Benchmarking Methodology: Considerations and Options

Executive Summary

In VBP, 'benchmarking' is the method by which the budget is set for a VBP arrangement. If total costs of care for the VBP arrangement come out lower than the benchmark, the providers involved may share in the savings; if costs end up higher than the benchmark, providers may have to share in the losses.

A benchmarking methodology contains four steps:

1. Establishing the *baseline* (i.e., the costs taken as the starting point for the determination of the benchmark)
2. Establishing the *growth trend* (i.e., the percentage of assumed yearly growth in costs per member/episode to be incorporated in the determination of the benchmark)
3. Establishing the *risk adjustment* (i.e., the method by which differences in co-morbidity and other member-factors are accounted for in determining the benchmark)
4. Establishing potential *value modifiers* (i.e., the method by which a benchmark is adjusted for the relative efficiency and/or the quality of the care delivered by the provider. Such adjustments can also be realized by selective *rebasing* and/or adjusting the shared savings/losses percentages).

The following key options are to be weighed by the Subcommittee:[1]

Per option, the Subcommittee should recommend whether the State should set a *Statewide Standard* or a *Guideline* for the methodologies employed between MCOs and the providers. The State will consistently employ a standard in its own approaches regarding methodologies and data dissemination to both MCOs and providers. The Subcommittee should recommend whether MCOs and providers should adopt the same standard or are free to vary, using the State's methods more as a guideline.

- A Standard is required when it is crucial to the success of the NYS Medicaid Payment Reform Roadmap that all MCOs and Providers follow the same method.
- A Guideline is sufficient when it is useful for Providers and MCOs to have a starting point for the discussion, but MCOs and Providers may deviate without that harming the overall success of the Payment Reform Roadmap.

In the case of the Benchmarking Methodology, a Guideline is minimally required. The State will utilize a standard in determining benchmark levels and the associated data views.

[1] This table includes the key options but is not exhaustive. The accompanying document includes some options that were not deemed relevant enough to be included here based on preliminary discussions with stakeholders.

#	Topic		Option i	ii	iii
1a	Baseline – Aggregation Level	What breadth of claims to incorporate	*Provider specific claims*	*Other*	
1b	Baseline – Look back Period	How many prior years of claims data to incorporate	*One year of claims*	*More than one year of claims (three)*	
2	Growth Trend	What Growth Trend to apply	*Provider-specific*	*Regional average*	*Other*
3	Risk Adjustment	What Risk Adjustment to use	*3M for (sub)populations and IPC; HCI3 for bundles*	*Other*	
4a	Value Modifier - Efficiency	How to adjust benchmark in the light of a provider's efficiency (based on standardized costs)	*Adjustment for high efficiency providers through adjusting the benchmark; adjustment for low efficiency providers through rebasing*	*Adjustment of shared savings/losses percentage*	*A combination of i and ii*
4b	Value Modifier - Quality	How to adjust benchmark in the light of a provider's quality	*Additional adjustment of the benchmark for high- or low quality providers*	*Adjustment of shared savings/losses percentage*	*A combination of i and ii*

Introduction

The Value Based Payments (VBP) initiative in New York State aims to increase the quality of care delivered to Medicaid patients while controlling the costs of providing services and strengthening the financial outlook of the healthcare delivery system as a whole. VBP arrangements incentivize the provision of high quality, low cost care by creating opportunities for providers to share in savings when efficiencies are generated. The determination of shared savings within an individual VBP arrangement involves defining the member population and the scope of services covered, and then setting an achievement target (usually called 'the benchmark'). In Level 1 and 2 arrangements, the actual expenditure of the providers for the care contracted in the VBP arrangement is compared to the target at the end of the contract period, and potential savings (or losses) are calculated and reconciliated. This brief focuses on the target setting, or benchmarking methodology, component of the shared savings calculation process. The State considers setting guidelines for providers and MCOs for creating these benchmarks. Also, it will use this benchmarking methodology for the cost, outcomes and potential savings information it will provide to providers and payers (as mentioned in the Roadmap).

