

An Evaluation of the Performance of the Part D Rx HCC Risk Adjustment Model

Stephen Calfo, Actuary, FSA

Good morning everyone. Today I will present an evaluation of performance of the Medicare Part D Risk Adjustment Model. First off I will provide an overview of Medicare Part D Risk Adjustment. I'll provide an overview and explain some of the attributes of the original and the revised CMS Rx HCC Risk Adjustment Model. I'll be describing how to measure the performance of the model, or the predictive accuracy. I'll describe the key steps in the analysis and also provide actual Medicare Part D plan liability along with other utilization measures through – during the period the calendar years 2007 to 2010. And I will illustrate that the revised Rx ACC model is predicting cost variations among subpopulations of Medicare Part D enrollees more accurately than the original model.

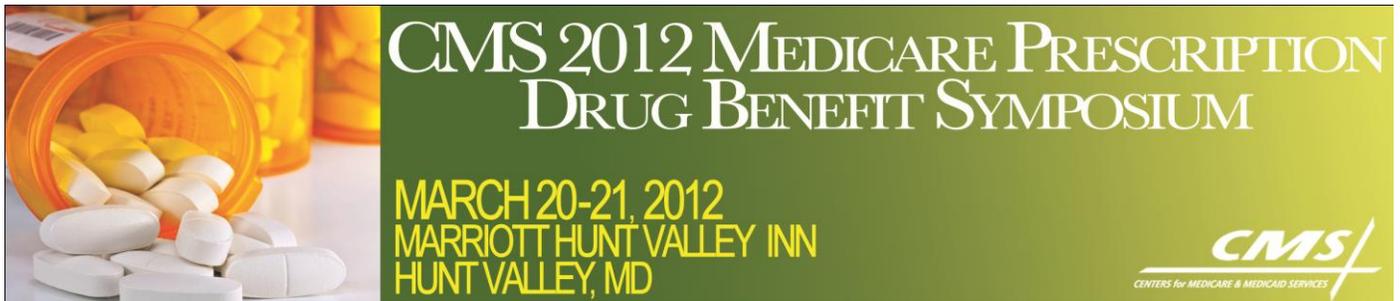
The goal of risk adjustment is to predict cost variations among or within Medicare subpopulations of Part D enrollees. Each model segment is developed using ordinary least square, which is a method for estimating unknown parameters within a linear regression model. The method minimizes the square differences between the predicted and the observed values. In the regressions, the dependent variables – or the dependent variable is the Medicare Part D plan liability and the independent variables are demographic and disease values. The – the development of a risk adjustment model is referred to as a calibration.

The risk adjustment models for Part D are prospective and they use demographic information and diagnosis information in a given year to predict plan liability or expenditures in the following year. For example, demographic and diagnosis data for 2007 are used to predict 2008 Medicare Part D plan liability. Then that model calibration is used, it's relativized and then it's used in payment – in a given payment year, for example, payment year 2010 is – the payments are predicted using 2009 demographic and disease information.

The risk adjustment methodology is designed to predict plan liability or cost variations for subpopulations of Medicare Part D enrollees. The explanatory power of the Rx HCC model is 25%, and that's – what that means is the model is explaining 25% of the variation in individual variation in medical – or Part D expenses. The model – the Risk Adjustment model for Part D is similar to the CMS HCC Part C model, however the Part D drugs are more predictable – the utilization is more predictable than the CMS HCC model, and that's why the R Squared is – is roughly double that of the Part C model.

Some of the commonly used terms in risk adjustment is the health risk or health status, and that's measured by risk factors. The risk factors are age, sex, Medicaid, disability, institutionalization and diagnosis data. The health status is measured – or estimated – using a proxy, and the proxy for health status is known as a risk score. Another few of the commonly used terms are calibration year and the data collection year. The data collection year is the calendar year prior to the payment year.

Okay. I'd like to briefly run through the steps in the development of a risk adjustment model and a risk score calculation.



The first – first step is to create a calibration which uses demographic and diagnosis data in given year X and Medicare Part D plan liability in year X-plus-one. Step two is to run the regression, and the result of the regression are coefficients. And what the coeff – there's a coefficient for each demographic category and each disease category, or HCC, which is a higher – hierarchical condition category. And the coefficients represent dollar amounts and are the incremental costs associated with having an HCC or a demographic factor.

