Circulation: Cardiovascular Quality and Outcomes

ORIGINAL ARTICLE

Opportunities and Challenges of Claims-Based Quality Assessment

The Case of Postdischarge β-Blocker Treatment in Patients With Heart Failure With Reduced Ejection Fraction

See Editorial by Heidenreich

Lauren Gilstrap, MD, MPH
Jonathan S. Skinner, PhD
Barbara Gladders, MS
A. James O'Malley, PhD
Amber E. Barnato, MD, MPH
Anna N. A. Tosteson, ScD
Andrea M. Austin, PhD

BACKGROUND: To combat the high cost and increasing burden of quality reporting, the Medicare Payment Advisory (MedPAC) has recommended using claims data wherever possible to measure clinical quality. In this article, we use a cohort of Medicare beneficiaries with heart failure with reduced ejection fraction and existing quality metrics to explore the impact of changes in quality metric methodology on measured quality performance, the association with patient outcomes, and hospital rankings.

METHODS AND RESULTS: We used 100% Medicare Parts A and B and a random 40% sample of Part D from 2008 to 2015 to create (1) a cohort of 295 494 fee-for-service beneficiaries with ≥1 hospitalization for heart failure with reduced ejection fraction and (2) a cohort of 1079 hospitals with ≥11 heart failure with reduced ejection fraction admissions in 2014 and 2015. We used Part D data to calculate β-blocker use after discharge and β-blocker use over time. We then varied the quality metric methodologies to explore the impact on measured performance. We then used multivariable time-to-event analyses to explore the impact of metric methodology on the association between quality performance and patient outcomes and Kendall's Tau to describe impact of quality metric methodology on hospital rankings. We found that quality metric methodology had a significant impact on measured quality performance. The association between quality performance and readmissions was sensitive to changes in methodology but the association with 1-year mortality was not. Changes in quality metric methodology also had a substantial impact on hospital quality rankings.

CONCLUSIONS: This article highlights how small changes in quality metric methodology can have a significant impact on measured quality performance, the association between quality performance and utilization-based outcomes, and hospital rankings. These findings highlight the need for standardized quality metric methodologies, better case-mix adjustment and cast further doubt on the use of utilization-based outcomes as quality metrics in chronic diseases.

Key Words: chronic disease ■ fee-for-service ■ heart failure ■ medicare ■ mortality

https://www.ahajournals.org/journal/circoutcomes

WHAT IS KNOWN

- Payers are increasingly relying on quality-based reimbursement to determine payment for hospitals and providers.
- Claims-based quality measurement has many advantages but may be sensitive to minor changes in quality metric methodology (ie, small changes in the specifications used to calculate the metric in claims data).

WHAT THE STUDY ADDS

- Quality metric methodology has a significant impact on measured quality performance.
- The association between quality performance and readmissions is sensitive to changes in quality metric methodology, but the association between quality performance and 1-year mortality is not.
- Changes in quality metric methodology have a substantial impact on hospital quality rankings and the interpretation of annual changes in hospital quality.

With the aim of improving the quality of medical care, payers are increasingly relying on measured quality performance to determine provider and hospital reimbursement. Unfortunately, the burden of measuring and reporting quality performance to various payers and regulators has become a multibillion dollar challenge.[1] In the hopes of minimizing the reporting burden, while simultaneously preserving the benefits of quality-based incentives, the Medicare Payment Advisory has advocated for claims-based quality assessment whenever possible.[2]

Claims-based quality assessment, which is the use of billing or claims data to calculate performance on predetermined metrics of quality, has many advantages. It is fast; it allows the assessment of quality performance for millions of providers, using data from millions of patients relatively quickly. It may also be less burdensome on providers and hospitals, less prone to error, and cheaper to calculate than methods that rely on data from nonstandardized electronic medical records or physician reports.

