maryland
health services
cost review commission

To:	**Commission Stakeholders**
From:	William Henderson, Principal Deputy Director- Medical Economics and Data Analysis
Re:	HSCRC Staff's Response to Stakeholders' Concerns About Benchmarking Methodology
Date:	January 26, 2022

Adam Kane, Esq
Chairman

Joseph Antos, PhD
Vice-Chairman

Victoria W. Bayless

Stacia Cohen, RN, MBA

James N. Elliott, MD

Maulik Joshi, DrPH

Sam Malhotra

....

Katie Wunderlich
Executive Director

William Henderson
Director
Medical Economics & Data Analytics

Allan Pack
Director
Population-Based Methodologies

Gerard J. Schmith
Director
Revenue & Regulation Compliance

I. Introduction

The Health Services Cost Review Commission (HSCRC) staff released their initial Total Cost of Care (TCOC) benchmarks in August 2020. Since then, stakeholders have raised questions and concerns about the use of these benchmarks in Commission methodologies and whether they present an accurate picture of Maryland's TCOC compared with that of the nation. In response to these points, and to provide the Commission additional comfort regarding the use of the benchmarks, HSCRC staff have completed additional research and analytics on the benchmarks.

Staff's stated goal for the benchmarking was to create a tool to allow incorporation of the Total Cost of Care (TCOC) benchmarks into appropriate methodologies at a hospital service area level, thereby guiding the State on areas of strength and weakness in terms of cost and quality.
The memo first describes the analytical principles that HSCRC staff used in establishing the benchmarks. It then reviews issues raised by stakeholders, describes additional analytics performed by staff, and concludes with an outline of changes that staff will make to the benchmarks and related methodologies based on this work. The memo assumes that readers are familiar with the benchmarking approach used; more background information can be found in the documentation available at this link: August 2020 benchmarking materials[1].

II. Analytical Principles

The benchmarks are intended to answer the following question: "How does the TCOC and utilization of services in each measured geographic area of Maryland (for example, hospital service areas) compared to a national selection of similar geographic areas?" In answering this question, HSCRC staff incorporated three analytical principles:

1. **The benchmarking is a comparison tool and not a policy, and thus the benchmarking results should not include value judgments.** HSCRC staff believe value judgments are a matter of Commission policy. If the

[1] https://hscrc.maryland.gov/Documents/August%202020%20Benchmarking%20Materials%2011-10.zip

To: Commission Stakeholders
From: William Henderson, Principal Deputy Director- Medical Economics and Data Analysis
Date: 1/26/2022
Page: 2

Commission believes Maryland should invest more in certain areas than the rest of the nation does, or address inequities in a way that the rest of the nation does not, then that should be incorporated into the appropriate policy, not assumed in the benchmarking methodology.

2. **Benchmarking is a complex process with many reasonable options on how to proceed. From these options, HSCRC staff must select the one they regard as optimal, balancing competing considerations in an unbiased and justifiable way.** In any market, health care cost is driven by a myriad of demographic, medical, economic, historical, policy, and other factors. Different stakeholders could argue for, and reasonably defend, different options. Staff's preferred approach, which was developed after testing many options and methods and consulting with nationally recognized health care cost experts, is well within the bounds of what is reasonable. Among many alternatives, staff selected an approach that balanced the complexity of the method with the reasonability of the results at the state and local level. The approach should therefore not be modified based on varying stakeholder preferences that could be associated with a particular perspective.

3. **The benchmarking should be based on demographic, social, and economic factors and answer the questions on comparative cost and utilization in similar geographic areas without adjusting for any artifacts of the Maryland Model or other unique aspects of the Maryland health care system.** Incorporating these artifacts would run the risk of adjusting away the very differences between Maryland and non-Maryland areas that the benchmarks are intended to capture. Staff applied this principle to all health system factors, not just those driven by Maryland's unique history of setting hospital rates.

III.   Stakeholders' Concerns

The following section addresses concerns raised in various forums about the benchmarking methodology:

1. **Concern: The matching and/or regression criteria should have included a wage factor, or specifically the Medicare Wage Index (assumed to mean the Medicare Hospital Wage Index).**

   Some commenters have said wage factors should have been used instead of, or in addition to, median income to select comparison areas and to make regression adjustments. HSCRC staff believe median income and wage factors are colinear and that one cannot be wholly inappropriate but the other wholly appropriate. To control for price differences in inputs, staff used a measure from the U.S. Department of Labor—regional price parities (RPP), which measure the cost of goods and services—in the matching algorithm but removed this measure from regression models, as models with this variable had worst model fit.