The next step in the regression model is to relativize the coefficients. And the way that's done is developing a denominator, which is also known as a – as a model mean. So all the coefficients are divided by the model mean, and relative factors are created. And the – so the way a risk score is calculated is by adding up all of the relative factors for the demographic and disease categories that a beneficiary has.

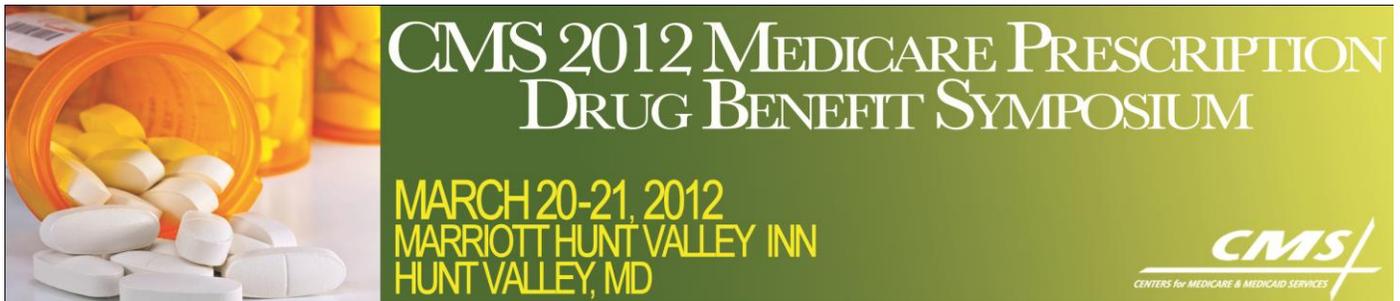
Okay, the next slide shows a – the matrix for the calibration of a risk adjustment model. Each row represents a Medicare beneficiary. The dependent variable is the – the – the furthest column on the left shows the Medicare Part D plan liability, and then the zero one values is a matrix which shows the demographics and the disease categories. You'll see this is just a representation. There's – I have four demographic categories and I'm showing eight – eight CCs, but in the actual model, the matrix is a lot – lot bigger. There are 32 demographics. There's some original disability codes. And there's also 78 ACCs and similar action variables, so the – the matrix would extend way out to the right. And the – and the number of benefit – beneficiaries in the sample is roughly 14.4 million used to calculate it, so, as you can see, the matrix would extend far down.

The relative factor calculation in the – the regression equation is listed here. The result of the – of the regression calibration run, or the coefficient, you'll see them across the bottom – second column from – or second row from the bottom. And then dividing by the overall model mean of all of the Medicare beneficiaries in the sample, dividing the coefficients by the model mean yields the relative – relative cost factor. So to calculate a risk, for example, for a Medicare beneficiary, I picked a 71-year-old male with three HCCs and you just add up all the individual relative factors for each – each of the characteristics of the Medicare beneficiary, so the resulting risk score is the sum of the demographic and disease factors, and it turns out to be .79.

I'd like to move on now to discuss some of the attributes of the original and the revised Rx HCC model. The original model was based on FEHBP and Medicaid data within the Medicare current beneficiary survey. It was actually used in payment for calendar years 2006 to 2010, and the – the old model used multipliers, additional multipliers for increased payments for Medicare beneficiaries who were low income subsidy or institutionalized. The new model, or the revised model, is based on actual Part D plan liability data, and it uses the – the model for this presentation uses 2007 diagnosis and estimates 2008 expenditures. This model was used in 2011, and then in 2012 and for 2013 we have a new calibration using 08 diagnoses and 09 expenditures.

As – as I said before, the goal – the goal of risk adjustment is to predict cost variations for subpopulations in Medicare Part D enrollees, and what we're doing here is we have separate model segments for non-low income subsidy beneficiaries, separate model segments for low income subsidy beneficiaries, and a separate segment for institutional Medicare beneficiaries.

