UMass Risk Adjustment Project for MassHealth Payment and Care Delivery Reform: Describing the 2017 Payment Model

Prepared by:

Arlene S Ash, PhD
Eric Mick, PhD

October 11, 2016

Table of Contents

Social Determinants of Health Modeling
Report of September 2016

The Social Determinants of Health Model was developed by UMass for MassHealth to improve its approach to risk adjustment and to incorporate, to the extent feasible, key Social Determinants of Health (SDH). The final model predicts a concurrent cost outcome that applies only to services that MassHealth expects its 2017 managed care payments and alternative payment models to cover. The model refines and enhances the State's previous DxCG medical-risk-based predictions, adding predictors for unstable housing, disability, agency relationships, serious mental illness and substance use disorders, plus a summary measure of "neighborhood stress" based upon residence in a census block group.

Risk adjustment models are designed to determine the *relative* resource needs of individual members, so that plans that enroll sicker members receive more money than plans that avoid them. It is not designed to determine the total amount of money allocated across all plans.

Methods:

We used concurrent models to predict costs only for those enrolled for at least one-half year in the CY2013 PCC plan (costs are annualized by the fraction of the year present and then top-coded at $125K). This reference sample was used to identify inherent differences in the need for services (and, therefore, expected cost) of individuals who enroll in the managed care (MCO) sector. This strategy enables us to rely on uniformly determined (administrative) rates that are not affected by differences in contracting costs, and is the same as CMS's approach to risk adjustment in their Medicare Advantage program. Our final model predicts the bundle of expenses that MassHealth currently intends its bundled payments to cover, extremely well. The model relies on the following predictors: DxCG v4.2 concurrent Medicaid RRS, selected medical conditions, social determinants of health (disability indicators, housing indicators, and medical/behavioral indicators), and neighborhood-based predictors of health.

Concurrent modeling allows for capturing the costs of several consequential subgroups of people that would be missed by a prospective model, including newborns and those with high end-of-life costs. A concurrent model is appropriate for making a payment in the future to an insured *group* of patients, so long as the underlying patient mix remains fairly stable.

The final risk-adjustment modeling proceeded as follows:

1. Only those medical expenditures from the CY2013 data that will be included in the State's 2017 MCO contract (see attachment) were used in constructing COST. This outcome captures 91% and 88% of oldCOST (mean reductions of $452 and $733) in the MCO and PCC plans, respectively.
   While COST was identical to oldCOST for 38% of MCO and 42% of PCC enrollees, more than 10% of members in each plan lost at least $700 in moving to the new outcome. For some people, reductions were sizeable; 1% of MCO members had drops of at least $4,700 and 1% of PCC members, over $20,500. See Appendix A for a list of inclusions and exclusions leading to COST.

2. Prior models (Basic and Full) were re-fit to the new outcome. The goal was to re-assess performance and recommendations from model building previously conducted to predict

total medical expenditures.
Note that top-coding at $125,000 eliminated $33.9 million (1.9%) of oldCOST dollars, incurred by members contributing 0.151% of all person-years, from the PCC plan and $91.9 million (3.8%), incurred by members contributing 0.177% of all person-years, from the MCO.

3. Model building steps

    - We defined COST2 by annualizing and top-coding COST at $125,000.
    - We used weighted least squares regression to predict COST2 in 2013 from 2013 patient characteristics for MassHealth members in the PCC plan.
    - We simplified the "neighborhood stress score" (NSS7) by basing it on the sum of each of the 7 previously-identified, standardized census block group variables (rather than a weighted sum from a principal components analysis). Each individual variable is first standardized by subtracting its mean and dividing by its standard deviation. Their sum is again standardized, so that the coefficient of NSS7 is the increment to expected cost associated with a 1 SD increase in NSS7. The minimum, median and maximum values of NSS7 are -1.75, -0.06 and +3.28.
    - We considered, but ultimately did not include "% living alone" as a neighborhood-level predictor. While "living alone" is likely an important predictor of an individual's cost, the "contextual" variable (percent living alone) would often reward providers for enrolling people in higher SES neighborhoods, such as Beacon Hill and Back Bay.
    - We also considered but did not include profound/severe DD (for which the COST2 coefficient would have been *negative* $1,800). Dropping this variable left the model R^2 unchanged to four digits of accuracy.
    - In a final round of edits, we also did not include any "stand-alone" disease indicators (such as, asthma/COPD, separately for kids and adults; diabetes; polyneuropathy; schizophrenia; and PTSD) that we had been considering because: 1) each is already recognized in the DxCG relative risk score which provides most of the final model's explanatory power, 2) these particular conditions had not emerged from a systematic search of all condition categories that might merit additional consideration, and 3) not explicitly including them as predictors only minimally reduced the model's predictive power (R^2 dropped by about 1/3 of 1 percentage point or less).