   However, staff acknowledge that wage levels might reflect a somewhat different dynamic than median income, and RPP might not be sensitive enough because it is measured at the level of the metropolitan statistical area (MSA).[2] Therefore, staff did consider wage measures in the additional analysis discussed in Section IV. We tested several broad-based wage indices available from the Bureau of Labor Statistics (BLS). The results indicated that our main model performed as well as, if

---

[2] RPP was used in the matching criteria but not the regression algorithm as models including RPP produced worse model fit statistics. We hypothesize that this is because RPP is reported at the MSA level, it results in poor model fit when used in a county-level regression. Models with RPP resulted in lower R-squares and changed the signs of coefficients for sociodemographic factors.

not better than, the models including wage indices, which is in line with the recommendations from a seminal report by the Institute of Medicine.[3]

Staff do not agree that the Medicare Hospital Wage Index is an appropriate or relevant metric to use because:

a. There is widespread concern about the Centers for Medicare & Medicaid Services' (CMS's) wage index.[4] The Institute of Medicine's report summarized the issues as problems with and inconsistencies in the definitions of payment areas and labor markets, concerns about the relevance and accuracy of the source data used to determine area wages and other input prices, questions about the occupational mix used to create the hospital wage and physician practice expense adjustments, and lack of transparency in the index construction. The main concern is that the wage index adjustment is circular: only hospital wages in a market are considered in the wage index, which then feeds back into hospitals' ability to pay those wages. The circularity is particularly acute in Maryland, where the model has undoubtedly affected the wages paid by Maryland hospitals. To use hospital wage index in adjustments and matching for benchmarking would clearly violate the analytical principle of avoiding building Maryland's uniqueness into the benchmark.

b. Hospital wage index cannot be used to adjust overall spending measures as hospital costs represents less than half of per capita spending. In addition, wages represent a decreasing portion of hospital spending as drugs, supplies, technology, and service contracts make up an increasing portion of spending.[5]

c. The Commission is focused on measuring per capita spending for both Medicare and commercial payers. Commercial payer prices are highly variable relative to Medicare prices nationally, and as a result, HSCRC staff are concerned about using Medicare's fee-for-service rate-setting factors to evaluate multi-payer per capita spending.

d. CMS does not use its wage index in setting rates for many hospitals, particularly in rural areas. Critical access hospitals are paid based on cost, whereas rural referral centers and sole community providers receive payments above the wage-indexed amount.

e. Medicare's assignment of hospitals to core-based statistical areas (CBSAs) leads to idiosyncratic results at the local level. For example, Christiana Care Union Hospital has the highest wage index of any Maryland hospital, more than hospitals in the District of Columbia market, due to its assignment to the Wilmington CBSA. In addition, Peninsula Regional Medical Center (PRMC) has cited Beebe Medical Center as a local peer, which seems reasonable given that they are 45 miles apart and both on the lower Delmarva Peninsula. But in 2020, Medicare assigned Beebe a

---

[3] Committee on Geographic Adjustment Factors in Medicare Payment, Board on Health Care Services, Institute of Medicine. *Geographic Adjustment in Medicare Payment: Phase I: Improving Accuracy.* Second edition. Edited by M. Edmunds and F.A. Sloan. Washington, DC: National Academies Press, 2011. Available at https://www.ncbi.nlm.nih.gov/books/NBK190074/. Accessed January 13, 2022.

[4] See discussion in Congressional Research Service. "Medicare Hospital Payments: Adjusting for Variation in Geographic Area Wages." March 3, 2021. Available at https://crsreports.congress.gov/product/pdf/R/R46702. Accessed January 13, 2022.

[5] Nonhospital wage costs could be adjusted using the three Geographic Practice Cost Indexes (GPCIs) used to adjust physician payments in Medicare. But GPCIs are based on different geographic boundaries, reflecting the historical artifacts of Medicare's adjustments, rather than on an empirical foundation—which exemplifies the challenges with using Medicare wage adjusters for analytical purposes.

wage index of 1.0549, whereas PRMC's is 0.9219, which suggests a 13 percent wage gap (and implicitly that PRMC should be 13 percent cheaper on the wage portion of hospital-related TCOC). Beebe's wage index is so much higher because Beebe is assigned to the Ocean City, New Jersey, CBSA, most of which is on the other side of the Delaware Bay,[6] and thus its wage index does not seem to be a real reflection of the labor market on the Delmarva Peninsula.

