Spending Carveouts Substantially Improve the Accuracy of Performance Measurement in Shared Savings Arrangements: Findings From Simulation Analysis of Medicaid ACOs

Derek DeLia, PhD[1]

Abstract

Accuracy of spending-based provider performance metrics is limited by random variation and components of spending that are uncontrollable by providers. Such components vary according to the care management focus and operational maturity of each provider group. This study uses data from New Jersey Medicaid accountable care organizations (ACOs) to examine how carving out uncontrollable components of spending affects the accuracy of performance measures in shared savings arrangements. Spending on injury care, custodial care in facilities (CCF), and amounts above $100 000 per patient are used as examples of potentially uncontrollable spending. Data from 7 applicant Medicaid ACOs are used to conduct Monte Carlo simulations examining the effects of carving out each type of uncontrollable spending under the assumption that controllable spending is reduced by 5%. The simulations show that failure to carve out uncontrollable injury care spending adds −3 to +1 percentage points of bias to the measurement of the true average savings rate (ASR) of 5% and can increase mean squared error (MSE) by a factor of up to 3. Failure to carve out uncontrollable CCF spending generates bias ranging from −4 to +9 percentage points and increases MSE by factors of 8 or more. Failure to carve out uncontrollable spending above $100 000 per person generates bias ranging from −5 to +5 percentage points and increases MSE by factors of 13 or more. Compared with the main modeling reported above, sensitivity analyses find even greater distortions in measured performance when uncontrollable spending is not carved out of the ASR calculation.

Keywords

risk sharing, financial, reimbursement, incentive, accountable care organization, Medicaid, Monte Carlo method, shared savings

Introduction

There is growing interest in the use of population-based payment methods to improve health outcomes and contain costs. These methods vary by the type and amount of financial risk assumed by provider organizations. The most comprehensive of these is full capitation where providers receive a fixed payment per patient per month to cover the total costs of care (TCOC). Although payments are typically risk adjusted, providers assume full financial risk for any spending that is not predicted by patient risk scores.

From the provider perspective, a somewhat less risky alternative to full capitation is partial capitation where certain types of spending, considered outside of provider control, are carved out of the capitation rate and paid separately. Such carveouts may include mental health services, out-of-area emergency care, specific high-cost illnesses (eg, end-stage renal disease), or high-cost outliers (eg, above the 99th percentile for a specific service or patient population). Still, many providers remain wary about entering into contracts

with any significant financial risk. This wariness stems from providers' own limited experience with managing risk combined with the financial failures under risk-based contracting in the 1990s.[1,2] Providers may also avoid risk-based payment because they lack the ability to measure and manage care provided outside of their own provider network.

Accountable care organizations (ACOs) paired with shared savings arrangements have been developed partly to give providers a more gradual way of taking on more financial risk. Under shared savings, ACOs are paid on

[1]Rutgers Center for State Health Policy, New Brunswick, NJ, USA

Corresponding Author:
Derek DeLia, Research Professor & Senior Health Economist, Rutgers Center for State Health Policy, 112 Paterson St., Room 540, New Brunswick, NJ 08901, USA.
Email: ddelia@ifh.rutgers.edu

a fee-for-service basis but are given incentives to improve quality and contain spending for an assigned patient population. If the ACO generates a positive average savings rate (ASR) by keeping spending below a projected benchmark level, it is rewarded with a share of the savings. The share given to the ACO is based on how well it meets predetermined quality standards. In a 1-sided shared savings model, the ACO is rewarded for savings but is not liable for any losses (ie, spending above the benchmark or a negative ASR). In a 2-sided model, the ACO would pay a penalty for generating losses but would be eligible for a greater share of any savings generated.

Alongside its role as a key measure of ACO performance, the calculation of the ASR is also a source of risk. Specifically, the observed ASR is subject to error due to random variation in annual per capita spending, which obscures true ACO savings performance.[3-6] For this reason, shared savings arrangements often require ACOs to meet a minimum savings rate (MSR) threshold to ensure that the observed ASR is a reflection of real performance and not random variation. Statistically, MSR thresholds protect payers from type I error (ie, rewarding false savings generated by random noise), while placing providers at greater risk for type II error (ie, inappropriately withholding rewards for true savings that are diminished by random noise). For a given ACO size (ie, number of assigned patients), the setting of an MSR threshold involves a clear trade-off between type I and II errors.

Early ACO experience reveals a great deal of provider risk aversion even in the relatively lower risk environment of shared savings. For example, when ACOs in the Medicare Shared Savings Program (MSSP) were given a choice between 1-sided and 2-sided shared savings, only 1% chose the latter.[7] Under the original MSSP, ACOs choosing the 1-sided model had to move to the 2-sided model within 3 years.[8] But in response to providers' ongoing concerns about taking on too much risk too quickly, the Centers for Medicare and Medicaid Services (CMS) created provisions for ACOs to extend their time in the 1-sided model, when it finalized rules for the next 3 years of the MSSP.[9]

Similar issues have arisen in the Medicare Pioneer ACO Program, which uses a combination of 2-sided shared savings and capitation payment. Although Pioneer ACOs are considered more advanced in their ability to manage patient risk, 16 of the original 32 Pioneer ACOs have left the program.[10] Typically, when ACOs dropped out of the Pioneer Program, they moved to the lower risk 1-sided model in the MSSP.[9] In recognition of providers' ongoing need to gain experience with managing patient risk, CMS redesigned the MSSP to provide a more shallow and flexible on-ramp to risk-based provider payment.[9]

This ongoing provider risk aversion, combined with the problem of random variation in spending, underscores the need to reduce statistical noise in ASR measurement. A recent study shows that this noise can be reduced somewhat through careful attention to the data collection strategy and methodology used to calculate the observed ASR.[3] An area that remains unexplored, however, is the use of carveouts in the calculation of ASRs, and in population-based payment more broadly.

Like partial capitation, shared savings arrangements usually include implicit or explicit carveouts in the determination of spending for which providers are held accountable. In the MSSP, savings calculations exclude prescription drug spending and spending for each patient is truncated at the 99th percentile, which is roughly $100 000, when calculating ASRs. This truncation strategy dates back to the Physician Group Practice (PGP) Demonstration, which began in 2005. The strategy of truncating at $100 000 was built into the Demonstration from the beginning to reduce random variation driven by small numbers of high-spending individuals, which it was feared could distort true savings performance.[11] Although this strategy is now common in the MSSP and other programs, its effects on savings performance measurement have not been evaluated empirically in the health services and policy literature. (The topic is treated only very briefly in the PGP Demonstration Design Report, which examines simple trends in Medicare A & B spending for 7 physician practices and their "state market areas" in 1993-1994.) We address this gap in the literature by conducting a more focused and detailed shared savings analysis, as described below.