There are, of course, drawbacks. Granular, clinical, and patient-specific data are not available, and this may limit adjustment for disease severity or social determinants of health. While this may be less important across large populations,[3] a limited ability to adjust for case-mix may disproportionately affect critical-access providers and hospitals that care for socioeconomically and medically disadvantaged patients such as tertiary referral centers. Additionally, claims-based quality assessment is limited to data about things that are billed. This means that certain things, such as a providers' trustworthiness or interpersonal skills, cannot be assessed.[4] Finally, it is important to recognize that claims-based quality assessment is, fundamentally, the use of billing data for something that it was not designed for. While claims-based quality assessment may be advantageous from a cost and efficiency standpoint, it does present limitations and challenges that merit careful consideration to avoid potential unintended consequences.[5,6]

This article explores how differences in quality metric methodology, which is the set of specifications used to calculate quality metric performance in claims data (eg, numerator and denominator definitions and exclusions), can affect measured quality performance in the sample of patients with heart failure with reduced ejection fraction (HFrEF), one of the most common and costly chronic diseases among Medicare beneficiaries.[7] Specifically, we explore the impact of the postdischarge exposure window on short-term β-blocker use after discharge rates and the impact of different adherence methodologies on longer-term β-blocker use over time metrics.[8–10] We then examine how changes in quality metric methodology affect the association between quality performance and patient outcomes. Finally, we also describe how changes in quality metric methodology influence hospital rankings and the interpretation of changes in hospital quality over time. A better understanding of these challenges should inform the interpretation of claims-based quality assessment and help mitigate potential unintended consequences.

METHODS

Study Population

We used a 100% national sample of patients enrolled in both Medicare Parts A and B and a random 40% sample of Part D enrollment to create a cohort of fee-for-service beneficiaries with at least one hospitalization (index admission) for HFrEF between 2008 and 2015. Only the patient's first hospitalization for HFrEF during the study period was included. HFrEF was defined using *International Classification of Diseases (ICD) 9 and 10* codes and methodology from previously validated studies (Data Supplement).[11,12] We required 1 year of fee-for-service coverage before the index HFrEF admission to determine heart failure type, preexisting comorbidities, and exclusions. We also required 1 year of fee-for-service coverage after discharge from the index admission to determine outcomes and 3 months of Part D drug coverage before the index admission and 1 year of Part D drug coverage after discharge. We limited the sample to patients discharged alive and excluded those who underwent cardiac transplant, placement of a durable mechanical circulatory support device during admission, and those admitted from or discharged to hospice or with home inotropes were excluded. Patients with a previously placed, durable mechanical circulatory support device or a prior cardiac transplant or who spent >100 days in hospital or rehab in the year after index hospitalization were also excluded.

To explore the impact of different quality metric methodologies on hospital rankings and the change in annual hospital quality ratings, we also constructed a cohort of 1079 hospitals with ≥11 HFrEF admissions (with all patients meeting the

criteria above) in both 2014 and 2015. We then created a subgroup of hospitals in the top decile of subsequent inpatient/postacute care days (ie, patients who received care at these hospitals had more inpatient and postacute care days over the following year than the other 90% of hospitals) to assess whether they were more or less susceptible to changes in quality metric methodology.

Quality Metric Methodologies

Using existing, claims-based quality metric methodologies from the Center for Medicare and Medicaid Services and the National Quality Forum, we examined the rates of β-blocker use immediately after hospitalization (β-blocker use after discharge) for acutely decompensated heart failure and then longitudinally over the following year (β-blocker use over time).[13]

β-Blocker Use After Discharge and Varying Exposure Windows

We considered 3 different exposure windows to determine β-blocker use rates after discharge. The most commonly used postdischarge exposure window is a 30-day exposure window.[14] Therefore, we explored a 50% shorter (15 day) and a 50% longer (45 day) exposure window. During each exposure window, an individual was considered to have received the drug if there was a fill in the Part D. Since discharge to a postacute setting may delay drug fills, we stratified results based on whether patients were discharged home or to a postacute setting.

β-Blocker Use Over Time and Varying Adherence Methodologies

We also considered 2 different methodologies to calculate medication adherence over time. Proportion of days covered (PDC) is one of the most commonly used methodologies.[15] Mathematically, PDC is the number of days in the period covered by medication divided by the number of days in the period. Because this methodology does not necessarily account for days in hospital/post-acute care (during which drug exposure is unknown), we also considered an adjusted PDC, subtracting the days in hospital/postacute care from the denominator of the calculation. Based on prior work, we considered adherent to be ≥80 PDC[16,17] and stratified results based on number of days in hospital/post-acute care: <30%, 30% to 60%, and >60% of days.