2. **Concern: The regression adjustment for median income creates a higher benchmark for wealthier areas, which results in strong performance in those areas. This is undesirable because it shifts resources to those areas.**

    HSCRC staff agree that the use of median income results in a higher benchmark for wealthier areas in Maryland and that these areas generally look better than their benchmarks. However, staff do not believe that either of these things reflects a flaw in the benchmark. Concerns related to adjustment for median income can be placed into two broad categories:

    a. **There are technical issues with HSCRC's statistical calculations.** Commenters said the R^2 for the Medicare TCOC regression is only 0.13 and is therefore not valid. They also noted that median income sometimes overwhelms the impact of the deep poverty percentage in the regression and given the importance of poverty in driving health care need, the implication is that this cannot be correct.

        HSCRC staff believe both points overlook the fact that the regression is the second step in a two-step TCOC benchmarking process. The first step is selecting peer jurisdictions that are similar to the Maryland jurisdictions. The selected peer counties are therefore already comparable to the Maryland counties, and the second regression step then adjusts for any *remaining* differences. We would expect a low R^2 because much of the variation in the outcome (TCOC) that can be explained by median income and deep poverty has already been removed in the peer-selection step. A low R^2 simply reflects the degree to which the relationship between income, deep-poverty and outcome is already accounted for during peer selection.

    b. **Median income is an inappropriate demographic factor because it results in higher benchmarks in wealthier areas and therefore implicitly moves resources to wealthier areas.** Removing median income from the analysis because it is favorable to wealthier areas is a value judgment that addresses the question about where we should invest. It is therefore outside the purview of the benchmarking methodology, which is designed to compare Maryland's TCOC against geographies with similar socio-economic characteristics (as described in the first Analytic Principle in Section II).

    HSCRC staff believe median income is an appropriate factor to consider, both because it is a proxy for wages and because it captures broader demographic factors; setting cost aside, median income is a better reflection than wages of local health care utilization because it reflects the income of the people who live in a market, rather than those who work in a market. Staff note that both Hierarchical Condition Categories (HCCs) and deep poverty are also included in the regression, and these should capture the impact of higher health care needs.

---

[6] Staff note that although the Cape May-Lewes Ferry provides regular service, the 85-minute cruise is likely not a viable commuter option for most health care workers living in New Jersey (https://www.cmlf.com/sailing-information).

To: Commission Stakeholders
From: William Henderson, Principal Deputy Director- Medical Economics and Data Analysis
Date: 1/26/2022
Page: 5

Some commenters argued for a more direct adjustment of wages via use of a wage factor. Staff recognize that direct and indirect measures of wages could result in different outcomes. Section IV describes additional analytics we conducted to evaluate the use of wage factors in addition to or instead of median income; in general, wage factors had little impact on the outcomes.

3. **Concern: The benchmarks fail to recognize the greater health care challenges in the Baltimore area.**

   One commenter cited data from the County Health Rankings and Roadmap website (https://www.countyhealthrankings.org/) to point out that Baltimore city faces significantly greater health challenges than the rest of the state. However, HSCRC staff's county-selection process means that Baltimore is being compared to similarly challenged areas, and thus any cost variation remaining after matching and regression adjustments is limited to that not measured by the selection factors used. Also, the County Health Rankings site measures health, which is a much broader concept than health care, and due to access and other considerations, factors that correlate with poor health do not necessarily correlate with higher-cost care. However, despite the differences in the County Health Rankings measurement approach, staff compared the rankings of Maryland counties with the average ranking of their peer counties in Appendix A. The rankings are similar, validating that the peer counties selected are generally aligned with the level of need in their peer Maryland county.

4. **Concern: The national 5 percent sample of Medicare claims might not accurately reflect actual spending, particularly in smaller counties; the Medicare Geographic Variation file might be a better source.**

   The 5 percent sample is provided by Medicare for research purposes within the Chronic Conditions Warehouse (CCW), which is the is the same environment and data standard under which Maryland's performance under the TCOC Model is evaluated. Therefore, by using the 5 percent data for the nation, HSCRC staff are using a data set that is comparable to the data used in Maryland's performance assessment. The 5 percent sample is widely used by researchers and, more important, by qualified entities to develop national benchmarks, indicating staff's use of the 5 percent sample is appropriate[7]. The sample has also been used by the Medicare Payment Advisory Commission and in numerous publications.