A broader variety of shared savings arrangements have emerged in states with Medicaid accountable care initiatives. For example, Vermont carves out long-term services and supports (LTSS) and some behavioral health services from their shared savings calculations but plans to revisit this issue in a future phase of implementation.[12] In Massachusetts, accountability for LTSS spending will be gradually carved into spending accountability measures but spending on home- and community-based services (coordinated by other state agencies) would remain carved out.[13] In contrast, Medicaid ACOs in New Jersey are responsible for TCOC for fee-for-service and managed care spending but with the opportunity to negotiate specific carveouts with managed care plans (N.J. P.L. 2011, ch.114). In Rhode Island, Medicaid Accountable Care Entities negotiate contracts directly with managed care plans and are not responsible for fee-for-service spending.[14]

The analysis below examines how carveout strategies affect the calculation of ASRs and derives implications for the financial risks faced by provider groups in shared savings arrangements. The analysis is built on 2 foundational assumptions developed more systematically below: (1) Specific services to be carved out are beyond the control of the contracting provider group and (2) the set of "uncontrollable" carveout services varies across provider groups with different levels of organizational sophistication and within provider groups as their sophistication evolves. The analysis uses health care spending data for patients in New Jersey Medicaid ACO regions to conduct simulations of ASR

calculations with and without the use of specific carveouts that are currently under consideration by ACOs and MCOs in New Jersey—specifically, spending on injuries, amounts above $100 000 for any individual, and custodial care in facilities (CCF), which includes nursing home care as well as intermediate care facilities for the intellectually and developmentally disabled. These carveouts, which are straightforward to calculate with administrative data, serve as exemplars for a broader conceptual framework outlined below. The simulations, which account for random variation in the observed ASR, show the effects of carveouts on various measures of statistical accuracy when using the observed ASR to measure ACO spending performance.

Methods

Conceptual Framework

In theory, health care providers are better equipped to manage some types of risks than others. Specifically, providers are well positioned to manage clinical risks, which are affected by the way in which services are provided and organized. For example, spending associated with hospital readmissions might be controlled through better coordination of transitions from inpatient care to community-based care. In contrast, providers are not well positioned to manage actuarial risks, which are generated by random events such as major trauma and injury episodes. Insurers rather than health care providers are generally better positioned to manage actuarial risk. Providers accepting full capitation for TCOC are liable for both clinical and actuarial risks, and in effect function as insurers. A nascent group of provider-sponsored health plans have formalized this process by vertically integrating into the insurance business.[15]

Most providers, however, are not likely to run their own health plans. As a result, effective use of shared savings and population-based payment will need clearer distinctions between the kinds of risks for which providers and insurers will be responsible. Such distinctions may vary by provider sophistication and the nature of the payment contract. A provider organization with little or no experience in population health management likely will require more spending carveouts than a more sophisticated provider group. In such cases, the shared savings or capitation contract may focus on carving in specific kinds of spending targeted by providers for focused care management (eg, avoidable inpatient or emergency department spending). All other spending would be covered under fee-for-service until the providers are better able to include more services in the population-based payment contract. Alternatively, providers may accept some portion of risk for TCOC after carving out specific service lines that they view as beyond the scope of their current ability or care management focus.

Although many types of spending can be beyond providers' control, those types that lead to very large and unpredictable swings in the costs of caring for assigned populations create the greatest financial risks for providers. As shown rigorously below, rapidly growing uncontrollable spending can offset savings achieved in controllable spending when performance is measured on the basis of TCOC, which lumps uncontrollable and controllable components together, creating a downward bias in the measure of true savings performance. Conversely, a large reduction in uncontrollable spending would give an ACO an unfair credit for savings in a TCOC measure even if controllable spending remains unchanged (or increases slightly), creating an upward bias in savings performance. Moreover, if uncontrollable spending is very volatile from year to year, then it would add substantial random noise, which diminishes the precision of the savings calculation. Thus, there is a need to identify empirically the types of spending that would cause the most distortions from true provider savings performance in the areas where providers have agreed to be held accountable.

Analytic Approach

The analytic approach focuses on an ACO that divides total health care spending into controllable and noncontrollable components based on what the ACO believes it can influence and its care management strategies. Building from prior work,[3-5] we model the ACO's reduction in controllable spending using the ASR defined below.

$$ASR = \frac{\left[\left(\mu_B + \mu_A\right) - \mu_P\right]}{\left(\mu_B + \mu_A\right)} \qquad (1)$$

The terms μ_B and μ_P represent risk-adjusted per capita amounts of controllable spending during a baseline and performance period, respectively. The term μ_A represents an adjustment to the baseline value to reflect expected or targeted growth in spending. If the ACO is successful at reducing controllable risk-adjusted spending by proportion s, then $\mu_P = (1-s)(\mu_B + \mu_A)$ or equivalently $ASR = s$. Importantly, Equation 1 assumes that the uncontrollable component of spending is carved out of the calculation. If uncontrollable spending is ignored (ie, not carved out), then ACO performance would be measured using a contaminated version of the ASR, which can be written as

$$ASR^C = \frac{\left[\left(\mu_B + \mu_A + \tau_B\right) - \left(\mu_P + \tau_P\right)\right]}{\left(\mu_B + \mu_A + \tau_B\right)} \qquad (2)$$

where τ_B and τ_P represent the uncontrollable component of risk-adjusted per capita spending in the baseline and performance periods, respectively. Defining the "uncontrollable savings rate" (USR) as $USR = (\tau_B - \tau_P)/\tau_B$, ASR^C can be written as a weighted average of the ASR and USR as follows:

$$ASR^C = w \cdot ASR + (1-w) \cdot USR \qquad (3)$$

where $w = \dfrac{(\mu_B + \mu_A)}{(\mu_B + \mu_A + \tau_B)}$.

