Ryan Andrew (Orcid ID: 0000-0002-2566-7763)

Improving Target Price Calculations in Medicare Bundled Payment Programs

Benjamin A. Y. Cher[1], MS, Baris Gulseren[2,3], MS, Andrew M. Ryan[2,3], PhD

Author Affiliations:
1. University of Michigan Medical School, Ann Arbor, MI
2. University of Michigan School of Public Health, Ann Arbor, MI
3. Center for Evaluating Health Reform, Ann Arbor, MI

Corresponding Author:
Andrew M. Ryan
M3124 SPHII
1415 Washington Heights
Ann Arbor, MI 48109
amryan@umich.edu
Phone: 734-936-1311

Acknowledgements: None.

Word Count: 3,211

Table: 1

Number of Figures: 3

Abstract

Word count: 200

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 $8,456 for the CMS approach versus $7,521 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.

Key Words

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

A. 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.

B. 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.

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 following years, and hospital spending that is unusually low in a particular year is likely to increase in 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.

Methods

Data Source and Definitions

We used inpatient and outpatient physician claims and 20% MEDPAR 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. Towards 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 5 clinical episodes.

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

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 **Appendix Figure A1.**

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: (1) one that is determined by a hospital's own risk-adjusted spending in the baseline period; and (2) another that is a hospital's expected spending, estimated by the hospital's characteristics. Throughout this paper we refer to appraisal (1) as "historical spending" and appraisal (2) 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 Appendix**. 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. Importance weights of variables in our model are presented in **Appendix Figure A2**. 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 **Appendix Figure A3**, where the median estimate is lower and extreme values are shrunk towards the mean. Further description of the methodology is provided in the **Technical Appendix** and **Appendix Figure A1**.

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 1,000 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 **(Appendix Figure A1)**. 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 **(Appendix Figure A4)**. 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).

Results

The study sample included 2,589 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 **(Appendix Table A1)**.

Allocation of weight between hospitals' historical spending versus expected spending was similar across episodes included in BPCI-A **(Appendix Table A2)**. 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, the difference in mean target price ranged from $343 for urinary tract infection to $2,757 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, Figure 1)**. The largest improvement was for cardiac valve (Δ=$11,716). For coronary artery bypass graft surgery (Δ=$2,757), major bowel procedure (Δ=$2,579), spinal fusion (Δ=$2,472), and sepsis (Δ=$1,752), the empirical Bayes estimator outperformed the CMS estimator by a wide margin. For 4 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 **(Appendix Figure A5)**. In sensitivity analysis including the year 2014, results did not differ substantially, and absolute and relative prediction errors were relatively similar **(Appendix Figure A6)**.

Across all clinical episodes, mean absolute prediction error was \$7,521 for the empirical Bayes approach versus \$8,456 for the CMS approach (p<0.001, **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<0.001 for all categories, **Appendix Figure A7**).

The traditional CMS methodology resulted in higher prediction error for large hospitals than small hospitals; mean absolute prediction error was \$9,042 for large hospitals versus \$8,437 for small hospitals **(Figure 3)**. The empirical Bayes methodology improved prediction accuracy for all hospital size categories. There were greater 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 \$7,982 for small hospitals versus \$6,846 for large hospitals. Lastly, improvements in accuracy for larger hospitals were generally higher for surgical episodes than medical episodes. Five of the 6 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 **(Appendix Table A3).**

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 ($8,470 for Sensitivity Analysis A vs. $8,456 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 ($7,681 for Sensitivity Analysis B versus $7,684 for the primary empirical Bayes analysis). When all information about peer-group spending trends was excluded, mean prediction error was similar $7,686, similar to Sensitivity Analysis B and the primary empirical Bayes analysis.

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 4 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 spending[16] and quality outcomes.[17-19] 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.[14] 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 programs[15] 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.[20,21] Both the Hospital Readmission Reductions Program[22] and Hospital Compare[23] use Bayesian Shrinkage to profile hospital readmission and mortality rates. The primary advantage of the 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,[24,25] high drop-out rates,[24,25] inequitable distribution of risk-sharing,[26] and substantial differences in hospital characteristics between participants and non-participants.[27,28] Savings associated with BPCI-A have been modest[1,29] 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. Additionally, even if BPCI-A participation were made mandatory – a policy solution suggested by many researchers[25] – 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.[30] 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.[13] 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.

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.

References

1. Centers for Medicare & Medicaid Services. *BPCI Advanced: Participant Resources*. 2020. https://innovation.cms.gov/initiatives/bpci-advanced/participant-resources.html
2. 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 (Millwood).* 2020 Jan;39(1):50–7.
3. Centers for Medicare & Medicaid Services. *BPCI Advanced Target Price Specifications – Model Years 1 & 2.* 2018. https://innovation.cms.gov/Files/x/bpciadvanced-targetprice-my1-2.pdf
4. Ash AS, Fienberg SE, Louis TA, Normand S-LT, Stukel TA, Utts J. Statistical issues in assessing hospital performance. 2011. https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/downloads/statistical-issues-in-assessing-hospital-performance.pdf
5. DeLia D. Monte Carlo analysis of payer and provider risks in shared savings arrangements. *Med Care Res Rev*. 2015 Nov 27;73(5):511–31. https://doi.org/10.1177/1077558715618320
6. Berlin NL, Gulseren B, Nuliyalu U, Ryan AM. (2020). Target prices influence hospital participation and shared savings in Medicare bundled payment program. *Health Affairs,* 39(9), 1479-1485. https://doi.org/10.1377/hlthaff.2020.00104
7. McClellan M, Staiger D. The quality of health care providers. Natl Bur Econ Res. 1999.
8. Silber JH, Satopää VA, Mukherjee N, Rockova V, Wang W, Hill AS, et al. Improving Medicare's Hospital Compare mortality model. *Health Serv Res*. 2016 Jun;51 Suppl 2(Suppl 2):1229–47. https://pubmed.ncbi.nlm.nih.gov/26987446

9. George EI, Ročková V, Rosenbaum PR, Satopää VA, Silber JH. Mortality rate estimation and standardization for public reporting: Medicare's Hospital Compare. *J Am Stat Assoc*. 2017 Jul 3;112(519):933–47. https://doi.org/10.1080/01621459.2016.1276021

10. 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):1699–718.

11. Chen LM, Staiger DO, Birkmeyer JD, Ryan AM, Zhang W, Dimick JB. Composite quality measures for common inpatient medical conditions. *Med Care*. 2013 Sep;51(9):832-7. doi:10.1097/MLR.0b013e31829fa92a

12. Dimick JB, Staiger DO, Baser O, Birkmeyer JD. Composite measures for predicting surgical mortality in the hospital. *Health Aff*. 2009 Jul 1;28(4):1189–98. https://doi.org/10.1377/hlthaff.28.4.1189

13. Markovitz AA, Ellimoottil C, Sukul D, Mullangi S, Chen LM, Nallamothu BK, et al. Risk adjustment may lessen penalties on hospitals treating complex cardiac patients under Medicare's bundled payments. *Health Aff.* 2017;36(12):2165–74.

14. 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 Aug;3(8):761–6.

15. 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):260–9. http://www.nejm.org/doi/10.1056/NEJMsa1801569%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/30021090

16. Linden A. Assessing regression to the mean effects in health care initiatives. *BMC Med Res Methodol*. 2013 Sep;13(1):119. http://dx.doi.org/10.1186/1471-2288-13-119

17. Dimick JB, Staiger DO, Birkmeyer JD. Ranking hospitals on surgical mortality: the importance of reliability adjustment. *Health Serv Res*. 2010 Dec 1;45(6p1):1614–29. https://doi.org/10.1111/j.1475-6773.2010.01158.x

18. Osborne NH, Ko CY, Upchurch Jr GR, Dimick JB. The impact of adjusting for reliability on hospital quality rankings in vascular surgery. *J Vasc Surg*. 2011 Jan 1;53(1):1–5. https://doi.org/10.1016/j.jvs.2010.08.031

19. Dimick JB, Ghaferi AA, Osborne NH, Ko CY, Hall BL. Reliability adjustment for reporting hospital outcomes with surgery. *Ann Surg*. 2012;255(4). https://journals.lww.com/annalsofsurgery/Fulltext/2012/04000/Reliability_Adjustment_for_Reporting_Hospital.14.aspx

20. Agency for Healthcare Research and Quality. *Quality Indicator Empirical Methods*. 2019. https://www.qualityindicators.ahrq.gov/Downloads/Resources/Publications/2019/Empirical_Methods_2019.pdf%0D%0A%0D%0A

21. Agency for Healthcare Research and Quality. *PSI 90 Fact Sheet Quality Indicators (QIs) Fact Sheet.* 2016. p. 1–8. https://www.qualityindicators.ahrq.gov/News/PSI90_Factsheet_FAQ.pdf

22. MedPAC. *Refining the hospital readmissions reduction program*. 2013. http://www.medpac.gov/docs/default-source/reports/jun13_ch04_appendix.pdf?sfvrsn=0

23. Centers for Medicare & Medicaid Services. *Methodology Resources*. 2018. qualitynet.org/inpatient/public-reporting/overall-ratings/resources

24. Holmgren AJ, Adler-Milstein J, Chen LM. Participation in a voluntary bundled payment program by organizations providing care after an acute hospitalization. *JAMA*. 2018 Jul;320(4):402–4.

25. Joynt Maddox KE, Orav EJ, Zheng J, Epstein AM. Participation and dropout in the Bundled Payments for Care Improvement Initiative. *JAMA*. 2018 Jan 9;319(2):191–3. https://doi.org/10.1001/jama.2017.14771

26. 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 Nov 6;2(11):e1914696–e1914696. https://doi.org/10.1001/jamanetworkopen.2019.14696

27. Liao JM, Martinez JR, Shan EZ, Huang JJ, Dinh CT, Huang EQ, et al. Medicare's new voluntary bundled payment program: episode selection and participant characteristics. *Healthc (Amsterdam, Netherlands)*. 2019 Jun;7(2):26–30.

28. Navathe AS, Emanuel EJ, Venkataramani AS, Huang Q, Gupta A, Dinh CT, et al. Spending and quality after three years of Medicare's voluntary bundled payment for joint replacement surgery. *Health Aff (Millwood).* 2020 Jan;39(1):58–66.

29. Barnett ML, Wilcock A, McWilliams JM, Epstein AM, Joynt Maddox KE, Orav EJ, et al. Two-year evaluation of mandatory bundled payments for joint replacement. *N Engl J Med.* 2019 Jan;380(3):252–62.