   The Medicare Geographic Variation file is a publicly available report intended to provide summary level statistics and does not align with datasets used in evaluating Maryland's performance and produces slightly different results for both Maryland counties and benchmark counties, owing to differences in the way CMS produces and calculates this report[8]. HSCRC staff do not believe it would be appropriate to compare Maryland data from one data set to national data from a different set, nor do staff want to evaluate Maryland with a data set that does not reflect Maryland's performance as measured by CMS. Also, unlike the 5 percent sample, the Geographic Variation file does not allow any drill-down beyond the fields provided, which limits staff's and stakeholders' ability to investigate possible interventions, research questions, and data anomalies at a more granular

---

[7] see https://www.qemedicaredata.org/apex/Data_Availability_and_Cost
[8] see https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/Downloads/Geo_Var_PUF_Methods_Paper.pdf

level. Finally, staff use a peer group of 50 counties each for small Maryland counties. This sample size was chosen partially to mitigate chance variation in any single county's 5 percent sample.

5. **Concern: There are issues with the commercial data used, and the national commercial data used are not freely available.**

Unfortunately, there is no comprehensive, high quality national database of commercial claims that is freely available to public. But HSCRC staff do not believe the absence of such data should prevent the Commission from using national data sets that are available for benchmarking. Specifically, the benchmarking relies on Maryland's Medical Claims Database (MCDB, often referred to as APCD), which is compiled by the Maryland Healthcare Commission based on data from insurers in the state and national commercial claims data acquired from a highly experienced national vendor (Abt, Inc., and its subcontractor Milliman using Milliman's Consolidated Health Cost Guidelines Sources Database[9]).

Some commenters highlighted concerns with the MCDB data in Abt's report. HSCRC staff evaluated these concerns with the contractors and determined that they did not invalidate the resulting benchmarks to any material degree. In addition, Milliman compared the MCDB data to the Maryland data from their national data set (which undergoes substantial vetting) and determined that data were comparable. Staff used MCDB data because these data allow for a much more geographically granular evaluation of Maryland, and overall, Maryland results extracted from the national data set would have yielded a similar result. Although more rigorous validation of the commercial data would be ideal, staff do not believe the lack of such a process should invalidate the outcomes. We did pursue a more rigorous process for the 2019 data, as discussed in Section V.
Some industry commenters also expressed concern that the commercial data are not publicly available, which limits hospitals' ability to replicate the methodology. The Commission has always strived to be as open as possible in making the details of methodologies available; however, staff do not believe that a hospital's ability to replicate a methodology should be the standard for accepting or rejecting an approach. And although we cannot release claim-level details, we have made extensive aggregated commercial data, on both Maryland and the nation, available to hospitals. Hospitals in the rest of the country would have to incur significant costs to replicate the data received by Maryland hospitals for free.

6. **Concern: The benchmarking methodology has not been sufficiently vetted or explained to the industry.**

The benchmarks were first used in a policy as part of the Medicare performance adjustment (MPA) in January 2021, followed by use in the fiscal year 2022 Inter-hospital Cost Comparison (ICC) methodology in July 2021. These uses occurred two years after HSCRC staff first began discussing the benchmarking approach. Staff began socializing the benchmarking process with stakeholders as early as January 2019. HSCRC staff shared a first draft of the selected peer counties with the TCOC work group in March 2019, released a revised selection of peer geographies in December 2019,

---

[9] Milliman's Consolidated Health Cost Guidelines Sources Database contains detailed enrollment, medical claim and pharmacy claim information for over one third of the employer-sponsored healthcare market (includes MarketScan data).

To: Commission Stakeholders
From: William Henderson, Principal Deputy Director- Medical Economics and Data Analysis
Date: 1/26/2022
Page: 7

shared a final list and methodology report with two education sessions in August 2020. Recordings of these sessions were also made available[10].

The final approach is substantially similar to what was described throughout the process, and the selected peer geographies have not changed from the list released in December 2019. HSCRC staff also updated the TCOC work group throughout the process and have met with a number of stakeholders who had specific questions. Staff followed this process/timeline to ensure stakeholders had sufficient time to familiarize themselves with the approach and provide feedback to determine the final methods. Finally, HSCRC has generally chosen to implement methodologies and then refine them over time (for example, ECMADs or MPA attribution). Staff are committed to continuing to refine the benchmarking process but believe the benchmarks are sufficiently valid for current use and stakeholders had multiple opportunities and sufficient time to vet key decision points.