It is important to emphasize that the definition of "uncontrollable spending" can vary across ACOs and within ACOs over time. For example, a very sophisticated ACO may include little or no spending in the uncontrollable category. In contrast, a newly formed ACO may include many categories of spending in the uncontrollable category at the beginning of its operations. As it gains more experience with care coordination and risk bearing, it can shift different types of spending from the uncontrollable to controllable category.

It can be shown algebraically that ASR^C will understate the true value of the controllable ASR if and only if $USR < ASR$. In particular, if uncontrollable spending is approximately constant over time (ie, $USR \approx 0$), then an ACO that produced positive savings for the controllable portion of spending would be clearly disadvantaged by not carving out the uncontrollable portion of spending. Conversely, if uncontrollable spending declines significantly (making USR very large), then the ACO would benefit from an inflated value of ASR^C.

In practice, the true values of ASR and ASR^C cannot be calculated, because the true underlying performance components μ_B, μ_P, τ_B, and τ_P are unobservable due to the problem of random variation described above. Thus, estimated quantities $\widehat{ASR}$ and $\widehat{ASR}^C$ must be calculated based on the observed (ie, estimated) values $\hat{\mu}_B, \hat{\mu}_P$, $\hat{\tau}_B$, and $\hat{\tau}_P$. The term μ_A might also require estimation depending on the design of the shared savings arrangement. To clearly isolate the effects of spending carveouts, we initially assume in the simulations described below that μ_A is a deterministic quantity set at the beginning of the arrangement. In sensitivity analysis (described below), we consider μ_A to also be a random variable with a corresponding observable value $\hat{\mu}_A$. This formulation would represent cases where the adjustment factor is based on a comparison group or secular trend in spending.

The formulas for $\widehat{ASR}$ and $\widehat{ASR}^C$ are very complex making it impossible to analyze their statistical properties with closed-form equations for their means, variances, or probability distributions. Therefore, we follow the methods used by DeLia[3] to conduct a series of Monte Carlo simulations to compare the statistical properties of $\widehat{ASR}$ and $\widehat{ASR}^C$ under alternative definitions of uncontrollable spending (ie, $\hat{\tau}_B$ and $\hat{\tau}_P$).

Data and Policy Context

Simulation input data come from the NJ Medicaid Management Information System (NJMMIS), which include all adjudicated Medicaid fee-for-service claims and managed care encounter records, for 2011-2014. These data are currently used for development and evaluation of the NJ Medicaid ACO Demonstration Program. The Demonstration began officially in July 2015 and is designed to last 3 years with the goal of providing an evidence base for subsequent Medicaid reform legislation. ACOs must meet a variety of conditions to be certified for participation in the Demonstration. Specifically, the ACO must include all area hospitals, 75% of Medicaid-participating primary care providers, and at least 4 behavioral health providers. It must be incorporated as a not-for-profit entity with a multistakeholder board, which includes health care, social service, and local community representation. Certified ACOs receive startup funding from the state (in addition to philanthropic funding obtained independently), anti-trust immunity, customized data feeds, and assistance with data analytics. Certified ACOs are held accountable for all Medicaid spending for all Medicaid recipients who live within a set of zip codes specified by the applicant and including at least 5000 recipients. Within this framework, however, the Demonstration includes substantial flexibility in how ACOs are held accountable and how shared savings arrangements are specified. ACOs must report on, and later improve upon, a combination of mandatory quality performance metrics as well as voluntary ones, which are chosen from a preset menu. ACOs can also negotiate shared savings arrangements with Medicaid managed care plans, which may include provisions for carving out particular types of spending and using different types of benchmarking strategies in the ASR calculation. (At the time of this writing, some arrangements are in place and others are still under negotiation. ACOs can also propose their own benchmarking strategies for fee-for-service spending.) This provision makes the NJ Demonstration an ideal environment for investigating the effects of carveout strategies. Additional details about the NJ Demonstration and its early development are found in previously published reports.[16,17]

In the simulations, we use 2013 data for the baseline period and 2014 data for the performance period for each ACO. Spending per person is prospectively risk adjusted (eg, data from 2013 are used to calculate risk scores for 2014) using the Chronic Illness and Disability Payment System (CDPS).[18] To test the robustness of the results, we run additional simulations using 2012 as the baseline and 2013 as the performance period. Since some beneficiaries are enrolled in Medicaid for only part of the year, all spending amounts are annualized by calculating spending per day of enrollment and multiplying this amount by number of days in the year (with each observation having equal weight in the analysis).

Only 3 out of 7 applicant communities met the state's criteria to participate in the Demonstration as certified Medicaid ACOs. But since all applicants represent a naturally occurring group of patients for understanding population-level spending variation, all 7 applicant organizations are included in this study to provide a wider variety of patient groupings

in terms of geography, ACO size, and within-ACO statistical parameters, which will enhance the robustness of the simulation analysis. As shown below, the simulation results are similar for all applicant communities, regardless of certification status.

Simulation Details

We conduct simulations where uncontrollable spending is defined alternatively as spending associated with injury care (including trauma but excluding injury due to medical care), CCF, and amounts exceeding \$100 000 per person. Injury care is identified using the International Classification of Disease, Ninth Revision, Clinical Modification (ICD9-CM) codes 800 to 959. CCF is identified using facility claims with a service category indicating services were delivered in a nursing facility or intermediate care facility for the intellectually disabled. (Within Medicaid claims data, rehabilitation cannot be distinguished from custodial nursing facilities stays. For Dual Eligibles, however, Medicare would be the primary payer for rehabilitation services.)

For each simulation scenario, we calculate 1000 Monte Carlo iterations by taking bootstrap samples from observed ACO spending distributions described above. The distribution of bootstrap samples from the observed data provides a close approximation to the probabilistic behavior of the underlying random process that generated the empirical distribution observed in the original data.[19] Bootstrap theory allows us to consider empirically observed parameters for each ACO as "true values" and parameters calculated in each bootstrap iteration as estimated values.

All simulations examine cases where the true ASR for controllable spending is 5% for each ACO and each ACO has no influence over uncontrollable spending. Using Equation 1, this level of true savings performance can be modeled by setting the true risk-adjusted per capita spending amount in the performance period μ_P. In sensitivity analysis, μ_A is assumed to be normally distributed with a mean of 500 and standard deviation of 50. Under this assumption, a random draw from this distribution is used at each iteration to generate a value for $\hat{\mu}_A$, which is used in place of μ_A.