4. Although weighted least squares regression was our core modeling tool, we "groomed" the regression outputs as follows:

    - The fitted model included variables for "homeless, by ICD code" and for "3 or more distinct addresses in a year." Our pooled category of "Unstable Housing" marks people with either condition, using the coefficient for "3+ addresses" for the pooled group. In doing this, we over-rode the unacceptably high coefficient for "coded homelessness," because – in the 2013 data – homelessness was almost never coded, and when it was, it was largely found on hospital bills (presumably to help explain a long stay due to not having a safe discharge location). We expect that paying for homelessness will encourage providers to code it more often among people who are not hospitalized. Note that creating a pooled variable (rather than setting both coefficients to have the same value) avoids paying double for the rare person who is both coded as homeless AND has 3 or more addresses during the year.
    - We set the coefficient for the indicator for "failure to GeoCode" (which applies to about 6% of members, and had been negative) to 0.

- We set the minimum prediction to ~$15 to pay for the cost of keeping track of (a minimal requirement for managing the care of) a member and set the maximum to ~$125,000. These numbers are approximate, since when models setting payments they will be normalized – that is, multiplied by a constant such as an inflation factor of 1.03 – to ensure that total predicted dollars are set to the total amount of money available for payments for the entire population. Such "trimming" has been imposed by other payers.
  Setting the minimum to $15 adds $8.50 to the PCC average prediction and $7 to the MCO average; the amount of this due to setting the minimum at $15 rather than 0 is around 50 cents or less in each program.
- Note that the outcome predicted, COST2 was top-coded (after annualizing) at $125,000, which means that it is not designed to pay out all the money actually used in caring for MassHealth members. Specifically, top-coding COST2 dropped 3.0% of all dollars spent: $104 per person-year (1.8%) from the PCC plan and $191 per person-year (3.9%) in the MCO plan. The state will need to decide how to address "covered" costs that were excluded from this model.
- Top coding *our predictions of COST2* at $125,000, on the other hand, is not very consequential. In our data, it affects just 54 and 78 people (in the PCC and MCO plans, respectively), and eliminates about $21,500 and $27,500 per person affected, but only $3 and $4 on average in each population.
- Our near-final prediction, for a patient with characteristics **X** is now
  PRED(**X**) = max {15, min (f(**X**), 125000)}, where f(**X**) is a sum of amounts associated with each patient characteristic, as described in more detail below.
- In testing the resulting model for its fit to the PCC and MCO 2013 populations we multiply by constants, k, (called "inflators"):
  $PRED_{pcc} = k_{pcc} * PRED$, and
  $PRED_{mco} = k_{mco} * PRED$,
  where each inflator is chosen to make mean predicted cost equal mean actual COST2 in that population. In these data, $k_{pcc} = 1.0058$ and $k_{mco} = 1.0827$.
- Finally, since top-coding COST2 at $125,000 removed 1.8% and 3.9% of dollars from the PCC and MCO plans, respectively, some mechanism (such as using a larger inflator) will be needed to account for these dollars that were "top-coded out."
- In practice, predictions from the final model will be used only for MCO contracting, and the State and its actuarial consultants will determine appropriate multipliers to reflect inflation, regional differences in labor costs and other inputs, while ensuring that the total number of dollars predicted matches the total dollars allocated for the entire MCO population.

The final product of this work is a **newRRS**, created by normalizing the final (2013 PCC) model to have mean its mean equal 1 in the full 2013 modeling population (PCC and MCO members). This score can be used in the same way that MassHealth has previously used DxCG RRS to make payments to plans intended to cover the first $125,000 of (included) dollars spent on members present for at least 183 days in a year.

