Supplemental Digital Content for

Risk-Adjusted Payment and Performance Assessment for Primary Care

Randall P. Ellis[1,2] and Arlene S. Ash[2,3]

February 2, 2012

[1] Boston University, Department of Economics

[2] Verisk Health Inc.

[3] University of Massachusetts Medical School, Department of Quantitative Health Sciences

Reprints: Randall P. Ellis, Department of Economics, Boston University, 270 Bay State Road, Boston MA 02215. Email: ellisrp@bu.edu

Acknowledgements

This research was supported by Verisk Health Inc. and a grant from The Commonwealth Fund. Dr. Ash was partially supported by UL1RR031982 from the National Center for Research Resources. The ideas here are preliminary and do not necessarily reflect the opinions of the University of Massachusetts, Boston University, Verisk Health, The Commonwealth Fund, or any other organization or entity.

Appendix A. Stability of Diagnosis-Based Risk Adjustment Model Over Time, Across Plans, and Across Data Sources

Appendix B. Further Results for PCAL Models

Appendix C. Implementation Steps Used to Guide CDPHP Implementation of the PCAL Model

Appendix D. Further Details about Performance Assessment Models

Appendix A. Stability of Diagnosis-Based Risk Adjustment Model Over Time, Across Plans, and Across Data Sources

In this appendix we examine the stability of the DxCG/Verisk Health Hierarchical Condition Category (HCC) predictive models over time, and across diverse plan and benefit types. Because the Primary Care Activity Level (PCAL) concept and bonus model described in the attached paper have only just been developed, they have not yet been subjected to extensive validation across time, plans, and datasets. However the DxCG/Verisk Health Hierarchical Condition Category (HCC) classification system on which these new models are based has been validated along all of these dimensions, which we document here.

Tables A1 through A3 present summary statistics from 2002 through 2007 for the concurrent HCC model using the Version 6.2 classification system. The only change in the DxCG/VH software over the six-year period is the addition of diagnostic codes, which Verisk Health updates annually in response to the addition of new ICD-9-CM diagnoses by the American Medical Association. All normalized risk scores were generated from the same regression model, using 2001 MarketScan data. Hence, regression weights on each HCC were held constant over this six-year period.

Table A1 shows that although the MarketScan sample size for enrollees with pharmacy benefit coverage grew from 5.6 million to 21.0 million enrollees from 2002 to 2007, the sample means of gender, age, numbers with coded diagnoses, and numbers of HCCs remained relatively stable over this time frame. Inpatient, outpatient, and pharmacy costs grew less rapidly in the MarketScan data than national averages, presumably because younger and healthier enrollees are dominating the growth in sample size.

Table A2 shows that rates of coded conditions over five years are increasing, but relatively stable over the six-year period. Rates have not been adjusted for the age and gender of the samples, which were becoming healthier, but reflect the common finding of more conditions being coded over time. Conditions are organized into 30 aggregated condition categories. Similar patterns are visible in the more detailed 184 condition categories (not shown).

Figures A1 and A2 illustrate that the concurrent and prospective risk scores generated using different years are remarkably stable, even as the MarketScan sample became gradually younger, but the prevalence of diseases overall grew modestly. The modest “sawtooth” pattern shown in each figure is largely an artifact of the feature that the size of the risk score intervals used on the horizontal axes are not constant, but become progressively wider (while encompassing fewer people) as the risk scores increase.

Table A3 shows that there have been significant changes in the composition of health plans in the MarketScan data, with large declines in comprehensive coverage plans, together with large growth in preferred provider (PPO) and point of service (POS) plans. Consumer-directed health

(CDHP) plans were nonexistent in the MarketScan data at the outset, growing to 3.83 percent of the sample in 2007.

The second panel of Table A3 reveals noticeable differences in mean age across plans, which also are associated with differences in diagnosis-based concurrent risk scores by plan type. Of particular note is that the health maintenance organizations (HMOs) and CDHP enrollees in the sample initially had average ages significantly lower than the sample average, which corresponds to risk scores being substantially below the sample mean. Because CDHPs did not exist in the 2001 data used to generate the concurrent DxCG/Verisk risk model, Version 6 models did not include any claims or information from CDHP models when calibrated using 2001 data. Despite this, the Version 6 models do well when applied to claims from 2007 CDHP plans. The bottom panel of Table A3 reveals that the model R^2 is similarly high across all plan types once the fitted risk scores are lightly recalibrated with the equation $Yhat_i = a + b*nRS_i$. (a and b are parameters specific to each plan type; nRS_i is the normalized risk score for person i using the commercial risk adjustment model calibrated using only indemnity, HMO, PPO, POS and POS with capitation data from 2001.)

One anomaly that stands out in Table A3 is the unusually low R^2 for PPO plans and All Plans in 2003 (19.9% and 25.0% respectively) versus means of 40% or better for other years and other health plans. Upon further investigation it was discovered that there was one individual in the PPO with total health spending in that year totaling $16,501,470, which is likely to be an error. Even though there are over 3 million people in the PPO, the squared error from this one person reduces the R^2 for the PPO regression in 2003 by .04 and has a similar impact on the overall R^2 for all plan types. Further examination identified 19 people with costs in 2003 exceeding $2 million, 18 of which were in the PPO, which collectively largely explain most of the difference. As a result of finding this problem back in 2005, DxCG/Verisk Health improved its data cleaning process by looking for high cost outliers when building further risk adjustment models. This sensitivity also highlights the value of using truly enormous datasets, in recent years over 17 million people, to reduce the sensitivity of the regression coefficients and measures of goodness of fit to a few extreme outliers.

TABLE A1. Summary Statistics Using MarketScan Private Insurance Claims Data, 2002-2007

Sample sizes	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
Total population	5,675,640	7,707,004	13,136,933	17,560,383	16,159,068	28,761,500
Population with Rx benefits	5,639,035	7,707,004	11,079,885	15,102,855	14,441,622	21,176,556

Note: All remaining statistics in this appendix refer to the population with Rx benefits.