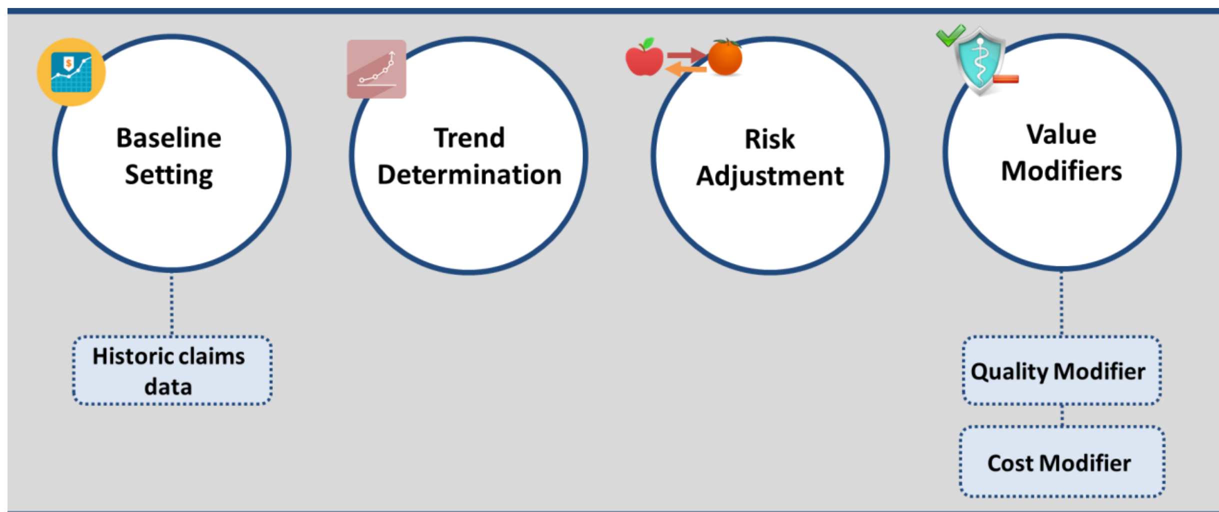

Figure 1: Overview of benchmarking steps. This figure does not include the step of determining actual shared savings based on actual performance of a provider against benchmark. This last step of shared savings determination may also include value modifiers. This will be discussed in the section "Value Modifiers" in this document.

The 'Technical Design I' Subcommittee is tasked with recommending a VBP benchmark methodology to the State by leveraging existing methods implemented in Medicare, State Innovation Models (SIM), commercial plans, or by investigating novel solutions. In order to facilitate this discussion, this brief discusses the components of the benchmark setting process, and describes the different options (with each their own pros and cons). A comprehensive benchmarking methodology includes four main components that should be taken in to consideration; (1) baseline setting, (2) trend determination, (3) risk adjustment, and (4) value modifiers. Value modifiers can either be introduced to modify the benchmark itself, or to determine the percentage of shared savings/losses for the provider.

During at least the first year of the VBP implementation, the State will use ***standardized costs*** *in its benchmark setting process. This standardization removes the effect of price on cost comparisons, leaving the differences observed between providers the result of either service mix and/or volume effects.*

Step 1
Baseline Setting
The foundation of the benchmark is the baseline expenditure: the aggregate of historic provider claims associated with a VBP arrangement prior to any adjustments. It enables a basic comparison of similar provider groups and serves as an initial point of reference at the end of the performance period. For providers that have not previously participated in VBP initiatives, the historic claims data is grouped to develop virtual episodic bundles, or the capitated payment baseline.[2]

Key considerations when defining the baseline include deciding the aggregation level of claims and the look back period. The aggregation level can be limited to an individual provider's claims history or span all claims in a state or region. Determining the look back period involves deciding how many prior years of claims data are incorporated into the baseline. In Medicare, both the Next Generation ACO Model (NGAM) and Pioneer ACO Model[4] aggregate provider specific baselines. NGAM develops the baseline using one prior year of claims data and Pioneer uses three years. BCBSMA leverages historic claims data to develop their provider specific baseline but has contemplated a move to a state standard or another fixed target.[3] The below table weighs the pros and cons of the baseline options for aggregation level and look back period:

#	Options	Pros	Cons
	Aggregation level		
1	**Provider Specific Baselines**	A baseline that is provider specific rewards providers for improvement against their prior performance. This incentivizes each individual provider to implement care redesign efforts to achieve savings.	It may be difficult for providers that are currently delivering care efficiently and with high quality to improve further.[4]
2	**Non-Provider Specific Baseline (Regional or Statewide data)**	This process of baseline setting rewards efficient providers and 'bends the cost curve' by penalizing inefficient providers.	For poor performers, this method of baseline setting may create a gap that is unrealistic in terms of being able to hit a benchmark and generate any form of shared savings. To that effect, it may deter participation in the benchmarking and shared savings process.

