Evaluating ACO Efficiency: Risk Adjustment within Episodes

Sponsored by the Society of Actuaries

Prepared by
Bill O'Brien
Rong Yi
Chun Yau
Yan Li
Milliman Inc.
December 2013

Table of Contents

Acknowledgments

Accountable care organizations (ACOs) have been heralded as a promising way to organize health care delivery and manage population health within an integrated and cost-effective environment. This research report is funded by the Society of Actuaries (SOA) to assist ACOs in developing tools to assess the likely financial performance of ACOs under common contractual arrangements.

We would like to thank the Project Oversight Group members for their valuable input, critique, and insight into the various issues surrounding the study design and the interpretations of study results; together they helped broaden the scope of our views and ensure the depth and relevancy of this research: Louise Anderson, Dan Bailey, John Bertko, Mark Bethke, Kristi Bohn, Dewayne Ullsperger, Rina Vertes (chairperson), Greger Vigen, and Jim Whisler.

We thank Steven Siegel, Sara Teppema, and Barbara Scott at the SOA for provided staffing and logistical support, and OptumHealth for providing a research license and technical support of the Symmetry Suite Episode Treatment Group product.

We expect that there will be a continuous stream of ACO-related actuarial research in the coming years. We welcome comments and feedback and look forward to refining the assumptions and analytic approaches taken in this research. Please contact Bill O'Brien at bobrien@evolenthealth.com and Rong Yi at rong.yi@milliman.com.

Executive Summary

Within an accountable care organization (ACO), a common starting point to evaluate cost efficiency is to examine patient "episodes of care." Common targets for review are episodes that are high volume, have significant implications for population health, and have varied utilization patterns and cost outcomes. For most episode types (e.g., diabetes care), certain patient episodes cost more and use more services than others of the same episode type. This study aims to answer the question of how much of an episode's cost and use of specific services can be attributed to the patient's own health status, as opposed to being driven by other factors such as the care setting, patient preference, and provider practice patterns. In other words, do episode costs vary so widely because of patient characteristics, including health status, or are they related to other, nonclinical factors?

Other studies have found that standard risk adjustment methods do not meaningfully explain episode cost differences *within* an episode type (as opposed to *between* episode types, where risk adjustment is highly explanatory of cost differences by design).[1] One study found that risk adjustment reduces dispersions in cost but exacerbates outliers.[2] The original hypothesis of this study was that risk adjustment factors could be specifically customized to each episode type to further explain cost variation. It was also hypothesized that risk adjustment application within an episode type may at least partially explain the variation in use of certain services within an episode type. For example, could high inpatient hospital use be related to the underlying health

[1] See, for example, J. William Thomas, "Should Episode-Based Economic Profiles Be Risk Adjusted to Account for Differences in Patients' Health Risks?", available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1702525/; Peter S. Hussey, M. Susan Ridgely, and Meredith B. Rosenthal, "The PROMETHEUS Bundled Payment Experiment: Slow Starts Shows Problems in Implementing New Payment Models," available at http://content.healthaffairs.org/content/30/11/2116.abstract; and the information and blog posts available at http://www.hci3.org/.

[2] Thomas MaCurdy et al., "Challenges in the Risk Adjustment of Episode Costs" (Feb. 2010), available at http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Reports/downloads/MaCurdy_ERA_2010.pdf.

status of the ACO's specific group of diabetic patients? Could higher use of advanced imaging among an ACO's lower back pain patients be similarly attributed to the patients' overall average health? Could the performance of surgery, when a nonsurgical treatment is available, be explained by the health status of the patients?

We found that this additional layer of risk adjustment within each episode still leaves meaningful portions of episode cost and usedifferences unexplained—and, thus, presents potential savings opportunities for ACOs.

We note that our analysis does not attempt to control episode costs or utilization for quality, patient outcomes, and patient experience. Careful identification and consideration of the contributors to higher quality and more successful health outcomes is critical to patient health, as well as to ACO success. The alignment of provider payments to the fulfillment of measurable quality and outcome standards is a cornerstone of ACO contracts. Further, patient experience is often an important performance metric that is reflected in provider payment and is commonly driving ACO infrastructure, including the increased availability of providers through expanded office hours and additional forms of communication, such as e-mail, telephone, group visits, and virtual online visits. This study is solely based on claims data and does not address the clinical and experience aspects that are also critical to the success of an ACO. Claims cost variation analysis is an important starting point for prioritizing financial savings opportunities; continuously taking advantage of cost savings opportunities will be an important determinant of an ACO's success.

Risk adjustment works best for large populations.[3] In order to evaluate our risk adjustment methods, we simulated an ACO population of 15,000 members to build a reasonable distribution of the episode occurrence and average episode costs. We also reviewed how the underlying health status of the members contributed to episode cost variation. The episode occurrence rate in an ACO's covered population depends on the underlying prevalence of those diseases and medical conditions that trigger an episode during the designated time interval required for a complete episode.

