Risk adjustment and shared savings agreements

Hans K. Leida, PhD, FSA, MAAA
Leigh M. Wachenheim, FSA, MAAA

Milliman

In a typical shared savings arrangement, claim costs during the measurement or experience period are compared with one or more targets, such as claim costs for the same group during a prior period or claim costs for a comparable population during the measurement period. Claim costs are *risk-adjusted* to take into account health status differences between the comparison and target populations.

The key drivers in this comparison between the measurement period and target period are claim cost levels, member months, and risk adjustment factors. Conceptually, we are comfortable with two of the key drivers in this comparison—our old friends "member months" and "claim costs." We know what they represent and how they are derived.[1] For many, however, the third driver, risk adjustment, remains a stranger, or acquaintance at best, that needs to be better understood given its impact.

In this paper, we will discuss the role of risk adjustment in shared savings agreements, the uncertainties involved in its potential impact, and steps that can be taken to maximize its performance in shared savings agreements. These steps can help both providers and payors have increased confidence in the process, optimizing participation and motivation by all players.

An example

While there are a variety of provider payment mechanisms that employ the comparisons described above, shared savings is one of the more common arrangements. In a shared savings arrangement, a provider and a payor agree upon a target, and if the claim costs come in lower than the targets, the savings can be split between the provider and the payor. This structure aligns the incentives between the two parties. However, if the population is healthier or sicker than expected, then the calculations could be distorted. Therefore, risk adjustment is often used to account for these population variations when doing this calculation, to ensure that the shared savings reflect actual savings, as opposed to differences in the population's health or demographics.

Because of their design, shared savings calculations are often highly sensitive to relatively small changes in experience, including risk scoring. Figure 1 illustrates the very significant impact that risk adjustment can have on the shared savings calculation. In this hypothetical example, the payor and provider have agreed on a risk-adjusted claim cost target of $360.50, which is a 3% increase over the average claim cost of $350.00 in the base period. If the average risk-adjusted claim cost in the experience period is lower than the target, the payor has agreed to pay the provider half the difference. This example assumes there are 10,000 members (120,000 member months) subject to the shared savings arrangement in the experience period.

As mentioned above, the underlying health of the population impacts the claim cost. The same claim cost can represent different actual savings. Figure 1 shows three scenarios. In the baseline scenario, the claim cost and risk score in the experience period are unchanged from the base period, which might be interpreted to mean that the provider has managed to hold costs at the base period level for a population with no discernible change in health status (risk adjustment factor of 1.000). This results in a shared savings payment of $630,000, or $5.25 per member per month (PMPM). The alternative scenarios show results that still assume no change in claim costs, but reflect risk adjustment factors of 0.98 and 1.02—a 2% variance in either direction.

1 Although, truth be told, the need to attribute members and their claims to providers can make even these old friends a bit unfamiliar at times.

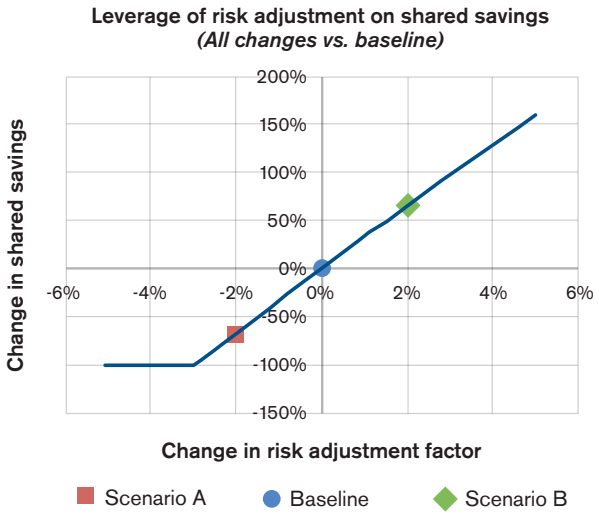
FIGURE 1: HYPOTHETICAL SHARED SAVINGS CALCULATIONS

SCENARIO		BASELINE	SCENARIO A	SCENARIO B
BASE PERIOD CLAIM COST PMPM	(a)	$350.00	$350.00	$350.00
TARGET CLAIM COST PMPM	(b)=(a) x 1.03	$360.50	$360.50	$360.50
EXPERIENCE PERIOD CLAIM COST PMPM	(c)	$350.00	$350.00	$350.00
RISK ADJUSTMENT FACTOR	(d)	1.000	0.980	1.020
CLAIM COST NORMALIZED FOR RISK ADJUSTMENT	(e)=(c) / (d)	$350.00	$357.14	$343.14
SAVINGS PMPM	(f)=(b) – (e)	$10.50	$3.36	$17.36
SHARED SAVINGS PERCENTAGE	(g)	50%	50%	50%
SHARED SAVINGS FOR 120,000 MEMBER MOS.	(h)=(f) x (g) x 120,000	**$630,000**	**$201,429**	**$1,041,765**

Because of the highly leveraged nature of the savings formula, this seemingly small change in the risk adjustment factor results in a large change in savings–increasing or decreasing the amount by approximately two-thirds. For example, in alternative Scenario A, applying the risk adjustment factor of 0.980 results in a risk-adjusted experience value of $357.14 PMPM and savings of just $3.36 PMPM.