Patient Outcomes

For the short-term, β-blocker use after discharge analysis, we examined the impact of exposure window duration on the association between quality performance and 90-day readmission or death. For the longer-term, β-blocker use over time analysis, we examined the impact of the adherence methodology on the association between quality performance and 1-year readmission or death.

Hospital Rankings and Annual Quality Change

To examine the association between changes in quality metric methodology and the quality rankings of hospitals, we calculated β-blocker use after discharge for all patients discharged from hospitals in 2014 and 2015 using both the 15-day and 30-day exposure window methodologies. We then calculated Kendall Tau correlation coefficient to summarize the relationship between the relative rankings of hospitals using each of the 2 quality metric methodologies. The Kendall Tau gives us the ordinal association between the 2 metrics at the hospital level. That is, whether the relative rankings of hospitals change between the 2 metrics. Next, we quantified the change in quality from 2014 to 2015, using both quality metric methodologies. Then, we determined if there was a change in the directionality of the quality change between the 2 approaches (ie, from 2014 to 2015 do both methodologies suggest an increase or decrease in quality or is the conclusion discordant different between the 2 methodologies). We did the same for the β-blocker use over time metric using both the standard PDC method and the adjusted PDC method. Finally, we separated out the hospitals with most subsequent inpatient and/or postacute care patient days (top 10%) to determine if their quality rankings were more or less susceptible to changes in quality metric methodology.

Statistical Methods

To begin, we determined the baseline rate of demographic, socioeconomic, and clinical variables using varying exposure windows and adherence definitions. To control for selection bias and confounding in the association between quality performance and readmissions/death, we adjusted for covariates reflecting: demographic characteristics, socioeconomic status, geography, medical comorbidities, presence of a previously placed implantable cardiac defibrillator, prior β-blocker exposure, number of hospitalizations in the year prior and number of days in hospital/post-acute care in year prior, for all models (full details in the Data Supplement). For the β-blocker use after discharge analysis, the exposure of interest was the receipt of a β-blocker prescription in the 15, 30, or 45-day window after discharge. Multivariable logistic regression was used to examine the impact of different exposure windows on measured drug use post-discharge (ie, quality performance) and patient outcomes. Similarly, for the β-blocker use over time analysis, we considered an exposure of ≥80 PDC versus <80 PDC, calculated using 2 different PDC methodologies as the exposure of interest. Cox proportional hazard models were used to compare the impact of different definitions of adherence on adherence rates (ie, quality performance) and patient outcomes. In the hospital-level analysis, we first calculated the Kendall Tau correlation coefficient to summarize the relationship between the relative rankings of hospitals using each quality metric methodology. We also determined if there was a change in the directionality of the quality change from 2014 to 2015 between the 2 methodologies. *P* values of <0.05 were considered significant. All analyses were performed using SAS version 9.4 (Cary, North Carolina). This study was approved by the Institutional Review Board at Dartmouth College. This article is compliant with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) reporting guideline for observational studies. Because of the sensitive nature of the data collected for this study, requests to access the dataset from qualified researchers trained in human subject confidentiality protocols

may be sent to the Research Data Assistance Center (ResDAC) available online at https://www.resdac.org.

RESULTS

The baseline characteristics of patients who filled β-blockers within 30 days of hospital discharge (most common exposure window) and those who did not are displayed Table 1. Overall, 49% percent of patients filled a β-blocker within 30 days of discharge. The mean age across both groups was 80 years. Older adults (age 85+) were less likely to have a β-blocker fill within 30 days after discharge. Just over 30% of those in both groups were dually eligible for Medicaid benefits. Sixty two percent of those with a β-blocker fill within 30 days had a β-blocker fill before admission, compared with 43% of those without a postdischarge β-blocker fill. Higher comorbidity counts were noted among those who filled a β-blocker within 30 days of discharge. In addition, the baseline characteristics of patients with

Table 1. Baseline Characteristics of Patients Admitted With HFrEF Between 2008 and 2015