The source data is a three million non-Medicare member sample from the Truven MarketScan® Commercial Claims Database for claims incurred in 2009 and 2010. We used the Symmetry suite of products, as provided by Optum Health, to assign individual claims to patient episodes and episode types. We used the Milliman Health Cost Guidelines® to assign claims to specific types of service. A common fee schedule was used for claim allowed amounts in order to remove price variances relating to geography and provider contracting. As a result, we were able to produce traditional actuarial values for utilization rates and average cost per service within the episodes studied.

Ten episode type groups (Symmetry "Base ETGs") were selected based on their overall frequency of occurrence, relevance to population health management, average costs, and variance of cost per episode. Within these episode types, industry-standard concurrent risk adjustment methods were used to develop separate risk adjustment factors for four utilization metrics and four service type cost categories. The methodology underlying this analysis is similar to the risk adjustment methodology used for commercial and Medicare programs. The key difference is that our episode risk factors are specific to the episode or illness and reflect only the *additional* cost or utilization *within* the specific episode type, as opposed to the episode type's impact on *total* allowed costs. The resulting table of risk factors can be applied to an ACO population to estimate the risk-adjusted expected average cost per episode. This adjusted cost per episode is compared to actual costs in order to determine the ACO's cost efficiency measure for that episode type, expressed as a percentage of expected costs.

[3] See Syed Muzayan Mehmud and Rong Yi, "Uncertainty in Risk Adjustment" (Sept. 2012), available at http://www.soa.org/research/research-projects/health/uncertainty-risk-adjustment.aspx.

By providing the technical details and risk factors used in this study (see Appendix A), ACOs will be able to evaluate their episode cost experience, after they have taken steps to normalize their own experience for provider payment schedules and health status differences from this report. ACOs can also use this study to evaluate the variation and potential impact of reducing the variance within episode types.

Major findings of the study include the following:

- *Patient diagnostic history explains relatively little of the differences in service type costs or utilization rates within an episode:* While patient age, gender, and recent diagnostic history tend to correlate relatively well with total episode cost for the episode types we studied, they do not explain the variation in different cost components at the episode level. We found that the correlation between service utilization and patient age, gender, and recent diagnostic history is even weaker than the correlation for cost components. This suggests that other factors, not examined in this study, account for the unexplained wide cost and utilization variance. These other factors may include patient preferences, physician practice patterns, and delivery system structure, among others.
- *New episode average values reported:* We believe that this new approach to risk adjustment within an episode adds significantly to the current ACO evaluation tools. In Appendix E we report episode-level averages, as well as percentile distributions for individual episodes and for an ACO with 15,000 members, including the following data:
    - Average total cost and service type cost per episode, including percentile distributions
    - Episode occurrence rates, expressed as episodes per thousand enrollees over the two-year experience period
    - Average utilization and unit costs, by service type within the episode (individual episodes only).
- *Four of the 10 episode types evaluated did not produce statistically significant correlation of episode costs to age, gender, and diagnostic history: asthma, congestive heart failure, high cholesterol, and high blood pressure:* For the remaining six episode types, we developed risk adjustment algorithms based on the CMS-HCC methodology. This risk adjustment is applied to service type costs and total cost only. The R^2 statistic values from the risk adjustment models we developed for service type costs are low but meaningful. They range from 0.11 to 0.38 for the selected episode types. Pregnancy episodes exhibited very high correlation, with an R^2 of over 0.80 for professional, inpatient, and total costs.
- *The risk-adjusted arithmetic mean exceeds the geometric mean for most episode types:* We find that the average cost per episode, after risk adjustment, exceeds the 50th percentile for most episode types, indicating a right-skewed, "long tailed" distribution. Our analysis suggests that large cost savings would be available to a 15,000 member ACO that is able to move the episode average costs to the median (50th percentile). A schedule of projected savings is included herein.

Additional Considerations

Actuaries reading this report should consider both the strengths and weaknesses of episode-based risk adjustment analyses. Traditional actuarial population studies seek to quantify the impacts of changes in cost or episode occurrence metrics, but we note that episodes do not fit neatly into the traditional actuarial framework and the concept of utilization rate:

- Nonchronic episodes are not time defined to match a financial reporting period. Rather, they are defined by the absence of any claims related to the episode for a predetermined amount of time (the "clean period"). A subsequent claim would start a new episode.
- Chronic episode types are defined as having a 365-day period. Since the episode can begin on any day of the year, two full years of claims experience are needed to ensure complete episode reporting.
- There can be multiple episodes within a calendar year or straddling across two years, and the length of an episode can also vary.
- Per member per month amounts may not be appropriate and must be carefully evaluated since the episodic data do not correspond to monthly or annual periods.
- Incomplete episodes exist because episode start or end dates fall outside the claims time period used to process the ETG grouper. Excluding incomplete episodes in an analysis of costs improves understanding of the costs per episode, but also results in an understatement of total claims costs at the population level. This is another example of the assertion stated above that episode analysis and episode occurrence rates are not perfectly consistent with the utilization rates employed in more traditional actuarial analysis.