An additional step is taken to ensure that the bootstrap samples are drawn from a probability distribution where the mean risk-adjusted per capita spending amount in the performance period is μ_P. Specifically, each individual's spending amount in the performance year is rescaled by multiplying this amount by the ratio $\mu_P/\hat{\mu}_B$. After this rescaling, bootstrap samples are generated by sampling individuals with replacement from within each ACO community using the "bsample" command in STATA 14.0.

We compare the accuracy of $\widehat{ASR}$ and $\widehat{ASR}^C$ in 2 ways. First, we display box-whisker plots showing the full distribution of each random variable. Then, we calculate and compare the mean squared error (MSE) for each random variable

viewed as an estimator of the true ASR of 0.05. Prior research has shown that random variation in the ASR increases with the coefficient of variation (CV) in per capita health care spending and decreases with the correlation of health care spending within patients over time.[3,4] To help frame the simulation results, we also show how the use of specific carveouts affects these key statistics.

Findings

For each ACO, the injury carveout generates small reductions in mean spending and has little effect on the spending CV and spending correlation within patients over time (Table 1). In contrast, the CCF spending and \$100 000 truncation carveouts generate large reductions in mean spending. The CCF spending carveout has little effect, while truncation has a large downward effect, on the spending CV. Also, the CCF spending carveout tends to decrease or maintain within-patient spending correlation, while the truncation carveout is more likely to increase it.

Figure 1 shows box-whisker plots for the ASR with and without the injury spending carveout (with ACOs 1-7 running downward). In the left panel (showing $\widehat{ASR}$), the distributions are centered (mean and median) on the true ASR of 0.05 with smaller variation exhibited for larger ACOs. In the right panel (showing $\widehat{ASR}^C$), none of the distributions have a median equal to 0.05 and most have a mean that is different from 0.05, indicating that $\widehat{ASR}^C$ is a biased estimator of the true ASR. This bias ranges from −3 to +1 percentage points (see Online Appendix).

Differences between $\widehat{ASR}$ and $\widehat{ASR}^C$ are much greater when they apply to the CCF spending carveout (Figure 2). $\widehat{ASR}^C$ is greatly biased with greater variation around the median. For some ACOs, all or nearly all of the distribution of $\widehat{ASR}^C$ lies completely to one side of the true ASR of 0.05. Here, the bias in $\widehat{ASR}^C$ ranges from −4 to +9 percentage points (see Online Appendix).

Contrasts are even greater for distributions involving the truncation carveout (Figure 3). When spending above \$100 000 per person is carved out, the ASR distribution is centered fairly tightly around 0.05, especially for larger ACOs. Without the carveout, ASR distributions exhibit much more variation and bias. For some ACOs, a large portion of probability mass lies completely to one side of 0.05. Here, the bias in $\widehat{ASR}^C$ ranges from −5 to +5 percentage points (see Online Appendix).

Table 2 shows the effects of the 3 carveout approaches on the MSE. Except for ACO-1 and the injury carveout, failure to carve out (assumed) uncontrollable spending increases the MSE by up to a factor of 3. For the CCF spending and truncation carveouts, the MSE can increase by extremely large multiples, reaching 20 or more.

**Table 1.** Effects of Carveouts on Per Capita Health Care Spending.

ACO	N	Mean spending with indicated carveout, 2014				CV in spending with indicated carveout, 2014				Patient-level correlation in spending from 2013 to 2014 with indicated carveout[a]			
		None	Injury	CCF	Truncation[b]	None	Injury	CCF	Truncation[b]	None	Injury	CCF	Truncation[b]
ACO-1	12009	$7333	$7034	$5977	$6592	3.36	3.43	3.81	2.51	0.71	0.71	0.67	0.82
ACO-2	29050	$7029	$6713	$6051	$6004	4.07	4.07	4.54	2.58	0.65	0.66	0.59	0.78
ACO-3	41110	$4923	$4766	$4785	$4492	3.63	3.69	3.61	2.46	0.58	0.60	0.55	0.63
ACO-4	48279	$6520	$6186	$5829	$5572	4.51	4.45	4.88	2.51	0.62	0.63	0.59	0.71
ACO-5	47987	$6325	$6014	$5671	$5254	9.98	10.15	11.04	2.60	0.93	0.96	0.93	0.71
ACO-6	53789	$5884	$5568	$5575	$5260	3.88	3.89	3.96	2.33	0.46	0.46	0.42	0.65
ACO-7	55791	$7495	$7177	$5810	$6077	4.12	4.21	4.29	2.57	0.81	0.82	0.63	0.80

