

METHODS BRIEF

Improving target price calculations in Medicare bundled payment programs

Benjamin A. Y. Cher MS¹ | Baris Gulseren MS^{2,3} | Andrew M. Ryan PhD^{2,3} ¹University of Michigan Medical School, Ann Arbor, Michigan, USA²University of Michigan School of Public Health, Ann Arbor, Michigan, USA³Center for Evaluating Health Reform, Ann Arbor, Michigan, USA**Correspondence**Andrew M. Ryan M3124 SPHII 1415
Washington Heights Ann Arbor, MI 48109.
Email: amryan@umich.edu**Abstract**

Objective: To compare the predictive accuracy of two approaches to target price calculations under Bundled Payments for Care Improvement-Advanced (BPCI-A): the traditional Centers for Medicare and Medicaid Services (CMS) methodology and an empirical Bayes approach designed to mitigate the effects of regression to the mean.

Data sources: Medicare fee-for-service claims for beneficiaries discharged from acute care hospitals between 2010 and 2016.

Study design: We used data from a baseline period (discharges between January 1, 2010 and September 30, 2013) to predict spending in a performance period (discharges between October 1, 2015 and June 30, 2016). For 23 clinical episode types in BPCI-A, we compared the average prediction error across hospitals associated with each statistical approach. We also calculated an average across all clinical episode types and explored differences by hospital size.

Data collection/extraction methods: We used a 20% sample of Medicare claims, excluding hospitals and episode types with small numbers of observations.

Principal findings: The empirical Bayes approach resulted in significantly more accurate episode spending predictions for 19 of 23 clinical episode types. Across all episode types, prediction error averaged \$8456 for the CMS approach versus \$7521 for the empirical Bayes approach. Greater improvements in accuracy were observed with increasing hospital size.

Conclusions: CMS should consider using empirical Bayes methods to calculate target prices for BPCI-A.

KEYWORDS

Bayesian shrinkage, bundled payments, health policy, regression to the mean, spending predictions, target prices

What is known on this topic

- The U.S. Centers for Medicare and Medicaid Services (CMS) implemented the voluntary Bundled Payments for Care Improvement-Advanced (BPCI-A) program in 2018.
- Prior work demonstrates that target price calculations used by BPCI-A do not account for regression to the mean over time in hospital spending.
- BPCI-A may lead to undue financial losses for CMS because hospitals are more likely to join the program if they are offered higher target prices—but hospitals offered higher target prices are more likely to experience decreases in spending, and therefore, achieve shared savings due to statistical artifact.

What this study adds

- Empirical Bayes estimation, which accounts for regression to the mean, can be used to predict hospital spending and set BPCI-A target prices.
- When applied to BPCI-A, empirical Bayes estimation improved target price accuracy for the majority of BPCI-A clinical episode types, and calculated target prices were generally lower.
- CMS should consider using empirical Bayes estimation to set BPCI-A target prices.

1 | INTRODUCTION

The Centers for Medicare and Medicaid Services (CMS) implemented the voluntary Bundled Payments for Care Improvement-Advanced (BPCI-A) program in 2018.¹ Bundled payment models seek to reduce spending by making providers responsible for spending that occurs throughout a predefined clinical episode.² For 29 inpatient clinical episode types, CMS defines target prices for each participating hospital for a particular measurement period. If hospital spending in the performance period is below the target price, a hospital earns shared savings. However, spending above the target price leads to penalties. Target prices are calculated for a particular hospital by applying a discount to that hospital's predicted spending for a particular episode.³ Predicted spending is based on risk-adjusted spending during prior years and peer-group spending trends. For BPCI-A to function appropriately, target prices should achieve a balance between incentivizing spending reductions and encouraging program participation. The ability for CMS to save money in voluntary programs like BPCI-A stems almost entirely from setting an appropriate target price.

However, the best way to set target prices under bundled payment is unknown. Predicting provider spending, while necessary for alternative payment models, is challenging.⁴⁻⁶ Hospital spending is susceptible to a statistical phenomenon known as regression to the mean, where hospital spending that is unusually high in a particular year is likely to decrease in the following years, and hospital spending that is unusually low in a particular year is likely to increase in the following years.⁶ In essence, random noise can obscure policymakers' ability to observe hospitals' true spending performance. Evaluating hospitals' expected spending trends, and incorporating them into predictions, is another challenge. Inaccurate predictions may lead to CMS failing to reward some deserving hospitals and rewarding some undeserving hospitals. Inaccurate predictions may also discourage program participation. Setting target prices that more accurately predict hospital spending has the potential to more appropriately balance incentives in BPCI-A.

In this context, we developed an alternative methodology to calculate target prices under BPCI-A. Specifically, we used empirical Bayes estimation to mitigate the effects of regression to the mean. Empirical Bayes estimation addresses regression to the mean by "shrinking" predictions of spending for any particular hospital to average spending across other similar hospitals.⁴ Using national Medicare data, we calculated target prices using the standard CMS approach and our alternative approach. We then compared the predictive

accuracy of target prices calculated using the standard CMS approach and our alternative method.