The – in – in developing some – some performance measures, there – there's three data sources that I used. The first is the Health Plan Management System, referred to as H – HPMS. It houses the – the



plan sponsor – the – I think so. Okay. The plan sponsor bid – it has all the bids for – for Part D for the given year. And the – so that's – there we go. It's over there. Okay. So one data source is the – is the Bid Information Plan Sponsor. There's roughly 4,000 bids – bids per year. The number's been coming down over time. The second data source is the Risk Adjustment System, and that – that's a beneficiary-level file. It's roughly 32 million records, and it includes risk – risk score information as well as beneficiary demographic, enrollee status, and also final risk scores. The third data source is the Payment Reconciliation System known as PRS, and that is the actual Part D expenditures and plan liability, and that – that summarizes the prescription drug event data, or the PDEs. It also includes rebates. And I – I've subtracted the rebates from the plan liability to get the net plan liability so I could better – better ref – compare predicted to actual costs. Okay.

The first step in allocation of beneficiaries is to assign beneficiaries into model segments. There's eight model segments. There's two for low income subsidy beneficiaries, two for non-low income subsidy Medicare beneficiaries. One model segment for institutionalized Medicare beneficiaries, and then there's three segments for new enrollees. The analysis is done in the aggregate and also separately by model segment, plan type, beneficiary type, organization and predicted risk score bands. And specifically, it – due to proprietary information I'm not going to be presenting the organizational level analysis.

The overall results of the predictive accuracy of the model is that the revised model is predicting plan liability more accurately than the original Rx HCC Risk Adjustment Model in the aggregate. And it's also performing well in standardizing payments across all levels of risk, so when we break beneficiaries down into predicted levels of risk score, either in DSALs or some other – some other banding or summarization into – into groups, the – the revised model is performing well and also better than the original model.

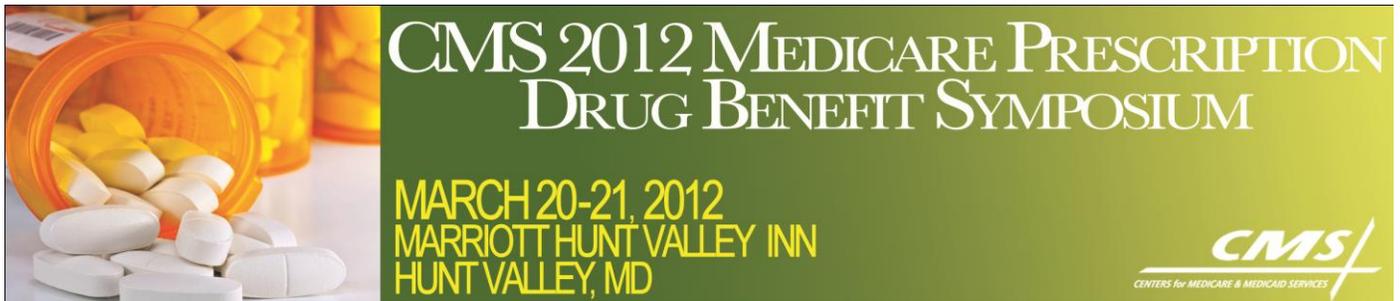
Okay. Next to – I'm going to provide an overview of the actual plan liability data for calendar years 07 to 2010. Have some charts and tables following, and we'll briefly cover those as well.

Any aggregate, and for the most segments, average plan line – liability is increasing from year to year. The segments are increasing at different rates over time. The low income subsidy beneficiary at plan liability are increasing a lot faster than the institutional and the non-low income subsidy. The institutionalized beneficiary group is – is increasing somewhat slower than the low income groups, but faster than the non-low income segment.

Byproduct of that is that the relative costs are decreasing for non-low income segments and the relative costs are increasing for low income and institutional, so that's a correction in the – in the bullet point. The sec – the second decreasing should be an increasing, should be the word increasing. So low income and institutional segments, the relative costs are increasing.

I was able to analyze some utilization measures for 2009 and 2010 including cost per script and scripts per beneficiary, and overall – also able to analyze for two years brand and generic utilization for both of those cost per script and scripts per bene measure. Average cost per script has increased for most segments with low income and institutionalized segments increasing faster than non-low income. And the average cost per script for brand are increasing significantly faster than generic.

Note that these – this is only two years of brand and generic and cost-per-script data, so we only have one – one trend data point, so keep that in mind.



The scripts per beneficiaries are just – the scripts per beneficiary are decreasing slightly in the aggregate and across most segments. The brand scripts per beneficiary are decreasing significantly, while generic scripts per beneficiary are increasing significantly.

The average risk scores are increasing over the period, with larger increases in MAPD relative to PDPs.