Source. NJ Medicaid Management Information System.
Note. ACO = accountable care organization; CV = coefficient of variation; CCF = custodial care in facilities.
[a]Across ACOs, 75% to 78% of the individuals in 2014 had Medicaid enrollment in 2013 to enable calculation of the correlation coefficient.
[b]Total health care spending above $100000 for any individual is truncated.
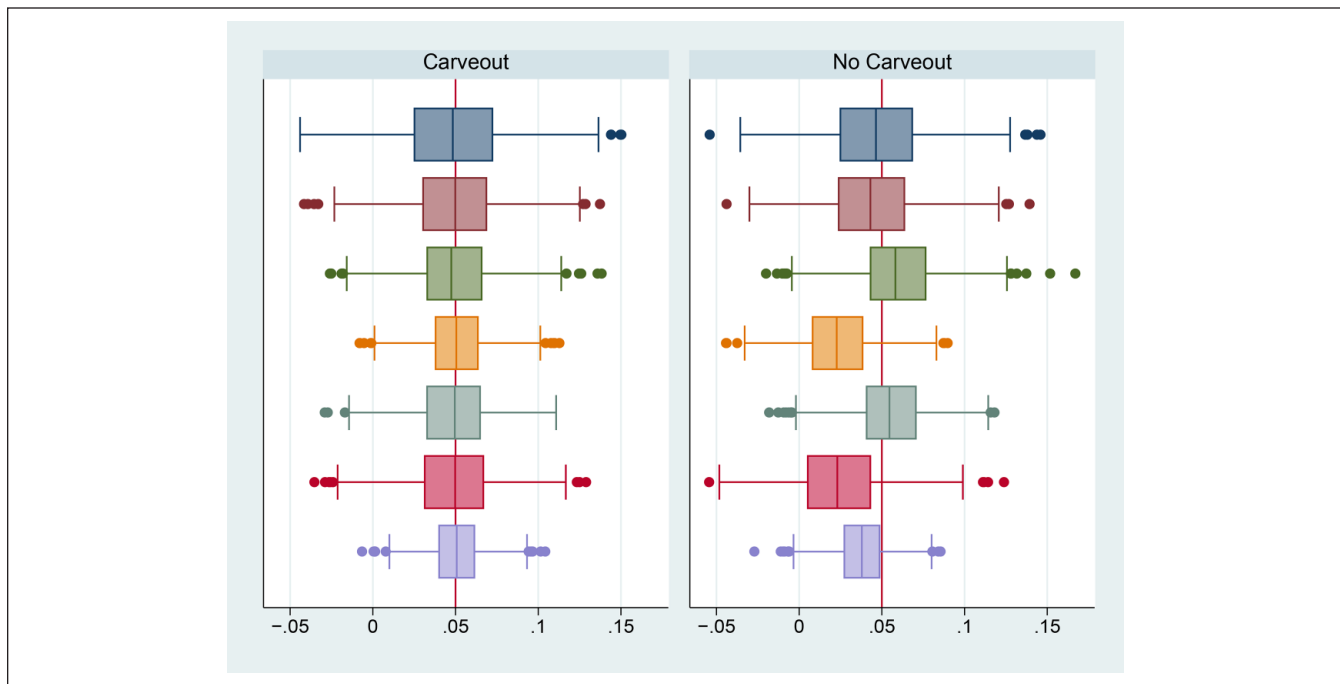
Figure 1. Distribution of the average savings rate with and without a carveout of injury spending.
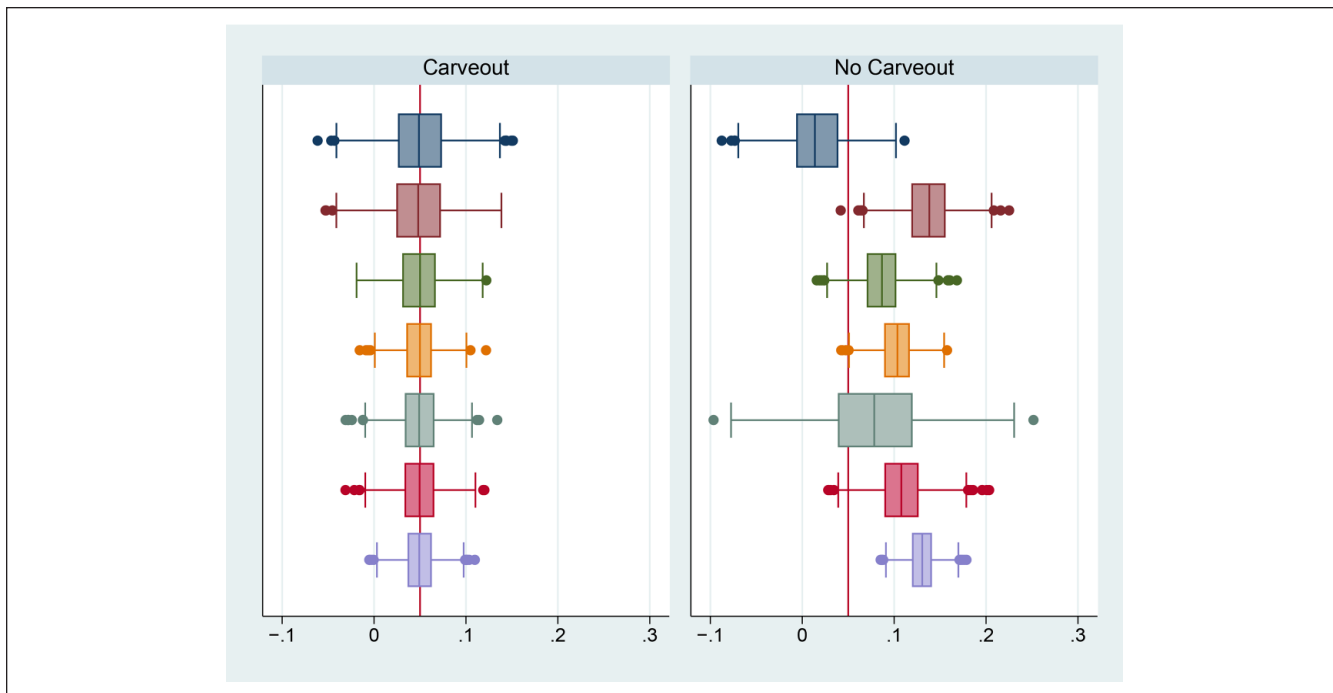*Source.* NJ Medicaid Management Information System.
Note. Box-whisker plots for ACOs 1 through 7 using data derived from 1000 iterations of the Monte Carlo simulation model described in the text. The model assumes that the average savings rate (ASR) for all spending excluding injury care is 0.05. ACO = accountable care organization.

The sensitivity analyses (found in Online Appendix) show that the findings are not sensitive to which years are used as the simulated baseline or performance periods. When the adjustment factor (μ_A) is modeled as a random variable, the simulations generate similar but much more exaggerated differences in the ASR with and without carveouts for uncontrollable spending. In these scenarios, all ASR estimates become much more variable and the ASR is systematically underestimated when uncontrollable spending is not carved out. This underestimation is especially pronounced in the case of spending on CCF and spending above $100000 where the mean and median are always negative without the carveout.