7. **Concern: Using a purely geographic approach to attribute beneficiaries is challenging for hospitals in areas with many hospitals and overlapping service areas. A primary care-based approach such as the MPA would be preferable.**

Because of the complexity of tracking relationships, the Commission has moved away from the primary care-based approach in the MPA. In addition, HSCRC staff did not use this approach for the benchmarking because it would be impossible to replicate with the commercial data, which represent a sample rather than the entire population of commercial beneficiaries.

IV. Additional Analytics

As discussed in Section II, Item 2, there are many reasonable methods for establishing benchmarks. HSCRC staff believe the selected approach strikes a balance between various considerations and produced results within the reasonable range. But there are certainly factors not included that could affect results, such as wage factors. It would also be helpful to know the size of the range of results different reasonable methods produce to understand the level of precision inherent in the selected approach.

HSCRC staff developed an approach to look at the variation in results produced with different methods and to assess if the results were biased against certain group of geographies such as rural counties.[11] Using this approach, they repeated the benchmark process 20 times, using alternative metric sets for peer selection and regression. Three different wage measures were included, both replacing and supplementing median income. Appendix B discusses the specific alternatives tested, the details of the process, and additional findings.

The alternatives all yielded very similar results to the selected approach in terms of rankings of Maryland counties' TCOC performance in TCOC versus their benchmark. The *lowest* Spearman correlation between the rankings under the selected approach and the rankings among the 20 alternatives was 0.87,

---

[10] Recording is available at https://attendee.gotowebinar.com/recording/3002307069132328450
[11] HSCRC staff only applied this approach to Medicare data because (1) Medicare benchmark results have a greater impact compared with commercial benchmarking on HSCRC policy; (2) the Medicare benchmarks are at the county level rather than MSA level, resulting in a greater opportunity for changes in the selected peer groups; (3) those benchmarks will inform future negotiations with CMS; and (4) the approach is resource intensive, and staff did not believe replicating it with the commercial data set was cost justified as the results would likely mirror the results with the Medicare data.

yielding an average absolute change in rank of 2.8 (which was driven by variations in the smallest counties). The Spearman correlation for most alternatives was well over 0.90. In addition, staff did not find any biases against types of counties (such as rural counties or counties in the Baltimore area) when using the selected approach versus the alternative approaches.

Compared with the rankings, staff found more substantial variation in the size of the average difference from benchmark in some of the alternatives (for example, one model changed the difference by 3.5 percentage points versus an overall average difference of 7.7 percent). Although some tests had individual outliers, these larger variations generally reflected most or all counties moving in concert, resulting in the smaller variations in rankings noted above.

Overall, HSCRC staff believe this analysis shows that the selected benchmarking approach yields results similar to those of a variety of alternative methods; thus, it is likely a valid approach within the reasonable range of possible methods. Our additional analysis indicated that the rankings were robust but absolute values were sensitive to approach therefore we believe the Commission should continue to be cautious about methodologies that use absolute values rather than ranked values. In Section V, we propose one change to current methodologies based on this caution.

V.   Proposed Revisions

Based on the stakeholder comments and additional work described in this memo, HSCRC staff are proposing three revisions to the benchmarking methodology and policy uses.

1. HSCRC staff believe, based on the work described in Section IV, that the selected benchmarking methodology is well within the range of reasonable approaches to assess Maryland's performance on TCOC by jurisdiction versus the nation's performance. But Section IV also illustrates that there is some sensitivity of absolute results among the various reasonable approaches to benchmarking.

   Staff therefore propose that the Commission revise its Full Rate Application methodology, adopted by the Commission in January 2020, to eliminate rate enhancements based solely on performance relative to Medicare and commercial TCOC benchmarks. Staff recommend either (1) eliminating any credit earned via the Full Rate Application methodology that is derived from TCOC performance relative to Medicare and commercial benchmarks or (2) providing credit to a hospital via the Full Rate Application methodology if the hospital shows positive performance relative to Medicare and commercial benchmarks *and* if the hospital's attributed TCOC is growing more slowly than the statewide average; credit will be equivalent to the delta between hospital growth and statewide average growth. This change will prevent hospitals from benefitting from their absolute benchmarking position, given that the absolute position is subject to some range of error. Staff do not believe any adjustment to the integrated efficiency policy is necessary because this policy relies only on relative ranking. Staff expect to bring this revision to the Commission for formal consideration during 2022.

2. HSCRC staff have begun a process with the Maryland Healthcare Commission and CareFirst to reconcile the data included for CareFirst in the benchmarking ties to CareFirst's own calculations. Although a similar process with other payers is not feasible, CareFirst represents a significant proportion of the data set, and the assurance provided by this reconciliation would increase confidence in the data received from other payers.