Okay. Now having summarized some of the actual data and some utilization measures, now can show – demonstrate – how to compare the predicted costs to the actual plan liability across the following dimensions and – and illustrate how – how the model is performing relative to individual segments.

To measure model Part D risk adjustment model performance, there's – there's two fundamental ways. One is to examine the predictive ratios based on plan liability. The ratio is defined as the predicted liability divided by the actual plan liability. The goal there is the one. One is perfect, means accurate payments. Above one means that we're over-predicting, below one means we're under-predicting.

The other way to – to evaluate performance is examining normalized costs, which is essentially the very same thing as looking at predictive ratios. Looking at normalized costs, we divide the predicted and the actual by the risk score and then compare the two. This – this presentation is going to focus on the predictive ratios.

Here's a benefit structure for 2010, and what we're trying to predict is the – is the light blue shaded area. So we're predicting costs for that area and we're comparing it to the actual plan liability in the blue shaded region.

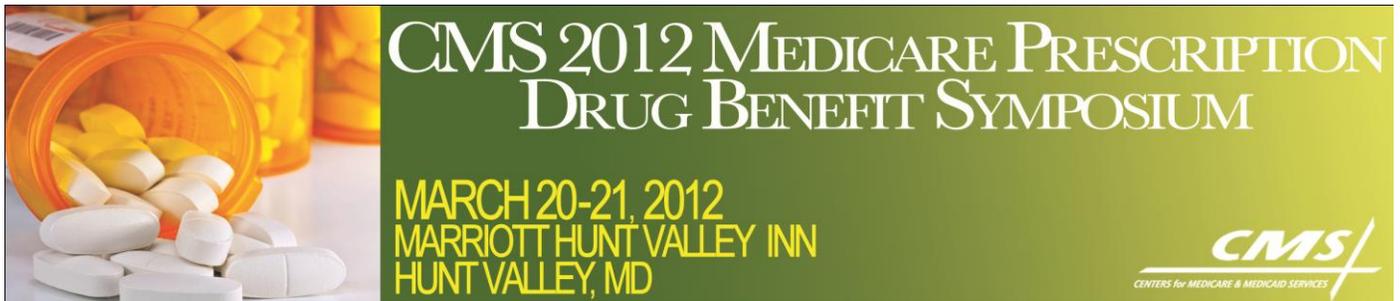
Okay. Here is a table – our first table that shows the actual plan liability for the four years. Note that these are – this – the all beneficiaries includes institutionalized, but the segments exclude the non-low income aged, non-low income disabled, low income aged, low income disabled, exclude institutionalized beneficiaries.

So what we're seeing is that average costs are increasing over the period. They're increasing but however there seems to be a slow down between 2009 and 2010, and we'll see later in the charts what – what the reasons for that are. It's – it's basically the – there's a huge decrease in scripts per bene utilizing brand drugs and there's a decrease in cost-per-script for generics, so better pricing and lower brand utilization resulting in a lower overall plan liability.

Looking at the different segments, we had non-low income aged and disabled are the lowest relative to the low income. The – the relationships of those costs relative to each other are similar across the years. However, the increases for low income relative to non-low income are much greater. We got 13% and 9% versus two and – two and three percent.

This is just a chart showing the – the levels in the plan liability.

The next shows the percentage changes in the plan liability. You'll see over there that the – the red going below zero is the non-low income aged, and that's a slight decrease there, relative to – for each year's overall mean, though, the – the segments – it's pretty much the same distance away from the overall average, so the difference between 07, 08 relative to all beneficiaries is – is – is the same across years when compared to the non-low income aged.



Okay this next chart shows a similar breakdown just for institutionalized beneficiaries. Similar – similar increases across the segments. You’ll notice that most of the institutionalized beneficiaries are low income aged, there’s 700,000. The other three groups there’s very few Medicare beneficiaries within the segment.

Cost per – per prescription. Two years’ worth of data. The cost increases are positive for – for all of the – for all of the segments, and also overall, and you’ll see significant increases for the brand cost per script. We’ll see some decreases for the generics for some of the segments and an overall decrease in cost-per-script for generic. And this is what’s feeding into the slowdown in the increase in the Part D plan liability from 09 to 2010.

This is just a – a graph of the cost-per-script percent changes.

