Mispricing in the Medicare Advantage Risk Adjustment Model

Jing Chen[*], Randall P. Ellis[†*], Katherine H. Toro[*] and Arlene S. Ash[‡*]

[*] EMC Corporation, Hopkinton, MA.

[†] Department of Economics, Boston University, Boston, MA.

[‡] Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA.

Corresponding author: Jing Chen, Ph.D, MBA, EMC Corporation, 176 South Street, Hopkinton, MA 01748. Tel: (508)293-6013. Fax: (757) 487-5983. Email: jchenstat@gmail.com

Acknowledgement: We thank Ron Russell, Matt Siegel, Deb Bradley, Jo Anne Lutz, Tim Layton and Wenjia Zhu for their comments and Timur Turkdogan and Punam Mahajan for preparing the data for this analysis.

Financial Disclosure: Chen is a former Research Scientist at Verisk Health, Inc. Ash and Ellis are senior scientists at Verisk Health, Inc., where they consult on developing health-based predictive models; neither has any ownership interest in Verisk Health, Inc. Toro is a former Director of Product Management at Verisk Health, Inc.

Complete Author Information:

Jing Chen, Ph.D, MBA, EMC Corporation, 176 South Street, Hopkinton, MA 01748. Tel: (508) 293-6013. Fax: (757) 487-5983. Email: jchenstat@gmail.com

Randall P. Ellis, Ph.D, Department of Economics, Room 442, Boston University, 270 Bay State Road, Boston, MA 02215. Tel: (617) 353-2741. Fax: (617) 353-4449. Email: ellisrp@bu.edu

Katherine H. Toro, M.A., Email: khtoro@hotmail.com

Arlene S. Ash, Ph.D, University of Massachusetts Medical School, Department of Quantitative Health Sciences, ASC 9.2071, 368 Plantation St., Worcester, MA 01605. Tel: (508) 856-8922/8999. Fax: (508) 856-8993. Email: arlene.ash@umassmed.edu

Abstract

The Center for Medicare and Medicaid Services implemented hierarchical condition category (CMS-HCC) models in 2004 to adjust payments to Medicare Advantage (MA) plans to reflect enrollees' expected health care costs. We use DxCG Medicare models, refined "descendants" of the same HCC framework with 189 comprehensive clinical categories available to CMS in 2004, to reveal two mispricing errors resulting from CMS' implementation. One comes from ignoring all diagnostic information for "new enrollees" (those with less than 12 months of prior claims). Another comes from continuing to use the simplified models which were originally adopted when capitated health plans did not submit the claims-like data that facilitate richer models. Even the main CMS model being used in 2014 recognizes only 79 condition categories, excluding many diagnoses and merging conditions with somewhat heterogeneous costs. Omitted conditions are typically lower cost or "vague" and not easily audited from simplified data submissions. In contrast, DxCG Medicare models use a comprehensive, 394-HCC classification system. Applying both models to Medicare's 2010 - 2011 Fee-For-Service five-percent sample, we find mispricing and lower predictive accuracy for the CMS implementation. For example, in 2010, 13% of beneficiaries had at least one higher-cost DxCG-recognized condition, but no CMS-recognized, condition; their 2011 actual costs averaged $6,628, almost one-third more than the CMS model prediction. Since MA plans must now supply encounter data, CMS should consider using more refined and comprehensive (DxCG-like) models.

Keywords: Medicare, CMS-HCC, DxCG, risk adjustment, payment models

Introduction

In 2013, the U.S. Medicare program provided health insurance coverage to 52 million beneficiaries entitled by age greater than 64, disability or end stage renal disease (ESRD)[1]. Medicare spending accounted for 16% ($536 billion) of the federal budget[2] and is projected to double by 2023 due to increasing numbers of beneficiaries and costs per person[3].

Medicare beneficiaries can enroll in a private sector option called Medicare Advantage (MA) rather than receive the traditional fee-for-service (FFS) benefit. In 2013, 28 percent were enrolled in MA[4]. Premiums paid to MA plans are calculated from age- , gender- and diagnosis-based hierarchical condition category (HCC) models[5]. The HCC framework requires classifying all coded diagnoses into condition categories (CCs) and using hierarchies to eliminate redundant recognition of a single underlying medical problem. The HCC modeling framework was selected "… largely on the basis of transparency, ease of modification, and good clinical coherence"[6]. First implemented in 2004, the CMS-HCC model is periodically updated.