	Percent of Population					
Categorical Variables	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
Male	47.51	47.42	48.19	48.28	48.48	48.51
Female	52.49	52.58	51.81	51.72	51.52	51.49
Child (0-17 yrs)	23.50	25.65	26.32	24.85	26.55	26.75
Young adult (18-44)	40.18	39.80	41.03	40.93	40.66	41.01
Older adult (45-64)	36.17	34.46	32.65	33.15	32.78	32.23
Senior (65+)	0.14	0.09	0.00	1.07	0.00	0.00
No diagnoses	23.60	25.28	26.36	25.78	23.94	24.22
No diagnoses mapping into HCCs	23.64	25.32	26.42	25.82	23.96	24.60
	Weighted Population Means					
Variable	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
Number of HCCs (V6)	2.78	2.71	2.69	2.74	2.83	2.89
Number of Diags (V6)	3.98	3.90	3.93	3.73	4.09	4.22
Age in years	34.87	33.99	33.32	33.75	32.83	32.58
Inpatient cost	$ 671	$ 698	$ 748	$ 717	$ 804	$ 809
Outpatient cost	$ 1,407	$ 1,423	$ 1,555	$ 1,508	$ 1,745	$ 1,785
Drug cost	$ 616	$ 613	$ 619	$ 612	$ 689	$ 644
Total cost	$ 2,693	$ 2,734	$ 2,921	$ 2,836	$ 3,238	$ 3,238
DxCG/VH software version	2.0.2	2.0.2	2.0.2	2.1.1	2.3.0	2.3.0
HCC version no.	6.2	6.2	6.2	6.2	6.2	6.2

Note: All results derived from processing the Thomson Reuters MarketScan Commercial Claims and Encounter database (formerly MEDSTAT) through the DxCG/Verisk health software, using only the sample of individuals with pharmacy benefit coverage.

TABLE A2. Rate per 10,000 of Aggregated Condition Categories (ACCs) Using DxCG/VH Model 6 on MarketScan Privately Insured Data, 2002-2007

ACC	Label	2002	2003	2004	2005	2006	2007
001	Infectious and Parasitic	870	884	827	858	910	915
002	Malignant Neoplasm	192	173	166	161	167	165
003	Benign/In Situ/Uncertain Neoplasm	914	821	810	801	842	838
004	Diabetes	385	378	386	398	415	420
005	Nutritional and Metabolic	1,277	1,240	1,324	1,340	1,389	1,473
006	Liver	118	115	116	117	116	120
007	Gastrointestinal	1,108	1,076	1,096	1,080	1,120	1,153
008	Musculoskeletal and Connective Tissue	2,188	2,093	2,132	2,105	2,196	2,182
009	Hematological	264	246	254	252	255	274
010	Cognitive Disorders	48	44	46	42	41	42
011	Substance Abuse	109	91	117	125	110	160
012	Mental	653	628	639	610	677	696
013	Developmental Disability	101	106	125	124	139	148
014	Neurological	389	383	391	390	406	411
015	Cardio-Respiratory Arrest	33	32	31	32	31	32
016	Heart	1,256	1,211	1,245	1,210	1,260	1,264
017	Cerebro-Vascular	83	79	78	76	78	75
018	Vascular	180	173	177	173	178	178
019	Lung	1,083	1,111	950	1,107	1,031	1,053
020	Eyes	777	810	794	813	879	877
021	Ears, Nose, and Throat	2,770	2,754	2,602	2,754	2,778	2,814
022	Urinary System	616	594	604	602	619	617
023	Genital System	1,144	1,024	995	955	989	997
024	Pregnancy-Related	193	208	205	211	208	208
025	Skin and Subcutaneous	1,432	1,333	1,339	1,339	1,416	1,428
026	Injury, Poisoning, Complications	1,450	1,391	1,379	1,356	1,381	1,368
027	Symptoms, Signs and Ill-Defined Conditions	2,893	2,853	2,886	2,940	3,063	3,125
028	Neonates	66	75	73	80	82	82
029	Transplants, Openings, Other V-Codes	16	15	17	17	16	18
030	Screening/History	4,032	3,961	3,870	4,058	4,296	4,385

Note: Numbers shown are rates per 10,000 eligible members, without adjusting for changes in age or gender over time.

TABLE A3. Summary Statistics by Plan Type and Year, MarketScan Commercial Data, 2002-2007

Plan Type	2002	2003	2004	2005	2006	2007
Frequency						
All Plan Types	5,639,035	7,707,004	11,079,885	15,102,855	14,441,622	21,176,556
Percent of Population						
Comprehensive	16.44	13.92	10.90	6.16	6.42	2.47
EPO	3.95	0.44	0.14	0.20	0.48	0.77
HMO	16.41	21.49	18.28	28.05	19.96	17.10
POS	13.39	15.98	11.94	9.90	11.69	9.93
PPO	37.01	42.06	55.11	52.26	56.23	62.40
POS with Capitation	12.78	5.98	2.64	1.76	0.84	0.46
CDHP	-	-	-	0.33	1.11	3.83
Unknown Plan Type	0.01	0.14	0.98	1.34	3.27	3.03
Mean Age						
Comprehensive	39.64	40.04	39.32	41.89	42.01	40.36
EPO	31.45	28.70	29.91	30.33	29.85	32.05
HMO	31.74	31.28	31.37	32.49	31.19	31.23
POS	34.23	33.01	32.88	33.45	32.81	32.81
PPO	35.81	34.05	32.98	33.63	32.53	32.76
POS with Capitation	31.74	32.25	30.72	33.24	32.52	32.34
CDHP	-	-	-	30.82	30.65	30.76
Unknown Plan Type	28.09	33.14	34.62	31.39	31.27	31.77
All Plan Types	34.87	33.99	33.32	33.75	32.83	32.58

