

Research Article

Payment Accuracy in Value-Based Care Contracts

Mackenzie A¹, Wang J¹, Teppema S¹ and Duncan I^{2*}¹Santa Barbara Actuaries Inc, 3221 Calle Mariposa, Santa Barbara, CA, USA²Department of Statistics & Applied Probability, University of California, Santa Barbara, CA, USA***Corresponding author:** Ian Duncan, Department of Statistics & Applied Probability, University of California, Santa Barbara, CA 93106, USA**Received:** August 30, 2021; **Accepted:** October 30, 2021; **Published:** November 06, 2021**Abstract**

Reimbursement for health care services is transferring more risk away from payers and toward health care providers in the form of Alternative Payment Models (APMs), also known as Value-Based Care (VBC) models. VBC models cover a wide variety of forms but all include guarantees by providers of services to improve quality of care and/or reduce cost. Types of risk include performance risk, contract design risk or stochastic risk (because of the random variation in health care services and costs). A form of contract risk that can be a significant driver of cost is model risk, defined as the probability that the savings calculated at contract reconciliation will deviate from the actual savings generated. To estimate the degree of risk we quantify the potential variance in outcomes in a naïve population prior to intervention and the components that could affect outcomes, using examples of maternity and type 2 diabetes. This analysis has implications for both participants in, and designers of value-based contracts.

Keywords: Alternative payment models; Value-based care; Health care management organizations

Introduction

The health care industry is undergoing a transformation as it focuses not only on health care treatment but also on population health management. Health care payers, such as insurance companies, employers, and the government (*via* Medicare and Medicaid), have developed new models of reimbursement to health care providers and Health Care Management Organizations (HCMs) who work to improve population health, that involve shifting some or all of the risk of a population's outcomes away from the payer and to the provider or HCM. This shift is amplified with activity from the Centers for Medicare and Medicaid Services (CMS) as it has implemented many new Alternative Payment Models (APMs), also known as Value-Based Care (VBC) models.

Financial risk arrangements and value-based care

Traditionally, healthcare services have been reimbursed with some type of transactional or fee-for-service payment arrangement. As payers have faced escalating costs and stagnating quality, interest has grown in transferring financial responsibility to the providers of healthcare services as a means of incenting improved outcomes and reduced cost [1,2]. Value-Based Care (VBC) arrangements represent "a path to achieving the aspirational goals of the Institute for Healthcare Improvement's "triple aim": improving the patient experience of care, improving the health of populations, and reducing the per capita cost of health care, as well as improving clinician experience, a fourth aim that others have proposed [3]. These VBC arrangements have been increasing in size (number of participants), volume (number of arrangements), and scope (proportion of payments linked to financial and quality performance, as well as the range of covered medical conditions) in recent years. Medicare and Medicaid have been leading the way in developing VBC arrangements, but commercial plans have been adopting and scaling VBC and financial risk arrangements as well [4-6]. According to Change Healthcare, the number of US States and Territories with

value based reimbursement programs has increased from only 6 in 2013 to 48 in 2018 [7]. Recent legislation, such as the Affordable Care Act, the Protecting Access to Medicare Act, and the Medicare Improvements for Patients & Providers Act, have contributed to VBC momentum [8].

Value-based care models take many forms. Figure 1 shows the terminology and relationship of different models along two dimensions: Services at Risk and Degree of Risk transferred in the contract. Any contracting entity must determine where on the spectrum it is comfortable contracting. Numerous studies have examined the mechanics and financial implications of different established payment models, for example [9,10]. The degree of risk of is denoted on the horizontal axis, from the traditional fee-for-service model (the lowest risk form of contract for a provider, and most risky for a payer) through gain- and loss-sharing models, to fully-capitated models, in which healthcare risk is completely transferred to the provider/HCM. The vertical axis indicates the extent of services whose risk is transferred ranging from costs incurred in managing a single condition to "total cost of care" where a provider is responsible for all of a patient's cost.

The Challenge of measuring VBC performance

While quality improvement and cost-reduction in healthcare are worthwhile goals, quantifying improved performance can be difficult. Numerous factors such as emerging technologies, new pharmaceutical and medical treatments, changing population risk profiles, the skewed distribution of expenditures (a minority of members driving a majority of costs), the randomness and severity of acute events, an aging population, healthcare price fluctuations, and the impact of systemic effects such as Covid-19 introduce significant uncertainty to the evaluation of VBC program performance. For the organizations administering these programs and the providers participating in them, this means that valid determination of the success or failure of VBC programs requires recognition of these confounding factors.