Verisk Health, a private for-profit health analytics firm, estimates and supports DxCG Medicare HCC models, originally relying on CMS' 189 condition categories[5]. For implementation, the CMS-HCC payment models have omitted and consolidated many CCs, now recognizing approximately 80 distinct conditions. In

contrast, today's DxCG Medicare (version 7) models exploit the full detail of a comprehensive classification system with 394 HCCs.

In this paper we use DxCG Medicare models to illustrate and quantify both lack of available precision and mispricing– that is, differences between actual and CMS-model-predicted costs – in CMS payment model implementation.

HCC Models

Both CMS-HCC and DxCG Medicare risk adjustment models are prospective and use demographic information (age, sex, Medicaid dual eligibility, and reasons for Medicare eligibility) as well as the profiles of major medical conditions in a base year to predict "costs" in the following, or target, year. Costs are payments for services covered by Medicare's hospital insurance (Part A) and supplementary medical insurance (Part B) benefit. Medicare has yet another model for its end stage renal disease program. Medicare Advantage CMS payments (but not DxCG predictions) also consider whether the beneficiary is institutionalized (e.g., living in a nursing home) or "new."

Both models first classify all (approximately 16,000) ICD-9-CM diagnosis codes into condition categories. Each CC contains clinically-related groups of diagnoses, such as colon cancer and rectal cancer, with similar cost implications. Hierarchies are imposed, so that a person is coded for only the most severe manifestation among related diseases (e.g., someone with cystic fibrosis would not also be coded for either "chronic obstructive lung disease" or cough). This converts CCs into

Hierarchical Condition Categories, or HCCs[7]. Both models also include interactions between disease groups (e.g., diabetes and congestive heart failure) and between diseases and disability status (e.g., disability and congestive heart failure) so long as they make sense to clinicians and strongly predict additional costs[7].

CMS-HCC Models

CMS implements distinct risk models for beneficiaries entitled by age, disability or ESRD, and for community-residing versus long-term institutional (nursing home) enrollees. Unlike the CMS-HCC models for "continuing" MA members, the payment formula for "new" Medicare enrollees (enrolled for less than 12 months in the base year) uses no diagnostic information.

Prior to the Affordable Care Act (2010), MA plans were exempted from submitting encounter records. Although CMS calibrated its models on FFS data with full ICD-9-CM coding, risk-adjusted MA payments use only a short list of the CCs present for each person. Model recalibration for 2014 used 2010 FFS claims data to predict 2011 costs[8].

The original CMS-HCC payment model included 70 HCCs; even now, CMS's model includes only 79 HCCs (87 HCCs for its ESRD models). New HCCs are either previously unrecognized conditions among the 189 HCCs available, or splits/collapses of previously included HCCs[8].

DxCG Medicare Models

As described, DxCG Medicare models share the same basic HCC structure as CMS' models, but rely on up to 394 HCCs. Altogether 138 of the DxCG HCCs have zero weight in the Medicare model, mostly lower cost or vague conditions, with the omitted set chosen using statistical criteria, clinical judgment and practical issues, balancing the desire for greater accuracy against the concern that difficult-to-verify distinctions in medical problems should not result in large distinctions in payment. Like the CMS models, the DxCG Medicare model excludes many vague and discretionary conditions – including hypertension and high cholesterol – in its desire to reduce opportunities for gaming payments by recording vague diagnoses. The DxCG models tested here use different payment weights for disabled (age <65) and elderly beneficiaries. Regression coefficients were estimated using FFS claims data for beneficiaries with both Parts A and B insurance in Medicare's 2005-2006 five-percent sample. Unlike CMS' implementation, prior-year diagnoses are used in all enrollees' predictions, not just continuing enrollees.

The Data

The data pertain to 1.5 million enrollees from Medicare's 2010-2011 FFS five-percent sample: enrolled exclusively in FFS; present and eligible for Parts A and B coverage for at least one month in each year, and not currently entitled to the ESRD program. (See Appendix Table A1.)

We use 2010 data to predict weighted annualized allowed cost (that is, Medicare allowed costs covered by the combined Parts A and B benefit) in 2011. This is the

dependent variable predicted by DxCG Medicare models. Following usual CMS practice, we downweight data for members eligible for only part of the target year, using the fraction of the year that they are eligible. Annualization (that is, multiplying by the inverse of the weight) makes the weighted average of the dependent variable match actual spending. One small distinction between the CMS and DxCG approaches is that CMS models predict "paid" (rather than "allowed") amounts. Paid amount equals allowed amount minus deductibles and copayments. The DxCG models predict allowed amounts because they are more closely linked to actual resource use and less subject to variation in plan-level cost-sharing features. Nonetheless, annual paid and allowed amounts are highly correlated (in our sample, ρ=0.998); thus, predictions evaluated with either outcome should perform similarly.