30. Schwartz AL. Accuracy vs. incentives: a tradeoff in performance measurement for public policy. 2018. https://scholar.harvard.edu/files/als/files/manuscript.pdf

31. Yount KW, Isbell JM, Lichtendahl C, Dietch Z, Ailawadi G, Kron IL, et al. Bundled payments in cardiac surgery: is risk adjustment sufficient to make it feasible? *Ann Thorac Surg*. 2015;100(5):1646–52. http://www.sciencedirect.com/science/article/pii/S0003497515007134

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		
BPCI-A Episode	Mean Target Price ($)	Mean Absolute Prediction Error ($)	Mean Absolute Prediction Error (%)	Mean Target Price ($)	Mean Absolute Prediction Error ($)	Mean Absolute Prediction Error (%)
Cardiac Valve	65,548.3	19,870.6	30.3	50,654.7	8,154.9	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	8,999.7	23.7
Spinal fusion (non-Cervical)	38,009.2	9,963.9	26.2	31,347.4	7,491.8	23.9
Hip and femur procedures except major joint	35,749.5	9,266.1	25.9	32,675.9	8,503.2	26.0
Major bowel procedure	34,506.3	12,328.4	35.7	28,861.9	9,749.9	33.8
Sepsis	28,812.0	8,951.2	31.1	23,858.5	7,199.4	30.2
Lower extremity and humerus procedure except hip, foot, femur	28,694.0	8,531.3	29.7	24,285.8	6,907.9	28.4
Stroke	26,588.7	8,879.3	33.4	23,169.0	7,844.5	33.9
Pacemaker	26,116.1	9,239.2	35.4	21,481.4	8,398.4	39.1
Cervical spinal fusion	26,046.7	8,271.7	31.8	22,202.2	7,358.3	33.1
Major joint replacement of the lower extremity	24,707.5	6,940.3	28.1	21,795.9	5,991.7	27.5
Acute myocardial infarction	23,415.3	9,322.4	39.8	20,042.7	8,417.9	42.0
Percutaneous coronary intervention	22,746.0	7,267.2	31.9	18,839.7	6,866.3	36.4
Renal failure	21,906.4	8,000.1	36.5	18,513.6	7,300.9	39.4
Congestive heart failure	21,582.6	8,208.0	38.0	18,256.4	7,646.3	41.9
Simple pneumonia and respiratory infections	19,586.9	8,504.0	43.4	16,971.5	7,892.0	46.5
Gastrointestinal hemorrhage	18,103.5	8,177.5	45.2	15,155.5	7,601.3	50.2
Cellulitis	17,892.6	9,309.8	52.0	15,351.7	8,966.4	58.4
Urinary tract infection	17,717.0	7,806.7	44.1	15,537.6	7,463.5	48.0
Chronic obstructive pulmonary	17,102.5	7,827.9	45.8	14,542.8	7,282.2	50.1

disease, bronchitis/asthma						
Gastrointestinal obstruction	15,810.1	7,591.9	48.0	13,325.5	6,826.2	51.2
Cardiac arrhythmia	15,371.7	7,046.5	45.8	12,893.9	6,634.3	51.5

Figure Legends

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

Figure 3. Mean prediction error across all clinical episodes, by hospital size, using traditional CMS estimation and empirical Bayes estimation

Appendix: Improving Target Price Calculations in Medicare Bundled Payment Programs

Technical Appendix

General Approach

As described in the manuscript, we extracted a 20% sample of Medicare fee-for-service claims for beneficiaries discharged from acute care hospitals during a baseline period (between 1/1/2010 and 9/30/2013) and a performance period (between 10/1/15 and 6/30/16). We used data from the baseline period to calculate target prices for the performance period. We evaluated prediction accuracy by comparing calculated target prices with observed spending during the performance period.

Traditional CMS approach to calculating target prices

We first extracted spending during the baseline period for each clinical episode. Let C_{Bj} be the cost per beneficiary for a particular clinical episode at hospital *j* during the baseline period *B*. We then estimated target prices for each hospital using the traditional CMS method, published by CMS[1] and summarized in **Appendix Figure A1.** For each clinical episode, let $\hat{P}_{j\_CMS}$ be the estimated target price at each hospital *j* calculated using the traditional CMS methodology.

Modified approach using Bayesian Shrinkage to account for mean reversion

We then estimated target prices for each hospital using empirical Bayes estimation. This technique is used to address random variation in health metrics and resultant mean reversion over time. The estimator is a weighted average of two quantities: (1) risk-adjusted spending and (2) expected spending conditioned on a variety of hospital-level factors.[2] Throughout this paper, (1) is referred to as “historical hospital spending” and (2) is referred to as “expected spending.” The estimator distributes weight between quantities (1) and (2) based on a measure of the reliability of risk-adjusted spending. The reliability measure depends on hospital volume and signal-to-noise measurements. When risk-adjusted spending is less reliable, less weight is applied to the estimate, so it is “shrunk” towards the conditional mean. Thus, for smaller hospitals, estimated spending depends to a greater extent on spending at other hospitals.

To implement the estimator, we first estimated patient episode spending as a function of hospital characteristics (Quantity 2 as described above). We selected independent variables using a random forest machine learning algorithm **(Appendix Figure A2)**. We implemented the estimator separately for each clinical episode. For each clinical episode, we estimated the following linear model for patient *i*, in hospital *j*, in quarter *t*:

$$C_{ijt} = \beta_0 + \beta_1 \, ln\left(Volume_j\right) + \beta_2 X_j +$$
$$\beta_3 \quad Timet + \quad \beta 4 \quad Timet \cdot \quad Xj + \quad \beta_5 \quad Seasont + eijt$$

Where C_{ijt} is episode spending, $Volume_j$ is the number of times the episode is performed at hospital *j*, *X* is a vector of hospital characteristics (academic, urban, safety net, census, bed size), *Time* is a quarterly time trend, and *Season* is a vector of dummy variables for each season. Explanatory variables were defined as in the traditional CMS methodology.[1]

We then calculated $\hat{C}_{j\_adjusted}$, the risk-adjusted spending at hospital *j* for each clinical episode during the baseline period (Quantity 1 as described above). To accomplish this, we took the ratio of predicted episode spending based on case mix (using the same HCCs and HCC interactions as in Step 3 of the traditional CMS process[1]) to observed spending at each hospital for each clinical episode. Then we multiplied this value by the average of C_j across all hospitals.

We used the empirical Bayes estimator to create shrunk estimates of the benchmark prices at each hospital:

$$\widehat{Empirial\ Bayes}\ \ j = \hat{C}_{j\_adjusted} W_j + (\hat{C}_j)(1 - W_j)$$

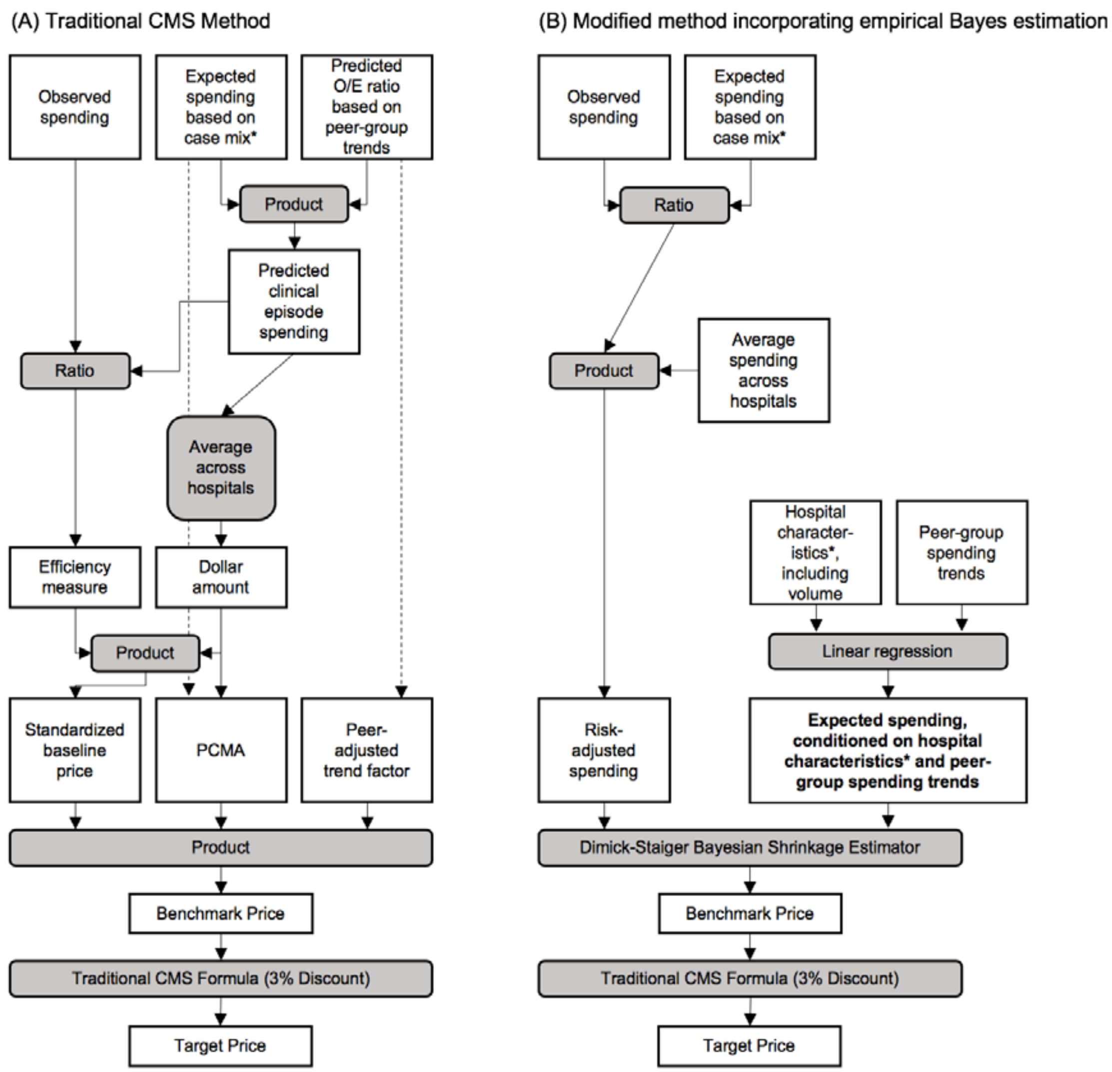
where $\widehat{Empirical}\ Bayes_j$ is the estimated benchmark price. In brief, W_j is the ratio of signal variance to total variance in residual spending. Signal variance is derived from a regression of hospital spending on hospital volume. Noise variance is derived from the mean-squared error of a regression of spending on hospital fixed effects to the number of observations for each hospital. This is described in detail in the statistical appendix of Ryan et al., 2012[3]. W_j is generally inversely associated with hospital volume.

We converted the benchmark price to the target price using the traditional CMS formula, which involved application of a 3% discount. For each clinical episode, let $\hat{P}_{j\_empirical\_Bayes}$ be the target price at hospital *j* calculated using the empirical Bayes estimator.