2 | METHODS**2.1 | Data source and definitions**

We used inpatient and outpatient physician claims and 20% Medicare Provider Analysis and Review files for patients discharged from acute care hospitals. We also used Provider of Service (POS), Academic Medical Center (AMC) list, Provider Specific Files (PSF), and American Hospital Association Annual Survey (AHA) for hospital characteristics.

For each inpatient clinical episode, BPCI-A determines target prices for a single year based on hospital performance during a prior period spanning multiple years. We mirrored this approach using index admissions between January 1, 2010 and September 30, 2013 to define a baseline period and index admissions between October 1, 2015 and June 30, 2016 to define a performance period. We evaluated these baseline and performance periods because they preceded the announcement of BPCI-A. As a result, our assessment of the accuracy of alternative approaches to set target prices would not be affected by hospitals' attempts to reduce episode spending as a result of the program. Toward this end, we also excluded hospitals that participated in the same episode under the BPCI program.

Following CMS methodology, we excluded hospitals with fewer than 40 cases during the baseline period for each clinical episode. This resulted in the exclusion of one clinical episode. We also excluded clinical episodes for which fewer than 20 hospitals met the case requirement, resulting in the exclusion of five clinical episodes.

Data on hospital characteristics came from the American Hospital Association Data Annual Survey between 2010 and 2013.

2.2 | Target price calculation using current CMS approach

We calculated target prices for each clinical episode using the current CMS approach. CMS calculates a benchmark price, which incorporates observed spending, expected spending based on case mix, and peer-group spending trends. Then, benchmark prices are converted to target prices using a formula that incorporates a 3% discount. The formula accounts for inflation; results are reported in 2013 dollars. The CMS approach is described in detail in Figure S1.

2.3 | Target price calculation using empirical Bayes estimation

We also calculated target prices for each clinical episode using empirical Bayes estimation. This approach derives two separate appraisals of hospitals' episode spending: (a) one that is determined by a hospital's own risk-adjusted spending in the baseline period; and (b) another that is a hospital's expected spending, estimated by the hospital's characteristics. Throughout this paper, we refer to appraisal (a) as "historical spending" and appraisal (b) as "expected spending."

A weight, based on the reliability of a hospital's risk-adjusted baseline (appraisal 1), is then derived and applied to each appraisal of spending. Generally, reliability increases as hospital case volume increases. If risk-adjusted spending is highly reliable, it will receive much of the weight. This approach was developed to profile hospital quality,⁷ has been shown to have greater predictive accuracy than other common approaches to measure quality,⁸⁻¹¹ and is used by agencies such as Leapfrog for quality reporting.¹² The formula for the weights is described in detail in the technical supplementary information. Essentially, weights are derived from a ratio of signal to noise. When hospital spending predictions are more reliable, they receive more weight.

To implement the empirical Bayes approach, we first employed random forest machine learning estimation to select independent variables to predict hospital spending. The goal of this approach was to develop the best possible predictive model of hospital spending during the performance period. Important weights of variables in our model are presented in Figure S2. These variables were then used to estimate linear models for each clinical episode. In contrast to the traditional CMS methodology, we incorporated peer-group spending trends as simply another factor that could predict future spending. The two separate appraisals of hospital episode spending were then developed and combined using the derived reliability weights. How the empirical Bayes approach affects target prices can be seen in Figure S3, where the median estimate is lower and extreme values are shrunk toward the mean. Further description of the methodology is provided in the technical supplementary information and Figure S1.

2.4 | Statistical analyses

Our analysis sought to compare the predictive accuracy of the CMS and empirical Bayes approaches. For each clinical episode type, at each hospital, we calculated the risk-adjusted spending in the performance period. This was our "gold standard"—the value that target prices sought to estimate. We then calculated the mean absolute prediction error, defined as the difference between risk-adjusted spending in the performance period and target prices. Mean absolute prediction error was calculated using both the CMS and empirical Bayes approaches. We compared the mean absolute prediction error between these approaches across all hospitals. We then conducted a sensitivity analysis where we evaluated hospitals separately by size,

categorized as follows: small (0-250 beds), medium (251-500 beds), large (501-850 beds), and extra-large (>850 beds).

We then created a measure of overall performance to compare the CMS and empirical Bayes estimators across all clinical episodes by calculating the unweighted mean absolute prediction error across all 23 episodes. We recalculated this value for 1000 bootstrap resamples of the data and compared the bootstrap distribution between the CMS and empirical Bayes approach. We then repeated this approach separately by hospital size, categorized as above. Standard errors were clustered by hospital.

Our empirical Bayes estimation differed in how it incorporated peer-group spending trends into target price calculations (Figure S1). To understand the extent to which changes in predictive accuracy were due to shrinkage itself versus the modifications to how peer-group trends were incorporated into the model, we conducted additional sensitivity analyses (Figure S4). First, we used the traditional CMS methodology with the peer-adjusted trend factor removed from the calculation (Sensitivity Analysis A). Second, we left the "peer-adjusted trend" as-is and excluded peer-group spending trends from the calculation of expected spending used by the empirical Bayes estimator (Sensitivity Analysis B). Third, we excluded all information about peer-group spending trends (Sensitivity Analysis C).