Both CMS and DxCG models generate relative risk scores (RRSs). Indeed, the CMS-HCC software automatically generates three RRSs for each person, and the user selects one, depending upon whether the enrollee is new, continuing and community-dwelling or institutionalized (usually in a nursing home). We have not examined the institutional model. Instead, we evaluate the CMS model in two ways: first, following CMS's approach of using the new enrollee model RRS for members enrolled for less than 12 months in 2010, and using the risk score from the community model for everyone else; second, we used the RRS from the community model for everyone. We will call the first method "CMS implemented" (as of 2014) and the second, "CMS improved."

Relative risk scores must be converted to payments by applying a dollar weight multiplier – the payment associated with next year's average cost – prior to

knowing next year's actual costs. We ignore the forecasting error associated with that unknown, and level the playing field among models, by choosing multiplicative factors to make each model's weighted mean predictions exactly match weighted mean actual cost in the 2011 sample, separately for new and continuing enrollees.

All models in this study: rely on the same basic HCC structure, use identical demographic predictors, and identify diagnoses from the same claims records. We call Verisk Health's DxCG Model 121 the "DxCG model"; its principal distinguishing characteristic is its reliance on a refined, comprehensive classification of up to 394 condition categories.

Results

Table 1 compares three "off-the-shelf" models: CMS Implemented and Improved 2014 models, and the DxCG model. By "off-the-shelf", we mean that each model was fit to other data and is applied here; thus, there is no concern about over-fitting. The three rows show, respectively, R^2s for: the full population; new enrollees (those with less than 12 months of eligibility in the base year); and continuing (that is, non-new) enrollees.

The contrast between the CMS Implemented and Improved columns shows that the CMS could increase its predictive accuracy (R^2) simply by applying its own CMS-HCC model to new enrollees. The R^2 improvement within the new enrollee population itself is huge (from 2.0% to 17.2%), but because few members are new,

this only increases the whole-population R^2 from 13.8% to 14.2%. The whole-population R^2 for the DxCG model is 16.5%.

Other performance measures involve comparing a model's payments for groups of people to their actual costs. For example, plans that enroll members with serious, high-cost-generating conditions should receive funds adequate to care for them; more generally, with a good model, most moderately large, "prospectively identifiable" subgroups will have similar model-expected and actual costs. We will examine mispayment, that is, differences between mean model-predicted payments and actual costs, and compute overpayment percentages (payment minus cost, divided by cost) for various subgroups and models.

1) **Mispayments by model-predicted risk quantiles**. First we evaluated model discrimination by sorting the population into quantiles of increasing CMS-model-predicted cost and calculating mean (observed) Medicare cost and percent mispricing for quantile-based groups. Table 2 shows the actual year-2 costs by prediction quantiles from the CMS implemented model, and associated overpayment percentages. Note that the model makes large distinctions among beneficiaries; average costs of those with the 1%-highest predictions are nearly 20 times as much as for the bottom 20%. However, we also care about "calibration" – do the plans get "right amounts" across the spectrum of expected costs? The last column of Table 2 shows the percent over or under payment within each subpopulation; CMS underpays both those in the top 5% *and* those in the bottom 20% of expected costs. For example, while those in the bottom 20% actually cost about $4000, the CMS model would have paid out 12% less, only about $3500.

2) **Mispayments by the presence/absence of various kinds of diagnoses**. We also examine means and mispricing separately for members who have and do not have any diagnoses recognized by the CMS classification system. Here we look at mispricing under both the CMS implemented and DxCG models, for everyone and separately among new enrollees. As shown in the top half of Table 3 (and in Figure 1), both models get the average payment right for the 66 percent of members with at least one clinical condition recognized by CMS' implementation. For the remaining 34 percent, with no HCCs identified by CMS, there is substantial mispricing within subgroups. For members with no HCC in either system (7%), CMS overpays by 44% while the DxCG model gets the average right. The remaining 27% of members can be split into the 13% of members with at least one higher cost DxCG condition not recognized by CMS, and those with only lower cost conditions. The CMS model underpays the higher-cost DxCG conditions by 25%, and overpays those with only low-cost DxCG HCCs by 17%. Table 3 examines the same information for new enrollees. Because CMS ignores diagnostic information for new enrollees, it underpays those with CMS conditions by 35%, and overpays the rest, especially the healthiest 25% of members costing less than $4000 each, for whom it pays over $8000. In contrast, the DxCG model's expected costs are close to observed costs across these subgroups.