Figure 2 extends the example graphically to show the impact of changes to the risk adjustment factor on a continuous basis.

FIGURE 2: CHANGE IN SHARED SAVINGS AS A FUNCTION OF RISK ADJUSTMENT



How confident are we in these calculations?

This sensitivity of shared savings calculations to small changes in risk scores naturally raises the question: "How confident are we that the risk scores being used accurately quantify the impact of a group's health status on cost?"

Risk adjusters are statistical models

The risk adjusters commonly used today are statistical models, which predict costs based on the presence of certain medical conditions as indicated in diagnostic codes and/or prescription drug codes in claim data.

For this reason, actuarial researchers have recommended that we consider the risk score to be a point estimate within a confidence interval.[2] In a recent report published by the Society of Actuaries, researchers provide a 90% confidence interval for risk scores for selected group sizes. The confidence intervals for a few group sizes that could credibly be subject to shared savings arrangements are reproduced in Figure 3.

FIGURE 3: 90% CONFIDENCE INTERVAL BY GROUP SIZE MEDICARE 5% SAMPLE AND CMS-HCC MODEL OBSERVED RISK SCORE = 1.000

GROUP SIZE (LIVES)	LOWER LIMIT	UPPER LIMIT
5,000	0.958	1.046
10,000	0.967	1.038
25,000	0.978	1.023

For example, according to these results, given an observed risk score of 1.000, we can be 90% confident that the interval from 0.967 to 1.038 contains the true population risk. Alternatively, there is a not insignificant 10% probability that the true risk is outside that range. This is a somewhat unsettling result given how sensitive shared saving amounts are to relatively small changes in risk scores.

The results in Figure 3 were generated using Medicare FFS 5% Sample data and the Centers for Medicare and Medicaid Services Hierarchical Condition Categories (CMS-HCC) risk adjustment software, along with extrapolation using a power law equation at high group sizes. We would expect similar results using other experience and models.

2 Mehmud, S. and Yi, R. (September 2012). Uncertainty in Risk Adjustment. Society of Actuaries. Retrieved November 21, 2014, from https://www.soa.org/research/research-projects/health/uncertainty-risk-adjustment.aspx.

Vendors, researchers, and other knowledgeable stakeholders frequently cite a model's "R^2" as an indicator of how well outcomes (or predictions) generated using the model match observed outcomes. The R^2—which ranges from zero to one—can be interpreted as the proportion of cost variance that is explained by the model. The typical range for a concurrent model is 30% to 60%. Therefore, even in a high-performing model where 55% of the variation in cost is accounted for, the other 45% is not.

Of course, we understand and expect that there are variables not captured in diagnostic data that impact healthcare utilization and cost. Some of them, such as the efficiency of care or reimbursement levels, are under the control of providers and are appropriate to reflect in the shared savings agreement. Others, such as benefit changes driven by regulation or market forces, are not normally impacted by providers, even though these variations are often not accounted for, by risk adjustment or otherwise, in shared savings agreements today—they are part of the 45%.

Improving accuracy

While any given measure of risk using a risk adjuster will only be an estimate of the relative health status of a population, subject to error, there are steps we can take to optimize its accuracy so the resulting savings settlements are as fair as possible.

Choose an appropriate model

Many risk adjustment models are available today from private vendors, academic researchers, and public entities. Many payors, if not most, are using more than one model within their companies for various purposes. Some of these purposes include revenue adjustment and allocation, member risk assessment and stratification (e.g., identifying potential high-cost patients), and provider network analysis.

Consider the following criteria when selecting a model to be used with shared savings arrangements, in addition to overall predictive ability, as described above.

- **Population:** Risk adjustment models are typically calibrated for specific populations—such as commercial, Medicare, or Medicaid. Avoid using a model developed for one population to calculate risk scores for a substantially different population without evaluating whether it still performs adequately. If an agreement includes more than one population (e.g., commercial and Medicaid), more than one model may be needed, although if multiple models are used, they must be normalized consistently.