Mean Concurrent Risk Score						
Comprehensive	1.554	1.643	1.652	1.820	1.803	1.751
EPO	1.052	0.965	1.050	0.946	0.960	1.092
HMO	0.891	0.973	1.093	1.106	1.130	1.205
POS	1.161	1.113	1.140	1.169	1.159	1.284
PPO	1.367	1.213	1.191	1.212	1.215	1.276
POS with Capitation	0.964	1.039	1.011	1.036	1.127	1.102
CDHP	-	-	-	0.819	0.793	0.946
Unknown Plan Type	0.318	0.681	1.247	1.206	1.118	1.154
All Plan Types	1.228	1.193	1.213	1.210	1.219	1.258
Eligibility-Weighted Concurrent R – Squared						
Comprehensive	0.454	0.443	0.456	0.439	0.438	0.451
EPO	0.415	0.348	0.123	0.502	0.454	0.475
HMO	0.386	0.280	0.437	0.329	0.360	0.363
POS	0.455	0.454	0.447	0.457	0.448	0.416
PPO	0.451	0.199	0.443	0.421	0.390	0.418
POS with Capitation	0.365	0.392	0.397	0.366	0.411	0.368
CDHP	-	-	-	0.467	0.376	0.431
Unknown Plan Type	0.310	0.434	0.463	0.322	0.376	0.414
All Plan Types	0.433	0.250	0.440	0.398	0.392	0.409

Notes: EPO = exclusive provider organization, HMO = health maintenance organization, POS = point of service, PPO = preferred provider organization, CDHP = consumer directed health plan. Results in this table were generated using the raw claims in the MarketScan database, without excluding certain plan types (EPO, POS and unknown) that were not used when generating the concurrent HCC model. All results reflect weighting by fraction of the year eligible. Each year of data was analyzed independently. R-squared values are from regressing total health spending in each plan separately on the HCC risk score, and hence allow two degrees of freedom that permit the mean payment by plan type to be predicted exactly.

FIGURE A1. Distribution of Concurrent DCG/HCC Normalized Risk Scores, MarketScan Private Claims Data 2002-2007 Using DxCG Model 6.2

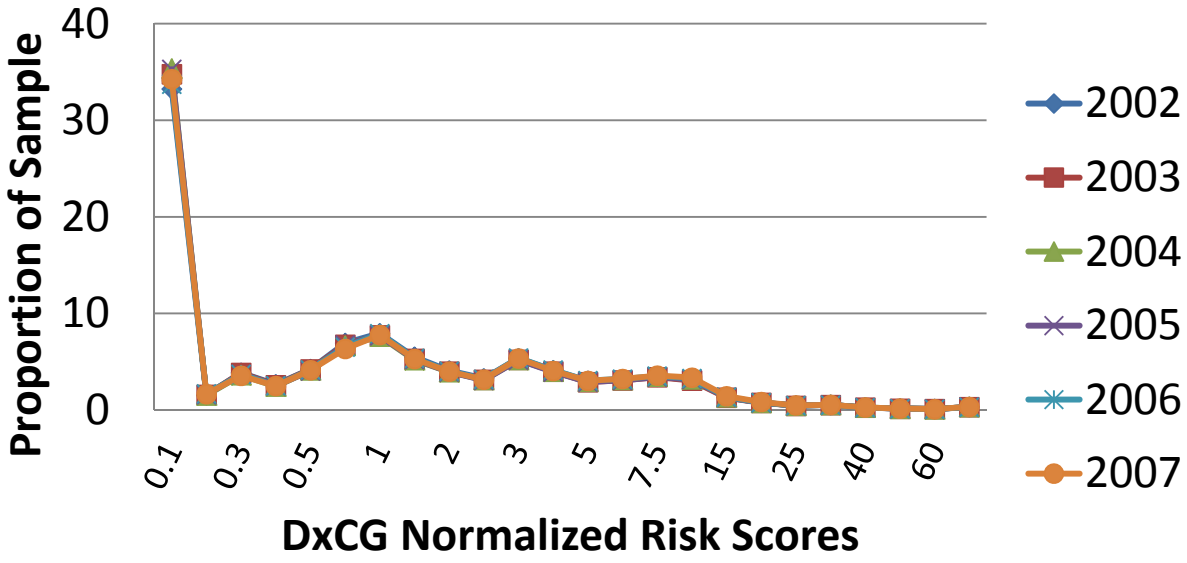
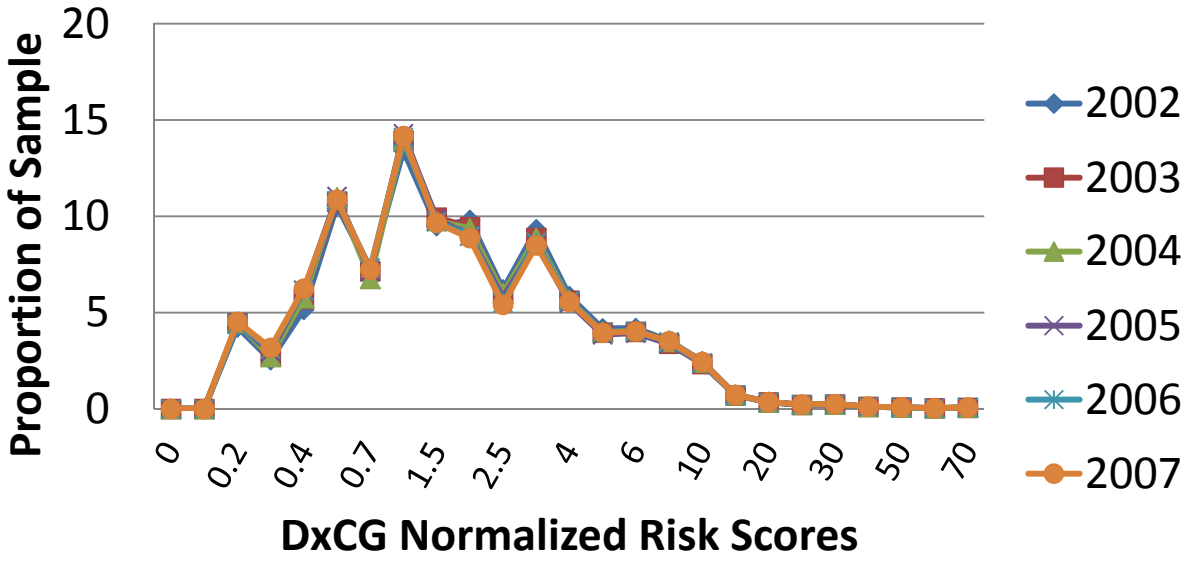


FIGURE A2. Distribution of Prospective DCG/HCC Normalized Risk Scores, MarketScan Private Claims Data 2002-2007 Using DxCG Model 6.2



Appendix B. Further Results for PCAL Models

Table B1 shows the information used to assign weights to each service spending component in the PCAL proxy spending variable. The first column shows the time allocated to providing primary care to people according to their service type from a small survey of practicing clinicians; the second shows the percent of spending consistent with this time allocation.