The Recommended model:

Our recommended model, based on predicting COST2 in 2013 data for PCC members with at least 183 days of enrollment has 28 predictors (27 degrees of freedom).

20 **age-sex category dummy variables**. These range from highs of about $1300 to $1650 for the youngest girls and boys, to some negatives for males, as low as -$400 to $650 for males aged 18 to 24, and those 55 and over.
1 DxCG **Relative Risk Score**. Each RRS unit adds about $3,600.
3 markers for **DMH clients**, (non-DMH) **DDS clients** and any other person entitled due to **Disability**. These factors add about $15,000, $2,800 and $1,500, respectively.
1 **Unstable housing**. Having 3 or more addresses and/or an ICD-code for homelessness adds about $600.
2 markers for **mental illness (SMI)** and **substance use (SUD)**, contributing $2500 and $2200, respectively.
1 standardized neighborhood stress score, NSS7. Each 1 SD increment of NSS7 adds about $50. Note that when addresses cannot be geocoded, we set NSS7 to 0.

The final prediction for the MCO population, PRED, is achieved by first bottom-coding at $15 and top-coding at $125,000 the outputs of the above formula, and then rescaling them (that is, multiplying them by the constant needed) to make their sum equal the sum of COST2 in the 2013 MCO population.

To make it easy to see how this model assesses *relative risk across MassHealth*, we use use **newRRS**, in which PRED is rescaled to have mean 1 in the 2013 PCC and MCO population.

Model Performance:

PRED has an R^2 of 57.2% for predicting COST2 in the 2013 PCC population, and 62.4% in the 2013 MCO. When used to predict *next* year's costs from this year's data, the model's explanatory power is estimated to be 38%, which is at the high end of best-performing prospective models in Medicaid populations.

Appendix Materials

Defining COST2

EXCLUDED are ALL costs in:

- Disbursement codes other than 0. This eliminated codes 1 (pass-through for other state agencies (DMH, DDS, DCF, DYS)), 2 (school-based Medicaid), and 6 (HSN); also, small dollars with Disbursement codes 3 and 4 (which, ideally should not be eliminated).
- Non-Emergency Transport (see code list in Section A2)
- Crossover (only 34 people, $8000 in MCO; 43 people, $7000 in PCC)
- Rest Homes
- ICFMR

Partly EXCLUDED are costs in:

- Dental claims (EXCLUSIONS are 89% of costs in MCO; 87% in PCC)
- NH/Chronic Inpatient (EXCLUSIONS are 8% in MCO; no such bills in PCC)
- Hospice (EXCLUSIONS are 4% in MCO; no exclusions in PCC)
- Early intervention (EXCLUSIONS are 5% in MCO; 9% in PCC)
- Chapter 766 (in MCO, 160K people, mean = $23; w 3% of costs EXCLUDED; in PCC, total cost is $171 for 5 people)
- Home Health Agency (only $2,000 of over $100 million in HHA costs are excluded)

INCLUDED are ALL costs in:

- Inpatient
- Outpatient
- Chiropractor
- Lab
- Abortion
- Pharmacy
- Any other cost that is not explicitly EXCLUDED

Excluded non-Emergency transport:

A0080 - NON-EMERGENCY TRANSPORTATION: PER MILE
A0090 - NON-EMERGENCY TRANSPORTATION: PER MILE
A0100 - NON-EMERGENCY TRANSPORTATION - TAXI, IN
A0110 - NONEMERG TRANSPRT & BUS, INTRA/INTER-ST
A0120 - NON EMERGENCY TRANSPORTATION - MINBUS/O
A0130 - NONEMERGENCY TRANSPORT (WHEELCHAIR VAN)
A0140 - NON-EMERGENCY TRANSPORT, AIR TRAVEL IN/
A0150 - NON-EMERGENCY AMBULANCE ONE WAY
A0160 - NON-EMERGENCY TRANSPORTATION: PER MILE
A0170 - NON-EMERGENCY TRANSPORTATION: ANCILLARY
A0180 - NON-EMERGENCY TRANSPORTATION: ANCILLARY
A0190 - NON-EMERGENCY TRANSPORTATION: ANCILLARY
A0200 - NON-EMERGENCY TRANSPORTATION: ANCILLARY
A0210 - NON-EMERGENCY TRANSPORTATION: ANCILLARY
A0215 - ALPHA-NUMERIC RESP. RELATED EQUIP & SUP