Now we’re moving on to the prescriptions per beneficiary. We see here the slight decrease in the aggregate scripts per beneficiary. We’re seeing a large decrease in brand scripts per beneficiary, and then we’re seeing an increase in scripts per beneficiary for generics. This is just a chart of the previous table.

Okay, this next chart shows the generic share of total – total dollars, that is generic share of total plan liability. That’s a chart of a generic share.

Okay, now we get into the predictive ratios, which is going to allow us to measure the accuracy of the – of the model – the new model versus the old model. So what we have here is we have the beneficiaries. The old model predicted costs. The new model predicted costs. And the actual, and the we take the ratio to – to calculate the predicted ratio. You’ll see that – the – this is for calendar year 2008, and the way the predicted costs are calculated, it’s the – it’s the model mean or the overall actual per member, per month plan – actual plan liability for 08 multiplied by the risk score for all the beneficiaries in a segment. They’re all summed and their divided by the sum of the actual plant liability for all those beneficiaries.

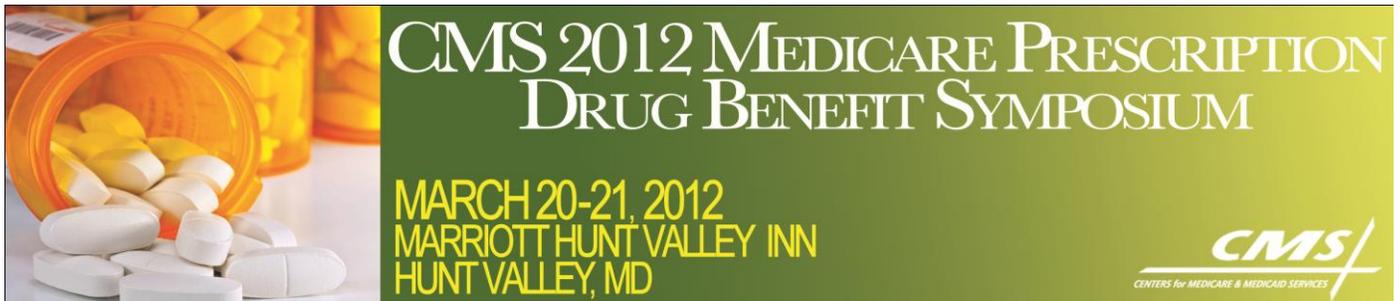
So what we’re looking at if you move all the way over to the right, the predicted ratios. As I stated earlier, the goal is 1.0. If it’s above one, we’re over-predicting, under one, we’re under-predicting. So, for the old model, you’ll see that for non-low income aged and disabled, we’re over-predicting. The ratios are 1.09 and 1.15, and we were under-predicting for low income aged and disabled.

The same pattern we’re seeing for the new, or the revised model, however the numbers are closer to one, that means that we are improving our predictive accuracy and predicting cost variations within Medicare subpopulations.

You’ll notice that the 04, the low income aged, the – the predictive ratio is lower by three points. And, however, when you combine the two segments, the low income aged and the disabled were actually improving the predictive accuracy over the entire low income group. We’re gaining seven points for the disabled, we’re losing three for the aged.

So this are the 2008 results. And that – that’s important because that’s the year in the calibration. And so these – when you take the weighted average of all these predictive ratios, you get back to one as it should be.

Now for 2010, a similar relationship holds as for 2008 across all the segments.



Okay, the next table shows the – similar as the two previous slides, but it shows only for the institutionalized beneficiaries. And we'll see the old model was significantly under-predicting the – the cost variation as their ratios were .83 all the way down to .78. However, the revised model, there's a – we're – we are – we drastically improved our predictive accuracy. We have 13 points for the non-low income aged. Eighteen for the low income aged. And 11 for the low income disabled. So, overall, for institutionalized bene – Medicare beneficiaries, we are predicting more accurately than the old model.

This is for 2010. Similar pattern as to 2008.

Okay, this next – next table shows a breakdown of Medicare beneficiaries into predicted risk score band. So all Medicare beneficiaries that we predict with a – with a risk score of .4 or below is in the first category. Fifty-seven percent of the Medicare beneficiaries are under the risk score of one, so half the people, and you'll see, are between zero and one, if you added those columns up. The – what's presented here are for 08, old model, new model predictive ratios and also the – for calendar year 2010, old model and new model predictive ratios.