Table B2 shows summary statistics for the PCAL model dependent variable actual and predicted values.

We used a series of exclusions and restrictions on parameters until the following conditions were satisfied:

(1) T ratio is at least 4.

(2) Product of frequency of group and t ratio is at least 2000.

(3) Main effect is positive (no negative predictions for healthy adults).

(4) Interactions are always nonnegative when combined with individual terms.

In a few cases we chose to include low frequency HCCs when the coefficients were plausible and conditions had high cost weights and statistical significance.

After running a first stage regression, we used the fitted values in a second stage regression on each of 22 age-sex categories to ensure that all intercepts are nonnegative and slopes are appropriate for each age and sex cell. For example, for a given set of HCCs, the risk score for a teenage male has less impact on either a teenage female or an adult male age 45.

Split sample results in the paper were generated using only the first stage regressions. Second stage regression models were used to generate predicted PCAL as described in the main papers. Results from the split sample test are shown in Table B3.

TABLE B1. Primary Care Time Allocation and Corresponding Cost Fractions Included in PCAL

Activity	Estimated Time Allocation by PCPs	Fractions of spending included in PCAL
Primary care core services	50%	100%
Specialty care related	5%	6%
Hospital care related	5%	6%
ED visit related	5%	17%
Prescription drug related	10%	12%
Administration, education, email, phone calls, etc.	25%	***

***Costs for education, email, and phone calls, even for fully healthy patients, were recognized by adding $65 to PCAL.

The first column shows average, rounded results from a survey of 5 primary care physicians asked how their time was allocated across the listed activities. The second column shows the percentage of each kind of spending needed to approximate these time allocations. Primary care core services were included as 100% of costs of all evaluation and management services plus low-cost tests typically done by PCPs. For example, 17% of ED visit spending yielded a dollar amount from ED visits that was approximately one tenth of the dollar amount for the primary care core.

TABLE B2. Patient Level Summary Statistics for Normalized Actual and Predicted PCAL Proxy

	Actual PCAL Proxy	Predicted PCAL
N	17,419,585	17,419,585
Mean	1.000	1.000
Std. Dev.	1.408	1.152
Skewness	6.061	5.207
Kurtosis	73.259	57.271
Maximum	74.551	58.581
Median	0.589	0.681
Mode	0.099	0.133
Minimum	-99.377	0.114
% Negative	0.005%	0%

Notes: Calculated using the Thomson Reuters MarketScan® 2007 Commercial Data. Predicted PCAL is based on regression analysis using 569 parameters on age, gender, and hierarchical condition categories (HCCs) and selected interactions. Results are after normalizing the PCAL proxy for the sample mean PCAL proxy of $659.

The 2007 estimation sample contained 985 people (.005% of the total sample) with negative PCAL proxy values. While we intended to include people with negative spending, we did not realize until writing this paper that one person had a truly exceptional negative value of $65,000. Fortunately, with 17.4 million people this outlier has negligible effect on results.

TABLE B3. Split Sample Regression Summary Statistics for Normalized PCAL Model with SAS Code

```
*Data is 2007 MarketScan data, N=17,419,585 Sample was randomly divided into two halves;
/*SAS uses y1 to create model equation, but preds are created for all members - even if y1 is
missing*/
*Estimation sample;
proc reg data = temp1;
title "Split Sample Preliminary Regression";
weight eligf07;
model y1 = as02-as34 hcc001-hcc394 khcc001-khcc394 ihcc001-ihcc394 d01-d35;
output out = temp2(keep=enrolid y1 pred1 eligf07 y2) pred=pred1;
```

Number of Observations Used 8709372
Weight: eligf07 = fraction of the year each enrollee was eligible for plan in 2007

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1165	11507604	9877.77158	15033.0	<.0001
Error	8.71E6	5721920	0.65707		
Corrected Total	8.71E6	17229524			

Root MSE	0.81060	R-Square	0.6679
Dependent Mean	1.00002	Adj R-Sq	0.6679
Coeff Var	81.05808		

Validation Regression using one degree of freedom on second half of data

```
/*validate y1 model using y2 actuals*/
proc reg data = temp2;
title "Split Sample Validation Regression";
weight eligf07;
model y2=pred1;
```

Number of Observations Used 8710213
Weight: eligf07 = fraction of the year each enrollee was eligible for plan in 2007

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	11533024	11533024	1.744E7	<.0001
Error	8.71E6	5760674	0.66137		
Corrected Total	8.71E6	17293699			

Root MSE	0.81325	R-Square	0.6669
Dependent Mean	0.99998	Adj R-Sq	0.6669
Coeff Var	81.32659		