Discussion and Conclusions

We examined CMS and DxCG Medicare models in Medicare FFS data, finding two changes that Medicare could implement to predict more accurately. These are: use whatever diagnoses are present to distinguish among "new" enrollees with less than

12 months of base year data, and adopt a more refined and comprehensive predictive tool, such as Verisk Health's DxCG model. The first change requires only an administrative decision to implement; the second takes more work. Now that MA plans are required to submit encounter data that include all diagnoses, additional research is warranted to explore the sensitivity of simplified versus refined, comprehensive models to aggressive, hard to audit upcoding. Reduced sensitivity to upcoding may require further refinements. Addressing behavioral responses to risk adjustment, as many researchers have discussed, is another area for further research.[9,10] Our suggested improvements to the current CMS-HCC models corrects some, although likely not all, of the troublesome payment and incentives problems related to under- and over-prediction of costs for large groups of prospectively identifiable people. Given that models similar to the CMS-HCCs for MA are also used for Part D, for health insurance exchanges and for diverse research evaluations, improving the classification and modeling approaches seems especially worthwhile.

Table 1: Off-the-shelf R^2 for predicting Medicare cost: CMS-HCC vs. DxCG models[1]

	CMS-HCC 2014 models		DxCG
	Implemented[2]	Improved[3]	model[4]
All enrollees	13.8%	14.2%	16.5%
New enrollees	2.0%	17.2%	19.0%
Continuing enrollees	14.1%	14.1%	16.4%

SOURCE: Medicare Fee-For-Service (FFS) 5-percent sample, present in both 2010 and 2011, excluding those with 2010 ESRD (N = 1,487,628). All models use 2010 information to predict 2011 Medicare cost.

Notes: 1. All models have 1 degree of freedom; each regresses cost on a formula-based risk score: cost = a + b*(risk score). The CMS-HCC 2014 models were calibrated on 100% FFS 2010-2011 data; DxCG models were calibrated on the 2005 – 2006 Medicare Fee-For-Service (FFS) 5-percent sample. Both models predict next year's costs from beneficiary age, sex, Medicaid dual eligibility, original reason for Medicare entitlement and diagnoses from the previous year's inpatient, outpatient and carrier-file claims.

2. "Implemented" means using the new enrollee model RRS for members enrolled for less than 12 months in 2010, and using the risk score from the community model for everyone else.

3. Improved means using the RRS from the community model for every enrollee.

4. DxCG version 7, Model 121.

Table 2: Mean Medicare cost and mispricing by 2014 CMS-HCC implemented model-predicted percentile groups

Percentile groups based on 2014 CMS-HCC predictions[1]		Mean Medicare cost in 2011	Percent overpayment by 2014 CMS-HCC model[2]
Top	1%	$78,584	-5
Next	4%	$44,371	-2
Percentiles	90-95%	$29,072	2
	80-90%	$19,831	4
	50-80%	$11,880	2
	20-50%	$6,457	0
Bottom	20%	$4,022	-12

SOURCE: Medicare Fee-For-Service (FFS) 5-percent sample, present in both 2010 and 2011, excluding those with 2010 ESRD (N = 1,487,628). All models use 2010 information to predict 2011 Medicare cost. N = 1,487,628.

Notes:
1. Using the "as implemented" algorithm – that is, ignoring all diagnoses for new enrollees.
2. Percentages are calculated as (payment - cost)/cost. For example, -5, means that what the model expects (and what a payment system based on it would pay) is 5 percent less than the actual cost.