- **Benefits:** The risk scores generated by a model assume some underlying set of benefits and scope of coverage. There are a few areas to focus on in particular:

  - **Carve outs:** Most medical (versus prescription-drug-only) models were developed assuming a broad set of benefits, including mental health. If specific services or conditions are excluded from the shared savings arrangement, it may be necessary to recalibrate the model.

  - **Prescription drugs:** In some cases, the payor does not have detailed prescription drug data for a material portion of the population, even when the benefit is included in the scope of coverage and subject to shared savings. This is often the case when prescription drug benefits are processed by a third party or pharmacy benefit manager (PBM). If prescription drug data is not available for a significant portion of the population, consider using a model that does not require drug data.

  - **Paid vs. allowed costs:** Risk adjusters are most commonly calibrated to predict variations in allowed costs—that is, costs including both the payor liability and member cost sharing. There are some risk adjusters, however, that are calibrated based on paid costs. For example, the U.S. Department of Health and Human Services-Hierarchical Condition Categories (HHS-HCC) model, used by CMS for calculating carrier-level risk adjustment settlements under the Patient Protection and Affordable Care Act (ACA), is calibrated based on estimated paid costs for five different cost-sharing levels. Many shared savings calculations are based on allowed costs, making a risk adjuster that is also based on allowed costs most appropriate. Allowed costs have the significant advantage of greatly reducing the need to adjust for changes in the average level and types of member cost sharing over time, although it is still important to consider whether changes in cost sharing might be impacting overall utilization over time. On the other hand, some of the "savings" on an allowed basis benefits a third party other than the payor or the provider—the insured member, who ends up paying less cost sharing in the form of deductibles, coinsurance, and copays. For populations with significant member cost sharing, consider whether an adjustment to the savings calculated on an allowed basis is needed to remove the portion that will impact member cost sharing. Otherwise, the payor may end up sharing "savings" that they never received.

- **Claim truncation:** Shared savings agreements frequently specify an upper limit on the amount of claims that will be counted toward cost for any given individual or episode. This is done to mute the impact of outlier (high-severity, low-frequency) claims on results. Likewise, claim "truncation" is often used in developing risk adjustment models for the same reason. Ideally, the truncation level behind the risk model being used to calculate shared savings will match the stop-loss level in the shared savings agreement. It may be possible to apply a truncation to risk scores after the fact to accomplish this as well.

  In choosing a claim truncation level, there is a trade-off involved. On one hand, dampening the volatility (and ultimately the risk to the provider) may call for a lower truncation point. On the other hand, a significant fraction of health spending is associated with outlier events, and those events are sometimes where effective management can make the most difference (although there are certainly extreme events that are not well suited for management efforts as well). It may be worth exploring other ways to reward providers for good management of outlier cases, as well as focusing on ways to help avoid them in the first place. The latter is a growing application of risk adjusters and other predictive models.

- **Prospective vs. concurrent:** Risk adjustment is performed on either a *prospective or concurrent* (aka *retrospective*) basis. Prospective risk adjustment is used to estimate costs in future periods. Concurrent risk adjustment is used to explain cost variations in a current period. In general, concurrent models are more appropriate for shared savings arrangements. Concurrent models have more predictive power and are better able to account for acute events such as accidents. Prospective models have lower predictive power and place more weight on chronic conditions.

- **Model changes:** Ideally, the same model versions and calibrations will be used to perform risk adjustment on all data sets used in the shared savings calculation for a given year. Otherwise, some of the difference in risk scores could be due to changes in the model and not differences in expected resource use that are due to health status. This means that risk adjustment may need to be rerun on the same data sets from one settlement calculation to the next to provide an apples-to-apples comparison of results. If it is necessary to use more than one model version or calibration within the same shared savings calculation, or if entirely different models are used on different population segments, then it is essential to adjust for that by renormalizing all the risk scores to some consistent basis.

Don't share the same gains twice

Risk-sharing arrangements are becoming more common and more complex–and not just between payors and providers. As this happens, it's becoming more and more likely that the same member's claims could be subject to multiple risk-sharing arrangements at the same time. For example, members in individual and small group plans subject to the ACA may be subject to risk sharing with the federal government through the risk corridors program, but may also be subject to risk-sharing arrangements between insurers and providers. Those same members may also generate settlements under ACA risk adjustment, transitional reinsurance, cost-sharing subsidy, and minimum loss ratio rules. To further complicate matters, the contract year of a provider risk-sharing arrangement may not line up with the benefit year for ACA settlement purposes.

All of these puzzle pieces need to fit together into a coherent whole so that the insurer doesn't share gains with a provider, only to realize that those gains turn into losses after all the ACA and other cash flows have settled. Gain-sharing payments to providers can affect settlements under some of these programs (for example, risk corridors and minimum loss ratios), so designing the shared savings agreement can become something of a chicken-and-egg problem.