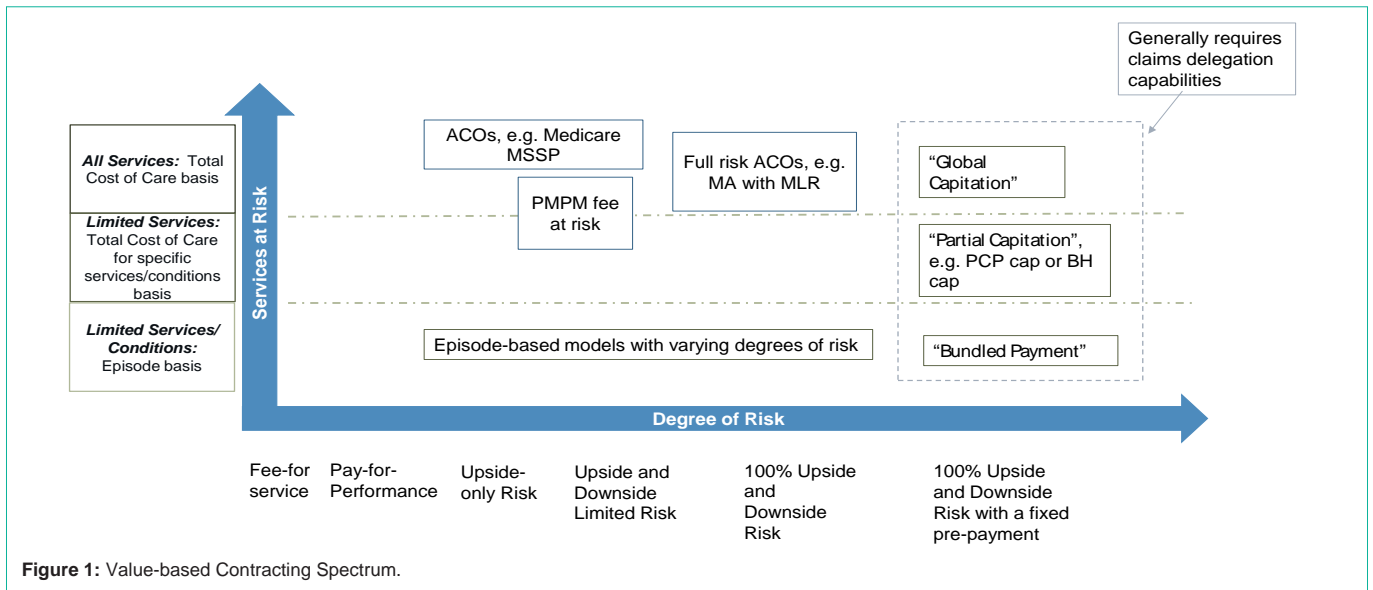


Figure 1: Value-based Contracting Spectrum.

Pricing a Value-based Contract

Any arrangement that seeks to reward a healthcare provider or HCM for the improved outcomes of its members by definition relies on a prediction of the counter-factual: what would have happened to those members without the program, intervention, or payment model in place? Otherwise, it is impossible to confirm or quantify actual improvement. In clinical trials, this type of outcomes-effectiveness study is performed through a randomized control study design where participants are separated into a control group that receives a placebo and a study group that receives the clinical intervention being tested. A sufficient sample size and an appropriate randomization process theoretically eliminate bias between the study group and the control group. In a VBC arrangement, the ideal environment of a pure randomized control study is rarely attainable. As such, we must pursue statistical approaches that allow for an accurate prediction of what would have happened to the study group without the VBC arrangement and healthcare intervention in place.

The optimal outcomes measurement approach will balance the need for measurement accuracy against simplicity of interpretation, ease of administration, and degree of predictability [11]. These elements are critical because both participants and payers need to be able to set budget expectations, have an understanding of what needs to be done to achieve desired outcomes, and have confidence in the accuracy and fairness of the outcomes reconciliation process. If participants feel the VBC arrangement lacks sufficient stability or fairness, many may choose not to opt into such an arrangement.

While value-based arrangements result in significant performance risk for participating providers, there is also a significant statistical risk. Many VBCs measure outcomes by comparing actual performance to a counter-factual, such as a predicted outcome. No outcomes measurement model will perfectly predict what would have happened without the intervention in place. We refer to the difference between predicted outcome and actual outcome, absent any intervention, as the pricing model error. Even without any type of intervention, there will be a difference between the predicted and actual outcomes

due to stochastic variability. To reduce the impact of pricing model error, risk mitigation approaches may be implemented such as risk corridors, contractual protections, or embedded reinsurance.