To: Commission Stakeholders
From: William Henderson, Principal Deputy Director- Medical Economics and Data Analysis
Date: 1/26/2022
Page: 9

3. As a formal part of benchmarking approach, HSCRC staff plan to include a periodic comprehensive review of their peer-selection and adjustment approach. Although continually changing the peer grouping and adjustment would create unnecessary instability and complexity, staff believe it is important to conduct a periodic reassessment to allow for evaluation of new data, incorporation of new use cases, and further input from the industry. Staff are proposing to complete this reassessment every five years.

Appendix A: County Health Rankings Comparison

Table A.1 shows states ranked from least to most challenged according to the social and economic dimensions of the County Health Rankings. The Maryland column indicates the Maryland county rank among the Maryland counties. The Benchmark Counties column is calculated as follows: (1) Average the absolute score of the selected peer counties for each Maryland county to generate a peer county average score for each Maryland county and (2) rank those average scores among peer averages.[12] Note this comparison only considers if the selected peers have similar relative rankings to the Maryland counties it does not compare absolute values. The second step of our benchmarking methodology adjusts for remaining differences in socio-economic conditions after ensuring we match Maryland counties to their best possible peers.

Table A.1. Comparison of County Health Ranking ranks for social and economic dimensions

County	Average rank	
	Maryland	**Benchmark Counties**
Howard	1	1
Carroll	2	2
Montgomery	3	6
Frederick	4	7
Calvert	5	4
Queen Anne's	6	5
Anne Arundel	7	8
St. Mary's	8	11
Harford	9	3
Charles	10	9
Talbot	11	10
Baltimore	12	14
Prince George's	13	16
Garrett	14	15
Washington	15	17
Kent	16	18
Cecil	17	13
Caroline	18	20
Allegany	19	22
Wicomico	20	19
Worcester	21	12

---

[12] For example, Howard County ranks least challenged county in Maryland. The peer county average for Howard County is also ranked the least challenged among the peer county averages across the peer groups selected for benchmarking.

County	Average rank	
	Maryland	Benchmark Counties
Dorchester	22	21
Somerset	23	24
Baltimore City	24	23

Appendix B: Description of Additional Analytics

Medicare benchmarking relies on a two-step process: (1) selecting similar peer counties based on matching demographic characteristics and (2) using a regression to adjust for remaining differences between counties. To evaluate both the level of variation resulting from selecting different metrics for the peer selection and for the regression, HSCRC consultants reran the benchmarking process 20 times using different sets of metrics. This process involved three steps: (1) select reasonable alternative metric sets for the peer selection and regression given the purpose of the benchmarking and analytical principles (e.g. no inclusion of health care market factors); (2) rerun the benchmarking process with the alternative metric sets, and compare the correlation between them and the selected approach; and (3) compare the result between tests, which produced the least-correlation with selected method, and the selected benchmarking set to evaluate the degree of variations.

Step 1: Select alternative reasonable metric sets for the peer selection and regression steps

Based on feedback from stakeholders and discussions with the consultants, HSCRC staff selected alternative metrics that could have been used in the benchmark approach. These alternatives were selected to be consistent with the analytical principles described in Section II, as well as other limitations such as data availability at the county level. We created 10 alternative metrics (test sets) with varying levels of overlap with the actual metric set selected (control set). First group of tests changed the metrics used in the regression while keeping the peer selection same as the control set. The second group of rests changed the metrics used in the peer selection, with the control set used in the regression. Owing to the many possible combinations, we did not combine test groups used in selection and regression together, after concluding that the initial results of the variation analysis did not diverge substantially from the results for the control sets.

Table B.1 shows the 10 variants used; the same variants were used in Test Groups 1 and 2. All other aspects of the process remained unchanged in all tests: the statistical method used, the direct adjustment for HCC, and the adjustment for medical education. For analytical simplicity, this process was conducted at the county level. Hospital service area results are used in most HSCRC policies, but these results are directly derived from the county results (a hospital's benchmark is a blend of the results from counties in its service area plus the regression adjustment). Accordingly, HSCRC staff believe the county-level process yields the same results as the process done at the level of hospital service area.

The sets in Table B.1 fall into three groups. Test Set 1 adds race, and Test Set 5 looks at metrics that cover a broader range of statistics. Test Sets 3 and 6 look at supplementing or replacing the deep poverty metric with the share of dual-eligible beneficiaries in the county, and Test Sets 2, 4, 7, 8, 9 and 10 look at supplementing or replacing the median income metric with various measures of wages. As discussed in Section III, Item 1, the Medicare Hospital Wage Index was not tested because HSCRC staff do not believe it is consistent with the analytical principles described in Section II.