We evaluated hospital performance by comparing the estimated target prices ($\hat{P}_{j\_CMS}$ and $\hat{P}_{j\_empirical\_Bayes}$) to risk-adjusted spending during the performance period. Let $\hat{C}_{Pj\_adjusted}$ be risk-adjusted cost per beneficiary for a particular clinical episode at hospital *j* during the performance period *P*. We calculated $\hat{C}_{Pj\_adjusted}$ using the same risk-adjustment procedure as $\hat{C}_{j\_adjusted}$, explained above.

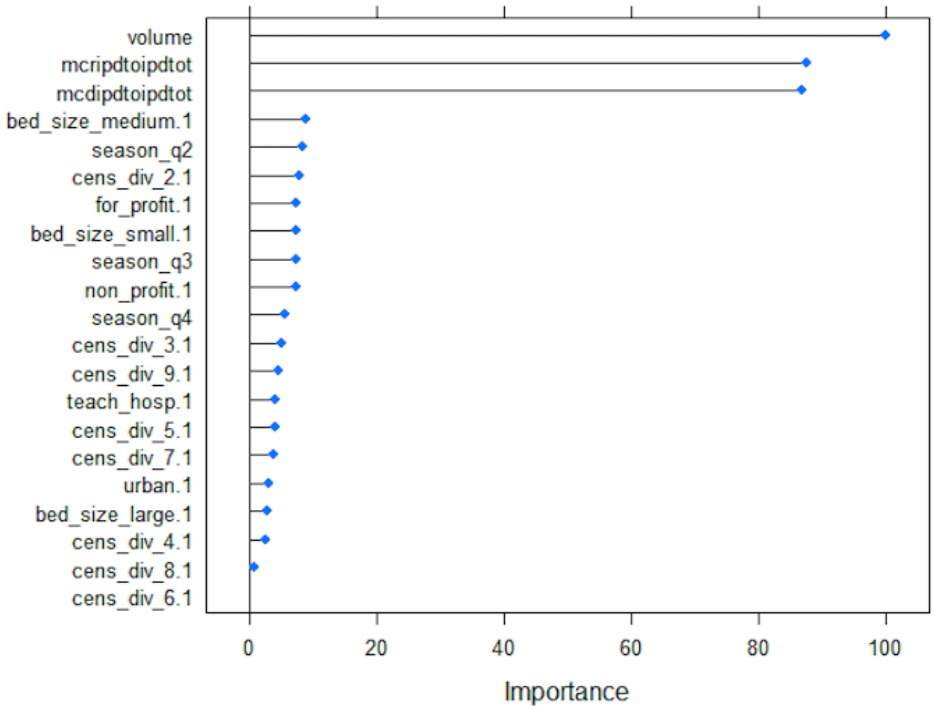
For each clinical episode, at each hospital, we determined the absolute value of the difference ("error") between cost/beneficiary during the performance period and the target price, using both traditional CMS methodology and the empirical Bayes methodology. Error using CMS methodology was $\hat{E}_{j\_CMS} = \left|\hat{P}_{j\_CMS} - \hat{C}_{Pj\_adjusted}\right|$. Error using the empirical Bayes estimator was $\hat{E}_{j\_empirical\_Bayes} = \left|\hat{P}_{j\_empirical\_Bayes} - \hat{C}_{Pj\_adjusted}\right|$.

Appendix Figure A1. Methodology for calculating BCPI-A target prices for a given clinical episode, comparing traditional CMS method (A) versus modified method incorporating empirical Bayes estimation to account for mean reversion (B)



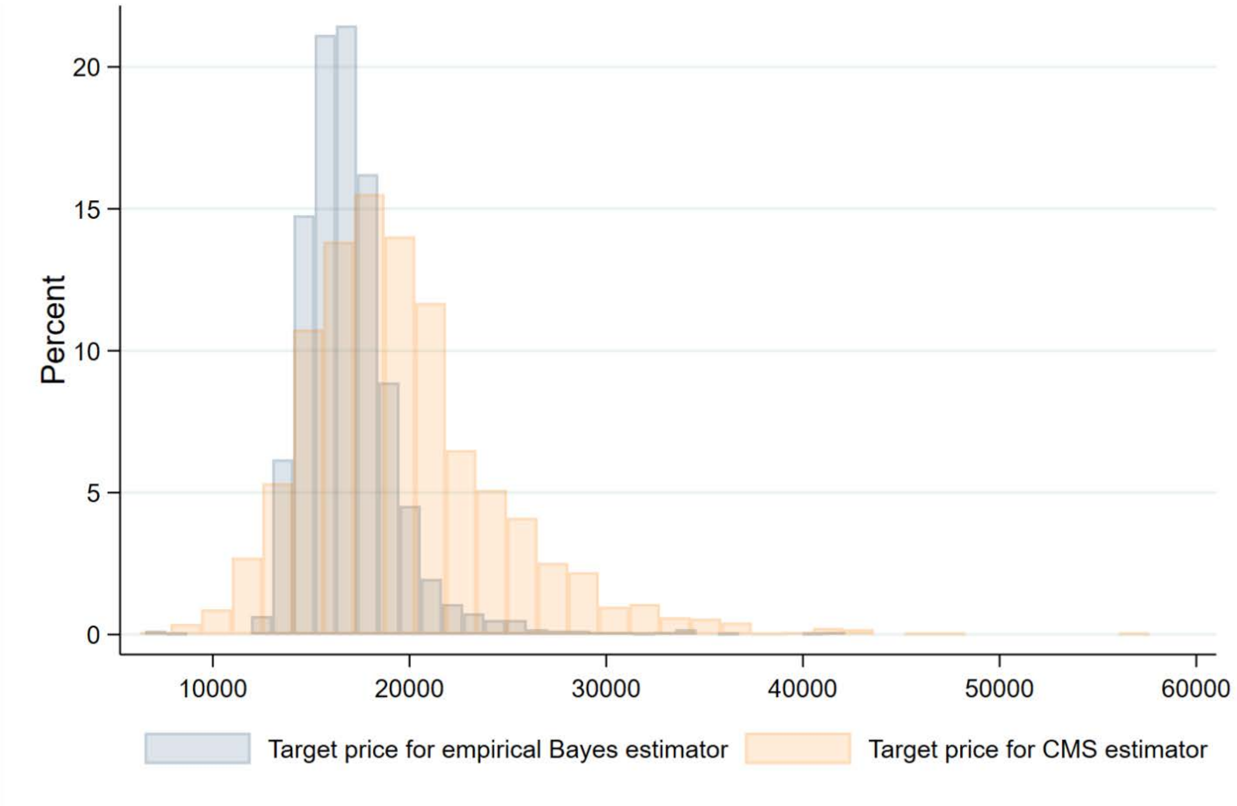
NOTES:
*Risk-adjustment based on age, sex, race, and HCCs.
**Hospital characteristics include volume (number of patients undergoing that particular clinical episode), academic vs non-academic medical center, urban vs rural, safety net hospital versus non-safety net hospital, census division (9 categories), proportion of Medicare days, proportion of Medicaid days, and bed size (small [0-250 beds], medium [251-500 beds], large [501-850 beds], extra-large [>850 beds])
***PCMA = patient case mix adjustment. This is based on realized case mix during the performance period.

Appendix Figure A2. Importance weights for random forest machine learning estimation used to model hospital expected spending



NOTES:
Volume is a continuous variable, representing the number of cases for a particular clinical episode at a particular hospital. Micripdtoipdtot is a hospital's proportion of Medicare days. Mcdipdtoipdtot is a hospital's proportion of Medicaid days. Bed size is categorized as follows: small [0-250 beds], medium [251-500 beds], large [501-850 beds], extra-large [>850 beds]. Cens_div is United States census division.

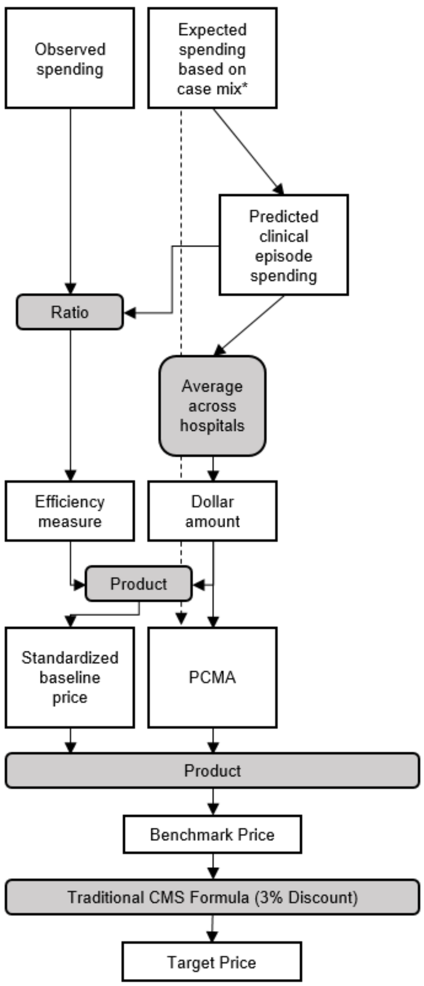
Appendix Figure A3. Distribution of target prices for simple pneumonia and respiratory infections, comparing traditional CMS methodology versus empirical Bayes estimation



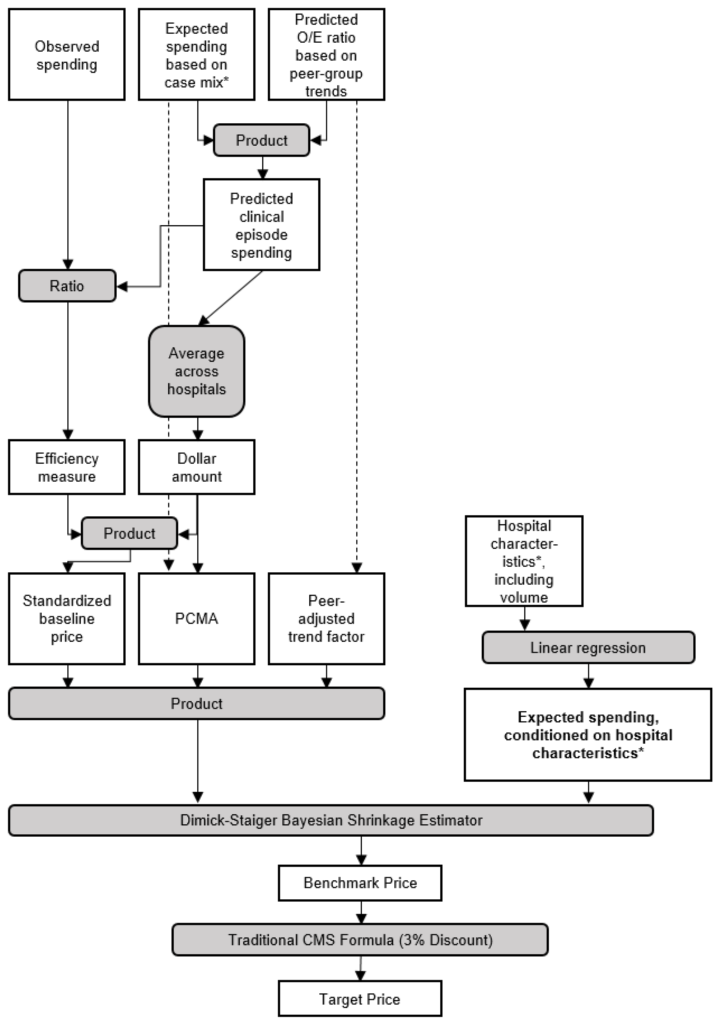
Note: simple pneumonia and respiratory infections is the most common episode in BCPI-A.