Finally, unless otherwise specified, all cost amounts are as of 2010 and are normalized for fee schedules.

Introduction

The Affordable Care Act encourages the development of accountable care organizations (ACOs) as a means to promote reductions in medical expenses in a clinically appropriate manner. A common method for measuring the cost of care within an ACO is to group services into patient episodes of care related to a specific condition or acute care episode. Episode methodologies use computer software to combine health care claims related to a common disease or condition into temporally and clinically defined groupings (episodes). This approach has the advantage of evaluating costs and service usage from the perspective of the attending physician, moving linearly along the condition progression and course of care provided. In this way, this approach provides deeper insights and more actionable results that augment the traditional actuarial method of comparing aggregate rates of service utilization and average unit costs by service category to benchmarks.

It has been widely shown that the average cost for a given episode type varies greatly, for reasons that are only partially understood. Many factors influence the cost and utilization within a patient care episode. These factors can be broadly categorized as the following:

- Provider-specific factors
- Market- or geography-specific factors
- Industry-wide factors
- Patient-specific factors.

Provider-specific factors include specialty, years of experience, medical school training, practice setting, geographical location, network arrangement, and payment arrangements. These impact issues such as a provider's treatment patterns, choice of referrals, and professional compensation.

The market- and area-specific factors include urbanicity, regional norms for unit prices, and concentration of health care providers and facilities. Industry-wide trends and factors include new drugs and technology for diagnosing and treating diseases, medical inflation, insurance, and payment reforms.

Patient-specific factors include the patient's age and gender, medical comorbidities, type of insurance, level of health insurance benefit coverage, functional status, care preferences, and other socioeconomic circumstances (such as language, education, and income) that influence how a patient seeks medical care and interacts with the health care system. Also impacted are the patient's compliance with physician instructions and their ability to self-manage care.

Risk adjustment has long been used for a variety of purposes, including setting program payment rates, analyzing cost and utilization trends, evaluating medical management programs, and conducting cost-effectiveness analysis. Diagnosis-based risk adjustment methodologies use membership information and the medical diagnosis codes on health care claims to predict or explain the variation in cost and utilization. Variations left unexplained by risk adjustment can be interpreted as resulting from factors that are not accounted for by the models, such as provider specific factors, market and area factors, and industry trends, as well as from inherent error within the risk adjustment methodology or data reporting.

This study aims to quantify the extent to which patient characteristics, including demographic and diagnostic history, impact the components of cost of a patient episode of care, as well as the utilization of specific services within an episode. We aim to answer the following question:

How much variance in patient episode experience (cost and utilization) can be explained by applying an industry-standard methodology of risk adjustment that incorporates patient demographics and diagnostic history?

We used a random sample of three million members from the 2009 and 2010 Truven MarketScan® Commercial Claims Research Database (hereafter MarketScan). Members in our sample were eligible for both medical and pharmacy benefits in both years and had 24 months of continuous coverage. To ensure that the sample is representative of the commercially insured population, we did stratified sampling using the age and gender distribution of the entire MarketScan 2010 dataset. Claims are grouped into clinical episodes using the Optum Symmetry Episode Treatment Groups (ETG) software (version 8.0). Requiring 24 months of continuous coverage helps reduce the number of incomplete episodes and the number of ungroupable claims and improves data integrity and episode sample size for the analysis.

We used the Milliman Health Cost Guidelines® Grouper (HCG) to categorize claims to type of service categories. The HCG assigns each line a unique HCG category, and there are more than 80 service categories in the HCG grouping system level we used. To neutralize provider-contracting reimbursement-level differences in the data, we repriced the allowed amount at the claim level using a common fee schedule, which generally reflects the average allowed amount for each procedure code. Refer to Appendix B for more details on the common fee schedule.

Further adjustments were made to include births and deaths during the period such that the overall distribution is consistent with the MarketScan database prior to our introducing the exclusions. MarketScan's commercial claims database is mainly from large employers and health plans and resembles a large group commercial health insurance pool. The Symmetry ETG grouper was used to further distinguish episodes with complications, surgery, and/or comorbidities.