Table B.1. Metrics used by test set (shaded rows = control variables; full definitions can be found in Appendix C)

Metric[13]	Test Set Number										
	Control	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
% 65 and older in 2016						x					
Diabetes prevalence in 2015						x					
% Non-Hispanic African American in 2016		x				x					
RPP	x*	x*	x*	x*	x*		x*	x*	x*	x*	x*
Median household income in 2016	x	x	x	x		x	x	x	x		
% persons in deep poverty in 2011–15	x	x	x	x	x			x	x	x	x
% of Medicare beneficiaries eligible for Medicaid in 2016				x			x				
Bureau of Labor Statistics (BLS) wage in 2016 (all industry, all ownership)								x		x	
BLS wage in 2016 (all industry, private ownership)			x		x						
BLS wage in 2016 (ambulatory health care service, private ownership)									x		x

*RPP is used only in peer selection, not in the regression.

Step 2: Rerun the benchmarking process with the alternative metric sets, and compare the correlation between them and the final approach used (control)

Table B.2 shows the correlation between each test set and the control measured on the rank (from 1 to 24) of the county-level percent variation from benchmark. We used spearman correlation to measure the correlation of the ranks estimated using selected approach vs. the ranks estimated using a given test.

---

[13] HSCRC staff used the 2016 statistics (or data from the most similar period available) to match the data used in the control peer selection. Typically, more recent data are used to conduct regression adjustment. But because the peer selection is held constant, additional modeling relies on the data used for the initial selection time frames. For this evaluation, staff determined that the most important factor was consistency between control and test groups and peer selection and regression, so we used older data from the final peer-selection period.

Table B.2. Spearman correlation coefficients between control and each test set

Description	Test Set Number									
	#1	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**	**#9**	**#10**
Test Group 1, variation in variables used in regression	0.95	0.95	0.96	0.94	0.93	0.96	0.96	0.95	0.95	0.94
Test Group 2, variation in matching variables used in peer selection	0.93	0.95	0.97	0.96	0.92	0.98	0.94	0.95	0.93	0.87

The lowest correlation on any of the 20 sets is 0.87, which still reflects a high degree of correlation. Most alternatives have correlations of 0.95 or higher. This shows that the approach selected by staff yields substantially similar rankings to a variety of reasonable alternatives.

Staff also reviewed the Maryland's statewide average cost variation from the benchmark and the R^2 produced for each model. In all cases, Maryland's average variation above the nation was within 1.5% points of the result from the control. The adjusted R^2 ranged from 0.05 to 0.19, meaning no version explained significantly more or less variation than the control (R^2= .13).

Step 3: Compare the result of the least-correlated tests and the final approach used (control)

Tests which produced the lowest correlation with the final approached used are the ones produced the most variation in results compared with the final approach. By focusing on the tests that have the lowest correlation with the control, we can evaluate the worst-case sensitivity to different metrics being used to select peer counties or calculate the regression adjustment versus the selected approach. As shown in Table B.2, the two least-correlated versions are Tests 5 and 10 in Test Group 2. Test 5 has the biggest shift in metrics used, whereas Test 10 substitutes an ambulatory health care wage index for median income. These least-correlated tests are in Test Group 2, meaning the test change is made during peer-group selection rather than changing the factors used for the regression.

To understand the impact of the least-correlated tests on the results at the county level, HSCRC staff calculated the simple average of the absolute value of:

- The change in ranking within Maryland of each Maryland county on the percent variation from benchmark between the control and each least-correlated test result
- The change in percent difference of the Maryland counties from benchmark between the control and each least-correlated test result

Staff used the absolute value so changes in a different direction would not offset each other once we calculate an average change between model results.

The average county moved 2.1 and 2.8 ranking slots, in Test 5 and Test 10 respectively. The largest changes in ranking are nearly all with small-population counties (such as Dorchester, Kent, and Talbot), where both the Maryland and national values are likely to be less stable and where the influence on any one hospital's results is smaller. Overall, staff believe these variations show that changing metrics within the benchmarking approach does not have a material impact on relative rank of a particular hospital,

To: Commission Stakeholders
From: William Henderson, Principal Deputy Director- Medical Economics and Data Analysis
Date: 1/26/2022
Page: 15

given that the least-correlated tests, including outlier results for small counties, generate only an average shift of two or three places in ranking. HSCRC methodologies often rely on quartile or quintile groups, for which this kind of move has no impact unless an entity is on the border of cut points for quartile or quintile groups.