Appendix Figure A4: Sensitivity analysis to address drivers of changes in prediction accuracy between traditional CMS methodology and empirical Bayes approach

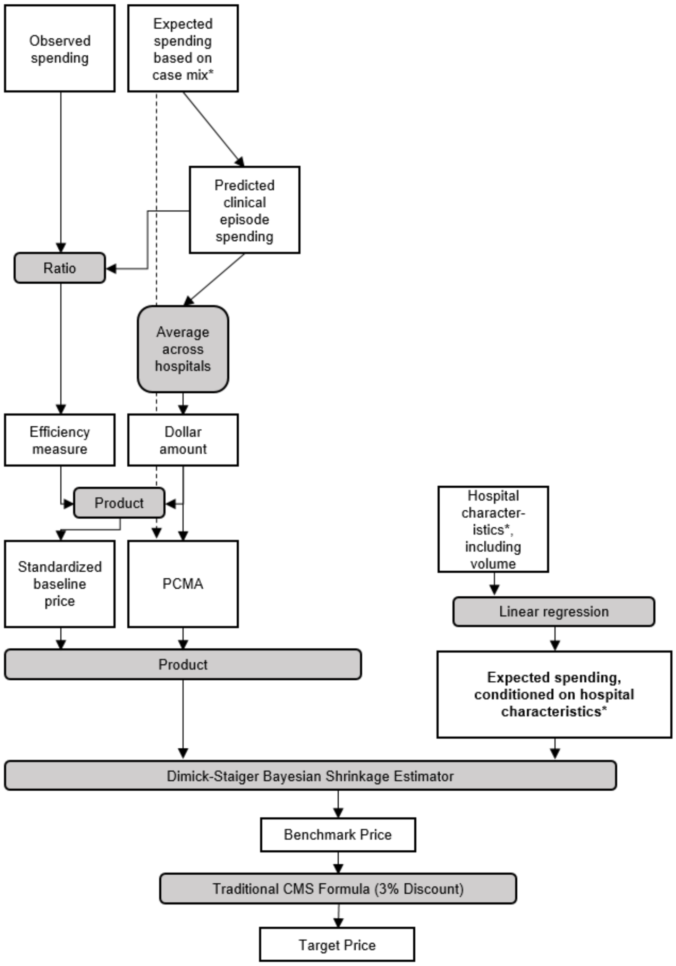
<u>Sensitivity Analysis A:</u> *Traditional CMS methodology with the peer-adjusted trend factor removed from the calculation*



<u>Sensitivity Analysis B:</u> *Leave the "peer-adjusted trend" as-is and apply the empirical Bayes estimation to the benchmark price as calculated using the traditional CMS methodology*



<u>Sensitivity Analysis C:</u> *Exclude all information about peer-group spending trends*



Appendix Table A1: Number of episodes and spending per episode, comparing the baseline period (2010-2013) and the performance period (2015-2016), for all clinical episodes

Episode	Number of hospitals	Baseline Period (2010-2013)		Performance Period (2015-2016)	
		Mean number of annual episodes (SD)	Mean spending per episode, $ (SD)	Mean number of annual episodes (SD)	Mean spending per episode, $ (SD)
All Episodes	2589	459465.3 (54418.25)	20,039 (18,209)	185527.0 (99198.6)	19,717 (17,782)
Acute myocardial infarction	523	8550 (1277.87)	20,587 (18,696)	3481 (1656.04)	19,178 (18,632)
Cardiac Valve	206	4454 (389.06)	52,283 (25,543)	2480 (1429.77)	47,001 (22,244)
Cardiac arrhythmia	1367	31453.5 (4280.71)	13,495 (14,540)	11479 (5648.37)	13,942 (14,817)
Cardiac defibrillator	79	1128.75 (337.48)	39,058 (19,813)	208.5 (89.8)	45,195 (24,181)
Cellulitis	472	7543 (1032.16)	16,007 (15,916)	2647.5 (1317.34)	15,945 (15,019)
Cervical spinal fusion	40	512.75 (25.53)	22,906 (16,259)	238 (121.62)	23,352 (18,139)
Chronic obstructive pulmonary disease, bronchitis/asthma	1800	46141 (7165.02)	15,406 (15,650)	14504 (8571.55)	15,131 (15,383)
Congestive heart failure	1822	50988.5 (7301.79)	19,127 (18,659)	21187.5 (11156.02)	19,073 (18,324)
Coronary artery bypass graft surgery	256	4247.75 (744.28)	39,056 (20,107)	1543.5 (801.15)	37,294 (17,668)
Gastrointestinal hemorrhage	1142	22018.75 (2574.33)	15,947 (15,563)	8152 (4091.32)	15,848 (15,415)
Gastrointestinal obstruction	610	4463 (528.37)	13,951 (15,149)	1627 (885.3)	12,779 (15,563)
Hip and femur procedures except major joint	436	9790.25 (1261.59)	33,378 (17,706)	4003 (1920.5)	32,307 (16,936)
Lower extremity and humerus procedure except hip, foot, femur	23	304 (53.97)	24,900 (18,109)	95.5 (51.62)	25,395 (15,595)
Major bowel procedure	436	7065 (773.08)	29,803 (22,993)	2804.5 (1529.47)	26,096 (20,355)
Major joint replacement of the lower extremity	1467	48990.5 (4916.42)	21,700 (12,686)	23147 (12440.84)	18,931 (11,752)
Pacemaker	360	5760 (1073.15)	22,180 (14,432)	1659 (975.81)	22,906 (15,080)
Percutaneous coronary intervention	892	25583.75 (4917.18)	19,539 (14,423)	7861.5 (4198.09)	21,818 (16,321)

Renal failure	1142	23037.5 (2855.65)	19,168 (18,019)	9675 (5170.36)	18,548 (16,953)
Sepsis	1569	42402.5 (3513.01)	24,700 (22,750)	26273 (14062.94)	22,569 (20,891)
Simple pneumonia and respiratory infections	2139	59388.25 (7433.74)	17,863 (16,905)	20866.5 (12304.37)	17,353 (16,770)
Spinal fusion (non-cervical)	238	4015 (435.84)	32,493 (15,900)	1803 (919.24)	31,588 (15,615)
Stroke	1066	23162.75 (2321.05)	23,863 (20,780)	10009 (5109.55)	22,456 (19,347)
Urinary tract infection	1357	28464.75 (4305.53)	16,282 (15,311)	9782 (4747.51)	16,094 (15,114)

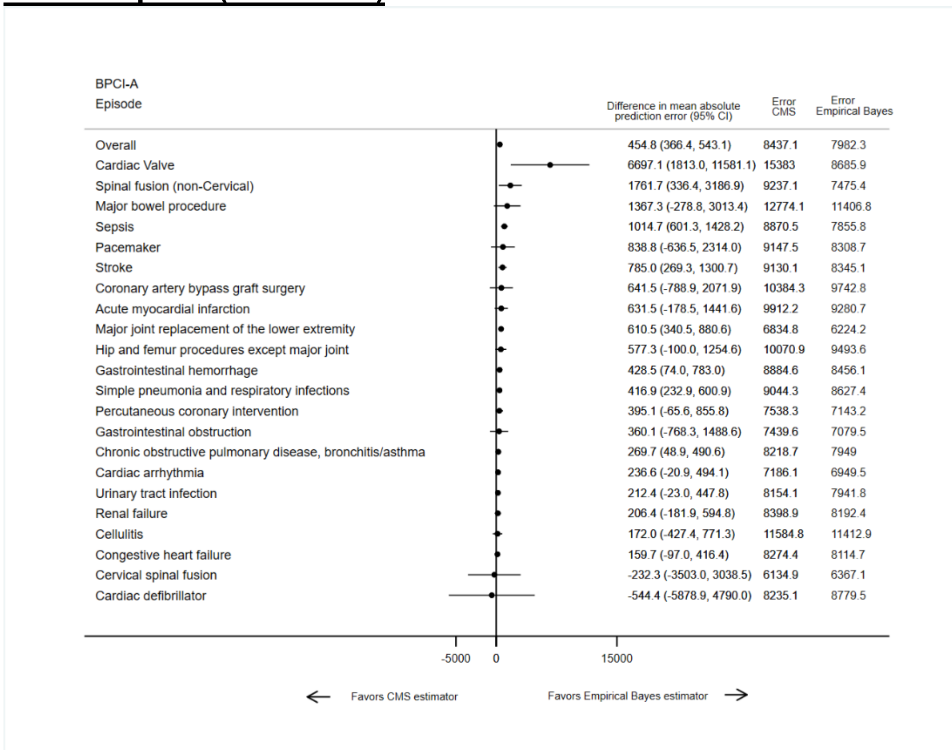
NOTES: SD = standard deviation.