A0222 - AMBULANCE SERVICE, RETURN TRIP, TRANSPO
A0300 - AMB SERV., BASIC LIFE SUPPORT, NON EMER
A0304 - AMB SERV., NO ADV LIFE SUPPORT, NON-EME
A0306 - AMB SERV., ADV LIFE SUPPORT, NON-EMERGE
A0320 - AMB SERV., BLS, MILES SEPARATE, NON-EME
A0360 - AMBULANCE (BLS) NON-EMERGENCY ONE WAY
A0382 - BLS ROUTINE DISPOSABLE SUPPLIES
A0384 - BLS SPECIALIZED SERV DISPOS SUPPLIES DE
A0426 - AMBULANCE SRVC ADVANCED LIFE SUPPORT NO
A0428 - AMBULANCE SRVC BASIC LIFE SUPPORT NON-E
Q3020 - ALS VEHICLE USED, NON-EMERGENCY TRANSPO
S0215 - NONEMERG TRANSPORTATION, MILEAGE (WHEEL
S9960 - AIR AMBULANC NONEMERG FIXED
S9961 - AIR AMBULAN NONEMERG ROTARY
T2001 - NON-EMERGENCY TRANSPORTATION; PATIENT A
T2002 - NON-EMERGENCY TRANSPORTATION; PER DIEM
T2003 - NON-EMERGENCY TRANSPORT, ENCOUNTER/TRIP
T2004 - NON-EMERGENCY TRANSPORT; COMMERCIAL CAR
T2005 - NON-EMERGENCY TRANSPORTATION; NON-AMBUL
T2006 - AMBULANCE RESPONSE AND TREATMENT, NO TR
T2007 - TRANSPORTATION WAITING TIME, AIR AMBULA
T2049 - NON-EMERGENCY TRANSPORTATION; STRETCHER

Answers to expected "Frequently Asked Questions" (FAQs)

How does the model recognize differences in the costs of infants vs. kids vs. adults?

- Use of coefficients within 10 age categories (0-1, 2-5, 6-12, 13-17, 18-24, 25-34, 35-44, 45-54, 55-59, 60+), separately for male and female
- Use of the DxCG model that both
  - Recognizes the costs of diseases, including empirically-identified differences in costs for diseases when they are differentially costly for kids (<18 years)
  - Uses second-stage "tuning" to ensure that average costs are right within age-sex categories that (among other things) distinguish infants (aged 0 or 1) from other young people and older adults

What fraction of people have (annualized) COST2 > $125,000? How many dollars are spent on people above that threshold?

- Among PCC members, 556 members (representing (0.151% of PCC person-years) had annualized COST2 greater than $125,000. In the MCO program, 991 such people contributed 0.177% of person-years. Thus, in the combined population, there were 1,547 people representing 0.166% of all person-years.
- Thus, truncation removes $33.90 million and $91.85 million dollars from the two plans, respectively.
- The removed dollars reduce the two population means by $104 and $191, representing 1.8% and 3.9% of original COST2 dollars in the two plans, respectively, and 3.0% overall.

How is the "neighborhood stress score" (NSS7) calculated?

- We first standardize (on the combined PCC and MCO population) each of the following 7 census block group (CBG) variables: % of families with incomes < 100% of the federal poverty level (FPL), % < 200% of FPL, % of adults who are unemployed, % of households receiving public assistance, % of households with no car, % of households with children and a single parent, and % of people age 25 or older who have no HS degree.
- We then add these variables and standardize their sum, so that the coefficient of NSS7 is the increment to expected cost associated with a 1 SD increase in NSS7.

What does the distribution of NSS7 look like?

- By design, the mean of NSS7 in the 2013 Medicaid modeling population is 0 and its SD is 1; its minimum, median and maximum values are -1.75, -0.06 and +3.28.
- Its means in the PCC and MCO populations are -.044 and +0.030.

Are risk scores ever negative?

- No. We "bottom code" all predictions at a value that translates into at least $15.
- Bottom coding is needed because some age-sex category coefficients (only among males, aged 18 and above) are negative; the most negative is about -$580 - it is for males between the ages of 18 and 24. Thus, without bottom coding, enrolling a 20-year-old man with no additional risk factors would lead to a loss of over $500!
- The only other predictor in the model that could subtract dollars is NSS7. Its coefficient is less than $50 and its lowest value, larger than -2, so it would never subtract as much as $100 – and, if this ever contributed to a prediction being smaller than $15, bottom-coding would be used to raise it.