Discussion

The analysis above is based on the premise that certain components of health care spending are uncontrollable by ACOs

Figure 2. Distribution of the average savings rate with and without a carveout of spending for CCF.
Source. NJ Medicaid Management Information System.
Note. Box-whisker plots for ACOs 1 through 7 using data derived from 1000 iterations of the Monte Carlo simulation model described in the text. The model assumes that the average savings rate (ASR) for all spending excluding CCF is 0.05. ACO = accountable care organization; CCF = custodial care in facilities.
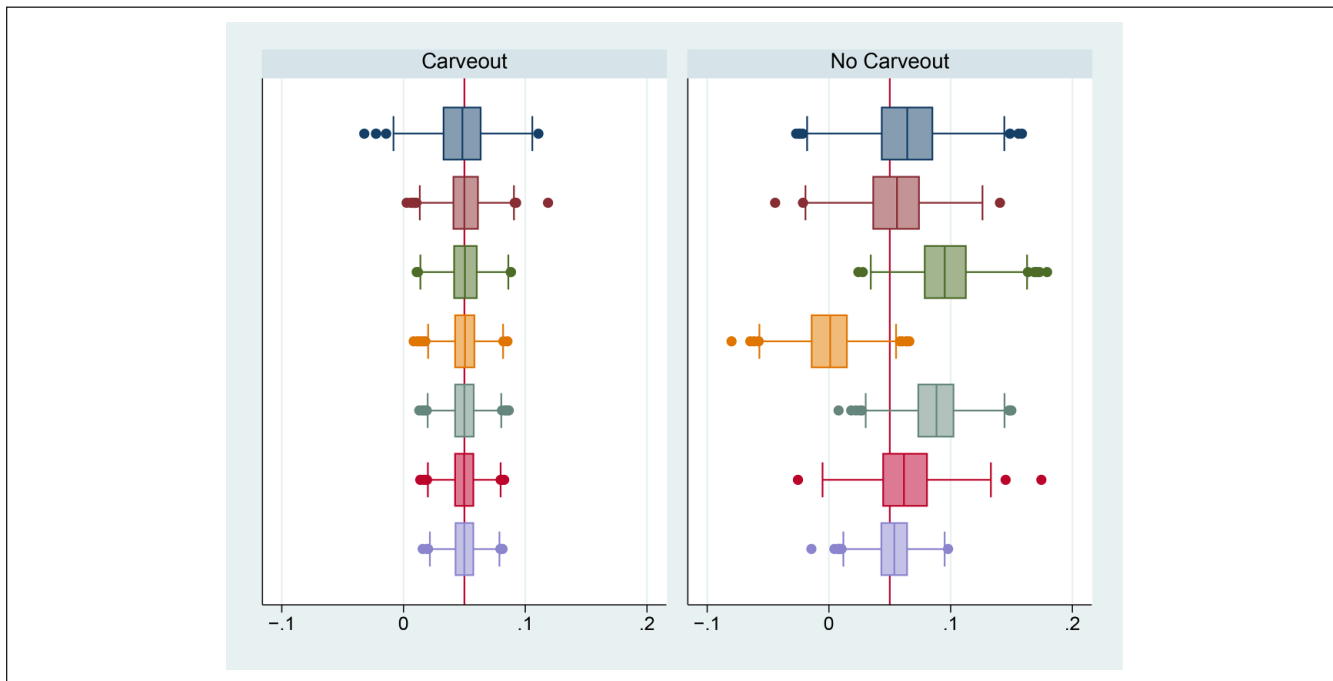
Figure 3. Distribution of the average savings rate with and without a carveout of individual spending amounts above $100 000.
Source. NJ Medicaid Management Information System.
Note. Box-whisker plots for ACOs 1 through 7 using data derived from 1000 iterations of the Monte Carlo simulation model described in the text. The model assumes that the average savings rate (ASR) for all spending excluding individual spending amounts above $100 000 is 0.05. ACO = accountable care organization.

Table 2. Mean Squared Error in Measured Savings With and Without Specified Carveout Approaches.[a]

	Carveout	No carveout	Ratio
Injury spending			
ACO-1	0.0011	0.0010	0.91
ACO-2	0.0008	0.0008	1.00
ACO-3	0.0006	0.0008	1.33
ACO-4	0.0004	0.0012	3.00
ACO-5	0.0005	0.0005	1.00
ACO-6	0.0007	0.0014	2.00
ACO-7	0.0003	0.0004	1.33
CCF spending			
ACO-1	0.0012	0.0023	1.92
ACO-2	0.0011	0.0083	7.55
ACO-3	0.0006	0.0019	3.17
ACO-4	0.0004	0.0032	8.00
ACO-5	0.0005	0.0004	8.00
ACO-6	0.0005	0.0041	8.20
ACO-7	0.0003	0.0067	22.33
Spending above $100 000			
ACO-1	0.0005	0.0012	2.40
ACO-2	0.0002	0.0008	4.00
ACO-3	0.0002	0.0027	13.50
ACO-4	0.0001	0.0029	29.00
ACO-5	0.0001	0.0019	19.00
ACO-6	0.0001	0.0008	8.00
ACO-7	0.0001	0.0003	3.00

Source. NJ Medicaid Management Information System.
Note. ACO = accountable care organization; CCF = custodial care in facilities.
[a]Derived from Monte Carlo simulations under the assumption that the average savings rate (ASR) for controllable (ie, carved out) spending equals 0.05 for each ACO. Simulation statistics are based on 1000 iterations for each ACO.

or other provider groups held accountable for spending metrics. The specific components will vary depending on the particular care management focus and operational maturity of each provider group. In this article, we examined spending on injury care, CCF, and amounts above $100 000 for any individual as exemplars of potentially uncontrollable spending, which are readily identifiable in claims data. The analysis shows that failure to carveout the uncontrollable portion adds substantial bias and significantly reduces precision in the measurement of the ASR, which we have assumed for illustrative purposes is a 5% reduction in controllable spending.

The spending carveout related to truncation at $100 000 leads to the greatest gains in accuracy of ASR measures. These gains reflect 3 features of the data when this truncation is implemented: (1) Mean spending is reduced, (2) the CV in spending is reduced, and (3) the within-patient correlation of spending over time is increased. The first of these indicates the overall potential for the carveout to affect savings measurement while the 2 remaining are associated with reduced

statistical variation in savings measures.[3,4] Although the CCF spending carveout has a large effect on mean spending, it has smaller effects on the CV and correlation statistics, making it somewhat less effective in improving the accuracy of the estimated ASR. The injury spending carveout has the smallest effect on the accuracy of the estimated ASR due to its comparatively limited impact on each of the 3 key spending statistics.

Sensitivity analyses (shown in Online Appendix) came to broadly similar conclusions about the distortions in the measured ASR when uncontrollable spending is not carved out. These distortions are especially large when the adjustment factor for baseline spending is a normally distributed random variable and uncontrollable spending involves CCF or amounts above $100 000 per person. In these simulations, failure to carve out uncontrollable spending leads to highly underestimated values of the ASR, which in many cases are expressed as large losses (ie, spending increases) even though controllable spending is assumed to have decreased by 5%. Thus, the added element of uncertainty from a random adjustment factor appears to penalize ACOs (through the underestimation of ASR) much more than it would penalize payers (through the potential to overestimate ASR).