Ten episode type groups (Symmetry "Base ETGs") were selected based on their overall frequency of occurrence, relevance to population health management, average costs, and variance of cost per episode. The ten included episodes are listed here with common shorthand in parentheses:

1. Asthma
2. Cerebral vascular disease
3. Chronic obstructive pulmonary disease (COPD)
4. Congestive heart failure (CHF)
5. Diabetes
6. Hyperlipidemia (high cholesterol)
7. Hypertension (high blood pressure)
8. Ischemic heart disease (coronary artery disease [CAD])
9. Joint degeneration, localized—back (lower back pain)
10. Pregnancy with delivery

Within these episode types, risk adjustment factors were developed for four utilization metrics along with four service type categories plus total costs:

Per Episode Utilization Metrics Include:

1. Medical/surgical inpatient hospital admissions
2. Avoidable emergency room visits (as defined by MediCal)
3. Advanced imaging services (CT/MRI/PET)
4. Physician office visits;

Per Episode Cost Metrics Include:

1. Inpatient hospital allowed per member per month (PMPM) cost
2. Outpatient hospital allowed PMPM cost
3. Physician and other allowed PMPM cost
4. Prescriptions drug allowed PMPM cost
5. Total allowed PMPM cost.

Episode Costs before Risk Adjustment

Detailed Costs by Service Category Reports

We include as a summary of observed cost (Table 1) and utilization rates (Table 2) by episode type and by detailed service type categories, where service types were defined using the Milliman Health Cost Guidelines groupings, and normalized fees. We believe that this detail can provide insight into the resource use patterns and intensity at the episode level, even though data are not risk adjusted. For instance, some episodes such as hyperlipidemia and hypertension have almost no inpatient utilization, and much of the care is managed with office visits and prescription medicine. On the other hand, episodes such as cerebral vascular disease, congestive heart failure, and ischemic heart disease have almost half of the expenditure in the inpatient setting. Detailed reports can be accessed in Appendix C. The average normalized costs per episode, for each episode type, and for each service type are shown below. Note that chronic episode types have episode lengths of one year, or 365 days, from the first related service date, and the total and service type costs therefore represent annual expenditures.

Table 1. Overall Average Cost per Complete Episode

Episode Type	Inpatient Cost	Outpatient Cost	Professional/ Other	Prescription Drugs	Total Episode
Asthma	$41	$155	$227	$501	$925
Cerebral vascular disease	2,976	1,346	1,467	190	5,979
Chronic obstructive pulmonary disease	415	426	489	552	1,882
Congestive heart failure	1,955	852	872	200	3,879
Diabetes	48	232	521	1,357	2,159
Hyperlipidemia	0	62	121	326	508
Hypertension	1	179	234	297	712
Ischemic heart disease	3,448	1,409	1,392	758	7,007
Joint degeneration, localized—back	418	861	1,283	294	2,856
Pregnancy with delivery	6,835	2,108	5,166	156	14,266

Table 2. Overall Utilization per 1,000 Episodes

Episode Type	Inpatient Admissions (Medical/ Surgical Only)	Avoidable ER Visits	Advanced Imaging	Office Visits
Asthma	6	2	11	1,107
Cerebral vascular disease	168	12	346	1,159
Chronic obstructive pulmonary disease	38	12	70	1,426
Congestive heart failure	108	2	17	1,248
Diabetes	6	3	16	1,972
Hyperlipidemia	0	0	0	770
Hypertension	0	2	13	1,303
Ischemic heart disease	134	1	23	1,652
Joint degeneration, localized—back	19	20	318	2,037
Pregnancy with delivery	1	3	6	1,664

Episode Cost Distributions before Risk Adjustment

Cost and utilization levels vary significantly from episode to episode. Table 3 illustrates the distribution of total episode costs for each episode type. Similar tables for each service type within each episode type are included in Appendix D.

Table 3. Episode Cost Distribution before Risk Adjustment

Episode Type	25%	50%	75%	90%	95%	Coefficient of Variation
Asthma	$144	$424	$1,172	$2,331	$3,386	4.93
Cerebral vascular disease	321	1,398	5,224	18,204	28,522	6.42
Chronic obstructive pulmonary disease	142	473	1,860	4,838	8,542	5.53
Congestive heart failure	193	730	2,290	10,797	19,284	5.77
Diabetes	371	1,130	3,035	5,604	7,620	4.46
Hyperlipidemia	107	248	733	1,368	1,716	40.25
Hypertension	159	398	943	1,659	2,389	11.27
Ischemic heart disease	548	1,853	4,994	20,145	36,525	3.46
Joint degeneration, localized—back	300	1,025	2,840	6,641	11,261	7.09
Pregnancy with delivery	10,480	13,239	17,137	22,067	26,205	0.94

We also include the coefficient of variation (CV) as a simple way to summarize the dispersion of episode cost. One can quickly observe that while the average episode costs for hyperlipidemia

and hypertension are not among the highest cost of the 10 episode types we examined, they have the highest CVs. As shown in Tables 2 and 3, most of the episode cost is associated with prescription medicine and office visits. Given that we have already normalized the fees, we hypothesize that most of the variation is driven by visit counts or the prescription of expensive drugs. Among the hypertension episodes, we observe advanced imaging tests at a rate of 13/1,000. This low incidence rate of a high cost service also contributes to the high CV.