Appendix Table A2. Weights applied to historical hospital spending and expected spending used by the empirical Bayes estimator, for each clinical episode

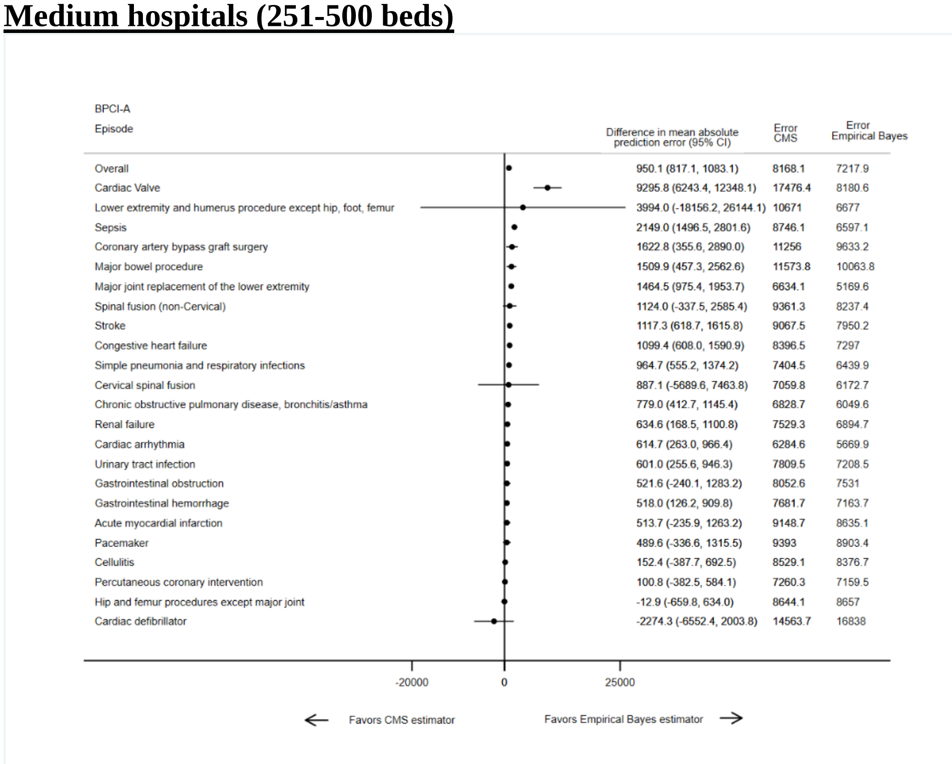
BPCI -A Clinical Episode	Weight applied to risk-adjusted spending, mean across hospitals (SD)	Weight applied to expected spending, conditioned on hospital characteristics and peer-group spending trends, mean across hospitals (SD)
Acute myocardial infarction	0.4529 (0.0456)	0.5471 (0.0456)
Cardiac Valve	0.323 (0.0292)	0.677 (0.0292)
Cardiac arrhythmia	0.2974 (0.0415)	0.7026 (0.0415)
Cardiac defibrillator	0.3252 (0.0274)	0.6748 (0.0274)
Cellulitis	0.2907 (0.0474)	0.7093 (0.0474)
Cervical spinal fusion	0.318 (0.0551)	0.682 (0.0551)
Chronic obstructive pulmonary disease, bronchitis/asthma	0.2928 (0.0469)	0.7072 (0.0469)
Congestive heart failure	0.2954 (0.0445)	0.7046 (0.0445)
Coronary artery bypass graft surgery	0.3249 (0.0303)	0.6751 (0.0303)
Gastrointestinal hemorrhage	0.2948 (0.045)	0.7052 (0.045)
Gastrointestinal obstruction	0.301 (0.0458)	0.699 (0.0458)
Hip and femur procedures except major joint	0.3155 (0.0385)	0.6845 (0.0385)
Lower extremity and humerus procedure	0.3075 (0.0474)	0.6925 (0.0474)
Major bowel procedure	0.3103 (0.0387)	0.6897 (0.0387)
Major joint replacement of the lower extremity	0.3286 (0.0461)	0.6714 (0.0461)
Pacemaker	0.3191 (0.0344)	0.6809 (0.0344)
Percutaneous coronary intervention	0.3181 (0.0315)	0.6819 (0.0315)
Renal failure	0.2996 (0.0445)	0.7004 (0.0445)
Sepsis	0.3004 (0.0512)	0.6996 (0.0512)
Simple pneumonia and respiratory infections	0.2903 (0.0501)	0.7097 (0.0501)
Spinal fusion (non-Cervical)	0.325 (0.0398)	0.675 (0.0398)
Stroke	0.2957 (0.0469)	0.7043 (0.0469)
Urinary tract infection	0.2884 (0.0465)	0.7116 (0.0465)

Appendix Figure A5. Difference in prediction error between traditional CMS methodology and empirical Bayes estimation, for all clinical episodes, by hospital size

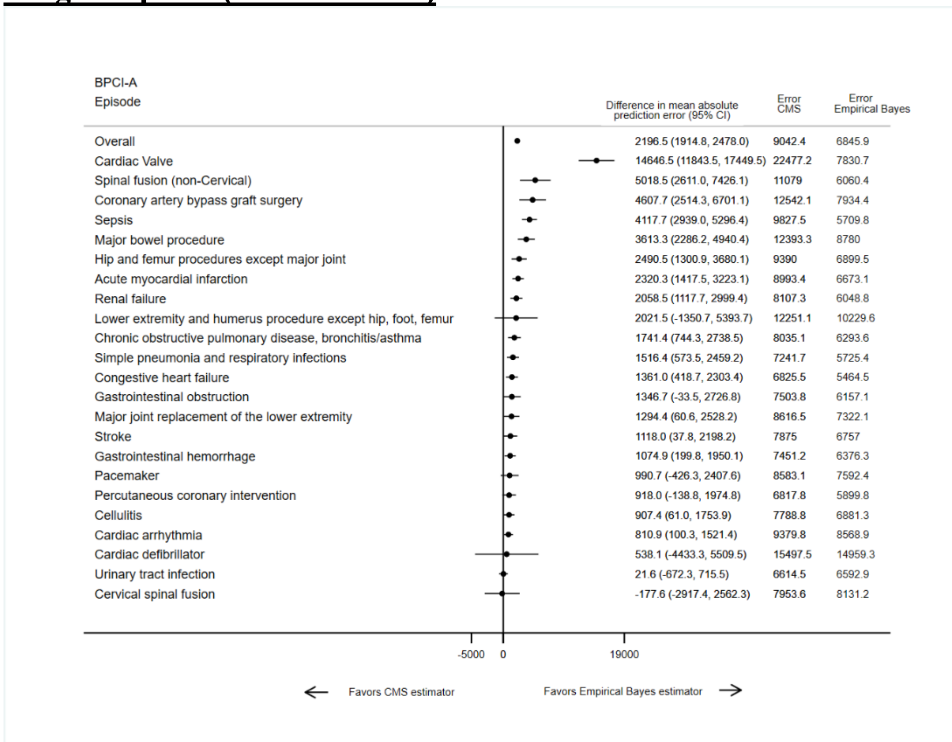
<u>Small hospitals (0-250 beds)</u>



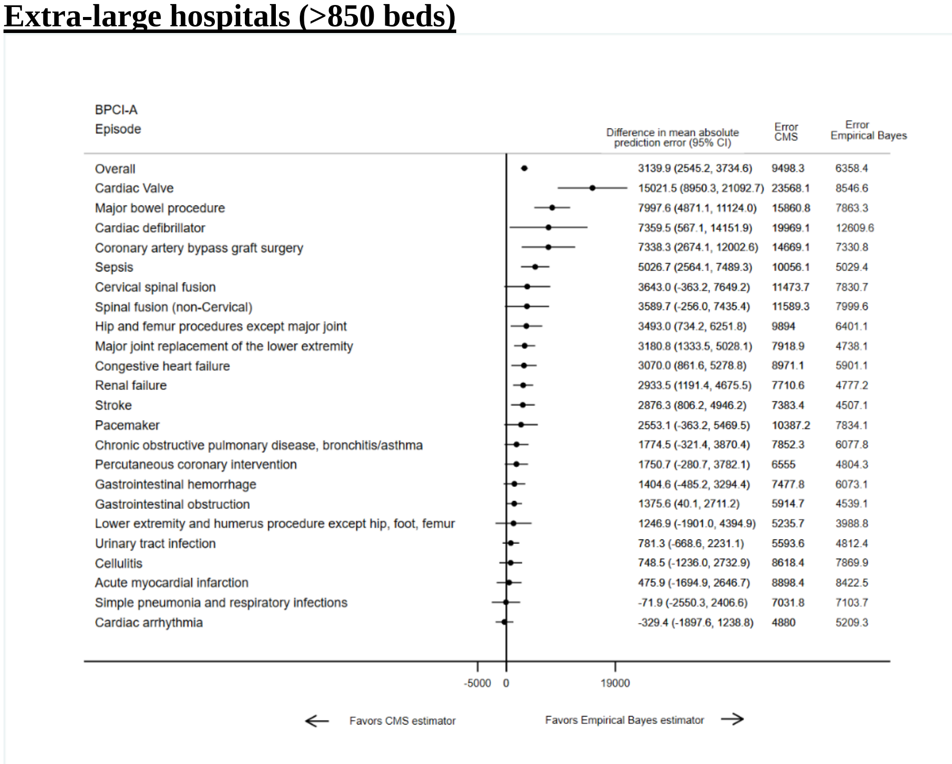
<u>Medium hospitals (251-500 beds)</u>



Large hospitals (501 - 850 beds)

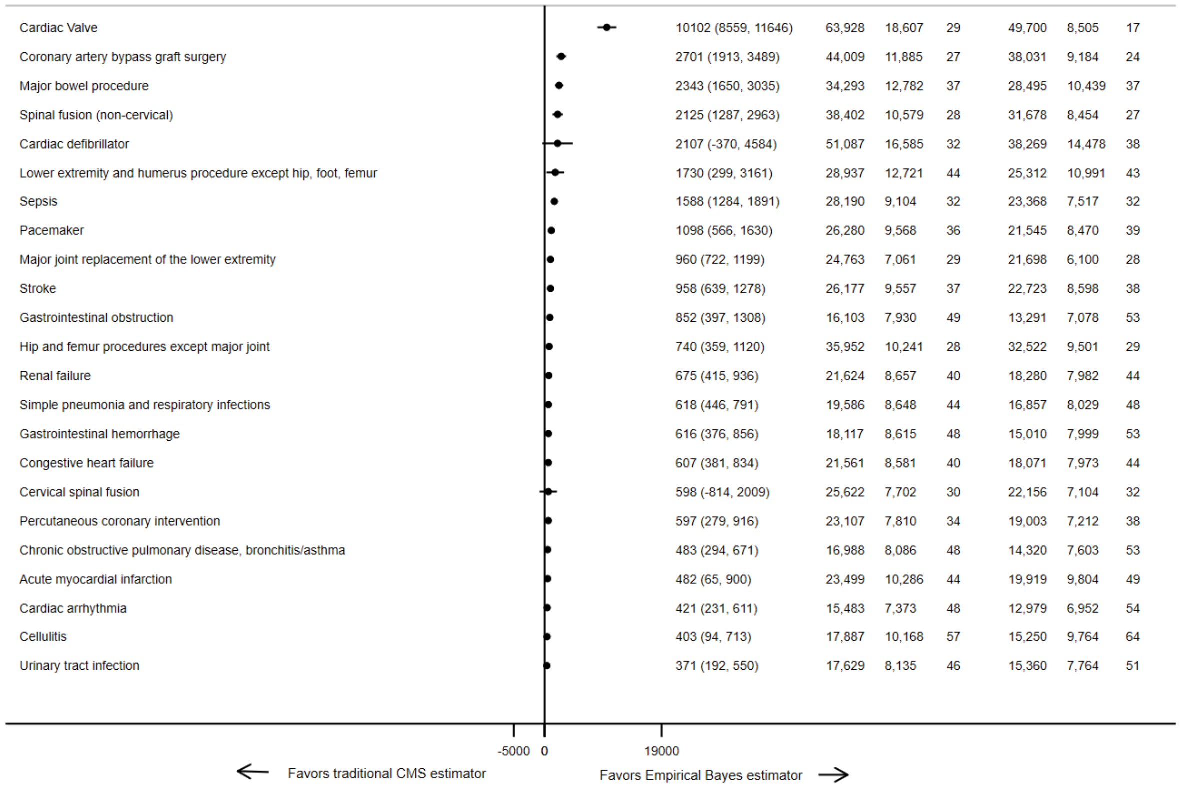


Extra-large hospitals (>850 beds)