	β-Blocker Use After Discharge Quality Metric		β-Blocker Use Over Time Quality Metric	
	Filled β-Blocker Within 30 d of Discharge	No Fill of β-Blocker Within 30 d of Discharge	≥80% of Days Covered* by β-Blocker	<80% of Days Covered* by β-Blocker
N (%)	144 133 (48.8)	151 361 (51.2)	107 011 (36.2)	188 483 (63.8)
Demographic characteristics				
Mean age	79.3	81.1	79.4	80.7
Age category, n (%)				
66–74	45 819 (31.8)	37 568 (24.8)	33 339 (31.2)	50 048 (26.6)
75–84	57 014 (39.6)	57 539 (38.0)	42 492 (39.7)	72 061 (38.2)
85+	41 300 (28.7)	56 254 (37.2)	31 180 (29.1)	66 374 (35.2)
Sex				
Male	68 035 (47.2)	69 046 (45.6)	149 886 (46.6)	87 195 (46.3)
Female	76 098 (54.4)	82 315 (54.4)	57 125 (46.3)	101 288 (53.7)
Race/ethnicity, n (%)				
Other	4258 (3.0)	3781 (2.5)	3081 (2.9)	4958 (2.6)
White	117 370 (81.4)	128 073 (84.6)	89 351 (83.5)	156 092 (82.8)
Black	14 371 (10.0)	12 626 (8.3)	9105 (8.5)	17 892 (9.5)
Hispanic	8134 (5.6)	6881 (4.5)	5474 (5.7)	9541 (5.1)
Socioeconomic characteristics†				
Dual eligibility, n (%)	50 053 (32.7)	51 362 (33.9)	35 928 (33.6)	65 487 (34.7)
Percentage bachelor's degree	26.7	27.2	27.2	26.8
Percentage below federal poverty line	16.0	15.5	15.5	15.9
Geography, n (%)				
Midwest	38 126 (26.5)	39 325 (26.0)	29 619 (27.7)	47 832 (25.4)
Northeast	30 164 (20.9)	32 495 (21.5)	23 856 (22.3)	38 803 (20.6)
South	56 936 (39.5)	59 830 (39.5)	39 968 (37.4)	76 798 (0.8)
West	18 907 (13.1)	19 711 (13.0)	13 568 (12.7)	25 050 (13.3)
B-Blocker use prior to index admissoon				
β-Blocker fill within 90 d of index admission (before), n (%)	65 060 (62.1)	89 563(43.0)	70 545 (65.9)	84 078 (44.6)
Number of hospital admissions in the year before index admission, n (%)				
0 admissions	67 842 (44.8)	73 278 (50.9)	55 788 (52.1)	85 332 (45.3)
1–2 admissions	61 201 (40.4)	53 359 (37.0)	39 280 (36.7)	75 280 (39.9)
≥3 admissions	22 318 (14.7)	17 496 (12.1)	11 943 (11.2)	27 871 (14.8)
Mean number of Elixhauser morbidities, n (%)				
≤3 comorbidities	66 679 (44.1)	73 727 (51.2)	55 202 (52.4)	85 204 (46.0)
≥4 comorbidities	84 682 (56.0)	70 406 (48.9)	51 809 (45.6)	103 279 (54.0)

HFrEF indicates heart failure with reduced ejection fraction.
*Calculated using standard proportion of days covered methodology.
†Percentage with bachelor's degree and % below federal poverty line are computed at the zip code tabulation area level.

and without good adherence (defined as ≥80 PDC, using the standard proportion of days covered methodology) to β-blockers during the year following admission are also displayed in Table 1. Sixty-four percent of patients had <80 PDC. The mean age was 80 across both groups and again, older adults (age 85+) were less likely have good adherence. Again, just over 30% were dually eligible for Medicaid. Sixty-six percent of those with good adherence had a β-blocker fill before their index admission while only 45% of those with <80 PDC did. Higher comorbidity counts were noted among those with <80 PDC.

The variation in postdischarge drug fill rates, calculated using different exposure windows, and separated by discharge destination (home versus post-acute care) is displayed in Figure 1. The rate of fills among those discharged home is higher than among those discharged to postacute care, but the rate of increase among those discharged to postacute care is faster, especially early on. By 15 days after discharge, 45% of those discharged home had filled a β-blocker compared with only 17% of those discharged to a postacute facility. However, between postdischarge days 15 and 30 days, the fill rate among those discharged home increased by 26% (to 57%) while the fill rate among those discharged to postacute care increased by 77% (to 30%). By 90 days, 71% of those discharged home had filled a β-blocker compared with 52% of those discharged to postacute care.