This sort of complication is not restricted to the ACA market. For instance, managed Medicaid, Medicare Advantage, and Part D populations are other examples where these sorts of complexities can arise.

Adjust targets for "code creep"

Risk scores tend to rise over time, which is due to improvements in coding practices. Where focused coding improvement initiatives are underway, resulting year-to-year increases in risk scores can be quite large, in excess of 5%. Even in cases where there are no active initiatives underway, the increase is likely to be in the range of 1% to 3% per year, which is due to systemic changes alone. Drivers of such systemic changes include the spillover effects of coding improvement initiatives introduced by other payors, as well as the ongoing adoption of electronic medical records and other technologies aimed at improving providers' recordkeeping. This means that risk scores are likely to go up over time, even when there is no real change in health status.

In Figure 1 on page 2, we said that the baseline scenario, where both the claim cost and risk score in the experience period are unchanged from the base period, might be interpreted to mean the provider has managed to hold costs level with no change in health status. However, taking into account the impact of coding improvement on risk scores, the health status of the population has likely improved and, as a result, we would expect claim cost to be lower for that reason alone.

We illustrated how a 2% increase in the risk adjustment factor could result in a 65% increase in shared savings–from $630,000 ($5.25 PMPM) to $1,041,765 ($8.68 PMPM) for our illustrative group of 10,000 members. This is well within the 1% to 3% cited above, meaning this level of overpayment could be expected if no adjustment is made for the expected inflation in risk scores. A larger increase in risk scores, in the 3% to 5% range, could completely offset any credited savings that would have existed absent an adjustment for the risk score inflation.

Estimating the impact of coding improvement on risk scores requires significant expertise, as well as knowledge of company coding improvement initiatives. The challenge stems from the fact that it can be hard to directly observe and distinguish changes in coding practices from true changes in health status. While the adjustment will likely always require significant professional judgment in addition to careful analysis of the available data, failure to adjust for risk score inflation is likely to result in overstated savings.

Understand and resolve data issues

Risk adjuster calculations are only as good as the underlying data. Whole businesses have been built around assisting payors and others to improve coding completeness and accuracy. In many cases, initiatives to improve coding are tailored to maximize the change in risk scores.

- **Data quality:** Risk adjusters generally rely on complete and accurate medical diagnosis and prescription drug claim data. Confirm that all the fields used by the risk adjuster are fully populated with properly formatted, valid data.

  Historically, capitation, carve outs, and other arrangements have been associated with missing or poor-quality encounter data, so consider these situations carefully.

  Inconsistent data quality across provider groups can also be a concern where a provider group is being compared with other groups or with a community average. For example, the experience of a provider with lower-quality data will likely be assigned a lower risk score than would otherwise be the case. That, in turn, will result in a higher risk-adjusted claim cost, which will generally lead to a lower settlement.

- **Partial year membership:** There are many reasons a member might have only partial year data–they may be born or die during the year, or simply change payors. In some cases, the attribution logic may assign a patient only partial year membership with a particular provider. (This may occur even though the payor has complete, or more complete, data available.) There are several established methods for addressing this issue, including using average risk scores for members with insufficient experience or using demographic factors in lieu of risk scores.

- **Run-out:** Risk adjustment should be performed using the same amount of run-out (e.g., three months or six months) on all data sets used in the shared savings calculation.

A final consideration

Risk adjusters throw off a considerable amount of information that may be useful to both payors and providers. This commonly includes a listing of the conditions of each member, the ability to identify members who are likely to be high-cost going forward (through prospective risk adjustment), and the disease profile of patient groups affiliated with specific providers. Individual risk adjusters may also include unique features of interest. For example, the Milliman Advanced Risk Adjusters™ (MARA™) software also breaks risk scores down by major service category (inpatient, outpatient, emergency room, professional, prescription drugs, and other).

This information is useful for a variety of purposes, including resource planning, better and earlier identification of patients who could benefit from care management programs, and assessment of individual providers and groups of providers (where a sufficient volume of data exists). These activities all have the potential to lower costs and improve quality, which is a primary goal of shared savings arrangements.

Hans K. Leida, PhD, FSA, MAAA, is a principal and consulting actuary in the Minneapolis office of Milliman.

Contact him at hans.leida@milliman.com.

Leigh M. Wachenheim, FSA, MAAA, is a principal and consulting actuary in the Minneapolis office of Milliman.

Contact her at leigh.wachenheim@milliman.com.

FOR MORE ON MILLIMAN'S HEALTHCARE REFORM PERSPECTIVE

Visit our reform library at www.milliman.com/hcr
Visit our blog at www.healthcaretownhall.com
Or follow us on Twitter at www.twitter.com/millimanhealth