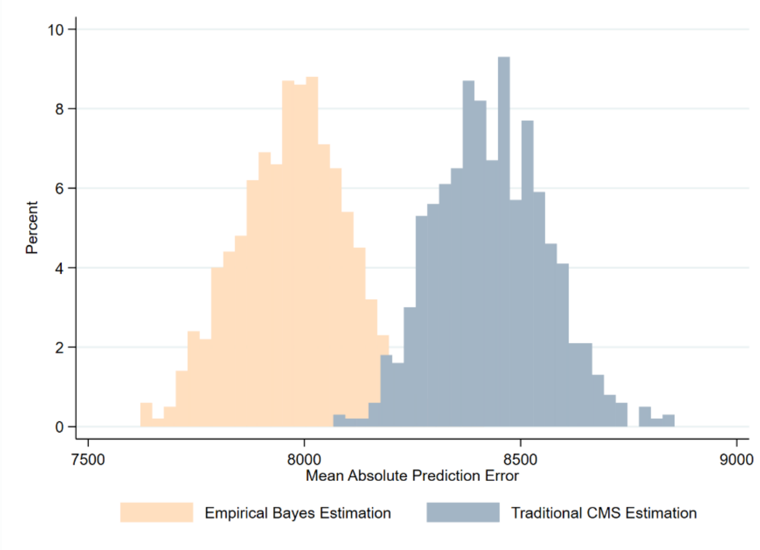
Appendix Figure A6: Difference in prediction error between traditional CMS methodology and empirical Bayes estimation, for all clinical episodes, including data from the year 2014

BPCI-A Episode	Difference in mean absolute prediction error in $ (95%CI)	Traditional CMS Methodology Target Price	Traditional CMS Methodology Error ($)	Traditional CMS Methodology Error (%)	Empirical Bayes Methodology Target Price	Empirical Bayes Methodology Error ($)	Empirical Bayes Methodology Error (%)
Cardiac Valve	10102 (8559, 11646)	63,928	18,607	29	49,700	8,505	17
Coronary artery bypass graft surgery	2701 (1913, 3489)	44,009	11,885	27	38,031	9,184	24
Major bowel procedure	2343 (1650, 3035)	34,293	12,782	37	28,495	10,439	37
Spinal fusion (non-cervical)	2125 (1287, 2963)	38,402	10,579	28	31,678	8,454	27
Cardiac defibrillator	2107 (-370, 4584)	51,087	16,585	32	38,269	14,478	38
Lower extremity and humerus procedure except hip, foot, femur	1730 (299, 3161)	28,937	12,721	44	25,312	10,991	43
Sepsis	1588 (1284, 1891)	28,190	9,104	32	23,368	7,517	32
Pacemaker	1098 (566, 1630)	26,280	9,568	36	21,545	8,470	39
Major joint replacement of the lower extremity	960 (722, 1199)	24,763	7,061	29	21,698	6,100	28
Stroke	958 (639, 1278)	26,177	9,557	37	22,723	8,598	38
Gastrointestinal obstruction	852 (397, 1308)	16,103	7,930	49	13,291	7,078	53
Hip and femur procedures except major joint	740 (359, 1120)	35,952	10,241	28	32,522	9,501	29
Renal failure	675 (415, 936)	21,624	8,657	40	18,280	7,982	44
Simple pneumonia and respiratory infections	618 (446, 791)	19,586	8,648	44	16,857	8,029	48
Gastrointestinal hemorrhage	616 (376, 856)	18,117	8,615	48	15,010	7,999	53
Congestive heart failure	607 (381, 834)	21,561	8,581	40	18,071	7,973	44
Cervical spinal fusion	598 (-814, 2009)	25,622	7,702	30	22,156	7,104	32
Percutaneous coronary intervention	597 (279, 916)	23,107	7,810	34	19,003	7,212	38
Chronic obstructive pulmonary disease, bronchitis/asthma	483 (294, 671)	16,988	8,086	48	14,320	7,603	53
Acute myocardial infarction	482 (65, 900)	23,499	10,286	44	19,919	9,804	49
Cardiac arrhythmia	421 (231, 611)	15,483	7,373	48	12,979	6,952	54
Cellulitis	403 (94, 713)	17,887	10,168	57	15,250	9,764	64
Urinary tract infection	371 (192, 550)	17,629	8,135	46	15,360	7,764	51

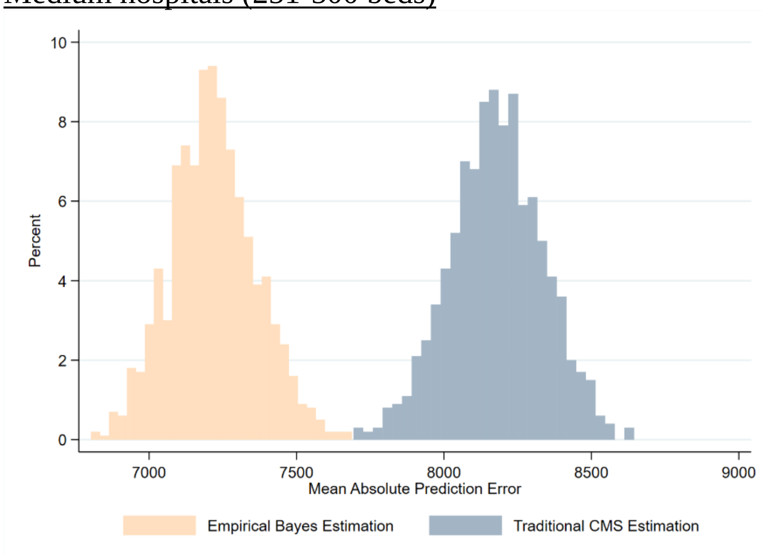


Appendix Figure A7. Mean prediction error for all hospitals averaged across all clinical episode types, across 1,000 bootstrap iterations, by hospital size

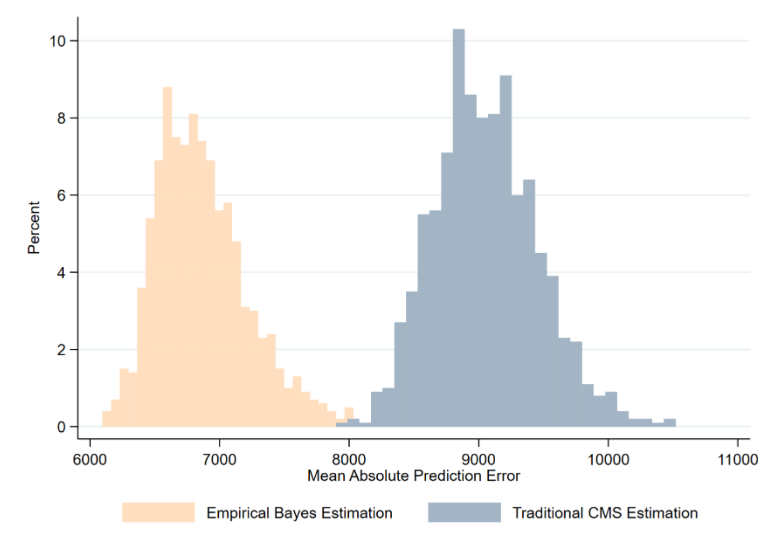
Small hospitals (0-250 beds)



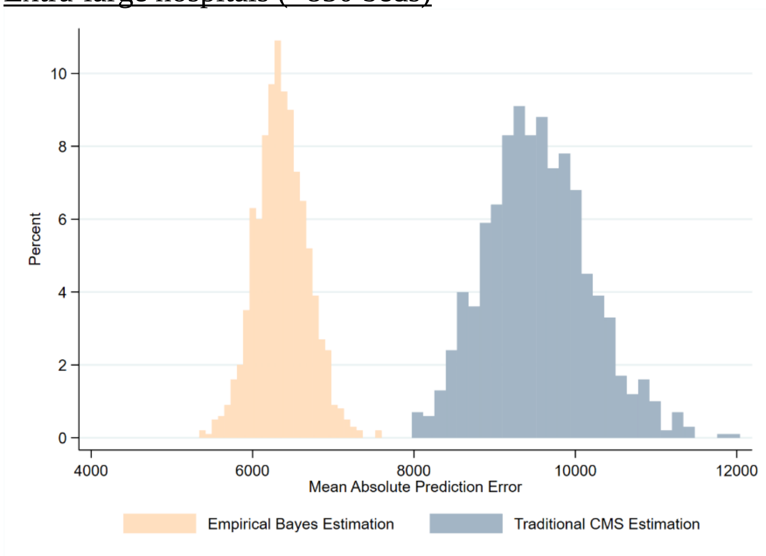
Medium hospitals (251-500 beds)



Large hospitals (501 - 850 beds)



Extra-large hospitals (>850 beds)



Appendix Table A3: Distribution of absolute prediction error, comparing traditional CMS methodology and empirical Bayes estimation, stratifying by hospital characteristics

Hospital characteristics	Percentile	Traditional CMS Estimation	Empirical Bayes Estimation
Hospital size			
Small (0-250 beds)	25p	3,039.1	2,706.8
	50p	6,210.6	5,560.1
	75p	10,387.9	9,294.8
Medium (251-500 beds)	25p	2,943.8	2,424.9
	50p	6,159.4	5,085.1
	75p	10,459.3	8,741.6
Large (501-850 beds)	25p	3,257.7	2,187.0
	50p	6,771.9	4,679.6
	75p	11,754.4	8,401.7
Extra-large (> 850 beds)	25p	3,693.5	2,090.1
	50p	7,291.7	4,455.8
	75p	13,163.6	8,175.5
Teaching status			
Teaching	25p	4,003.3	2,369.2
	50p	8,231.6	5,069.9
	75p	13,616.6	8,926.4
Non-teaching	25p	2,928.2	2,533.3
	50p	6,018.1	5,280.7
	75p	10,130.2	8,952.4
Profit status			
For-profit	25p	2,982.8	2,862.2
	50p	6,162.6	5,916.0
	75p	10,773.5	9,726.3
Not-for-profit	25p	3,070.4	2,472.1