Because some hospitals may use more recent data to inform their decisions related to alternative payment models, we conducted a sensitivity analysis where we extended the baseline period until December 31, 2014. To examine possible distributional effects related to the accuracy of target price predictions, supplemental analysis also examined differences in the accuracy of target prices across hospital size, teaching status, profit status, urban versus rural, and region.

All *P* values were two-sided, and $\alpha = 0.05$ was set as the threshold for significance. Analyses were performed using Stata version 16.0 (Stata Corp, College Station, TX).

3 | RESULTS

The study sample included 2589 hospitals across 23 BPCI-A clinical episodes. During the baseline period (2010-2013), there were 1 837 861 clinical episodes with average spending of \$20 039 per hospital-episode (Table S1).

Allocation of weight between hospitals' historical spending versus expected spending was similar across episodes included in BPCI-A (Table S2). For 22 of 23 episodes, between 28% and 33% of the weight was applied to hospitals' historical spending. For acute myocardial infarction, 45% of the weight for the empirical Bayes approach was based on the historical spending, and 55% was based on expected spending.

The empirical Bayes approach had a lower mean target price for all 23 clinical episode types (Table 1). For cardiac valve, there was a very large difference in mean target price—mean target price was \$11 716 higher under the traditional CMS methodology than under the empirical Bayes methodology. For the remaining clinical episodes,

TABLE 1 Target price, mean absolute prediction error, and percent error comparing traditional CMS methodology and empirical Bayes methodology, for all clinical episode types

	Traditional CMS methodology			Empirical Bayes methodology		
	Mean target price (\$)	Mean absolute prediction error (\$)	Mean absolute prediction error (%)	Mean target price (\$)	Mean absolute prediction error (\$)	Mean absolute prediction error (%)
BPCI-A episode						
Cardiac valve	65 548.3	19 870.6	30.3	50 654.7	8154.9	16.1
Cardiac defibrillator	50 770.2	15 716.5	31.0	37 706.9	14 454.0	38.3
Coronary artery bypass graft surgery	44 005.8	11 756.2	26.7	37 936.4	8999.7	23.7
Spinal fusion (non-Cervical)	38 009.2	9963.9	26.2	31 347.4	7491.8	23.9
Hip and femur procedures except major joint	35 749.5	9266.1	25.9	32 675.9	8503.2	26.0
Major bowel procedure	34 506.3	12 328.4	35.7	28 861.9	9749.9	33.8
Sepsis	28 812.0	8951.2	31.1	23 858.5	7199.4	30.2
Lower extremity and humerus procedure except hip, foot, femur	28 694.0	8531.3	29.7	24 285.8	6907.9	28.4
Stroke	26 588.7	8879.3	33.4	23 169.0	7844.5	33.9
Pacemaker	26 116.1	9239.2	35.4	21 481.4	8398.4	39.1
Cervical spinal fusion	26 046.7	8271.7	31.8	22 202.2	7358.3	33.1
Major joint replacement of the lower extremity	24 707.5	6940.3	28.1	21 795.9	5991.7	27.5
Acute myocardial infarction	23 415.3	9322.4	39.8	20 042.7	8417.9	42.0
Percutaneous coronary intervention	22 746.0	7267.2	31.9	18 839.7	6866.3	36.4
Renal failure	21 906.4	8000.1	36.5	18 513.6	7300.9	39.4
Congestive heart failure	21 582.6	8208.0	38.0	18 256.4	7646.3	41.9
Simple pneumonia and respiratory infections	19 586.9	8504.0	43.4	16 971.5	7892.0	46.5
Gastrointestinal hemorrhage	18 103.5	8177.5	45.2	15 155.5	7601.3	50.2
Cellulitis	17 892.6	9309.8	52.0	15 351.7	8966.4	58.4
Urinary tract infection	17 717.0	7806.7	44.1	15 537.6	7463.5	48.0
Chronic obstructive pulmonary disease, bronchitis/asthma	17 102.5	7827.9	45.8	14 542.8	7282.2	50.1
Gastrointestinal obstruction	15 810.1	7591.9	48.0	13 325.5	6826.2	51.2
Cardiac arrhythmia	15 371.7	7046.5	45.8	12 893.9	6634.3	51.5

the difference in mean target price ranged from \$343 for urinary tract infection to \$2757 for coronary artery bypass graft surgery.

The empirical Bayes approach had significantly lower mean absolute prediction error than the CMS approach for 19 out of 23 clinical episodes (Table 1 and Figure 1). The largest improvement was for cardiac valve ($\Delta = \$11716$). For coronary artery bypass graft surgery ($\Delta = \$2757$), major bowel procedure ($\Delta = \$2579$), spinal fusion ($\Delta = \2472), and sepsis ($\Delta = \$1752$), the empirical Bayes estimator outperformed the CMS estimator by a wide margin. For four clinical episode types (lower extremity and humerus procedure, cardiac defibrillator, cervical spinal fusion, and cellulitis), there was no significant difference in mean absolute prediction error between both approaches. The fact that target prices were generally both lower and more accurate under the empirical Bayes methodology suggests that the CMS methodology was systematically over-predicting spending during the performance period.