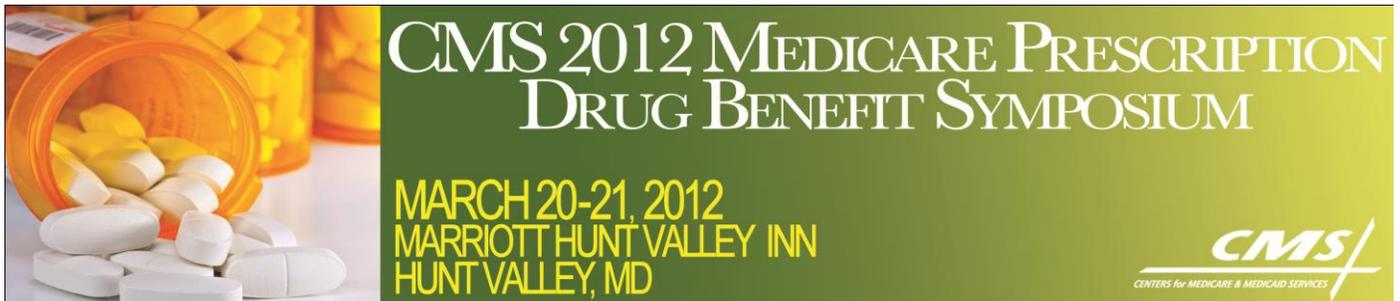
So what's the interpretation of this – this table is for the – for the Medicare, the healthiest, or Medi – the group of Medicare beneficiaries were predicting the – the – the – the lowest risk score for over-predicting the most. You see the 1.71. However, for where the majority of the people are, which are the middle – the middle risk score bands, we're predicting closer to one, using the old model. If you evaluated for each of the – the risk score bands, the old model versus the new model, for more segments the – the – the new model is – is predicting – there's a better risk score – a better predictive ratio for the – for the – using the new model. And in 2010, the new model is – is doing fairly better than the old model as well.

This chart shows just a graph of each of the two years, old model and new model, and we're just comparing all those values relative to one when the points above one, we're over-predicting, below one we're under-predicting. So in the middle – the middle risk score bands, we're doing a pretty good job. And in the – the details of the distribution, we're overpaying for the low end, the healthier people, we're underpaying for the sicker.

Okay, this – this next chart shows, for 2008 and 2010, the – just a chart of the predictive accuracy for MAPDs, and you'll see if you look over to the right, the far right-hand column compared to the – which is the 2010 new model, compare that to the column to the left, which is the old model. You will see that starting in risk score band .6 to .8, the predictive accuracy is improving. We – in the old model we were over-predicting by more than we are now in the new model, so the predictive accuracy is improving for – for the MAPDs, but we are still substantially over-predicting across risk score bands for MAPDs.

The purpose of this is to show the change in predictive accuracy across risk score bands over time. And the data calibration was 08. And the data calibration reflected the relative cost differences in 08. And as I mentioned earlier, the – the actual plan liability at a segment level was increasing differently for the – for the subpopulations. And – and over time, as that happens, you'd lose some predictive acc – some predictive – predictive accuracy, and you'll see that reflected in the middle risk score bands from the scores – from the predicted ratios going from, like, .97 to .96 to .95 or .98 to .96 to .95. So there is a – a – a slight degree of precision lost as relative cost factors change over time.

This is a chart of the predictive ratios across – for several – several segments. You'll see that it varies across segments. We're overpaying for the – this is a breakout of some of the previous charts – charts I



showed, and this – this is illustrating that, you know, we're still over-pre – over-predicting for non-low income and we're under-predicting for low income. And we're also under-predicting for institutional.

This is a chart that shows the predictive ratios by segment.

And here's the last pretty compelling chart. And this illustrates the predictive ratios for all segments combined separately for MAPDs and PDPs, and what we're seeing here is that for the majority of risk score bands, we're over-predicting for the MAPD, however, we are under-predicting for the PDPs. If you divide the predictive ratios, predictive – predictive costs divided by actual costs, and so when we're going across an individual risk score band, when you divide the predictive ratios, you get the ratio of the PDP costs and the MAPD costs. And so, what the ratio on the right is – represents is the actual PDP cost relative to the MAPD cost. And for every level – for virtually every level of risk score that we're predicting, so for the very healthy and also the very sick, if you enroll in a PDP, your costs are going to be roughly 14 to 16% higher than in the MA – MAPD world. And what that – what that shows is that the PDPs can do a – a better job in utilizing control – utilization control, and also encouraging and generic utilization.