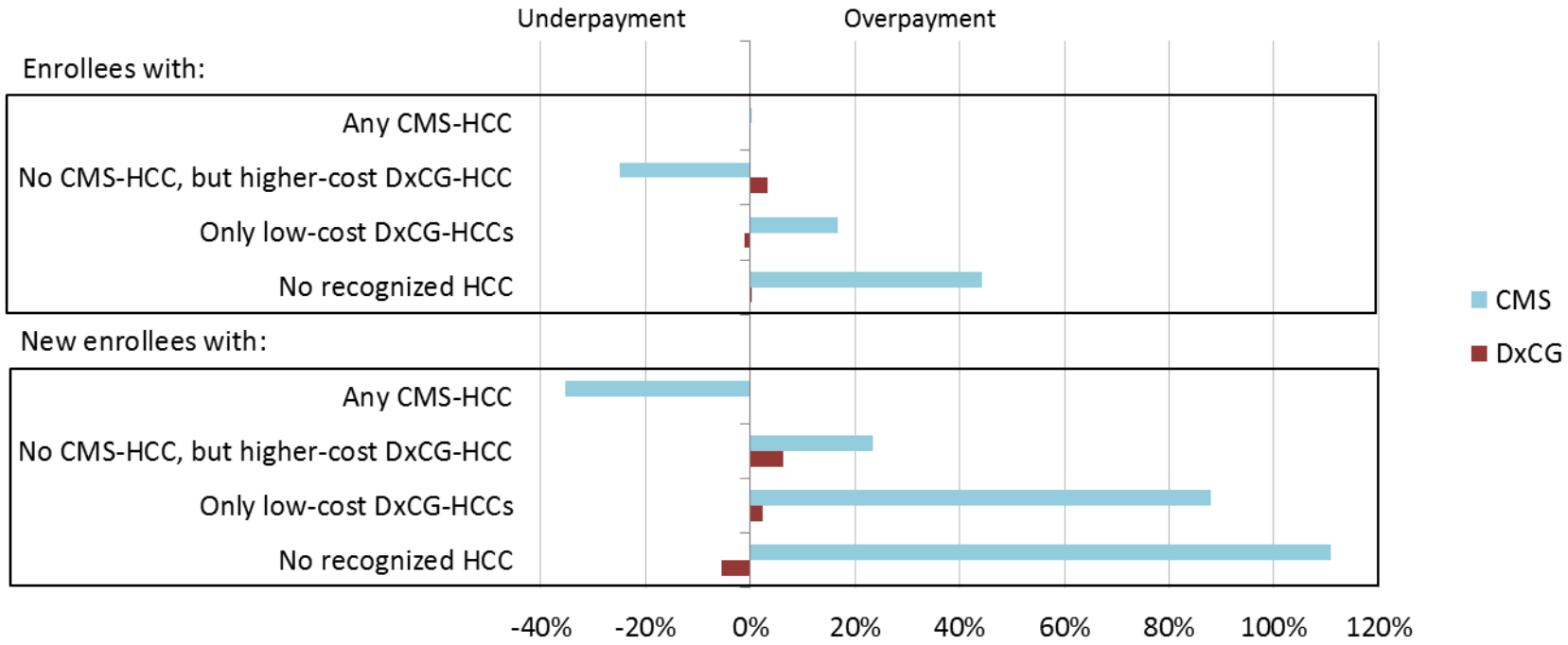
Table 3: 2011 mean costs, model-based payments and percent over- and underpayments for subgroups of people by types of conditions

Groups	Subgroups	Percent	Mean Actual Costs	Model-Based Payments: CMS, as implemented		Model-Based Payments: DxCG, as recommended	
				Mean	Error[1]	Mean	Error[1]
ALL enrollees (n = 1,487,628)	Any CMS-HCC	66	$15,715	$15,743	0%	$15,677	0%
	No CMS-HCC:	34	$4,886	$4,833		$4,957	
	Any higher-cost DxCG-HCC[2]	13	$6,628	$4,975	-25%	$6,852	3%
	Only low-cost DxCG-HCCs[3]	14	$3,997	$4,665	17%	$3,955	-1%
	No recognized HCC	7	$3,403	$4,906	44%	$3,416	0%
	Total	100	$11,943	$11,943		$11,943	
New enrollee subgroup (n = 68,671)	Any CMS-HCC	41	$14,346	$9,263	-35%	$14,307	0%
	No CMS-HCC:	59	$4,385	$7,823		$4,411	
	Any higher-cost DxCG-HCC[2]	11	$6,355	$7,843	23%	$6,761	6%
	Only low-cost DxCG-HCCs[3]	23	$3,989	$7,502	88%	$4,083	2%
	No recognized HCC	25	$3,846	$8,115	111%	$3,634	-6%
	Total	100	$8,405	$8,405		$8,405	

SOURCE: Medicare Fee-For-Service (FFS) 5-percent sample, present in both 2010 and 2011, excluding those with 2010 ESRD (N = 1,487,628). Both models use 2010 information to predict 2011 Medicare cost. The CMS model uses its 2014 update calibrated on 2010-2011 data; the DxCG model version 7 was calibrated on 2005 – 2006 data.