[2]BCBSMA, Alternate Quality Contract:https://www.bluecrossma.com/visitor/pdf/aqc-harvard-study.pdf
[3]BCBSMA, Payment Reform from on the Ground: http://www.bluecrossma.com/visitor/pdf/avalere-lessons-from-aqc.pdf
[4] The inclusion of value modifiers may reduce this disadvantage.

			This method also removes the ability of providers to work towards a goal that is specific to them and, therefore, more within their span of control given past performance.
	Look Back Period		
1	**One Year of Historic Claims**	This method will capture the most recent state of actual expenditures without being influenced by historic factors that may no longer be relevant.	A period of one year may not be representative of the overall financial and quality position of the provider.
2	**More than One Year of Historic Claims**	Capturing more than one year of claims data may allow for a more complete profile of the provider and guard against annual fluctuations. Additionally, it provides several years of data to calculate the growth trend.	By taking a longer period, the baseline calculation may include influences that are no longer relevant, therefore not giving an accurate picture of the provider's current position on cost and quality.

NOTE: Given the strong desire of the State and stakeholders to give all providers the incentive of potential shared savings (including those who are currently relatively inefficient), using a regional or Statewide benchmark is not included as an option in the Executive Summary.

Step 2
Trend Determination

Once a baseline has been established, the annual increases in healthcare costs per member/episode between the baseline period and the performance period must be incorporated into the benchmark evaluation. Medicare NGAM utilizes a regional healthcare cost growth trend based on the Medicare Advantage methodology.[5] Alternatively, the California Pay for Performance Program (P4P) incorporates the average change in the Consumer Price Index over the previous three years to account for costs.[6] The table below reviews the advantages and disadvantages of determining growth with historic increases, additional factors, or fixed standards:

[5] CMS Innovation Center, NGAM RFA: http://innovation.cms.gov/Files/x/nextgenacorfa.pdf ;
CMS, Trends Report: http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and Reports/ReportsTrustFunds/Downloads/TR2014.pdf
[6] IHA, Value Based Pay Performance Design (p. 7): http://www.iha.org/pdfs_documents/p4p_california/VBP4PDesign032513.pdf

#	Options	Pros	Cons
1	**Provider Specific Historic Rate**	Utilizing the provider's own historic growth rate for trend determination will allow for specific target setting per provider and removes the influences of factors to the benchmark that are not related to the provider at stake.	This method 'rewards' providers with higher growth in costs and 'punishes' providers with a lower cost growth, regardless of the value of their care.
2	**Non-Provider Specific Historic Rate** (Regional)	A regional trend approach removes the dependence on individual provider fluctuations and holds all providers in the same region to the same growth trend expectations.	This trend setting method does not allow for relevant changes to the situation of individual providers to translate through to the trendsetting process.
3	**Industry Growth Trend** (*MEI, CPI, Global Cap Limit, Zero, or Other*)	This method allows for the translation of industry expectations (e.g. Medicaid Global Cap) into the benchmark setting process.	Because these growth trends are national or Statewide averages, they may not be a true reflection of the rise of healthcare costs an individual region or provider is experiencing for a particular type of care.

NOTE: Given the overly generic nature of using an Industry Growth Trend parameter, this option is not included in the Executive Summary.

Step 3
Risk Adjustment

Prior to the calculation of the shared savings payments, risk adjustment is necessary to ensure a fair comparison between baseline and performance year financial performance. Risk Adjustment allows for an "apples to apples" comparison of the member populations over the two periods of time by adjusting the benchmark to account for the relevant risk factors that influence the cost of providing care. Currently within the Medicaid Managed Care rates, risk scores are calculated using 3M's Clinical Risk Grouping (CRG) model, and cost weights are developed by DOH. On an annual basis DOH and its actuary incorporate changes in case mix, utilization, and cost of care into MMC premium development. The use of 3M's CRG model in MMC rate setting makes it the preferred risk adjustment method for the Total Care for the Total Population VBP arrangement, as well as for Integrated Primary Care (which has a similar focus on a total population and considers the 'downstream costs' as an important factor in determining potential shared savings/ losses).

For Total Care for specific Subpopulations, a similar alignment between the risk adjustment used for the rate setting for the special needs plan would be preferable. These details will be discussed in the individual Advisory Groups for these subpopulations.