In sensitivity analysis by hospital size, we observed similar results for hospitals of all sizes (Figure S5). In sensitivity analysis including the year 2014, results did not differ substantially, and absolute and relative prediction errors were relatively similar (Figure S6).

Across all clinical episodes, mean absolute prediction error was \$7521 for the empirical Bayes approach versus \$8456 for the CMS approach ($P < .001$, Figure 2). There was not a single bootstrap iteration in which the CMS approach outperformed the empirical Bayes approach. For all four hospital size categories, mean absolute prediction error was higher when using the CMS estimator than when using the empirical Bayes approach ($P < .001$ for all categories, Figure S7).

The traditional CMS methodology resulted in higher prediction error for large hospitals than small hospitals; mean absolute prediction error was \$9042 for large hospitals versus \$8437 for small hospitals (Figure 3). The empirical Bayes methodology improved prediction accuracy for all hospital size categories. There were greater

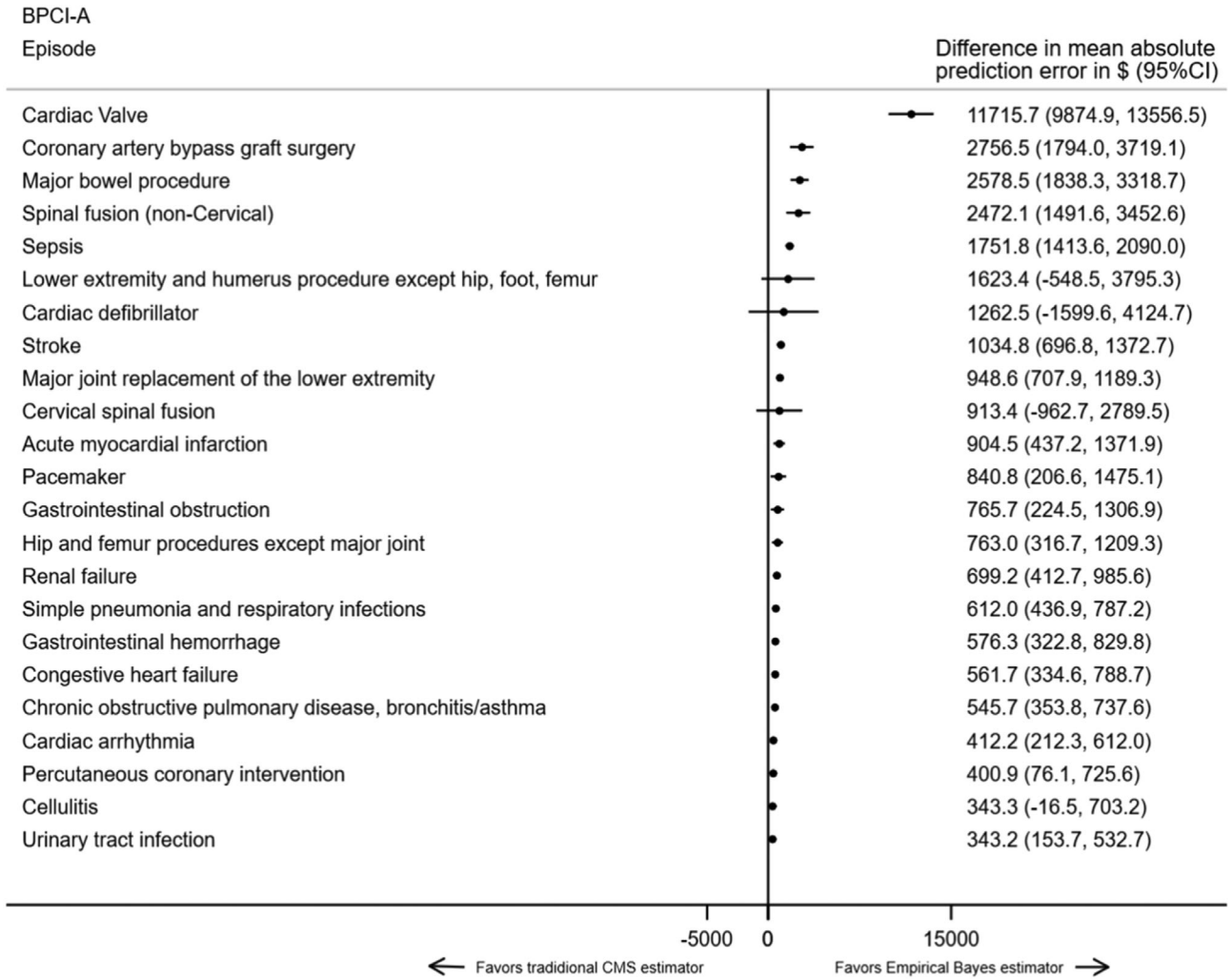
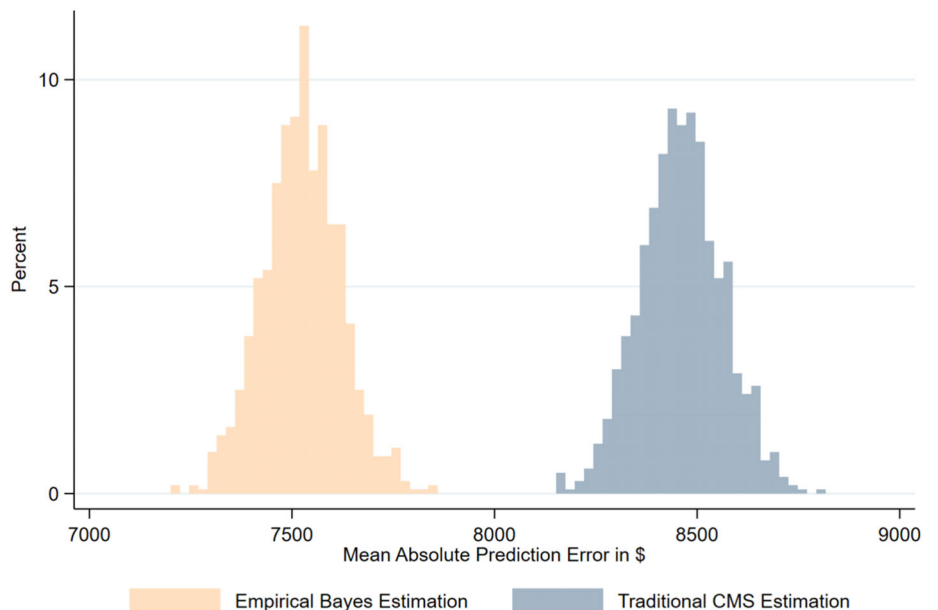


