consists of patients within 6 to 12 months of end of life. This class consumes considerable resources, without much affecting the ultimate outcome.

Service Category	Complex Population	End-of-life Sub-Population
Inpatient Admissions	1,000 per 1000	2,500 per 1000
IP Admissions Medical	750 per 1000	2,150 per 1000
IP Admissions Surgical	250 per 1000	350 per 1000
IP 30 day readmissions	18%	25%
30-day Readmits Medical	20%	25%
30-day Readmits Surgical	12%	20%
Emergency Room visits	1,200 per 1000	200 per 1000
Specialist visits	6,500 per 1000	5,000 per 1000
MRI services	400 per 1000	500 per 1000
Primary care visits	4,500 per 1000	3,000 per 1000

Example 4: Case Finding and Opportunity Analysis

Fortunately, there is plenty of peer-reviewed literature about, and good outcomes from programs that aim to manage patients at the end of life. We developed a program and constructed a predictive model to predict those patients at risk of death in the next 6 to 12 months. Not all patients, however, are targets for the program. We combine the predictive model and program with an economic model, ranking members by likelihood death.



Example 4: Case Finding and Opportunity Analysis

Combining the predicted probability of the event, the likelihood of the member engaging, and the expected outcome (earlier transfer to hospice, for example) with the cost of intervention, we are able to determine which members represent an economic return for a particular program.

Depth of Dive Target

ROI Target

Member #	Dive %	Intervention Cost	Net-Net Opportunity	Cum Savings	Cum ROI
001	0.1%	$500	$5,000	5,000	10.0
002	0.2%	$500	$4,800	9,800	9.8
003	0.3%	$500	$4,300	14,100	9.4
…	….	$500		100,000	2.0
		$500	$0		
		$500	-$100	…	…
		$500	-$500		
1000	100%	…	…		

Max. Absolute Savings

Selected references

This is not an exhaustive bibliography. It is only a starting point for explorations.

- Duncan, I. Healthcare Risk Adjustment and Predictive Modeling. 2011 Actex Publication.
- Shapiro, A.F. and Jain, L.C. (editors); *Intelligent and Other Computational Techniques in Insurance;* World Scientific Publishing Company; 2003.
- Dove, Henry G., Duncan, Ian, and Robb, Arthur; *A Prediction Model for Targeting Low-Cost, High-Risk Members of Managed Care Organizations*; The American Journal of Managed Care, Vol 9 No 5, 2003
- Berry, Michael J. A. and Linoff, Gordon; *Data Mining Techniques for Marketing, Sales and Customer Support*; John Wiley and Sons, Inc; 2004
- Montgomery, Douglas C., Peck, Elizabeth A., and Vining, G Geoffrey; *Introduction to Linear Regression Analysis*; John Wiley and Sons, Inc; 2001
- Kahneman, Daniel, Slovic, Paul, and Tversky (editors); *Judgment under uncertainty: Heuristics and Biases;* Cambridge University Press; 1982.

Selected references (contd.)

- Winkelman R. and S. Ahmed. *A comparative analysis of Claims Based Methods of health risk assessment of Commercial Populations*. (2007 update to the SOA Risk-Adjuster study.) Available from the SOA; the 2002 study is on the website at: http://www.soa.org/files/pdf/_asset_id=2583046.pdf
- Duncan I: *Managing and Evaluating Healthcare Intervention Programs* (2013). Actex Publications. www.actexmadriver.com
- Lewis, GL, Duncan I, Kirkham H, Vaithianiathan R: *How Health Systems Could Avert "Triple Fail" Events that are Harmful, Are Costly and Result in Poor Patient Satisfaction."* Health Affairs 32 No. 4 (2013) 669-676.
- Duncan I: "*Mining Health Claims Data for Assessing Patient Risk*" In *Data Mining: Foundations and Intelligent Paradigms* Volume 3: *Medical, Health, Social, Biological and other Applications.* D.E. Holmes and L. Jain (eds.), Springer Verlag 2012.

Ian Duncan FSA FIA FCIA MAAA
Adj. Assoc. Professor
Dept. of Statistics & Applied Probability
University of California, Santa Barbara

(860) 614-3295
duncan@pstat.ucsb.edu
iduncan@sbactuaries.com

Developing your own model

Model Evaluation:

Most of us are familiar with techniques for evaluation of the accuracy of a regression model (R^2 anyone?)

For evaluating Logistic Regression models we need other techniques. One which is useful is the Receiver Operating Curve. The area under the ROC curve, often termed the ***c*-statistic** or ***c*-index c**orresponds to the likelihood that an event will have a higher predicted Probability than an event at random.

ROC Curve for Fitted Logistic Regression Model I Using the Sample Data