What codes are included in "serious mental illness" (SMI)?

HCC	Chronic	Description
160	-	PSY.15 Acute Paranoid Reaction and Confusion
161	C	PSY.20 Schizophrenia
162	C	PSY.30 Other Nonorganic Psychosis
163	C	PSY.40 Delusional Disorder and Paranoid States
166	C	ANG.20 Bipolar Disorder
168	C	ANG.40 Major Depression

What codes are included in "substance use disorders"?

HCC	Chronic	Description
148	-	SAD.15 Drug Induced Hallucinations, Delusions, and Delirium
149	C	SAD.20 Withdrawal and Other Specified Drug-Induced Mental Disorders
150	C	SAD.30 Drug Dependence
151	C	SAD.40 Drug Abuse without Dependence, Except Alcohol and Tobacco
152	C	SAA.20 Alcohol Psychosis
153	C	SAA.30 Alcohol Dependence
154	C	SAA.40 Alcohol Abuse, Without Dependence

Table 1 Population Characteristics, COSTS, COST2, and Risk Scores for MCO and PCC Members in CY2013

	PCC CY2013					MCO CY2013				
Members	357,660					524,607				
Member years	326,501					480,389				
Population statistics	**Mean**		**SD**		**Median**	**Mean**		**SD**		**Median**
Enrollment duration (months)	11.2		1.5		12.0	11.2		1.4		12.0
Age in years	26.1		18.6		22.0	21.6		17.0		22.0
COST	6,371		12,913		2,038	5,179		11,004		1,811
COST2	5,590		11,684		1,719	4,694		10,395		1,475
Relative Risk Score (RRS)	1.16		2.29		0.42	0.89		1.88		0.33
NSS7	(0.04)		1.03		(0.10)	0.03		1.06		(0.03)
Categorical data	**Member Years**	**%**	**Mean**			**Member Years**	**%**	**Mean**		
			COST	**COST2**	**RRS**			**COST**	**COST2**	**RRS**
Age < =18	148,396	45.5	3,643	3,195	0.53	263,317	54.8	3,123	2,725	0.44
Female	164,619	50.4	6,624	5,804	1.21	270,862	56.4	5,552	5,063	0.97
Race/ethnicity										
White/non-Hispanic	122,968	37.7	8,071	6,960	1.46	165,566	34.5	6,294	5,711	1.10
Black/non-Hispanic	33,744	10.3	6,264	5,545	1.16	48,682	10.1	4,915	4,453	0.83
Hispanic	32,004	9.8	6,119	5,391	1.13	60,494	12.6	4,642	4,157	0.81
Other non-Hispanic	20,331	6.2	3,645	3,182	0.62	21,828	4.5	3,579	3,202	0.59
Missing/unknown	117,454	36.0	5,162	4,638	0.96	183,819	38.3	4,613	4,194	0.78
Unstable housing										
Homeless, by ICD-9 coding	2,191	0.7	33,171	32,647	6.19	76	0.0	30,326	29,745	5.93
3 or more addresses in a year	37,694	11.5	9,010	8,441	1.67	55,250	11.5	7,178	6,734	1.25
Additional risk factors										
Family income ≤ 86% of FPL	243,290	74.5	7,247	6,342	1.33	334,027	69.5	5,780	5,240	1.00
Client of DMH	4,817	1.5	29,929	28,150	3.39	2,110	0.4	21,631	20,419	2.79
Client of DDS (not DMH)	8,022	2.5	22,047	11,647	2.42	5,067	1.1	16,377	9,648	1.89
Other Disabled	58,802	18.0	14,092	12,503	2.82	51,294	10.7	14,551	13,132	2.66
Serious mental illness (SMI)	46,962	14.4	17,050	15,570	3.02	49,038	10.2	14,240	13,277	2.67
Substance use disorder (SUD)	31,401	9.6	17,457	16,488	3.46	29,697	6.2	16,469	15,667	3.15
Persistent LTSS use	10,955	3.4	42,042	26,943	4.54	7,426	1.5	28,654	26,231	3.68
MassHealth Standard	257,678	78.9	6,807	5,883	1.21	410,743	85.5	5,223	4,704	0.89
Could not be geocoded	19,650	6.0	6,937	6,267	1.33	24,454	5.1	6,087	5,643	1.07