	50p	6,299.5	5,161.7
	75p	10,625.2	8,838.8
Other	25p	3,008.8	2,304.8
	50p	6,257.4	4,984.4
	75p	10,625.6	8,658.5
Urban/Rural status			
Urban	25p	3,046.2	2,499.4
	50p	6,290.6	5,254.9
	75p	10,686.4	8,950.9
Rural	25p	3,073.4	2,596.3
	50p	5,830.4	5,146.7
	75p	9,504.4	9,111.3
Region Category			
Midwest	25p	2,965.2	2,532.7
	50p	5,992.8	5,217.3
	75p	9,964.0	8,831.6
Northeast	25p	3,327.5	2,565.3
	50p	6,973.3	5,337.9
	75p	12,044.7	9,141.1
South	25p	2,701.8	2,436.1
	50p	5,586.3	5,138.4
	75p	9,551.9	8,952.4
West	25p	4,165.6	2,591.5
	50p	8,346.7	5,460.2
	75p	13,086.4	8,915.7

References

1. Centers for Medicare & Medicaid Services Innovation Center. *BPCI Advanced Target Price Specifications Model Years 1 and 2.*; 2018. https://www.cms.gov/Research-Statistics-.
2. Dimick JB, Staiger DO, Baser O, Birkmeyer JD. Composite measures for predicting surgical mortality in the hospital. *Health Affairs*. 2009;28(4):1189-1198. doi:10.1377/hlthaff.28.4.1189
3. Ryan A, Burgess J, Strawderman R, Dimick J. What is the best way to estimate hospital quality outcomes? A simulation approach. *Health Services Research*. 2012;47(4):1699-1718. doi:10.1111/j.1475-6773.2012.01382.x

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

REQUIRED INFORMATION AND FORM FOR ACCEPTED MANUSCRIPTS AT HSR [SEPT., 2009]

HSR Authorship Responsibility, Disclosures, and Acknowledgments

General instructions:
The corresponding author is responsible for coordinating with each author to obtain all information disclosed on this form. This information is required to be submitted in the section on Manuscript Central labeled for the editors only, i.e., as supplementary files not for review, at the time of the first revision and should be updated as necessary for subsequent revisions or in response to queries from the editor-in-chief.

Each author must read and provide the information to the corresponding author on:
(1) Authorship Responsibility, Criteria, and Contributions,
(2) Financial Disclosures, and
(3) Other Disclosures,

The **corresponding author** must complete and submit on behalf of all authors the summary form, affirming the information provided is accurate. The submitted form should contain the answers for the corresponding author on items 1-3 and summarize all authors' answers in the (4) Author and Other Contributors Section and (5) Acknowledgment/Disclosure Statement.

This form is available online at HSR.org *and* http://mc.manuscriptcentral.com/hsr. If necessary, it may be photocopied and distributed to coauthors or other contributors.

NOTE: We require this form to be completed before publication. The joint acknowledgement/disclosure statement and author matrix will be published electronically; the joint acknowledgement/disclosure statement [including any contributors to be acknowledged] will also appear in the print version. Each form will be approved by the editors to make sure that all necessary and relevant information that readers need to know to evaluate the work published in HSR has been reflected in the acknowledgement/disclosure paragraph; detailed disclosures will not be published.

The corresponding author should collect the disclosure information from each co-author and provide that information in the appropriate box below. The corresponding author should then prepare a draft joint acknowledgement/disclosure statement along with the author matrix. Each author should review and approve of the final joint statement and authorship information.

The wording of the acknowledgement/disclosure statement and author matrix may undergo some negotiation with the editors, either to eliminate some material that is not needed or to include some that should be disclosed. As only the corresponding author can upload information to Manuscript Central, he or she will coordinate the process. All authors should submit drafts of their disclosures electronically to the corresponding author. The corresponding author may submit them and a draft of the joint statement electronically at any time during our processing of the manuscript. This material will be reviewed for consistency and completeness when the manuscript is close to acceptance.

Health Services Research
Certification and Disclosures by Authors

Manuscript Number: HSR- 20-0636.R1
[After submitting your manuscript online, your manuscript is assigned a number that must be entered in the space above].

Corresponding Author: Andrew M Ryan

Please fill in the following table: AUTHOR ORDER AND CONTACT INFORMATION

Authors, in preferred order for publication	**Telephone Number**	**Fax Number**	**e-mail address**
Benjamin Cher			bcher@med.umich.edu
Baris Gulseren			baris@umich.edu
Andrew Ryan	734-936-1311		amryan@umich.edu

INSTRUCTIONS: EACH AUTHOR MUST READ SECTIONS 1-3 AND PROVIDE FULL DISCLOSURE TO THE CORRESPONDING AUTHOR. THE SUBMITTED FORM SHOULD PROVIDE INDIVIDUAL INFORMATION FOR THE CORRESPONDING AUTHOR AND SUMMARIZE ALL AUTHORS' INFORMATION.

1. Authorship Responsibility, Criteria, and Contributions. Each author should meet all criteria below (A, B, C, and D) and should affirm his or her general and specific contributions to the corresponding author. The corresponding author should then complete the information for him/herself by checking the appropriate boxes below and summarize the information for all authors in section 4 and 5. This summary statement should be consistent with and reflect all significant disclosures of each of the individual authors. Research and manuscripts may depend critically on various individuals who do not meet all the criteria for authorship; they should be acknowledged as contributors in section 4 below.

☑ A. I certify that

- the manuscript represents valid work and that neither this manuscript nor one with substantially similar content in which I am an author has been published or is being considered for publication elsewhere, except as described in an attachment, and that copies of closely related manuscripts have been provided to the editors or HSR; and
- [for papers with more than 1 author], I agree to allow the corresponding author to serve as the primary correspondent with the editorial office, to review the edited typescript and proof, and to make decisions regarding release of information in the manuscript to the media, federal agencies, or both; or,
- [for papers for which I am the only author] I will be the corresponding author and agree to serve in the roles described above.

☑ B. I have given final approval of the submitted manuscript and each of its revisions, if any.

☑ C. I have participated sufficiently in the work to take public responsibility for (check 1 or 2 below)

☑ 1. the whole content.

☐ 2. part of the content: (fill in as appropriate)

limited to

all but

☑ D. To qualify for authorship, you must check at least 1 box for *each* of the 2 categories of contributions listed below. For examples of contributions that qualify for acknowledgment but not for authorship, see section 4 below.

I have made substantial contributions to the intellectual content of the paper as described below.

1. (check at least 1 of the 3 below)

☑ a. conception and design
☑ b. acquisition of data
☑ c. analysis and interpretation of data

2. (check at least 1 of 2 below)

☑ a. drafting of the manuscript
☑ b. critical revision of the manuscript for important intellectual content

2. Financial Disclosure. For the purposes of this disclosure and the one below, "conflicts of interest" include not only situations in which decisions and judgment have been influenced, but also those that may, if subsequently disclosed or uncovered, lead readers to question whether decisions or judgment may have been influenced by such situations. We anticipate that the vast majority of "conflicts," if disclosed in advance, will have little bearing on a reader's assessment of the research and no bearing on our decision to publish. Except in cases in which all the authors certify they have nothing to disclose, the corresponding author will submit a brief acknowledgement/disclosure of the conflicts of all the authors in Section 4.

Please check the appropriate boxes below (applies to the past five years and foreseeable future):

☐ a. I have no relevant financial interests pertaining to this manuscript.

☑ b. I certify that all my conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (e.g., employment / affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are disclosed in an attachment.

☑ c. I certify that all financial and material support (including those providing data or access to patients, interviewees, etc.) for this research and work are clearly identified in an acknowledgement/disclosure section to be published with the manuscript.

☑ d. I certify that I agree with the description in the Acknowledgements/Disclosures statement in Section 5 of any support for the project or conflicts of interest as they pertain to myself.

3. Other Disclosures. Aside from financial interests, there may be other real or potential conflicts that need to be disclosed. In particular, we are concerned about situations in which an external sponsor, provider of data, or other entity may be able to influence either an author's ability to publish or to shape what he or she seeks to publish. We are also concerned that an author, or organizations with which he or she is affiliated, may have taken certain positions relevant to the research that should be disclosed. We may require that such information be publicly acknowledged. For the purposes of this disclosure, the term "contractual right" includes not only formal contracts and memoranda of understanding with outside parties, but also the rights of your employer to review and or approve your publications. Please check the appropriate boxes below and provide the information requested:

a. Sponsors and/or supporters of this research, e.g., organizations providing data, or supervisors within my own organization have:
   - ☐ i. a contractual right to review and approve the manuscript before submission or publication.
   - ☐ ii. a contractual right to review and comment on the manuscript within days, after which it can be submitted without constraint.
   - ☐ iii. no contractual rights to review the manuscript before submission, but there is a requirement that the sponsor/supporter be given a copy of the accepted manuscript prior to publication.
   - ☐ iv. no requirement for prior approval or notification, but I solicited feedback and/or plan to provide an advance copy as a courtesy.
   - ☐ v. the manuscript has not been reviewed or commented upon by the sponsor(s) and I have no plans to provide advance notification.
   - ☑ vi. this research did not have a sponsor as defined above.

If you checked any of the boxes i through iv, please indicate the organizations involved and the circumstances. (For example, some organizations require that manuscripts undergo internal peer review for comment, but an author may publish any findings he or she wishes.)

IN THE SUBMITTED FORM, THE CORRESPONDING AUTHOR SHOULD PROVIDE DETAILS FOR EACH AUTHOR [NAMED] WHO HAS CHECKED ANY BOX i THROUGH iv.

Please describe:

A contractual right by a sponsor, supporter, or employer to review and approve (and hence censor) a manuscript makes it appear to be a "work for hire" rather than research. On occasion such work may warrant publication in HSR, e.g., because it illustrates novel methods, but we will be reluctant to accept a manuscript with findings or conclusions that might be influenced by the ability of the sponsor to withhold permission to publish.

b. I have taken public stands (e.g., in print, media, expert witness, legislative testimony or other venues, with or without compensation) that are identified with a particular advocacy position relevant to the manuscript.

IN THE SUBMITTED FORM, THE CORRESPONDING AUTHOR SHOULD PROVIDE DETAILS FOR EACH AUTHOR [NAMED] WHO HAS CHECKED BOX b.

Please describe:

c. The organization with which I am currently (or was, at the time of the manuscript's preparation) affiliated (as a spokesperson, board member, or similar prominent position) is identified with a particular advocacy position relevant to the manuscript.

IN THE SUBMITTED FORM, THE CORRESPONDING AUTHOR SHOULD PROVIDE DETAILS FOR EACH AUTHOR [NAMED] WHO HAS CHECKED BOX c.

Please describe:

d. I certify I agree with the description in the Acknowledgements/Disclosures statement in Section 5 of disclosures as they pertain to myself.

4.A. Author matrix. Please complete the authorship matrix affirming which roles each author has fulfilled. To be listed as an author, persons should have made contributions in both of the shaded areas; all others should be acknowledged as contributors. In the box in 4.B., we invite you to name other contributors. In addition, you may use the box in 4.B. if you wish to name other kinds of contributions by authors.