FIGURE 1 Difference in prediction error between traditional CMS methodology and empirical Bayes estimation, for all clinical episode types

FIGURE 2 Mean prediction error for all hospitals, averaged across all clinical episodes [Color figure can be viewed at wileyonlinelibrary.com]



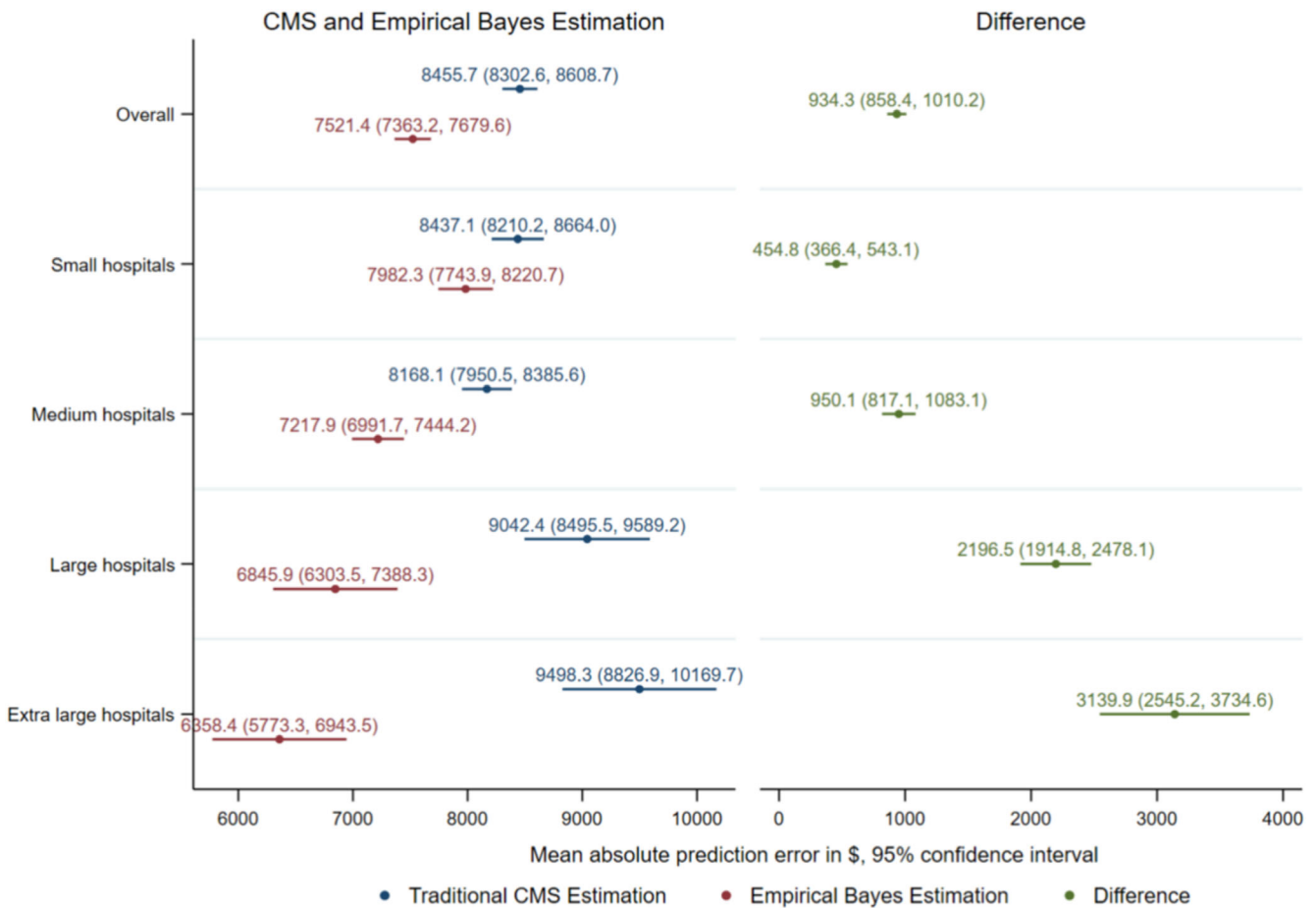


FIGURE 3 Mean prediction error across all clinical episodes, by hospital size, using traditional CMS estimation and empirical Bayes estimation [Color figure can be viewed at wileyonlinelibrary.com]

improvements for large hospitals than for small hospitals, so that compared with the traditional CMS methodology, the relationship between hospital size and prediction error was reversed. Using empirical Bayes estimation, prediction error was higher for small hospitals than for large hospitals; mean absolute prediction error was \$7982 for small hospitals versus \$6846 for large hospitals. Lastly, improvements in accuracy for larger hospitals were generally higher for surgical episodes than medical episodes. Five of the six episodes with greatest improvements in prediction accuracy were surgical episodes. Hospital size was the only hospital characteristic for which the accuracy of target prices varied substantially between the traditional CMS methodology and the empirical Bayes methodology (Table S3).

Decreases in mean absolute prediction error were due to the shrinkage aspect of the empirical Bayes model to a greater extent than modifications of how the peer-adjusted trend factor was incorporated into the predictive methodology. When the peer-adjusted trend factor was removed from the traditional CMS methodology (Sensitivity Analysis A), mean error did not decrease substantially (\$8470 for Sensitivity Analysis A versus \$8456 for traditional CMS methodology). When the peer-adjusted trend factor was left as-is and peer-group trends were excluded from the calculation of expected spending used by the empirical Bayes estimator (Sensitivity Analysis B), mean absolute prediction error decreased

substantially and was similar to the empirical Bayes estimator used in the primary analysis (\$7681 for Sensitivity Analysis B vs \$7684 for the primary empirical Bayes analysis). When all information about peer-group spending trends was excluded, mean prediction error was similar \$7686, similar to Sensitivity Analysis B and the primary empirical Bayes analysis.

4 | DISCUSSION

In this national study comparing the accuracy of target prices for BPCI-A between the current CMS approach and a modified approach using empirical Bayes estimation, we report three main findings. First, there was substantial prediction error in BPCI-A target prices calculated using the traditional CMS methodology, and target prices were generally too high. Second, the empirical Bayes estimator statistically outperformed the CMS estimator for 19 of 23 clinical episodes. Performance was not statistically different for the remaining four episodes, and there were no episodes where the CMS estimator outperformed the empirical Bayes estimator. Third, the empirical Bayes estimator outperformed the CMS approach for hospitals of all sizes, and improvements were greatest for larger hospitals. Together, these findings suggest an empirical Bayes approach could improve the