Source: MassHealth Medicaid Management Information System

Notes: Analyses are weighted (WGT = fraction of the year during which the member was eligible) and include only those enrolled for at least 183 days, representing 91.6% and 78.0% of all enrolled months for PCC and MCO members, respectively; **COST** and **COST2** are top-coded at $125,000 and annualized = minimum (dollars spent/WGT, 125000); **RRS** is the DxCG v4.2 concurrent model 312 risk score, normalized to have mean = 1 in the combined 2013 PCC/MCO population and used to identify medical conditions (**SMI**, **SUD**); **Disabled** = Eligible for aid due to disability, as flagged in MMIS; **Persistent LTSS use** = at least $500 (included in COST, but not necessarily in COST2) of long-term supportive services in each of 3 consecutive months.

Table 2 Coefficients and R2 for Models Predicting Top-Coded CY2013 COST2 Separately for PCC & MCO Members

Models	PCC alone (N = 357,660)				
		(1) Base	(3) Modified		(3') Final
	%*	Coeff	Coeff	t-stat	Coeff
Intercept	100		1105	19.7	
Age-sex categories					
0-1 Female	1.8		104	1.0	**1208**
2-5 Female (Referent)	5.2				**1105**
6-12 Female	7.9		-276	-3.8	**829**
13-17 Female	5.7		-178	-2.3	**926**
18-24 Female	5.2		-393	-4.9	**711**
25-34 Female	7.3		-346	-4.7	**758**
35-44 Female	6.3		-826	-10.8	**279**
45-54 Female	6.0		-728	-9.2	**377**
55-59 Female	2.6		-611	-6.1	**494**
≥ 60 Female	2.4		-624	-6.1	**481**
0-1 Male	1.9		182	1.7	**1287**
2-5 Male	5.4		431	5.5	**1535**
6-12 Male	8.6		157	2.2	**1262**
13-17 Male	6.3		-386	-5.1	**718**
18-24 Male	4.8		-1683	-20.7	**-578**
25-34 Male	6.0		-1148	-14.8	**-43**
35-44 Male	5.5		-1098	-13.9	**6**
45-54 Male	6.4		-1277	-16.5	**-172**
55-59 Male	2.6		-1555	-15.7	**-451**
≥ 60 Male	2.0		-1457	-13.4	**-352**
RRS (normalized)	116.4	4804	3337	514.7	**3337**
DMH client	1.5		13804	124.0	**13804**
Not DMH but DDS client	2.5		2569	30.4	**2569**
All other disabled	18.0		1428	35.6	**1428**
Homeless, by ICD-9 coding	0.7		7089	44.1	**555** *
3+ addresses in a year	11.5		555	13.6	**555**
Serious mental illness (SMI)	14.4		2295	54.2	**2295**
Substance use disorder (SUD)	9.6		2014	40.6	**2014**
NSS7s	-4.4		43	3.4	**43**
Not able to geocode (flag)	6.0		-126	-2.3	**-**
Model degrees of freedom		**1**	**29**		**27**
Model R-squared (x 100)		**53.48**	**57.29**		**57.22**
R-squared after top- and bottom-coding predictions					**57.244**

Source: MassHealth Medicaid Management Information System (MMIS)

Notes: Analyses are weighted (WGT = fraction of the year during which the member was eligible) and include only those enrolled for at least 183 days; Cost is top-coded at $125,000 and annualized = minimum (dollars spent/WGT, 125000); RRS is the DxCG v4.2 Percent (%) columns for continuous variables RRS and NSS7 = mean x 100. Specifically, mean RRS is 1.16 in the PCC and 0.89 in the MCO; mean NSS7 is -0.044 in the PCC and +0.030 in the MCO. R2s in the final columns of each block represent what would be achieved if the predictions from the formulas were normalized (that is, multiplied by the required constant, an inflation factor of < 1%) to ensure that mean PRED = mean COST2.

For more information, please contact Arlene Ash at (508) 856-8922 or arlene.ash@umassmed.edu