The Care Bundles are derived using the HCI3 (Health Care Incentives Improvement Institute) Evidence-informed Case Rate (ECR) Analytics. For these bundles of care, HCI3 has developed a risk adjustment method that has different parameters per bundle, and calculates the expected cost of a specific bundle given the history and acuity of the member.

Total cost of care of (sub)populations / Integrated Primary Care	
Method	Summary
3M CRG Risk adjustment methodology	This method of risk adjustment is already used in the state of New York for Medicaid Managed Care premium development. Employment of this methodology would enable full alignment with a previously tested, tried and accepted method.
Other	Any other methodology would encompass the introduction and development of completely new risk adjustment methods.
Bundles of care	
Method	Summary
HCI3 Risk adjustment methodology	This method of risk adjustment is an integrated aspect of the HCI3 Evidence-Informed Case Rate (ECR) methodology.
Other	Other risk adjustment methodologies may be preferable if providers and MCOs decide to adopt Medicare BPCI (Bundled Payments for Care Improvement) bundles.

4. Value Modifiers[7]

When provider-specific baselines are incorporated into the benchmark (see above), providers that already deliver highly efficient care at a high quality level have little opportunity to generate shared savings. Simultaneously, relatively inefficient, low quality providers will have a large opportunity for shared savings, which may inadvertently result in 'rewarding' providers for their historic inefficieny and/or poor quality. An approach for addressing these challenges is to introduce *value modifiers* into the benchmark calculation. In this document, we distinguish between two types of value modifier:

1. Cost (or efficiency) modifiers[8]; and
2. Quality modifiers

Value modifiers would increase or decrease a provider's benchmark according to that provider's previous cost and/or quality performance as compared to a regional or statewide average (based on standardized costs). A regionally adjusted *cost* modifier compares a provider's cost performance with the performance of its peers, and the *quality* modifier will compare a provider's outcome measures with those of its peers. In the Medicare NGAM's methodology, efficiency and quality modifiers change the 'discount' that CMS applies to the ACO specific benchmark, creating a possible range of 0.5% (for high quality and low cost ACOs) to 4.5% (for low quality and high cost ACOs). In practice, this

[7] This section assumes that a provider-specific baseline is used.

[8] As said above, during at least the first year of the VBP implementation, the State will use **standardized costs** in its benchmark setting process. This standardization removes the effect of price on cost comparisons, leaving the differences observed between providers the result of either service mix and/or volume effects.

means that the counter for 'shared savings' start at either .5% or 4.5%
below the historical baseline.[9] In CMS' model, in other words, the value modifiers adjust the 'haircut' that providers receive before savings become shared.[10] An option for New York State VBP would be to have value modifiers increase or decrease the benchmark in a similar manner to the NGAM Model, without applying a discount. The result would be to increase the benchmark of a high performing provider (of maternity care, the chronic care bundle, Total Care for a Subpopulation etc.) as a reward for their current efficiency and quality with 2-3%. Poorly performing providers, for example, could get a 2-3% haircut; highly performing providers could get a 2-3% 'raise', increasing the opportunity of shared savings.

Value modifiers can also be introduced to adjust the amount of savings/losses to be distributed to the provider by a percentage based on quality performance. High performers receive higher shares of savings and lower shares of losses, and vice versa. In theory, a similar adjustment could be made for efficiency. (A separate option paper will be presented to discuss the details of how value modifiers could be applied to shared savings/losses).

Finally, value modifiers could impact how to calculate new baselines in future years ('rebasing'). In any future year, a historical provider-specific baseline would include years when the shared savings/losses mechanisms were put in place. If the actual costs would be taken into account, providers with significant shared savings would see their baseline decline rapidly – as occurred in the Medicare MSSP and Pioneer programs. Finding the right balance between reasonable rebasing (to prevent inefficient providers from long-term rewards) and unwanted downward adjustment (pushing already highly efficient providers further downward; taking away shared savings so fast that the motivation to participate evaporates) is essential. For efficient and/or high quality providers, for example, the baseline could be calculated *including* the savings realized.