ability of BPCI-A to set accurate target prices that balance incentivizing spending reductions with encouraging program participation.

Our results are consistent with other research showing the benefits of empirical Bayes estimation for profiling hospital spending¹³ and quality outcomes.¹⁴⁻¹⁶ However, ours is the first to apply empirical Bayes estimation to the problem of setting target prices under BPCI-A. We also provide insight into where improvements in the predictive accuracy of target prices are most likely to be observed. We found greatest improvements for larger hospitals, who are more likely to participate in voluntary bundled payment programs than smaller hospitals.¹⁷ We still found improved spending predictions for smaller hospitals, whose spending is more susceptible to regression to the mean. Improvements were generally larger for surgical conditions, which are more susceptible to influence by bundled payment programs¹⁸ than medical conditions.

CMS should consider incorporating empirical Bayes estimation into target price setting for BPCI-A. This may be especially helpful for particular episode types, such as cardiac valve and coronary artery bypass grafting, where we observed the highest improvements in predictive accuracy when employing empirical Bayes estimation. There is a precedent for using empirical Bayes estimation in other CMS incentive programs, including the construction of the PSI-90 for the Hospital Acquired Conditions Reduction Program.^{19,20} Both the Hospital Readmission Reductions Program²¹ and Hospital Compare²² use Bayesian Shrinkage to profile hospital readmission and mortality rates. The primary advantage of using the empirical Bayes approach for BPCI-A is that it addresses the issue that hospitals with high target prices may join the program and experience unwarranted financial gains through regression to the mean.⁶ More accurate target prices could also address issues such as low participation rates,^{23,24} high drop-out rates,^{23,24} inequitable distribution of risk-sharing,²⁵ and substantial differences in hospital characteristics between participants and non-participants.^{26,27} Savings associated with BPCI-A have been modest^{1,28} in prior years; lower target prices resulting from empirical Bayes estimation would further encourage hospitals to lower spending and achieve shared savings with CMS. Lastly, our finding that current target prices are too high suggests that CMS may be losing money both because hospitals are more likely to join the program if they are offered higher target prices and because CMS is paying unnecessarily high target prices to hospitals who are already participating in the program. In addition, even if BPCI-A participation were made mandatory—a policy solution suggested by many researchers²⁴—the program would continue to result in financial loss for CMS if there are no substantial changes in the target price formula. Of note, while our analysis suggests how the accuracy of spending predictions may be improved, an additional policy question is whether 3% is the appropriate discount factor between the benchmark price and target price. Further research can explore the implications of different discount rates for hospital behavior and reconciliation payments under bundled payment programs.