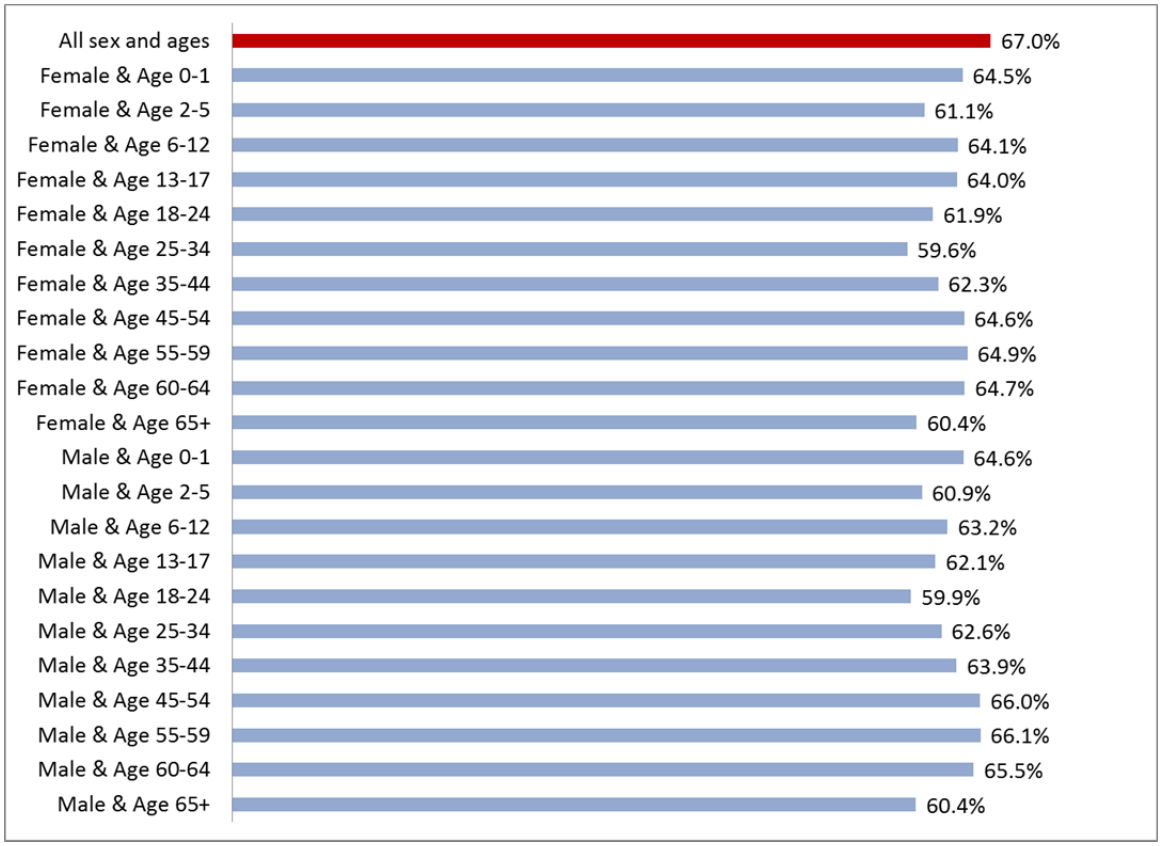
TABLE B4. Comparing PCAL Model Performance by Plan Type

		Mean PCAL Proxy		Standard Deviations		
Plan	**N**	**Actual**	**Predicted**	**Actual**	**Predicted**	**R^2**
All	17,419,585	1.000	1.000	1.408	1.152	67%
Comprehensive	447,956	1.209	1.197	1.756	1.477	66%
Health Maintenance Organization (HMO)	3,266,406	0.951	0.967	1.349	1.104	67%
Non-Capitated Point-of-Service (POS)	1,858,966	1.042	1.000	1.418	1.105	68%
Preferred Provider Organization (PPO)	11,154,794	1.005	1.013	1.408	1.169	68%
Consumer-Directed Health Plan (CDHP)	691,463	0.912	0.811	1.376	0.938	68%

PCP: primary care provider; PCAL: primary care activity level; HCC: hierarchical condition category.

Notes: All calculations use Thomson Reuters MarketScan 2007 Commercial Data. Predicted PCAL is based on regression analysis using 653 parameters on age, gender and HCCs and selected interactions. The mean PCAL proxy in the full data is $659; mean and SD values reported in the table are normalized to average 1.0. R^2 values for each plan type were calculated by regressing the PCAL variable (in this subpopulation) on the predicted PCAL from the full data set, thus allowing each plan type to have its own constant and intercept term. (Eg, for the 3.3 million people enrolled in HMOs, we fit $PCAL_{HMO} = a + b\ PCAL_{All}$.)

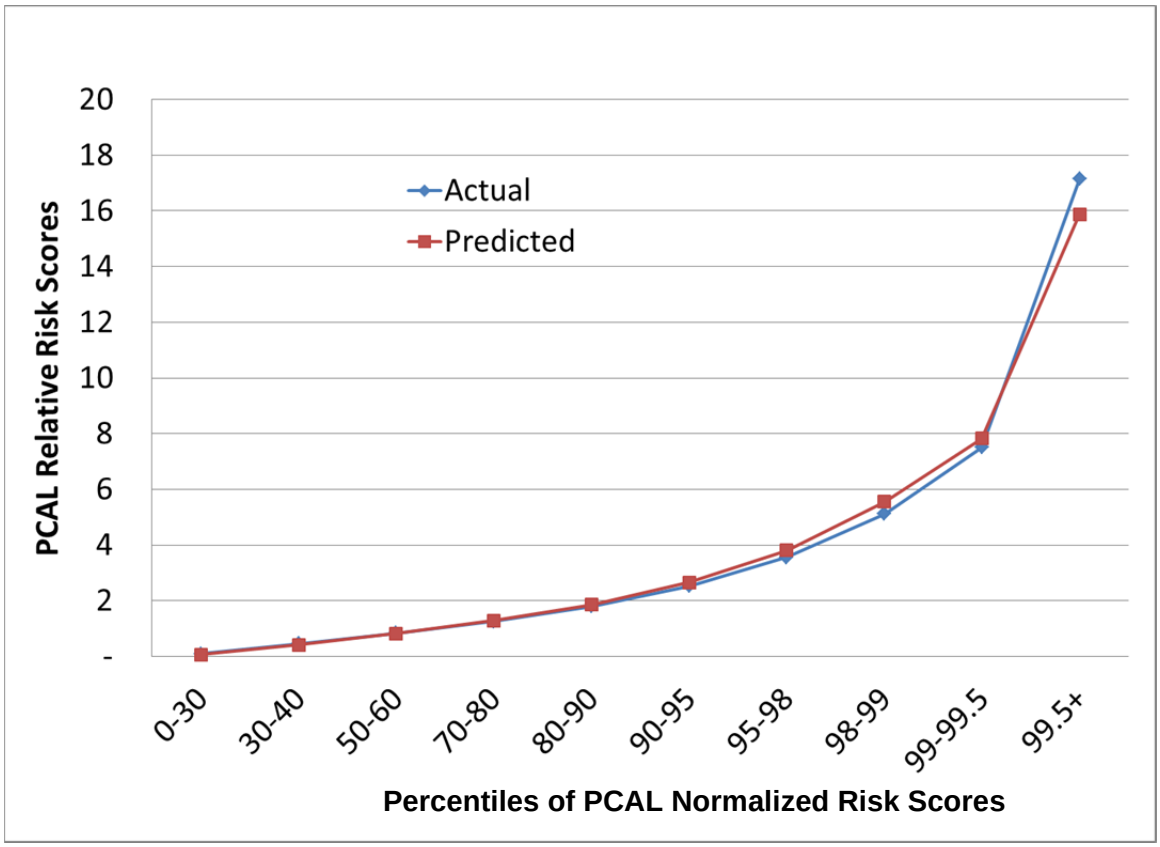
FIGURE B1. PCAL Model R^2 Calculated for Each Sex and Age Group



Notes: All calculations use Thomson Reuters MarketScan 2007 Commercial Data, N=17,419,585.