Notes:
1. Error is calculated as (payment - cost)/cost. For example, -6%, means that what the model expects (and what a payment system based on it would pay) is 6 percent less than the actual cost.
2. The conditions with the highest 100 coefficients in the DxCG model from the subgroup after excluding people with any conditions classified in CMS-HCC.
3. All DxCG-HCC conditions not previously classified.

Figure 1: Percent overpayment (underpayment) by presence of HCC type: CMS versus DxCG models



Appendix

Table A1: Characteristics of the 2010 – 2011 Medicare Fee-For-Service, Non-ESRD[1] 5% Sample

	Mean	SD
Annualized 2010 Medicare cost	$10,153	22,907
Annualized 2011 Medicare cost	$11,943	29,453
Age in 2010	71.4	12.6
	N	%
Aged 65+ on December 31, 2010	1,229,140	82.6
Female	831,378	55.9
Continuing (enrolled for 12 months in 2010)[2]	1,418,862	95.4

SOURCE: Medicare Fee-For-Service (FFS) 5-percent sample, present in both 2010 and 2011, excluding those with 2010 ESRD (N = 1,487,628).

Notes:

[1] Even after removing members with ESRD as their current reason for entitlement in 2010, the study sample still contains 10,428 members with an ESRD diagnosis in 2010, probably those newly diagnosed with ESRD who are not yet eligible for this program, or with diagnoses or renal disease durations that do not meet ESRD program eligibility criteria.

[2] "New" members, enrolled for < 12 months, account for ~0.4% for each number of months of eligibility, from 1 to 11.

References

1. 2013 SMI Trend Table. Centers for Medicare & Medicaid Services Web site. http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MedicareEnrpts/Downloads/SMI2013.pdf. Accessed January 14, 2015.
2. State Health Facts. Kaiser Family Foundation Web site. http://kff.org/medicare/state-indicator/total-medicare-beneficiaries/. Accessed January 14, 2015.
3. Medicare Spending and Financing Fact Sheet. Kaiser Family Foundation Web site. http://kff.org/medicare/fact-sheet/medicare-spending-and-financing-fact-sheet/. Published July 28, 2014. Accessed January 14, 2015.
4. Medicare Advantage 2013 Data Spotlight: Enrollment Market Update. Kaiser Family Foundation. http://kaiserfamilyfoundation.files.wordpress.com/2013/06/8448.pdf. Published June, 2013. Accessed January 14, 2015.
5. Pope GC, Ellis RP, Ash AS, et al. Diagnostic Cost Group Hierarchical Condition Category Models for Medicare Risk Adjustment Final Report. Centers for Medicare & Medicaid Services Website. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Reports/downloads/pope_2000_2.pdf. Published December 21, 2000. Accessed January 14, 2015.
6. Risk Adjustment Methods and their Relevance to “Pay-or-Play”. Institute for Health Policy Solutions Web site.

http://www.ihps.org/pubs/2005_Apr_IHPS_SB2_ESup_Risk_Adj.pdf. Published March, 2005. Accessed January 14, 2015.

7. Pope GC, Kautter J, Ingber MJ, et al. Evaluation of the CMS-HCC Risk Adjustment Model Final Report. Centers for Medicare & Medicaid Services Web site. https://www.cms.gov/MedicareAdvtgSpecRateStats/downloads/Evaluation_Risk_Adj_Model_2011.pdf. Published March, 2011. Accessed January 14, 2015.
8. Advance Notice of Methodological Changes for Calendar Y ear (CY) 2014 for Medicare advantage (MA) Capitation Rates, Part C and Part D Payment Policies and 2014 Call Letter. Centers for Medicare & Medicaid Services Web site. http://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/downloads/Announcement2014.pdf. Published April 1, 2013. Accessed January 14, 2015.
9. McGuire TG, Glazer J, Newhouse JP, et al. Integrating risk adjustment and enrollee premiums in health plan payment. *J Health Econ.* 2013 Dec;32(6):1263-77. doi: 10.1016/j.jhealeco.2013.05.002.
10. Newhouse JP, McGuire TG. How Successful Is Medicare Advantage? *Milbank Q*. 2014, 92(2): 351-94. doi: 10.1111/1468-0009.12061.