A detailed analysis of the episode cost distributions, service type amounts, and comorbidity levels could lead to many interesting conclusions. For instance, an increasing rate of Cesarean sections (C sections) has been noted in recent literature, reaching almost 33% on average in 2011. A recent paper in *Health Affairs*[4] found that the rate of C sections varies from 7.1% to 69.9% across hospitals and suggests that much of the variation is due to practice patterns. In the population used for this study, where members are mostly from the commercial large group market with rather rich benefits, about 30% of completed pregnancies did not have complications or comorbidity, yet more than 22% of these uncomplicated pregnancies were C sections (see Table 4). This suggests some degree of overuse of C sections and potential savings opportunity for ACOs.

Table 4. Pregnancy with Delivery Episodes Distribution

ETG Label	Episode Count	Average Allowed Amount	Coefficient of Variation
No complication, no comorbidity	8,509	$8,958	0.90
No complication, no comorbidity, with C section (22.9%)	2,521	13,526	0.44
No complication, with comorbidity	581	8,239	0.87
No complication, with comorbidity, with C section (24.1%)	184	14,189	0.46
With complication, no comorbidity	13,671	12,018	0.80
With complication, no comorbidity, with C section (41.8%)	9,817	18,320	0.93
With complication, with comorbidity	1,319	13,365	1.00
With complication, with comorbidity, with C section (46.2%)	1,133	21,935	0.93

Risk Adjustment

A typical approach to explain a wide variance in claims cost is to apply risk adjustment factors that adjust the cost of each episode for the underlying expected claims cost differences due to differing patient demographics and existing medical conditions as reported in claims data as diagnosis codes.

Diagnosis-based risk adjustment methodologies use membership information and the medical diagnosis codes on health care claims to predict or explain the variation in cost and utilization.

[4] Katy Backes Kozhimannil, Michael R. Law, and Beth A. Virnig, "Cesarean Delivery Rates Vary Tenfold among US Hospitals; Reducing Variation May Address Quality and Cost Issues," Health Affairs, April 2013, available at http://content.healthaffairs.org/content/32/3/527.abstract?sid=9f9abe9f-0215-498b-82ee-5942bbc98f93.

Variations left unexplained by risk adjustment can be interpreted as resulting from factors that are not accounted for by the models, such as provider-specific factors, market and geography factors, and industry-wide factors, as well as from inherent error within the risk adjustment methodology or data reporting.

We developed risk adjustment models based on the Hierarchical Condition Categories (HCC) risk adjustment system used by Medicare Advantage. The HCC system is primarily based on an individual member's age, gender, and medical diagnoses as coded in standard health care claims. It begins by selecting claim records coded by clinicians during a face-to-face encounter, then grouping the ICD-9-CM diagnosis codes into clinically homogenous groups called condition categories (CCs), imposing clinical hierarchies to create hierarchical condition categories (HCCs), and assigning risk factors (or coefficients) to the HCCs based on restricted multivariate regression analysis. A risk score at the individual level is derived by summing up the risk factors across all of an individual's conditions. HCC risk adjustment is used in Medicare and Medicaid managed care for program payment and will be used in the commercial individual and small group markets to help stabilize premiums after 2014, though with different risk factors and a few additional HCCs.

We developed separate HCC risk adjustment models for inpatient hospital, outpatient hospital, prescription drug, physician, and all other costs at the episode level, as well as models that predict cost for the entire episode. We used the allowed amount to mitigate the impact of benefit design and patient responsibility on cost. To reduce the impact of high-cost outliers, we excluded the top 0.5% of the episodes by cost. Linear regressions were then used to estimate total cost and service type cost for each episode type. The R^2 statistics (in percentage terms) are reported for these measures in Table 5.

Episode-based risk adjustment has two primary applications to ACOs. First, we examine the explanatory nature of episode-based risk adjustment to individual episodes. Then we turn our attention to the implications and simulated outcomes of applying episode-based risk adjustment to large populations, such as those served by ACOs.

Evaluating the Model

The R^2 statistic has been used as one of a few key measures of "goodness of fit" of risk adjustment models. The closer the statistic is to 100%, the more accurate the model is. We can see that asthma, congestive heart failure, hyperlipidemia, and hypertension episode types have very low R^2 values throughout the five cost measures, suggesting that patient age, gender, and medical comorbidity do not explain the cost variation at the episode level. In other words, factors such as practice patterns and patient preferences that are not included in the models are driving the variations instead. The other episode types have relatively higher R^2 values. The highest R^2 value is observed in the pregnancy with delivery episode type, in which the total episode cost model has an R^2 of 86.8%. While this suggests that practice patterns and patient preference do not seem to be driving most of the cost variation for completed pregnancies, it is not the same to say that care itself is efficient.