The empirical Bayes approach may have disadvantages. Shrinkage may reduce incentives for small hospitals to change behavior, since target prices are less dependent on their own spending.²⁹ In addition,

empirical Bayes estimation is limited by the ability of hospital characteristics to explain spending. Contrary to other applications of empirical Bayes estimation,¹⁰ such as profiling hospital mortality, we found greater improvements in accuracy for larger hospitals than for smaller hospitals. This was likely because of stronger relationships between hospital characteristics and spending for larger hospitals than for smaller hospitals. Even though the empirical Bayes estimator was designed to help smaller hospitals specifically, there was more room for improvement in spending predictions for larger hospitals than for smaller hospitals. Nevertheless, the substantial errors observed for our application of empirical Bayes estimation suggests that hospital spending predictions could be improved further, enhancing target prices set under BPCI-A and other alternative payment programs.

Our study had limitations. First, we used a 20% sample of Medicare claims rather than the 100% sample used by CMS to determine target prices. However, the 100% sample is only available to researchers working under contract for CMS. In addition, sensitivity analysis found that the empirical Bayes approach outperformed the CMS approach for all hospital size categories, suggesting that it would similarly outperform the CMS approach when using 100% files. Second, we used data between 2010 and 2016, which are older than the data that will be used for BPCI-A, and hospitals may have changed their clinical operations between the baseline and performance period because of the influence of other value-based purchasing programs. To address this, we excluded hospitals that participated in similar clinical episodes in BPCI, the precursor program to BPCI-A. Additional limitations derive from minor differences in our replication of the CMS approach to calculating target prices. For instance, we used generalized linear models instead of compound lognormal regression. We also did not include spending on home health and durable medical equipment, which are a small component of episode spending.³⁰ These minor differences are unlikely to materially affect our conclusions. Finally, we were not able to observe the “true spending” of hospitals, instead relying on the ability of alternative estimators to predict future spending as a proxy for relative accuracy. While imperfect, this strategy allowed us to examine estimator accuracy using actual data (rather than simulated data) under the plausible assumption that an estimator that is better able to predict observed future spending provides a more accurate estimate of true spending, which is unobserved.

5 | CONCLUSIONS

Effective alternative payment programs depend on the ability of program sponsors to set accurate and appropriate targets for quality and spending. Empirical Bayes estimation has the potential to enhance BPCI-A by improving target price setting under the program.