Testing the statistical accuracy of the pricing model

Many VBC measurement methodologies depend on comparing a predicted with an actual outcome. As noted, the results of these studies are subject to measurement error. Participants in VBC contracts should understand the degree of potential error and allow for it in their contracting. We test the extent of model error in VBC measurement in a specific population (people with Type 2 diabetes and maternity).

Data

We used a five million-life sample of the IBM MarketScan database spanning the years 2016-2019 to measure the extent of model error and the relationship between pricing model error and several pricing variables such as sample size, claims truncation thresholds, enrollment duration, cost inclusion (total cost of care vs. condition specific cost), as well as various trend assumptions. We modeled the effect of these variables on VBC contracts (defined as the set of services provided to treat a clinical condition or procedure) for maternity episodes of care and for members with type 2 diabetes.

Methodology

One measure of accuracy that is often applied in assessing model error is Mean Absolute Error (MAE), which is defined as the mean of the absolute differences between each predicted value and the actual value over all observations. The formulaic representation is as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| = \frac{1}{n} \sum_{i=1}^n |e_i|$$

Where $|e_i| = |f_i - y_i|$ is the absolute error for observation i and n is the total number of observations.

Confidence intervals are another way to quantify model accuracy. We built our confidence intervals using bootstrapping [14]. We selected samples of members of two example VBC populations (for a type 2 diabetes model and a maternity episode-based model)

Table 1: Confidence Intervals and Mean Absolute Errors for Diabetes and Maternity VBC Examples.

Population Example	Diabetes Example 1,000 Members		Maternity Example 200 Members	
Average Per-Patient-Per-Month (PPPM) Cost	\$1,703		\$2,351	
Confidence Interval	Lower Bound	Upper Bound	Lower Bound	Upper Bound
95%	-15.00%	14.60%	-14.50%	13.90%
90%	-12.40%	12.50%	-11.60%	11.70%
75%	-8.90%	8.50%	-8.10%	8.50%
Mean Absolute Error	6.10%		5.80%	

Table 2: Model Error Risk Mitigation Techniques.

Treatment Technique	Description	Example
Risk corridors	A risk corridor defines the minimum threshold that the savings or loss must exceed for a payment to be made. Since there is a degree of confidence around the actual outcomes versus the measured outcomes, risk corridors will limit the potential for financial reimbursement to only those outcomes that fall outside the corridor.	The Medicare Shared Savings (MSS) ACO program applies a risk corridor in their models [14]. A study by the authors demonstrates that while a corridor eliminates some risk there are still observations outside the corridor [15].
Retrospective trend	Retrospective adjustments can be made to account for unanticipated emerging expenses that would affect the population at risk without the intervention being in place. We have modeled the specific impact here and included that in our base case example for the maternity episode.	CMS's BPCIA program introduced a retrospective trend adjustment [16].
Stop-loss (outlier) Reinsurance	While reinsurance is primarily used to protect against adverse intervention experience, it can also be used to protect against adverse measurement/pricing experience	Medicare Part D includes embedded reinsurance for plans. Many risk bearing entities will also choose to purchase reinsurance
Risk pools	Like reinsurance, providers/HCMs can partner to share risk which provides protection against both model risk as well adverse intervention experience	Provider groups can reduce model error and actual performance risk by pooling their value-based gains and losses together with other provider groups in the model
Other contractual protections	In designing a risk arrangement numerous steps can be taken to reduce model error by reducing variance, outlier, and selection biases	It is common to remove outliers, high-cost conditions, or novel therapies from a risk arrangement.

at random from the MarketScan data set. Specific value based measurement design assumptions that applied to each population are provided in Appendix 1. For the purpose of this study, our normalization process in the diabetes example applied a diabetes-specific risk adjustment model [12] while the maternity example used the HHS HCC risk adjustment model [13]. After normalization, a perfect measurement model in every simulation should result in a \$0 MAE between the control group cost and the study group cost. Because none of the typical VBC risk-modification techniques (e.g. stop-loss insurance; risk-corridors) will completely eliminate model error, we observe the range of model error in our simulations resulting from the specific VBC design and inputs. To understand the magnitude of the pricing model error, we rank-ordered by size 1,000 simulations of the difference in the predicted and actual allowed amounts for randomly sampled study groups. The difference between the 87.5th percentile and the 12.5th percentile is defined as the 75% confidence interval with 75% of simulated outcomes falling within this range and 25% falling outside the range. Since we want to understand the relationship between pricing model error and various model elements for our two sample populations, we repeated this process for different population sizes, truncation thresholds, study durations, cost inclusion rules, and trend assumptions. Definitions of the Type 2 diabetes and maternity VBC models are provided in Table A1 and A2 in the Appendix.