It is not apparent why a random adjustment factor would cause uncontrollable spending to asymmetrically disadvantage ACOs and not payers. This finding might be driven by unique patterns in New Jersey Medicaid data or possibly the specific assumptions about the randomness involved in the adjustment factor. Although these issues may be sorted out with different datasets and modeling assumptions, it is clear that the random adjustment factor raises complex issues for measuring savings. This finding reinforces prior work, which showed that a random adjustment factor adds substantial riskiness to shared savings contracts for both ACOs and payers that could be avoided by specifying a deterministic adjustment factor at the beginning of the agreement period.[3,4]

Here, it should be emphasized that the ASR formulas represent a measured performance standard that is agreed upon by a payer and an ACO rather than an evaluation method to determine the ultimate effect of an intervention. In the latter case, a deterministic adjustment factor would be inappropriate as a "counterfactual" measure of spending trend, because it would not be known with certainty and would need to be estimated statistically. However, in the context of setting a performance target for the ACO, a deterministic adjustment factor removes statistical noise from the ASR calculation, as it is known with certainty what level of spending growth the ACO must achieve to earn a savings payment, leaving nothing to chance.

This difference in methods is an example of a broader distinction between an "administrative formula" and a "research-based evaluation" for provider performance measurement. As noted elsewhere, these approaches are designed to serve different purposes and, as a result, can sometimes produce divergent conclusions about provider performance.[20]

Administrative formulas, such as those analyzed in this article, provide a consistent standard that is known to all participants upfront and can be implemented fairly easily but are not necessarily designed to identify causal relationships. Assessment of causality (ie, the extent to which provider behavior truly caused the observed data patterns) is the ultimate goal of research-based evaluation. But as it requires substantial time and effort to develop model specifications, thorough causal research designs are typically not used to implement pay-for-performance contracts such as shared savings arrangements. Moreover, research evaluations often involve sensitivity analyses, which would confuse the implementation of contracts, which require one final number to determine final agreed-upon performance.

A fundamental assumption in this article is that certain spending components are completely unalterable by ACO activities. This assumption would be directly applicable to ACOs that have no focus on particular care domains such as injury and CCF, as investigated in this article. It would be invalid, however, for ACOs that have a focus on optimizing injury care or preventing patient admission to nursing homes.

This assumption is more complicated when applied to the truncation of spending above $100 000. Some ACOs focus directly on high users where they seek to generate substantial savings above the truncation point and prevent individuals with spending near that point from crossing over in subsequent years. Even if high users are not specifically targeted, care management activities for individuals with multiple chronic illnesses could potentially reduce an individual's costs significantly, but the costs still remain above the truncation point. Thus, the use of a truncation carveout to improve the accuracy of savings measurement must be done with the full understanding that it would discount savings achievements above the truncation threshold.

Although this article focuses specifically on ACOs in shared savings arrangements, it also outlines concepts and provides some guidance for a broader set of arrangements where providers assume responsibility for population-based spending, but some components of spending are not within providers' control. For example, under a capitation arrangement, the provider group will be paid an amount per member per month that reflects the average cost of care for the assigned population. The group would then have to manage the risks associated with random fluctuations around this average. As shown above, the inherent riskiness can be made smaller with spending carveouts that significantly reduce the mean and CV and increase the within-patient correlation as these apply to health care spending among the assigned population.

Development of a suitable carveout strategy would require 2 steps. First, identify categories of spending that are agreed to be beyond the current scope of the provider group's care management activities. Second, examine how carving out the identified categories affects the 3 key descriptive statistics described above. Those categories with the greatest potential to distort measured performance would be priority categories for carving out.

This article focuses on 3 examples of carveout approaches that could be implemented in a straightforward way with available claims data. In practice, many other carveout strategies could be considered. Conceptually, the most relevant carveouts would involve categories of spending that are outside of the provider group's current focal areas for care management and are large enough to potentially obscure actual performance within the focal areas. Additional carveout strategies not examined in this analysis may include pediatric cancers and expensive brand name drugs where providers have limited discretion over their use (eg, for hepatitis C or HIV-AIDS). Similarly, widely used and costly drugs that exhibit large price increases are beyond the ability of providers to control but could distort spending totals for patients who are the focus of care management activities.

As mentioned above, the definition of uncontrollable spending can vary across provider groups and may evolve over time. Thus, spending carveouts could be a negotiated feature of shared savings and other population-based payment contracts. For example, in a shared savings contract, a payer might be willing to accept fairly generous carveout features if the provider group is willing to accept a more stringent minimum savings rate threshold to ensure that the providers do not receive inappropriate rewards for a random decrease in spending. Payers, however, may be wary of carving out too many components of spending if the remaining components account for a fraction of total spending that is too small to make a meaningful difference in their TCOC.

Shared savings with carveouts can be considered within a broader menu of options that may be negotiated to improve the effectiveness and fairness of performance-based payment arrangements. One such option is episode-based payment where providers are paid one lump sum for an entire episode of care (eg, surgery, post–acute service, outpatient follow-up). This option provides strong incentives to coordinate and deliver care efficiently across multiple providers within a well-defined medical episode. But because it is narrowly focused and self-contained, it does not address broader concerns about prevention, screening, well care, and population health that are often included in the care improvement and cost-efficiency strategies of ACOs.

Another fairly common option is reinsurance where a separate entity assumes liability for expenditures above a certain threshold. In theory, an ACO responsible for TCOC could contract with a reinsurance firm to cover expenses above a certain threshold, or possibly, for expenditures in certain classes viewed as beyond its control. Such a contract, however, would still create a financial liability for the ACO to purchase the reinsurance and likely add new administrative costs by involving an additional organization to the broader payer-provider arrangement. Conceptually, a TCOC arrangement with carveouts acts as a form of reinsurance for the providers, where the reinsurance is provided by the

original payer (who could then in turn transfer this risk to a reinsurance firm, if desired). Thus, a key difference between carveouts and reinsurance has to do with who bears the financial liability for specific segments of health care spending and at what administrative cost.