Section 4 A. and B. will be made available in the electronic version of accepted manuscripts.

*You may wish to copy the author's names from the box on page 2.

Manuscript Number: HSR-20-0636.R1	**Each author must check at least one blue box showing important contributions to:**			**& check contributions in at least one orange box:**	
Name of Each Author*	**Conception and design**	**Acquisition of data (arranging for or obtaining)**	**Analysis and interpretation of data**	**Drafting the manuscript**	**Critical revision for important intellectual content**
Benjamin Cher	☑	☑	☑	☑	☑
Baris Gulersen	☐	☑	☑	☐	☑
Andrew M Ryan	☑	☑	☑	☑	☑
	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐

Section 4.B. Other Contributions

Some research projects involve very extensive teams with many people who do not meet the criteria for authorship but who have contributed importantly to the work. The box below is designed to acknowledge other types of contributions. As appropriate, use this box to name other people who have made important contributions to the work reported in this manuscript. Please also name people who have played a significant role in preparing the manuscript but who are not listed as authors. NOTE: If you wish to acknowledge any of these contributors in print, please include their names and roles in the joint acknowledgment section below.

Please name contributors to be acknowledged for:

Statistical analysis

Data collection (under supervision)

Interviewing (under supervision)

Programming and data management

Editing for presentation or style

Obtaining funding

Administrative, technical, or material support

Supervision of research staff

Other (specify)

INSTRUCTIONS: THE CORRESPONDING AUTHOR MUST FILL OUT SECTION 5.

5. Joint Acknowledgment/Disclosure Statement. Please enter a 1-3 sentence statement that acknowledges all forms of financial and material support for the project, the roles of key individuals who should be recognized; (this may or may not extend to everyone listed as contributors). Please also include summary statements about what you feel are the conflicts of interests and disclosures without which some readers may feel that important relevant information is being withheld. The editors will determine whether these disclosures are sufficient or excessive and may return a revised version. All authors will need to approve the final version. As appropriate, please add at the end of your statement: 'Disclosures: None' or 'No Other Disclosures'.

Section 5 will be made available in both the electronic and print version of accepted manuscripts.

Joint Acknowledgement/Disclosure Statement:

Dr. Ryan is supported by his grant from the National Institute of Aging (R01-AG-047932).

The corresponding author should obtain permission to name all individuals named in an Acknowledgment or the Contributorship Section because readers may infer their endorsement of data and conclusions.

The corresponding author must check the box below to affirm his/her certification that:

- all persons who have made substantial contributions to the work reported in this manuscript (e.g., data collection, analysis, or writing or editing assistance) but who do not fulfill the authorship criteria are named with their specific contributions in the Contributorship Box associated with the manuscript.
- all persons named in the Contributorship box have provided me with permission to be named.
- no other persons have made substantial contributions to this manuscript.
- all authors have approved the Joint Acknowledgement/Disclosure Statement intended for publication and the author matrix.

☑ I, the corresponding author, certify that the above statements are true.

(Adapted, with permission, from the Journal of the American Medical Association, 2006)

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

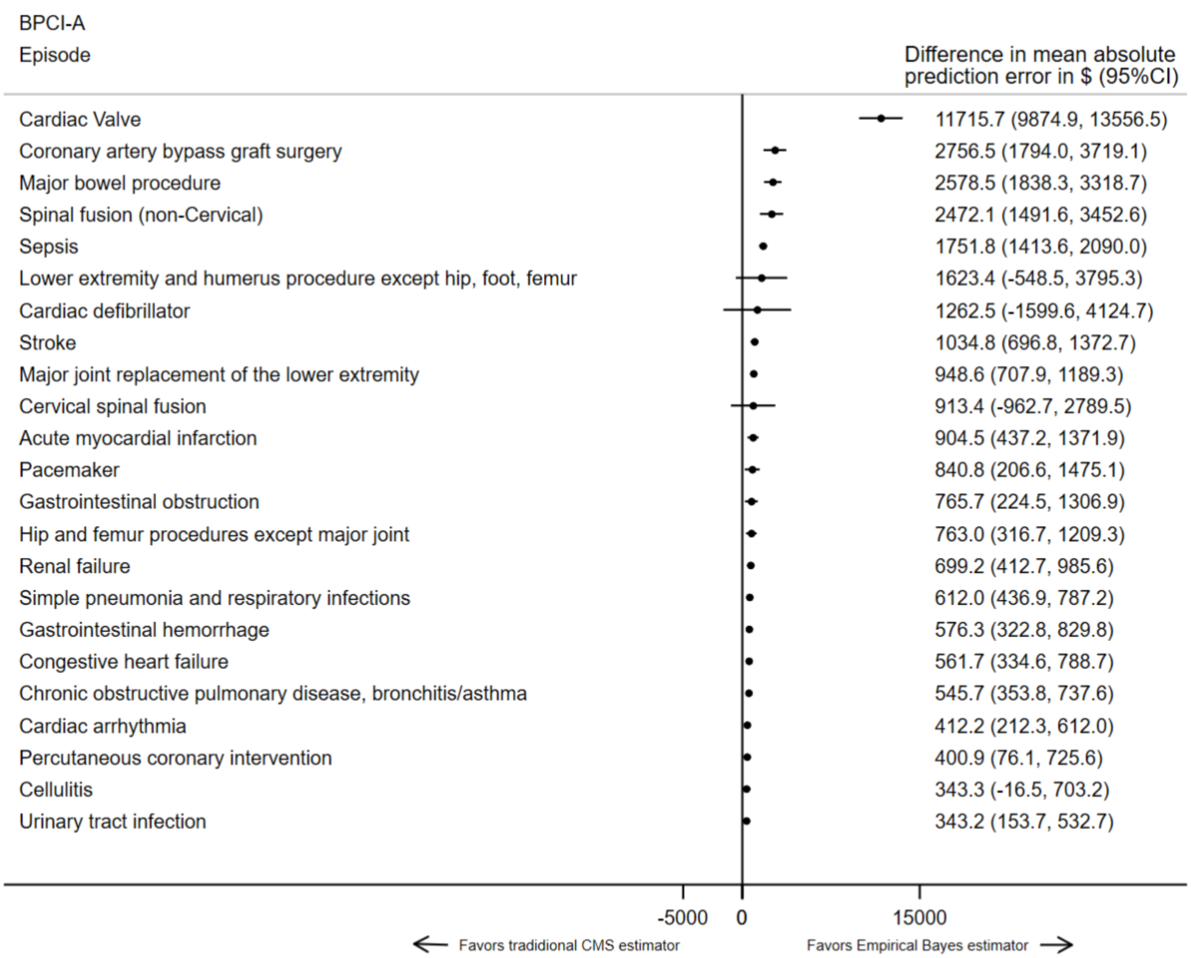
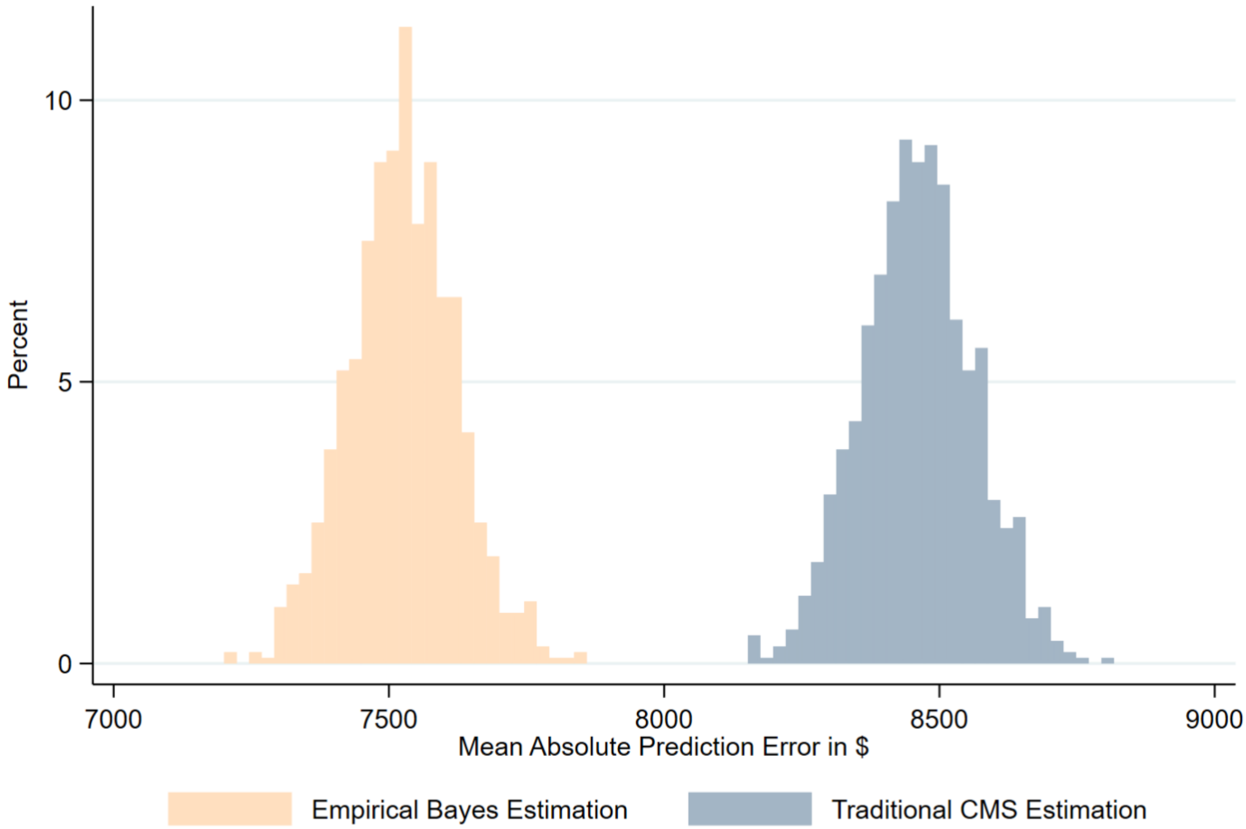


Figure 2. Mean prediction error for all hospitals, averaged across all clinical episodes



NOTES: Figure based on 1,000 bootstrap iterations. Mean absolute prediction error is unweighted mean error across all episodes. Mean prediction error for traditional CMS estimator = $8,455.7. Mean prediction error for empirical Bayes estimator = $7,521.4.

Figure 3. Mean prediction error across all clinical episodes, by hospital size, using traditional CMS estimation and empirical Bayes estimation

NOTES: Figure based on 1,000 bootstrap iterations. Mean absolute prediction error is unweighted mean error across all episodes. Hospital size defined as follows: small (0-250 beds), medium (251-500 beds), large (501-850 beds), and extra-large (>850 beds).