The key practical differences between the three ways of applying a modifier are explicated in the table below:

Modifying the Benchmark	Modifying the benchmark increases or decreases the potential for shared savings/losses. If a Total Cost of Care for the Total Population VBP arrangement would have an $8,000 PMPY benchmark, a 2% uptick for a low cost and/or high quality provider would reset the benchmark at $8,160 PMPY. If this group of providers would serve 50,000 beneficiaries and end the year at an average of $8,000 PMPY, the total savings to be shared would be 50,000 * (8,160 – 8,000) = $ 8M. (With an end of year average of $7,800, the total savings to be shared would be $18M). Alternatively, a high cost/low quality provider group with a 2% haircut, a $10,000 PMPY benchmark and a $10,000 end of year PMPY average would be looking at a $10M loss. (With an end of year average of $10,200 PMPY the total losses to be shared would be $20,2M).
Modifying the Shared Savings/Losses	Here, value modifiers do not impact the benchmark, but they increase or decrease the percentage of savings/losses that the provider will receive. In the first examples, without the modifiers impacting the benchmark, the provider groups' end of year result would be at the benchmark (8,000 resp. 10,000 PMPY). They would therefore have no shared savings or losses. In the second examples, the provider groups' end of year result would be below resp. above benchmark (7,800 resp. 10,200 PMPY), resulting in savings resp. losses of $10M. A high quality and/or efficiency provider would receive a

[9] Risk adjusted and appropriately trended.
[10] In NGAM, as in all Medicare ACO models, CMS reduces the benchmark by a 'discount', ensuring a minimum level of savings for CMS. CMS Innovation Center, NGAM RFA (p. 12): http://innovation.cms.gov/Files/x/nextgenacorfa.pdf

	high percentage of the savings and would only be held accountable for a low percentage of the losses, and vice versa. Note that in this methodology, the resulting impact on the providers is less than in a methodology that directly impacts the benchmark itself.
Rebasing	The key difference between rebasing and modifying the benchmark is that 'rebasing' impacts the benchmark *after* the first shared savings/losses are realized. To prevent attrition, downwards rebasing should be limited to those providers with standardized costs above the statewide average.

The following options (or combinations thereof) can be considered:

#	Options	Pros	Cons
1	**Inclusion of Cost Modifier in Benchmark Setting**	Providers that are already operating efficiently will be rewarded through the application of this modifier. Similarly, poorly performing providers are not 'rewarded' for previous inefficiencies. Without a downward adjustment, also, poor performers may quite easily hit their targets.	For 'high cost' providers, including a 'haircut' may reduce the appetite to participate in VBP arrangements.
2	**Inclusion of Quality Modifier in Benchmark Setting**	Providers that do not have a cost advantage may be providing exceptional quality of care. The introduction of a quality modifier would allow these providers to be rewarded for their performance even if they would otherwise not realize savings. Similarly, poor quality providers would incur a 'haircut'.	For low quality providers, including a 'haircut' may reduce the appetite to participate in VBP arrangements
3	**Inclusion of Cost Modifier in Shared Savings/Losses Adjustment**	This could reward efficient providers and reduce potentially 'unfair' gains for low efficiency providers	For 'low cost' providers with little opportunity to save, modifying shared savings may have little impact. For 'high cost' providers, the 'haircut' is still in place – only now through a reduction of the share of any savings.
4	**Inclusion of Quality Modifier in Shared Savings/Losses Adjustment**	Rather than adjusting the benchmark, here a high quality provider would receive a higher share of savings (or a lower share of losses). A low quality provider would receive little to no savings and would have to incur a larger share of the losses.	For high quality providers, this generates more income only when savings are realized.

5	**Inclusion of Cost Modifier in Rebasing**[11]	This allows poorly performing providers to initially obtain the 'full' shared savings for one or more years. After that, the baseline would move downwards to incorporate the new cost levels, which prevents long-term pay-outs to previously inefficient providers.	Although upwards rebasing is possible, providers operating at high efficiency would likely prefer a modification of the benchmark in year 1 (with immediate impact). The limits of downward rebasing need to be clearly delineated to prevent attrition.

Important consideration in weighing these options:

1. It may be preferable to incentivize high efficiency & high quality providers through a benchmark modifier, while some downward rebasing can be an effective way to reduce expenditures of the low efficiency/low quality providers over time while allowing them to reap full shared savings in the first year(s). Downward rebasing could be limited to those providers whose standardized costs are above average. Also, all funds that would become available through this mechanism would be directly invested in rewarding higher value providers and MCOs.

[11] Although it is theoretically possible to include a quality modifier in the Rebasing methodology, this is not a method that is used in practice.