Table 5. Risk Adjustment Models R^2 Statistics by Service Type Costs

Episode Cost Measures	Inpatient Cost	Outpatient Cost	Professional/ Other Cost	Prescription Drugs Cost	Total
Asthma	2.40%	3.06%	7.22%	1.56%	4.96%
Cerebral vascular disease	19.38	20.59	28.39	0.91	32.25
Chronic obstructive pulmonary disease	9.91	3.93	13.30	3.84	13.42
Congestive heart failure	8.77	2.60	6.96	0.39	5.31
Diabetes	5.19	11.59	39.42	11.13	21.29
Hyperlipidemia	0.00	0.08	0.80	5.39	3.31
Hypertension	0.11	11.12	7.71	2.15	5.53
Ischemic heart disease	27.58	15.91	39.27	1.74	37.84
Joint degeneration, localized—back	2.63	18.42	5.72	3.18	23.74
Pregnancy with delivery	83.57	34.30	85.80	12.46	86.83

We also attempted to develop risk adjustment models for utilization counts—counts of medical and surgical admissions, counts of avoidable emergency room visits, counts of advanced imaging test, and counts of office visits—using a variety of linear and nonlinear modeling techniques. We also trimmed outliers where applicable. Some members on our research team had developed robust risk adjustment models for utilization counts at the population level using similar techniques. However, despite our best efforts, we found that the same explanatory variables—age, gender, medical comorbidity—did not present much predictive power for the utilization counts at the episode level, with the possible exception of inpatient admissions for ischemic heart disease. For this reason, we are not including risk adjustment models for utilization counts in this report.

It is important to point out that our analysis evaluates the correlation between patient characteristics and medical costs, not the extent to which delivered care is efficient. If the care provided within our base database is inefficient, a strong correlation of another population's costs to this baseline would not indicate efficiency. In this way, our outcomes are dependent on the input data.

Given the low R^2 values for many of the evaluated cost metrics, we eliminated the episode types and service types where the risk adjustment model R^2 fell below 10%. Table 6 shows the six episode types and the applicable service types that are risk adjusted for the rest of the study.

Table 6. Average Cost per Complete Episode, R^2 >10%

Episode Type	Inpatient Cost	Outpatient Cost	Professional /Other Cost	Prescription Drugs Cost	Total
Cerebral vascular disease	$2,976	$1,346	$1,467		$5,979

Chronic obstructive pulmonary disease			489		1,882
Diabetes		232	521	$1,357	2,159
Ischemic heart disease	3,448	1,409	1,392		7,007
Joint degeneration, localized—back		861			2,856
Pregnancy, with deliver	6,835	2,108	5,166	156	14,266

After risk adjustment, we recalculated the distribution of episode cost and service type costs by the six episode types and compared it to the distribution before risk adjustment. The results are presented in Table 7. We can see that having normalized by age, gender, and medical comorbidity, the episode cost distribution is much flatter than before. Said differently, if risk adjustment were not performed, using the raw cost distribution could lead to unrealistic potential savings estimates.

Table 7. Episode Cost Distribution before and after Risk Adjustment

	Episode Type	25%	50%	75%	90%
Before	Cerebral vascular disease	$321	$1,398	$5,224	$18,204
After	Cerebral vascular disease	2,484	5,069	7,744	12,001
Before	Chronic obstructive pulmonary disease	142	473	1,860	4,838
After	Chronic obstructive pulmonary disease	939	1,566	2,094	3,358
Before	Diabetes	371	1,130	3,035	5,604
After	Diabetes	1,375	1,595	2,564	3,912
Before	Ischemic heart disease	548	1,853	4,994	20,145
After	Ischemic heart disease	2,953	4,872	8,631	15,429
Before	Joint degeneration, localized—back	300	1,025	2,840	6,641
After	Joint degeneration, localized—back	2,249	2,632	3,743	4,739
Before	Pregnancy with delivery	10,480	13,239	17,137	22,067
After	Pregnancy with delivery	12,705	13,907	15,598	17,477

Risk Adjustment Applied to Large Populations

The previous sections dealt with application of risk adjustment to individual episodes. We now turn our attention to applications of episode-level risk adjustment to large populations, including those of an average size served by most ACOs. Generally ACOs contract for services to be provided to the ACO's primary care physicians' "attributed populations" for a given payer, employer, or government-sponsored health program. Methods for attribution vary, but generally they attempt to identify the population that currently accesses the ACO's network of physicians for the majority of their primary care needs. The ACO assumes some form of financial accountability for the total cost of care for these attributed members. Often, shared savings arrangements allow for some portion, or all, of the savings from agreed-upon expected costs to be shared with the ACO subject to satisfying quality requirements.