ORCID

Andrew M. Ryan  <https://orcid.org/0000-0002-2566-7763>

REFERENCES

- Centers for Medicare & Medicaid Services. BPCI Advanced: Participant Resources; 2020. Available from <https://innovation.cms.gov/initiatives/bpci-advanced/participant-resources.html>.
- Agarwal R, Liao JM, Gupta A, Navathe AS. The impact of bundled payment on health care spending, utilization, and quality: a systematic review. *Health Aff.* 2020;39(1):50-57.
- Centers for Medicare & Medicaid Services. BPCI Advanced Target Price Specifications—Model Years 1 & 2; 2018. Available from <https://innovation.cms.gov/Files/x/bpciadvanced-targetprice-my1-2.pdf>.
- Ash AS, Fienberg SE, Louis TA, Normand SLT, Stukel TA, Utts J. Statistical issues in assessing hospital performance; 2011. Available from <https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/downloads/statistical-issues-in-assessing-hospital-performance.pdf>.
- DeLia D. Monte Carlo analysis of payer and provider risks in shared savings arrangements. *Med Care Res Rev.* 2015;73(5):511-531. <https://doi.org/10.1177/1077558715618320>.
- Berlin NL, Gulseren B, Nuliyalu U, Ryan AM. Target prices influence hospital participation and shared savings in Medicare bundled payment program. *Health Aff.* 2020;39(9):1479-1485. <https://doi.org/10.1377/hlthaff.2020.00104>.
- McClellan MB, Staiger D. The Quality of Health Care Providers (August 1999). NBER Working Paper No. w7327, Available at SSRN: <https://ssrn.com/abstract=196391>.
- Silber JH, Satopää VA, Mukherjee N, et al. Improving Medicare's hospital compare mortality model. *Health Serv Res.* 2016;51 (Suppl 2): 1229-1247. <https://pubmed.ncbi.nlm.nih.gov/26987446>.
- George EI, Rocková V, Rosenbaum PR, Satopää VA, Silber JH. Mortality rate estimation and standardization for public reporting: Medicare's hospital compare. *J Am Stat Assoc.* 2017;112(519):933-947. <https://doi.org/10.1080/01621459.2016.1276021>.
- Ryan A, Burgess JF, Strawderman R, Dimick J. What is the best way to estimate hospital quality outcomes? A simulation approach. *Health Serv Res.* 2012;47(4):1699-1718.
- Chen LM, Staiger DO, Birkmeyer JD, Ryan AM, Zhang W, Dimick JB. Composite quality measures for common inpatient medical conditions. *Med Care.* 2013;51(9):832-837. <https://doi.org/10.1097/MLR.0b013e31829fa92a>.
- Dimick JB, Staiger DO, Baser O, Birkmeyer JD. Composite measures for predicting surgical mortality in the hospital. *Health Aff.* 2009; 28(4):1189-1198. <https://doi.org/10.1377/hlthaff.28.4.1189>.
- Linden A. Assessing regression to the mean effects in health care initiatives. *BMC Med Res Methodol.* 2013;13(1):119. <https://doi.org/10.1186/1471-2288-13-119>.
- Dimick JB, Staiger DO, Birkmeyer JD. Ranking hospitals on surgical mortality: the importance of reliability adjustment. *Health Serv Res.* 2010; 45(6p1):1614-1629. <https://doi.org/10.1111/j.1475-6773.2010.01158.x>.
- Osborne NH, Ko CY, Upchurch GR Jr, Dimick JB. The impact of adjusting for reliability on hospital quality rankings in vascular surgery. *J Vasc Surg.* 2011;53(1):1-5. <https://doi.org/10.1016/j.jvs.2010.08.031>.
- Dimick JB, Ghaferi AA, Osborne NH, Ko CY, Hall BL. Reliability adjustment for reporting hospital outcomes with surgery. *Ann Surg.* 2012; 255(4):703-707. <https://doi.org/10.1097/SLA.0b013e31824b46ff>.
- Oseran AS, Howard SE, Blumenthal DM. Factors associated with participation in cardiac episode payments included in Medicare's bundled payments for care improvement initiative. *JAMA Cardiol.* 2018;3(8):761-766.
- Joynt Maddox KE, Orav EJ, Zheng J, Epstein AM. Evaluation of Medicare's bundled payments initiative for medical conditions. *N Engl J Med.* 2018;379(3):260-269. <https://doi.org/10.1056/NEJMsa1801569>.
- Agency for Healthcare Research and Quality. Quality Indicator Empirical Methods; 2019. Available from https://www.qualityindicators.ahrq.gov/Downloads/Resources/Publications/2019/Empirical_Methods_2019.pdf%0D%0A%0D%0A.
- Agency for Healthcare Research and Quality. PSI 90 Fact Sheet Quality Indicators (QIs) Fact Sheet; 2016. pp. 1-8. Available from https://www.qualityindicators.ahrq.gov/News/PSI90_Factsheet_FAQ.pdf.
- MedPAC. Refining the Hospital Readmissions Reduction Program. 2013. Available from http://www.medpac.gov/docs/default-source/reports/jun13_ch04_appendix.pdf?sfvrsn=0
- Centers for Medicare & Medicaid Services. Methodology Resources; 2018. Available from qualitynet.org/inpatient/public-reporting/overall-ratings/resources
- Holmgren AJ, Adler-Milstein J, Chen LM. Participation in a voluntary bundled payment program by organizations providing care after an acute hospitalization. *JAMA.* 2018;320(4):402-404.
- Joynt Maddox KE, Orav EJ, Zheng J, Epstein AM. Participation and dropout in the bundled payments for care improvement initiative. *JAMA.* 2018;319(2):191-193. <https://doi.org/10.1001/jama.2017.14771>.
- Kim H, Meath THA, Dobbertin K, Quiñones AR, Ibrahim SA, McConnell KJ. Association of the mandatory Medicare bundled payment with joint replacement outcomes in hospitals with disadvantaged patients. *JAMA Netw Open.* 2019;2(11):e1914696-e1914696. <https://doi.org/10.1001/jamanetworkopen.2019.14696>.
- Liao JM, Martinez JR, Shan EZ, et al. Medicare's new voluntary bundled payment program: episode selection and participant characteristics. *Healthcare.* 2019 Jun;7(2):26-30.
- Navathe AS, Emanuel EJ, Venkataramani AS, et al. Spending and quality after three years of Medicare's voluntary bundled payment for joint replacement surgery. *Health Aff.* 2020;39(1):58-66.
- Barnett ML, Wilcock A, McWilliams JM, et al. Two-year evaluation of mandatory bundled payments for joint replacement. *N Engl J Med.* 2019;380(3):252-262.
- Schwartz AL. Accuracy vs. incentives: a tradeoff in performance measurement for public policy; 2018. Available from <https://scholar.harvard.edu/files/als/files/manuscript.pdf>.
- Markovitz AA, Ellimoottil C, Sukul D, et al. Risk adjustment may lessen penalties on hospitals treating complex cardiac patients under Medicare's bundled payments. *Health Aff.* 2017;36(12): 2165-2174.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Cher BAY, Gulseren B, Ryan AM. Improving target price calculations in Medicare bundled payment programs. *Health Serv Res.* 2021;56:635–642. <https://doi.org/10.1111/1475-6773.13675>