Each bar presents the R^2 (in percent) from regressing the PCAL spending proxy on the predicted PCAL from the full sample model.

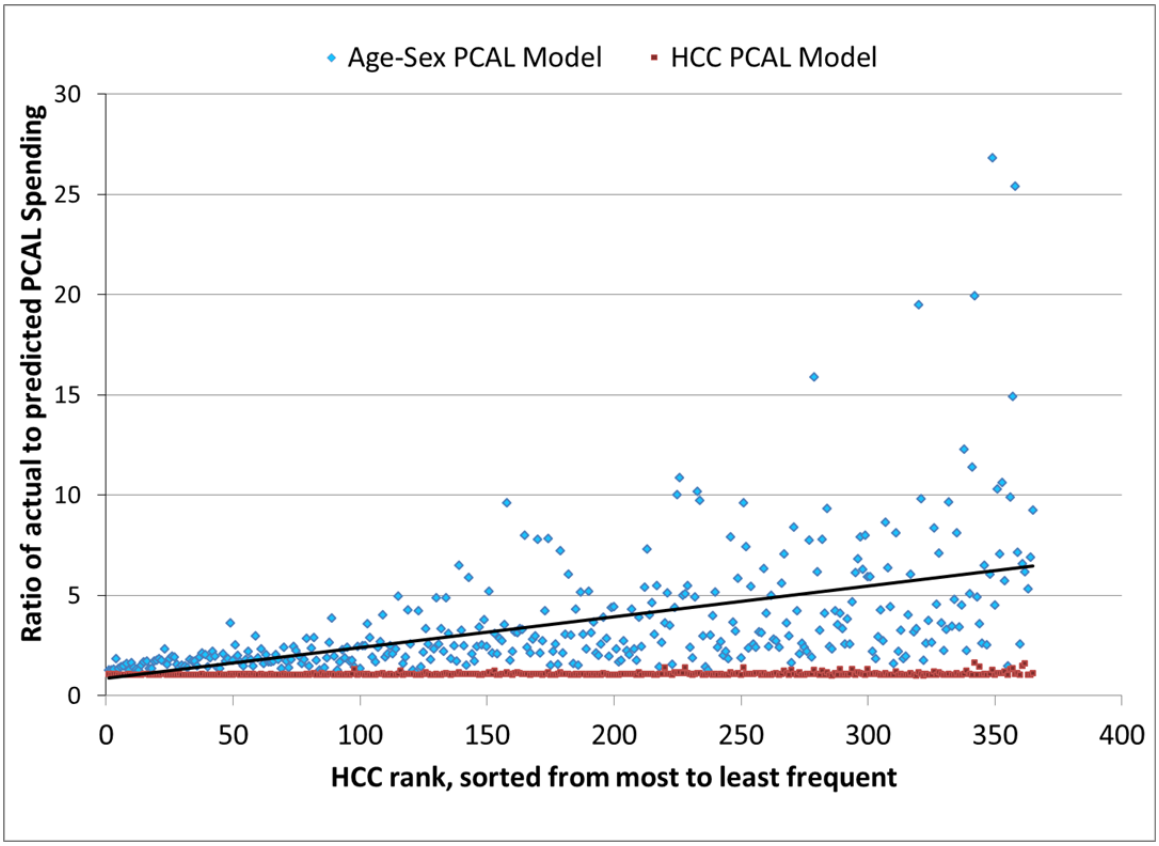
FIGURE B2. Values of the PCAL Proxy Outcome Y (labeled Actual) and of PCAL $\hat{}$ (labeled Predicted) Versus Percentiles of PCAL Normalized Risk Scores



PCAL: primary care activity level

Notes: Using the full estimation sample, patients were sorted and classified (on the horizontal axis) by increasing value of the PCAL prediction, then plotted against the means of the PCAL proxy (Y) and its prediction, both expressed as normalized risk scores. Categories were formed to show greater detail at the top of the distribution, where PCAL values increase most rapidly.