It is important within these financial arrangements to understand the impact that episode occurrence rates within the attributed population have on the actual PMPM total population costs, generally the basis for determining sharing savings amounts. Higher than average occurrence rates for higher cost episode types can lead to poor financial performance by the ACO, even if the average cost per episode, for all episode types, is cost efficient. Similarly the average cost per episode impacts the actual PMPM total population costs. Inefficient episode costs can be masked by lower than average episode occurrence rates, in which case savings may ultimately be shared with inefficient ACOs (as measured by average cost per episode after risk adjustment), in opposition to the primary goals of ACO creation.

Episode Occurrence Rates

We used Monte Carlo simulations to estimate episode occurrence rates at the population level for an ACO of 15,000 members. Specifically, we took 10,000 random draws of 15,000 unique members from the aforementioned three million member dataset. Note the following points:

1. These episodes occurred over a two-year period.
2. Episode occurrence rates include only *complete* episodes, as defined by the Symmetry grouper.
3. By design of the simulation, we assume that members join the simulated ACO randomly and ignore the systematic differences among ACOs. In practice, however, the benefit design, provider network, marketing approach, and other factors create nonrandom enrollment and may produce adverse or positive selection that affects the occurrence rates for some episode types.

Table 8. Occurrence Rates of Complete Episodes over a Two-Year Period

Episodes per Thousand Enrollees	Mean	25%	50%	75%	90%
Asthma	79	77	79	81	83
Cerebral vascular disease	7	6	7	7	8
Chronic obstructive pulmonary disease	11	10	11	12	12
Congestive heart failure	4	3	4	4	5
Diabetes	78	76	78	80	82
Hyperlipidemia	133	130	133	135	138
Hypertension	218	215	218	222	224
Ischemic heart disease	26	24	26	27	28
Joint degeneration, localized—back	65	63	65	66	68
Pregnancy with delivery	15	14	15	15	16

ACO Savings Estimation

Most ACO contracts base sharing of population cost savings on a target of either total cost of care per member per month (TCOC PMPM) or a percentage of revenue. Savings beyond a predetermined threshold are then shared between the ACO participants. While TCOC PMPM is a standard measure that actuaries use to evaluate the financial performance of an organization, providers in an ACO often find it difficult to relate to such measures because their clinical practice

is mostly organized by treating patient episodes. To bring the population-based view and the episode-based view to a common ground, the episode costs need to be considered together with the episode occurrence rates at the population level.

In practice, both the episode occurrence rates and episode costs vary. An ACO could potentially get a windfall of "savings" simply from a random one-time reduction in episode occurrence rates without any change to the way they practice medicine. The opposite could also happen, which would put the ACO at adverse risk due to higher occurrence rates of episodes.

To understand how episode occurrence rates and episode cost variation impact TCOC PMPM, we ran two scenarios. In one scenario, we created 10,000 randomly formed ACOs of 15,000 members assuming fixed episode costs and varying occurrence rates. In the other scenario, we created 10,000 randomly formed ACOs of 15,000 members assuming fixed occurrence rates and varying episode costs. We then compared the TCOC PMPM in the two scenarios. In both scenarios, the 50th percentile TCOC PMPM is the same for each episode type. The results are presented in Table 9.

For most episode types, the cost per episode causes more risk-adjusted variation than occurrence rates. Take, for example, ischemic heart disease. The potential TCOC PMPM impact could range from $7.20 (25th percentile) to $8.19 (90th percentile) simply due to variation of occurrence rates. Assuming a constant episode occurrence rate but allowing episode costs to vary according to the risk-adjusted distribution, TCOC PMPM can vary more widely, from $7.00 to $8.62. Except for pregnancy with delivery, all episode types have a similar pattern as ischemic heart disease. This suggests that contracts that base gain sharing amounts on risk-adjusted PMPM amounts are providing incentives for episode cost reduction and not just rewarding low occurrence rates.

Pregnancy with delivery presents an opposite pattern; that is, occurrence rates cause a wider spread in TCOC PMPM than average cost per episode. This is intuitive because pregnancy and delivery rates fluctuate quite a bit from year to year for a population this small, more than the other episode types, which are mostly chronic in nature.