In theory, risk adjustment of health care spending might mitigate the concern about expenditures that providers would be interested in carving out of their shared savings or other population-based payment arrangements. But even the most sophisticated risk adjustment algorithms are limited in how much person-level expenditure they can predict, especially when based on administrative data with limited clinical detail. In addition, expenditure categories that might be carved out include random events that by their nature are not amenable to prediction in a risk adjustment model (eg, injury, high-cost outliers).

The carveout approach studied in this article begins with TCOC and removes components of spending that are thought to be beyond providers' ability to control. An alternative approach could do the opposite—that is, begin with nothing and add only those components of spending that would be actively and explicitly targeted by providers at the beginning of the contracting period. Common spending targets include inpatient admissions or emergency department visits. Such an approach would minimize risk and provide upfront clarity to providers. But it would significantly limit the scope of the arrangement along with its spending reduction potential and could preclude important synergies in care management that are not obvious or fully considered at the time of contracting. Such synergies include links between mental and physical health and redesigning the use of post–acute care after inpatient discharge. Performance measures that "carve in" specific components of utilization, instead of spending, would be even less amenable to exploiting these potential avenues for efficiency gains through spending reductions.

This study is subject to some limitations. First, to implement the simulations, it examines a specific set of scenarios where each ACO achieved a true ASR of 5% and the allowable growth factor was fixed at (or normally distributed around) $500. Thus, the findings here may be viewed as illustrative of how carveout strategies can affect the riskiness of shared savings arrangements rather than a definitive account of all possibilities. Second, the study focuses on only 3 potential carveout strategies. Although it is beyond the scope of this study, other carveout strategies such those involving high-cost/nondiscretionary drugs (eg, for hepatitis C, HIV-AIDS) or end-of-life care, as well as combinations of carveout strategies, are important for future investigation. Third, the study examines only Medicaid ACO communities in New Jersey. Other arrangements with different payers, patient risks, and geography could potentially produce different results. Finally, the study does not address the potential trade-offs between the use of various carveout strategies and the administrative complexity of implementing them.

Despite these limitations, this article develops a rigorous framework for analyzing the effects of carveout strategies on the accuracy of provider spending performance. It also provides empirical assessments of 3 specific strategies that can be applied in a variety of settings.

References

1. Burns LR, Pauly MV. Accountable care organizations may have difficulty avoiding the failures of integrated delivery networks of the 1990s. *Health Aff (Millwood)*. 2012;31(11): 2407-2416.
2. Burns LR, Pauly MV. Integrated delivery networks: a detour on the road to integrated health care? *Health Aff (Millwood)*. 2002;21(4):128-143.
3. DeLia D. Monte Carlo analysis of payer and provider risks in shared savings arrangements. *Med Care Res Rev*. 2016;73(5):511-531.
4. DeLia D. Leaving it to chance: the effects of random variation in shared savings arrangements. *Health Serv Outcomes Res Method*. 2013;13(2):219-240.
5. DeLia D, Hoover D, Cantor JC. Statistical uncertainty in the Medicare Shared Savings Program. *Medicare Medicaid Res Rev*. 2012;2(4). doi:10.5600/mmrr.002.04.a04
6. Pope GC, Kautter J. Minimum savings requirements in shared savings provider payment. *Health Econ*. 2012;21(11): 1336-1347.
7. Centers for Medicare & Medicaid Services. *All Medicare Shared Savings Program ACOs and Pioneer ACOs (fast facts)*. Baltimore, MD: Centers for Medicare & Medicaid Services; 2015.
8. Centers for Medicare & Medicaid Services. Medicare program; Medicare Shared Savings Program: Accountable Care Organizations. Final rule. Fed Regist. 2011;76(212):67802-67990. Codified at 42 CFR §425.
9. Centers for Medicare & Medicaid Services. Medicare program; Medicare Shared Savings Program: Accountable Care

Organizations. Final rule. Fed Regist. 2015;80(110):32692-32845. Codified at 42 CFR §425.

10. McCluskey PD. Steward, Mount Auburn Quit Medicare Pilot Program. The Boston Globe. https://www.boston-globe.com/business/2015/11/03/accountable/h6vx7520a-CYs5VgxNy9h5L/story.html. Published November 3, 2015. Accessed October 28, 2016

11. Pope GC, Trisolini M, Kautter J, Adamache W. *Physician Group Practice (PGP) Demonstration Design Report* (Prepared for the Centers for Medicare and Medicaid Services). Waltham, MA: Health Economics Research; 2002.

12. Department of Vermont Health Access. *Accountable care organization in DVHA's next generation model: request for proposals* (Requisition Number 03410-175-16). Waterbury, VT: Department of Vermont Health Access; 2016.

13. Boozang P, Woda S, Codner SS. Manatt on Medicaid: Massachusetts releases details on forthcoming DSRIP waiver proposal. Manatt. https://www.manatt.com/insights/newsletters/medicaid-update/manatt-on-medicaid-massachusetts-releases-details. Published May 9, 2016. Accessed October 28, 2016.

14. Rhode Island Executive Office of Health and Human Services. *Accountable entity Coordinated Care Pilot Program: program description and application*. Providence, RI: Rhode Island Executive Office of Health and Human Services; 2015.

15. Herman B. Medicare's new ACO model preps providers to become insurers. *Mod Healthc*. 2016;46(3):10.

16. DeLia D, Yedidia M, Lontok O. *Year 1 of the New Jersey Medicaid Accountable Care Organization Demonstration Project: assessment of operations and care management strategies*. New Brunswick, NJ: Rutgers Center for State Health Policy; 2017.

17. Thompson F, Cantor JC. *The New Jersey Medicaid Accountable Care Organization Demonstration: lessons from the implementation process*. New Brunswick, NJ: Rutgers Center for State Health Policy; 2016.

18. Kronick R, Gilmer T, Dreyfus T, Lee L. Improving health-based payment for Medicaid beneficiaries: CDPS. *Health Care Financ Rev*. 2000;21(3):29-64. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4194678/pdf/hcfr-21-3-029.pdf. Accessed October 28, 2016.

19. Efron B, Tibshirani RJ. *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall/CRC; 1994.

20. DeLia D. Calculating shared savings: administrative formulas versus research-based evaluations. Health Affairs Blog. September 26, 2016. http://healthaffairs.org/blog/2016/09/26/calculating-shared-savings-administrative-formulas-versus-research-based-evaluations/