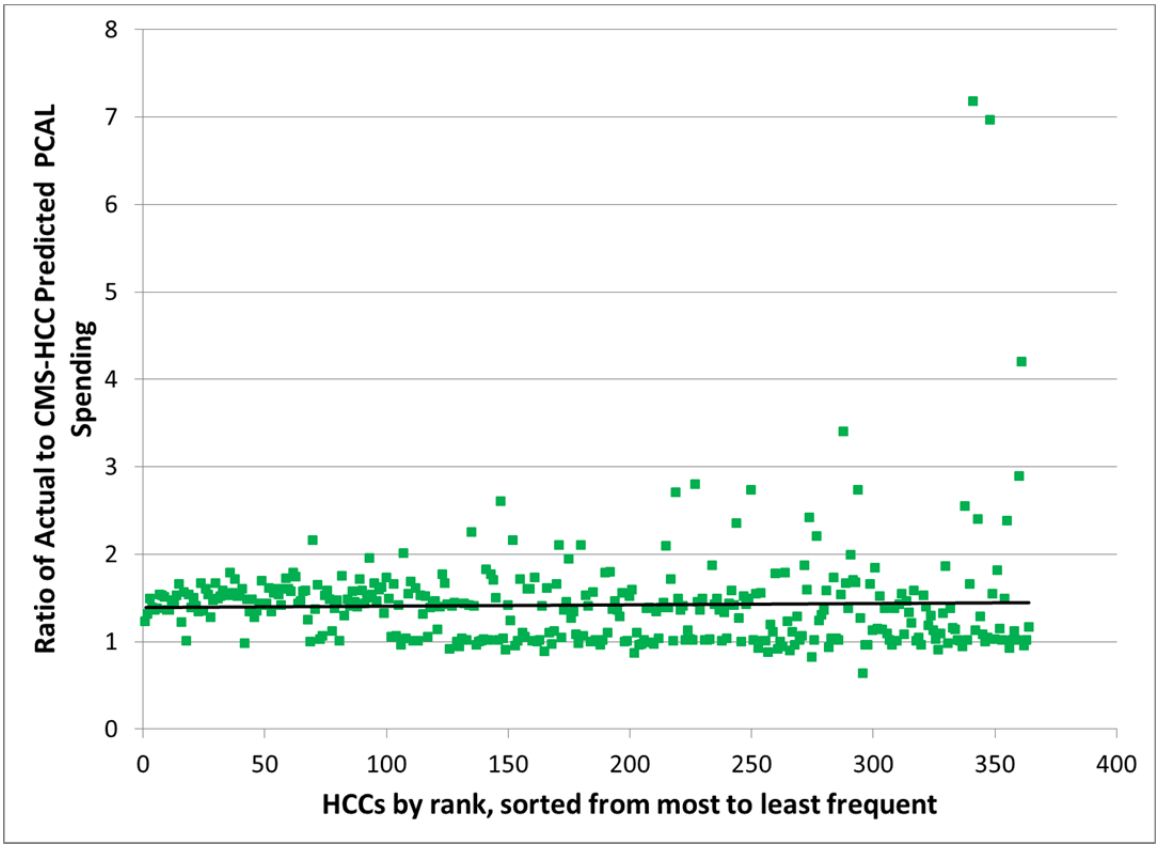
FIGURE B3. Ratio of Actual to Predicted PCAL for Those 365 Condition Categories With Over 500 Persons Each, Using the Age-Sex and HCC Models to Predict the PCAL Proxy



PCAL: primary care activity level; HCC: hierarchical condition category.

Notes: Regression models predicting the PCAL proxy variable (*Y)* were estimated on all 17.4 million people: 1) using only age and sex (◆) and 2) using age, sex, and HCCs (■). Per capita averages were calculated for each model for each of the 394 HCCs, based on actual and predicted PCAL costs for all people with at least one diagnosis in that HCC. HCCs were sorted from most common to least common; each data point is the ratio of actual to predicted spending, shown for 365 HCCs with more than 500 cases, ranging from HCC383 = Screening/ Observation/Special Exams with 750,471 people at the far left to HCC213 = Heart Transplant Complications with 561 people, at the far right. Line shown is an unweighted OLS trend line for the age-sex predicted (◆) data points.

FIGURE B4. Ratio of Actual to CMS-HCC Predicted PCAL for Those 365 Condition Categories with Over 500 Persons Each, Using the CMS-HCC Model to Predict the PCAL Proxy



PCAL: primary care activity level; HCC: hierarchical condition category.

Notes: Regression models predicting the PCAL proxy variable (*Y)* were estimated using the full sample of 17.4 million people using the 70 CMS-HCCs with 22 age and sex dummy variables. Per capita averages of actual and predicted PCAL were calculated for each of the 394 HCCs, for all people with at least one diagnosis in that HCC. HCCs were sorted from most common to least common; each data point plots the ratio of actual to predicted spending, shown for 365 HCCs with more than 500 cases, ranging from HCC383 = Screening/ Observation/Special Exams with 750,471 people at the far left to HCC213 = Heart Transplant Complications with 561 people, at the far right. Line shown is an unweighted OLS trend line.

Appendix C. Implementation Steps Used to Guide CDPHP Implementation of the PCAL Model and Verification that Models Developed in MarketScan's Commercial Data Predict Well for CDPHP's Medicare and Medicaid Patients

In support of the Capital District Physician's Health Plan's (CDPHP's) Risk-Based Comprehensive Payment Model pilot, Verisk Health and CDPHP modified the PCAL model to better approximate distributions of spending across CDPHP physician practices. During the pilot's first year, a preliminary PCAL model with different proportions of spending on various services was used, but, using 2007 MarketScan data, the earlier and current PCAL models were found to have correlations of .83 at the individual level and .99 at the practice level.

First, we established that the PCAL model achieved an R^2 of 52% at the individual level and 73% when predicting CDPHP PCAL payments to PCPs in the 13 treatment practices in 2006 and 2007. Predictions were also relatively stable from 2006 to 2007.

Second, even though the PCAL model was calibrated using only data for privately insured individuals age 0 to 65, it was found to also work well with CDPHP Medicare Advantage and Medicaid HMO and PPO enrollees, and applied to those groups. Because of the variable and high costs of newborns, fee-for-service payments were used for babies.

Third, due to CDPHP concerns about extreme values for PCAL for some patients, predicted PCAL values were top-coded at $25,000 and bottom coded at $70. The subsequent PCAL model described in the text built in these features by adding on a minimum dollar amount ($65) to all individuals and top-coding (at the 99.9^{th} percentile) each summand used to create the dependent variable. In the current version of the PCAL, top-coding is done prior to estimation.

Fourth, separate cost numbers for the appropriate value of a standardized PCAL risk of 1.00 were calculated for private HMO, private non-HMO, Medicare, and Medicaid health plans in CDPHP, and were shown to capture actual variation at the practice level.

Fifth, results using the HCC-based PCAL model were compared to models using only age and gender and the latter were shown to generate significantly biased results at the individual level. We also generated results leading to a scatterplot like Figure B3.

Sixth, we found that PCAL payments are fairly insensitive to alternative specifications of the PCAL outcome and levels of aggregation to the practice level (unpublished work).

Finally, given concerns that models built on commercially insured populations might work poorly for other payer groups, we regressed total health spending on the MarketScan-developed concurrent nRS separately for CDPHP's private, Medicare and Medicaid patients. R^2 values for these 1-degree-of-freedom models are satisfyingly high: 59.5%, 65.4% and 56.4%, respectively, confirming that, at least for predicting total spending, the "relatives" established in the commercial data also work well for Medicare and Medicaid populations. (Table C1.)