The association between quality performance (on the β-blocker after discharge metric) and patient outcomes (death and readmissions) and how these relationships vary depending on how β-blocker use after discharge is calculated is shown in Table 2. Regardless of exposure window duration and discharge destination, more β-blocker fills were consistently associated with a lower odds of death. In contrast, the association between postdischarge β-blocker fills and all-cause readmission was weaker. Filling within 15 days was associated with an increase in all-cause readmissions among those discharged home (OR, 1.029 [95% CI, 1.008–1.051]; *P*<0.007) while a first, postdischarge fill between days 30 and 45 was associated with higher odds of all-cause readmission among those discharged to postacute care (OR, 1.088 [95% CI, 1.044–1.131]; *P*<0.001 for 30 days and OR, 1.195 [95% CI, 1.143–1.249]; *P*<0.001 for 45 days). Postdischarge β-blocker fills within 30 and 45 days were not significantly associated with all-cause readmission rates among those discharged home.

The variation in drug adherence over time, calculated using both the standard and alternative PDC methodologies, and separated by number of subsequent days inpatient or in postacute care, is displayed in Figure 2. Using the standard PDC methodology, adherence rates (defined as ≥80 PDC) necessarily decreased (from 39% to 13%) as the number of days inpatient or in postacute care increased. Using an adjusted PDC methodology, adherence rates remained stable, around 45%, regardless of the number of days spent inpatient or in postacute care.

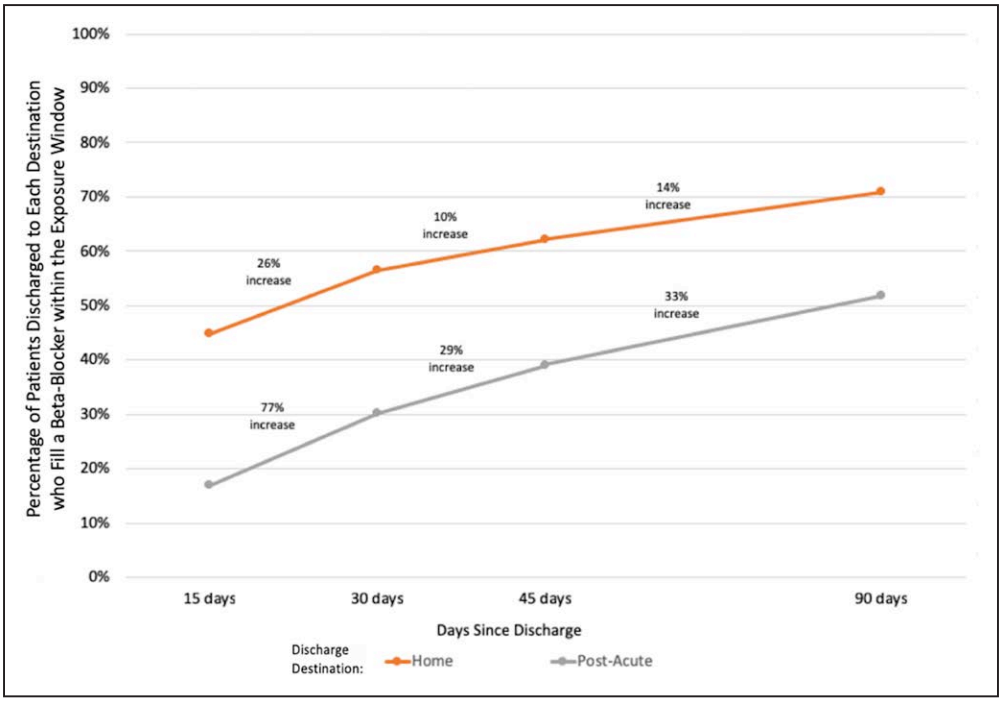

Figure 1. Variation in quality performance rates by exposure window duration and discharge destination.
This figure shows measured performance on β-blocker use after discharge quality metrics and how this varies over time and how that varies with different exposure window durations. While those discharged to postacute care have lower overall drug fill rates, the rate of increase over time (days out from discharge) is higher.