Table 9. Impact of Occurrence Rate and Episode Cost on Total PMPM Cost

	Cost per Complete Episode Percentile				
	Episode Type	**25%**	**50%**	**75%**	**90%**
Fixed episode cost	Asthma	$3.01	$3.08	$3.16	$3.23
Fixed occurrence rate	Asthma	2.99	3.08	3.18	3.27
Fixed episode cost	Cerebral vascular disease	1.53	1.67	1.82	1.93
Fixed occurrence rate	Cerebral vascular disease	1.44	1.67	1.91	2.15
Fixed episode cost	Chronic obstructive pulmonary disease	0.80	0.86	0.92	0.97
Fixed occurrence rate	Chronic obstructive pulmonary disease	0.76	0.86	0.96	1.07
Fixed episode cost	Congestive heart failure	0.51	0.58	0.65	0.71
Fixed occurrence rate	Congestive heart failure	0.44	0.58	0.73	0.88

Fixed episode cost	Diabetes	6.90	7.09	7.28	7.46
Fixed occurrence rate	Diabetes	6.88	7.09	7.30	7.50
Fixed episode cost	Hyperlipidemia	2.80	2.85	2.91	2.96
Fixed occurrence rate	Hyperlipidemia	2.79	2.85	2.91	2.97
Fixed episode cost	Hypertension	6.46	6.56	6.66	6.75
Fixed occurrence rate	Hypertension	6.45	6.56	6.67	6.77
Fixed episode cost	Ischemic heart disease	7.20	7.54	7.87	8.19
Fixed occurrence rate	Ischemic heart disease	7.00	7.54	8.08	8.62
Fixed episode cost	Joint degeneration, localized—back	7.56	7.77	7.99	8.18
Fixed occurrence rate	Joint degeneration, localized—back	7.42	7.77	8.15	8.51
Fixed episode cost	Pregnancy with delivery	8.44	8.84	9.28	9.64
Fixed occurrence rate	Pregnancy with delivery	8.67	8.84	9.02	9.17

Concluding Comments

While risk adjustment models have been quite successfully used to estimate total annual member cost at the individual member level using age, gender, and medical comorbidity indicators, our research suggests that they do not work very well at the episode level or at the service type level within an episode. Factors that are not included in risk adjustment or observable to us in empirical data, such as practice patterns, patient preference, or patient socioeconomic status, may be key drivers of cost at the episode level. Some of these cost drivers might be managed in order to reduce cost without lowering care quality.

Although we presented some PMPM savings estimates in the paper, these estimates are for illustrative purposes only. It is more important for the reader to understand the methodology than the estimates themselves. We also point out in the paper that both the episode occurrence rates and episode costs affect the financial outcome of an ACO at the population level. Episode occurrence rates depend on the underlying prevalence of the diseases and medical conditions within the ACO's attributed population that trigger an episode during the designated time interval.

We believe there are several follow up studies that would expand upon the findings of this study and improve its meaningfulness to real populations. Among them are:

1. Evaluate risk-adjusted episode costs PMPM, by service category, for specific providers, delivery system types, geographic regions or government programs. This would help to establish benchmarks for episode costs and the range of current performance within different organization types and populations.

2. Evaluate the model's predictive ability if the basis for episode grouping is the "ETG" groups (instead of the higher level "Base ETGs" which include complications and comorbidities). Perhaps a risk adjustment algorithm that is confined to episodes without complications would produce more satisfactory results. While this would diminish the value of the methodology for use as a provider payment mechanism (since complications

can be influenced by provider claims coding or procedure selection), it may create better ACO retrospective methods for evaluation of cost efficiency.

3. Compare results for different episode risk grouper software programs. Since the start of this study, several groupers have either emerged or expanded the coverage of their programs. Among these are the PROMETHIUS grouper and Clinical Risk Groups (CRGs) by 3M. In addition, Optum and Thompson Reuters (MEGS) continue to update their software. Our study is heavily dependent on the grouper software to recognize and categorize clinically relevant patient characteristics into different episode types.

Many commercial payers and provider service organizations are attempting to address the needs of ACOs by delivering robust clinical analytics and actuarial services. These services often include strategic advice about risk assumption in general and a market-based approach to progressive episode quality and cost management. An ACO may be able to replicate the analyses presented in this paper using their historical data. This involves the following steps:

1. The ACO's actual claims experience needs to be grouped into episodes using the Symmetry ETG software. Given the wide use of ETGs in the industry, especially among commercial payers, we expect that this would be a rather easy step, if not already taken by the payer. Specifics for software configuration are included in Appendix F.
2. The claims would also need to be repriced. Appendix A includes the complete schedule of allowed costs by procedure code used in the analysis. An ACO can use these fees to reprice their claims experience so that the resulting costs per episode are comparable to the averages shown in the paper.
3. The ACO can probe further into their episode experience by comparing actual service utilization and average costs per procedure by service type to the values shown in Appendix C. This provides a more granular comparison of costs by service type, though the results have not been adjusted for the risk of the underlying populations.
4. It is important to note that this retrospective use of risk analysis after a set of interventions has been introduced within the population will lead to a bias to underestimate the underlying risk that would have been present had the interventions not been done. Nonetheless, we believe that the analysis will prove meaningful in understanding opportunities for further increases in cost efficiency, despite this limitation.

It is our hope and belief that our approach to evaluating and risk adjusting episodes of care will add to the collective understanding of population health management and contribute to more actuarially sound ACO evaluations and provider risk-sharing contracts.