TABLE C1. Predicting Observed Total Health Spending From Concurrent nRS (top-coded model) in CDPHP Private, Medicare and Medicaid Data

	Full Sample Means (St. Dev.)	Weighted least squares coefficients (standard errors)					
		Pooled Sample			Private	Medicare	Medicaid
Model		**1**	**2**	**3**	**4**	**5**	**6**
Intercept		89***	-40***	3	-40***	-695***	-255***
		(13)	(15)	(18)	(15)	(89)	(21)
Risk Score	1.98	1721***	2070***	2118***	2070***	1525***	1309***
	(4.13)	(3)	(4)	(5)	(4)	(8)	(5)
Medicare	0.07		-655***	-694***			
	(0.26)		(55)	(56)			
Medicaid	0.18		-215***	-259***			
	(0.38)		(35)	(36)			
Risk Score X Medicare	0.54		-545***	-593***			
	(2.94)		(6)	(7)			
Risk Score X Medicaid	0.24		-761***	-810***			
	(1.49)		(9)	(9)			
PPO dummy	0.03			-238**			
	(0.18)			(70)			
POS dummy	0.07			-457***			
	(0.26)			(50)			
FFS+EPO dummy	0.09			92			
	(0.29)			(46)			
Risk Score X PPO	0.06			-356***			
	(0.69)			(19)			
Risk Score X POS	0.12			-106*			
	(0.92)			(14)			
Risk Score X FFS+EPO	0.17			-129***			
	(1.24)			(11)			
R-Square		59.7%	62.2%	62.3%	59.5%	65.4%	56.4%
No. Of Obs.	248,413	248,413	248,413	248,413	184,674	18,908	44,831
Degrees of freedom		1	5	11	1	1	1
Dependent mean		3,511	3,511	3,511	3,263	9,544	1,679
Standard Deviation		12,204	12,204	12,204	11,607	20,394	8,461
Standard Error of the Regression		5,831	5,650	5,641	5,491	9,528	3,800

Notes: ***, ** and * denote statistical significance at 1%, 5% and 10% levels respectively.

Results are based on the Capital District Physicians' Health Plan 2008 claims from patients in four counties, plus all patients attributed to four PCMH pilot practices. All results are for weighted least squares, weighting annualized total costs by the fraction of the year eligible. The first three models are for the pooled sample for all plan types, while the final three are for subsamples of Private, Medicare, and Medicaid enrollees respectively. The single nRS used in all models is the Verisk Health "out of the box" model 88, calibrated using MarketScan privately insured claims data from 2007, with the dependent variable top-coded at $250,000. PPO=preferred provider organization, POS = point of service plan, FFS=fee for service, EPO = exclusive provider organization.

Appendix D. Further Details about Bonus Models

TABLE D1. Drugs Included in the Prescriptions of Antibiotics of Concern Model

Table ABX-B: Antibiotics of Concern by NCQA Drug Class			
Description	Prescription		
Quinolone	• ciprofloxacin	• levofloxacin	• norfloxacin
	• gatifloxacin	• lomefloxacin	• ofloxacin
	• gemifloxacin	• moxifloxacin	• sparfloxacin
Azithromycin and clarithromycin	• azithromycin	• clarithromycin	
Cephalosporin (second, third, fourth generation)	• cefaclor	• cefotetan	• ceftizoxime
	• cefdinir	• cefoxitin	• ceftriaxone
	• cefditoren	• cefpodoxime	• cefuroxime
	• cefepime	• cefprozil	• loracarbef
	• cefixime	• ceftazidime	
	• cefotaxime	• ceftibuten	
Amoxicillin/clavulanate	• amoxicillin-clavulanate		
Ketolide	• telithromycin		
Clindamycin	• clindamycin		
Miscellaneous antibiotics of concern	• aztreonam	• dalfopristin-quinupristin	• vancomycin
	• chloramphenicol	• linezolid	

Source: HEDIS NCQA measure, accessed on June 6, 2011 at http://www.ncqa.org/tabid/892/Default.aspx

TABLE D2. Other Drugs Not Among the Antibiotics of Concern

Table ABX-C: All Other Antibiotics by NCQA Drug Class			
Description	**Prescription**		
Absorbable sulfonamide	• sulfadiazine	• sulfasalazine	
	• sulfamethoxazole-	• sulfisoxazole	
Aminoglycoside	• amikacin	• kanamycin	• tobramycin
	• gentamicin	• streptomycin	
Cephalosporin (first generation)	• cefadroxil	• cephalexin	
	• cefazolin	• cephradine	
Lincosamide (other than clindamycin)	• lincomycin		
Macrolide (other than azithromycin and clarithromycin)	• erythromycin	• erythromycin lactobionate	
	• erythromycin esto	• erythromycin stearate	
	• erythromycin ethy	• erythromycin-sulfisoxazole	
Penicillin (other than amoxicillin/ clavulanate)	• ampicillin	• penicillin G potassium	
	• ampicillin-sulbacta	• penicillin G procaine	
	• amoxicillin	• penicillin G sodium	
	• carbenicillin	• penicillin V potassium	
	• dicloxacillin	• piperacillin	
	• nafcillin	• piperacillin-tazobactam	
	• oxacillin	• ticarcillin	
	• penicillin G benzat	• ticarcillin-clavulanate	
Tetracyclines	• doxycycline	• minocycline	• tetracycline
Miscellaneous antibiotics	• daptomycin	• nitrofurantoin macrocrystals	
	• fosfomycin	• rifampin	
	• metronidazole	• trimethoprim	
	• nitrofurantoin		

Source: HEDIS NCQA measure, accessed on June 6, 2011 at http://www.ncqa.org/tabid/892/Default.aspx

TABLE D3. Top-coding values used on performance measures before analysis

Performance measures	Top-coded value
Total health spending, in dollars (Y_0)	$250,000
Total prescription drug spending, in dollars	$37,866
Number of prescriptions for antibiotics of concern	10
Number of prescriptions for antibiotics	15
Hospital admissions, all types	6
Hospital admissions, not related to childbirth, pregnancy, or behavioral health	4
Hospital admissions for ambulatory-care-sensitive conditions	4
Advanced imaging tests in relative value units (RVUs)	200
Emergency department visits	6

Notes: top-coding values were chosen to correspond to the 99.9th percentile for each measure, including zeros.