Results

Table 1 shows the results of simulating the pricing model error on the two populations discussed above using 1,000 members for both the diabetes example and 200 maternity members T2D, a 12-month experience period for the diabetes example and an episode of care

for the maternity example. Confidence intervals and mean absolute errors are expressed as a percentage of the underlying population's total per-patient-per-month allowed cost (PPPM). Average PPPM costs are measured across the entire sample. In both examples, the pricing model error is not insignificant. The maternity episode is subject to less pricing error than the diabetes example, which is not surprising since maternity cost components are specific to the maternity condition than total cost of care for members with type 2 diabetes.

In Table 1, the total cost of care for 1,000 type 2 diabetes members under the assumptions in Table A1 has an average model error of 6.1% while the maternity example (defined under the assumptions in Table A2) with 200 members has an average model error of 5.8%. For the diabetes example, the 90th percentile confidence interval spans 12% in both directions. In other words, if the benchmark price were set at \$1,700 PPPM, 90% of the time the true baseline cost without intervention would be between \$1,489 and \$1,912 PPPM (based on a random sample of 1,000 members) and 10% of the time it would be either higher or lower than this range by at least \$212. This may not seem like a very large difference, but when performance is paid on the margin, even one or two percentage points can result in the success or failure of a VBC program.

For example, suppose that under a two-sided shared risk contract, 50% of the savings (the difference between the actual average cost and the \$1700 benchmark) is to be paid as a bonus to the provider (or 50% of the difference above \$1,700 is to be refunded to the payer for poor financial outcomes). Our results show that there is a 10% chance that without any intervention on the part of the provider, the VBC contract would recognize more than \$212 PPPM in total savings

(or \$212 in losses), split 50/50 between the plan and the provider. The implication of these savings (losses) generated without any intervention is that the provider (and payer) is at considerable risk of over- or under-payment under the contract design due to model risk, independent of actual performance.

Effect of different contract features

We tested the impact that various evaluation design parameters have on the resulting model error. Inquisitive readers can find detailed table results in the Appendix. We summarize the conclusions as follows:

- **Population size:** As the population grows, model error decreases due to the law of large numbers;
- **Duration of the study (for the diabetes example):** Longer evaluation time periods decrease model error as we are in effect increasing the sample size with more member months;
- **Application of a cost truncation threshold:** Reduction in the maximum claim size threshold per member per evaluation period decreases model error due to a reduction in outlier bias;
- **Redefining outcomes metrics as diabetes- or maternity-specific costs as opposed to the total cost of care:** A reduction in total expenditure liability results in reduced financial risk due to model error; and
- **Application of a retrospective trend for the maternity episodes versus a prospective trend:** Maternity results in Table 1 were based on a prospective trend; applying a retrospective trend reduces model error because it eliminates the model error associated with setting an inaccurate prospective trend rate.
- **Removal of certain high-cost conditions:** The base scenario exclusion criteria removed members with hemophilia, organ transplant or end-stage renal disease in order to limit the impact of outlier payments. We also tested the effect of removal of conditions that have no association with diabetes: multiple sclerosis, rheumatoid arthritis, Crohn's disease or ulcerative colitis, migraine and active cancer (defined as HCC 8 through HCC 12). Removal of these conditions does not have a significant effect on Table 1 results: for example, T2D MAE declines from 6.10% to 5.88%, with similar minor declines in confidence intervals.
- **Effect of larger sample sizes:** As the size of the patient sample increases the confidence interval decreases, as one would expect. Appendix 2 shows the effect of increased sample sizes on confidence intervals. Limited frequency of pregnancies in the dataset restricts our scenarios to a maximum of 6,000 maternities.

Discussion

Model risk in a value-based contract is common and can be significant, as the results show. The issue arises: how can a participating provider (or payer) mitigate this risk? Although model risk may not be eliminated completely, Table 2 provides several techniques that may be applied that partially address this risk. The first and most important requirement before introducing risk mitigation techniques, however, is for contracting parties to understand the extent of the model risk inherent in their proposed contract. We believe that modeling of the type illustrated above should be conducted before any contractual

terms are determined. Once the extent of the risk is understood, the effect of mitigating techniques can be estimated and incorporated into the contract.