The average county's *absolute* change in the difference from the benchmark is 2.9 percentage points in Test 5 and 3.5 percentage points in Test 10. These "worst-case" changes are more material than the movements in ranking discussed above , given a statewide average difference from the benchmark of about 8 percent. However, the fact that the change in the relative rankings is benign indicates that much of the variation is simply all of Maryland moving in concert in relation to the benchmark, resulting in worse or better performance for all Maryland counties but similar positions relative to each other (which is how the benchmarks results will affect the Commission's policies).

Another approach to evaluating the significance of the changes is to consider whether they systematically change the results of different groupings of counties. For example, if all rural counties improve their results relative to the control, the test or the control might have a bias. If there is no such pattern, it is more likely that the change in results is caused by random selection within each approach rather than by any inherent bias.

To evaluate potential bias, HSCRC staff ran the same calculation as above for two sets of subgroups:

- Three density subgroups: urban and suburban, exurban, and rural
- Five regional subgroups: Baltimore area, northern DC suburbs, southern Maryland, western Maryland, and eastern Maryland

In this case, staff did not use the absolute value because it was important to maintain the directionality of the change[14] in order to evaluate biased effects on specific subgroups.

In nearly all subgroups in both least-correlated tests, the direction of the change in percent variation matched the overall result for the state, suggesting all subgroups of counties were moving in generally the same way. In cases in which a group moved in a direction different from the state, the absolute size of the movement was minute.

In the least-correlated tests, the largest change in any single subgroup occurred in rural counties in Test 10, which showed a 3.6 percent increase in the difference from benchmark and a 1.9 average worsening in rank. Although the degradation versus the benchmark is somewhat significant, once again, the change in average rank is not significant. So even when considering the largest change in a subgroup in the test that is least-correlated with the selected approach, staff still found only a small change in rank.

---

[14] In measuring the instability of the rankings, HSCRC staff cannot consider the direction of changes in rankings, or the average change will be 0; therefore, the absolute value of the changes must be used. For example, if County A moves from Rank 6 to 1 and County B moves from Rank 7 to 12, that is a change of five slots in both cases. If staff took directionality into account, the average change would be 0. But in measuring bias, staff must consider directionality because the goal is to evaluate whether a subgroup of counties all moved in the same direction, which would indicate that different models produce a bias against certain groups of counties. In this case, the subgroup containing Counties A and B shows no bias because they moved equally in opposite directions.

Appendix C: Full Definition of Metrics Used in Sensitivity Tests and Data Sources

Table C.1 shows the metrics used in additional analysis and the source of the data.

Table C.1. Metric definitions and sources

Metric	Full definition	Data source
Average HCC score in 2016	Average Hierarchical Condition Category score in 2016	2016 CMS Geographic Variation Public Use File[15]
% 65 and older in 2016	Percentage of population ages 65 years and older in 2016	2016/17 Area Health Resource file[16]
Diabetes prevalence in 2015	Percentage of adults ages 20 and above with diagnosed diabetes in 2015	County Health Rankings (2018 file)
% Non-Hispanic African American in 2016	Percentage of non-Hispanic African Americans in 2016	2016/17 Area Health Resource file
RPP	Regional price parities	Bureau of Economic Analysis[17]
Median household income in 2016	Median household income in 2016	2016/17 Area Health Resource file
% persons in deep poverty in 2011–15	Percentage of people in deep poverty in 2011–2015	
% of Medicare beneficiaries eligible for Medicaid in 2016	Percentage of Medicare beneficiaries eligible for Medicaid in 2016	2016 CMS Geographic Variation Public Use File
Bureau of Labor Statistics (BLS wage in 2016 (all industry, all ownership)	Annual wage for all industry (all ownership) in 2016	Bureau of Labor Statistics, Quarterly Census of Employment and Wages (QCEQ) North American Industry Classification System (NAICS) -based data files[18]
BLS wage in 2016 (all industry, private ownership)	Annual wage for all industry (private ownership) in 2016	
BLS wage in 2016 (ambulatory health care service, private ownership)	Annual wage for ambulatory health care service (private ownership) in 2016	

---

[15] https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html (State_county_2016 table)

[16] https://datawarehouse.hrsa.gov/data/datadownload.aspx#MainContent_ctl00_gvDD_lbl_dd_topic_ttl_0

[17] https://www.bea.gov/itable

[18] https://www.bls.gov/cew/downloadable-data-files.htm