And this is a chart of that, and that distance between the two lines is roughly the 14%.

Okay, so there's – there's – as I mentioned earlier when we looked at the – the 2008, 2009, 2010 predictive ratios for the new model, that the – the year-to-year changes in the actual per member per months are changing at different rates, and that can have an impact on the predictive accuracy of the model. And also, the predic – the effect of the – the calibration to payment year impacts the predictive ratios.

And one final thing is that the – the model that – that we use is calibrated using PDP data only, however in application and payment, the – the – the PDP calibrated model is used to pay both MAPDs and PDPs.

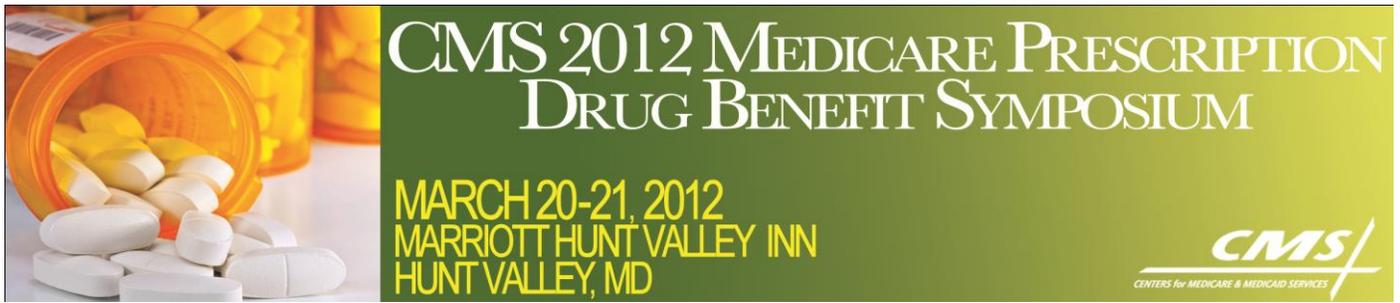
The results are that the actual plan liability for bene – beneficiaries in the model segments are increasing at different rates. And the final illustration is that the analysis shows the revised risk adjustment model is more accurately predicting plan liability than the original Rx HCC model.

And the results are for, you know, three different ways of looking at the predictive ratios. It's the old model versus the new model. We – we've done that, and in most segments the new model has predicted costs more accurately for Medicare subpopulations. It's also predicting costs more accurately for MAPDs and PDPs than the old model.

In looking at both the old and the new model over time, the old model is predicting costs less accurately over time than the new model. And in – as – as time goes by and the – there's greater utilization and greater – greater utilization control in PDPs and generic substitution, we're going to see that – that the new model is going to be predicting costs more accurately over time.

Okay. I'm going to go to the assessments. It is time to conduct the assessment. Please get out your ARS response cards. We would encourage all of you to participate. As a reminder, if you are seeking CPE credit, you must respond to all assessment and evaluation questions. After the questions and responses are read, you will have ten seconds to respond. You will see the timer on the screen.

Which of the following statements best describes CMS's Rx HCC risk adjustment model? One, a, predicts cost variation among subpopulations of Medicare Part D enrollees. Two, b, predicts which Part D



plan is best for each Medicare enrollee. Three, c, determines how many Part D plans should be in a specific region. Four, d, determines the effectiveness of special needs plans. Five, e, all of the above. Please vote now. You have ten seconds.

The poll is now closed. Let's look at the results. Okay. All right. Let's move on.

Number two. Question number two. Okay. The three data sources used by CMS to measure the performance of the Rx HCC Risk Adjustment Model include one, a, Part D pharmacists' self-reports, Part D claims data, CMS annual reports. Two, b, CMS's payment reconciliation system, Part D sponsor bid data, and CMS's risk adjustment payment system. Three, c, Part D claims data, diagnosis codes, Part D sponsor bid data. Four, d, Part D sponsor self-reports, diagnosis codes, Part D claims data. Five, e, all of the above. Please vote now. You have ten seconds.

The poll is now closed. Let's look at the results. All right. The correct answer was two, b. we got 73%.