Limitations

This study is based on a single national database comprising members of multiple different health plans in different regions. In practice, confidence intervals produced through similar bootstrapping techniques may be narrower if contract members are more homogeneous, for example being drawn from the same plan, provider group or geographic area. Because of data limitations, we were unable to test model error over a period longer than one year. It is possible that over a longer period of time variability is less. However, contracts that accumulate results over 2 or 3 years (which would be required in order to reduce variability) are unusual. More usually, contracts are assessed on an annual basis, as modeled in this study.

Conclusion

This paper evaluates the significance on outcomes of model variability in disease-specific value-based contracts. Contracts may demonstrate gains and losses that even without interventions will be magnified once interventions are applied to a population. Contracting parties should understand the effect of model risk and should apply features that mitigate this risk before entering into a value-based contract.

Declaration

IBM Disclaimer: Certain data used in this study were supplied by International Business Machines Corporation. Any analysis, interpretation, or conclusion based on these data is solely that of the authors and not International Business Machines Corporation.

Funding: This study was provided in part by Signify Health Inc. Opinions are those of the authors and do not necessarily represent those of Signify Health Inc.

References

1. Kaplan RS and Porter ME. The big idea: how to solve the cost crisis in healthcare. *Harvard Business Review*. 2011.
2. Porter ME. What Is Value in Health Care? *NEJM*. 2010; 363: 2477-2481.
3. Teisberg E, Scott Wallace ES and O'Hara S. Defining and Implementing Value-Based Health Care: A Strategic Framework. *Academic medicine: Journal of the Association of American Medical Colleges*. 2020; 95: 682-685.
4. Muhlestein D and McLennan M. Accountable Care Organizations in 2016: Private and Public sector growth and dispersion. *Health Affairs*. 2016.
5. David Muhlestein, Robert Saunders, Mark McClellan. Growth of ACOs and Alternative Payment Models in 2017. *Health Affairs*. 2017.
6. David Muhlestein, Paul Gardner, William Caughey, Katelyn de Lisle. Projected Growth of Accountable Care Organizations. *Leavitt Partners*. 2015: 1-9.
7. Change Healthcare. Value-Based Reimbursement State-by-State: A 50-State Review of Value-Based Payment Innovation. *Change Healthcare: Nashville TN*. 2017.
8. Centers for Medicare and Medicaid Services, Medicare Program; Merit-Based Incentive Payment System (MIPS) and Alternative Payment Model (APM) Incentive under the Physician Fee Schedule, and Criteria for Physician Focused Payment Models, Health and Human Services, Editor. *Federal Register: Washington D.C*. 2016.
9. Laschober M, Rich E, Lake T, Merrill A and Natoli C. Developing Alternative

- Payment Models: Key Considerations and Lessons learned from Years of Collaboration with CMS, in Issue Brief, Mathematica Policy Research, Editor. 2015.
10. Spector JM, Stuebaker B, Menges EJ. Provider Payment Arrangements, Provider Risk, and Their Relationship with the Cost of Health Care. 2015. Society of Actuaries: Schaumburg IL.
 11. Williams DV, Liner DM and Norris C. Building a successful value-based payer contracting strategy. 2017. Milliman Inc.: Denver CO.
 12. Mackenzie AJ, Zhao E, Duncan I and Wang J. Comparing Model Error Between a Standard Risk Adjustment Model and a Disease-Specific Risk Adjustment Model. Health Watch. 2020.
 13. John Kautter, Gregory C. Pope, Melvin Ingber, Sara Freeman, Lindsey Patterson, Michael Cohen, et al. The HHS-HCC Risk Adjustment Model for Individual and Small Group Markets under the Affordable Care Act. Medicare and Medicaid Research Review. 2014; 4.
 14. Centers for Medicare and Medicaid Services, Shared Savings Program Participation Options for Performance Year 2022, CMS, Editor. 2020. Health and Human Services: Washington DC.
 15. Duncan I, Liao X, Bonfiglio E, Mackenzie A and Wrigley T. Shared Savings Model Risk in the Medicare Shared Savings Program. Forthcoming in the North American Actuarial Journal.
 16. Centers for Medicare and Medicaid Services, BPCI Advanced Model Year 4 Fact Sheet, Health and Human Services, Editor. 2020: Washington DC.