Table 2. Association Between Quality Performance and 90-Day Patient Outcomes, by Exposure Window Duration and Discharge Destination (Short-Term Quality)

		Low 95% CI	High 95% CI	*P*-Value
	Odds ratio for death			
15-d exposure window	0.864		0.888	<0.001
Discharged home	0.936	0.905	0.968	<0.001
Discharged to postacute care	0.751	0.716	0.788	<0.001
30-d exposure window	0.808	0.785	0.831	<0.001
Discharged home	0.868	0.836	0.902	<0.001
Discharged to postacute care	0.742	0.71	0.776	<0.001
45-d exposure window	0.859	0.831	0.888	<0.001
Discharged home	0.909	0.864	0.946	<0.001
Discharged to postacute care	0.819	0.780	0.860	<0.001
	Odds ratio for readmission			
15-d exposure window	1.019	1.000	1.038	0.052
Discharged home	1.029	1.008	1.051	0.007
Discharged to postacute care	0.991	0.951	1.034	0.685
30-d exposure window	1.038	1.016	1.059	<0.001
Discharged home	1.019	0.995	1.044	0.128
Discharged to postacute care	1.088	1.044	1.131	<0.001
45-d exposure window	1.064	1.038	1.091	<0.001
Discharged home	1.011	0.982	1.041	0.467
Discharged to postacute care	1.195	1.143	1.249	<0.001

The association between quality metric performance (on β-blocker use after discharge metric) and patient outcomes (death and readmissions) and how these relationships vary depending on how β-blocker use over time is calculated is shown in Table 3. We found a consistently lower risk of death among those adherent to β-blockers, regardless of the quality metric methodology used and/or number of days in hospital/postacute care. The association with all-cause readmissions was weaker, but good adherence appeared to be associated with a lower risk of readmission. Notably, however, among patients with >60 days in hospital/postacute care, this association barely achieved significance with the standard PDC methodology (hazard ratio, 0.955 [95% CI, 0.912–1.000]; *P*=0.045) and did not reach significance using the adjusted PDC methodology (hazard ratio, 0.9 [95% CI, 0.954–1.017]; *P*=0.365).

The comparison of hospital rankings using 2 different β-blocker use after discharge exposure windows, 15- and 30-days, yielded a Kendall Tau of 0.63 (*P*<0.001). Among hospitals whose patients had the most subsequent inpatient and postacute care days (top decile of hospitals), the Kendall Tau was 0.61 (*P*<0.001). This suggests only a modest relationship between the rank order determined using the 2 different methodologies and a weakening of the relationship among hospitals whose patients require more inpatient or in postacute care days. The Kendall Tau comparing hospital rankings on the β-blocker use over time metric using the 2 PDC methodologies was 0.67 (*P*<0.001) and for hospitals in the top decile for subsequent inpatient and postacute care days, the Kendall Tau was 0.66 (*P*<0.001) again suggesting only a modest relationship and a slight weakening of the relationship among hospitals whose patients require high levels of subsequent inpatient or postacute care.

Between 2014 and 2015, we found that 20% of hospitals had discordant quality change, that is, measured quality improved using one quality metric methodology but declined using the other (Table IA in the Data Supplement). A comparison of the within-hospital change in quality between 2014 and 2015 using a 15- and 30-day exposure window is displayed in Figure 3A. The discordant hospitals are evident in the upper left and lower right-hand quadrants. For hospitals that fall into these 2 quadrants, the year to year change in their quality metric performance changed direction between the 2 quality metric methodologies. For the upper left quadrant, these hospitals improved in quality from 2014 to 2015 when the 30-day exposure window was used and declined in quality when the 15-day exposure window was used. The reverse is true for the hospitals in the lower right quadrant. Similarly, between 2014 and 2015, 18% of hospitals had discordant quality changes on the β-blocker use over time metric when different PDC methodologies were used (Table IB in the Data Supplement) A comparison of the change in quality between 2014 and 2015 using the standard PDC method and the adjusted PDC method is displayed in Figure 3B. Hospitals in the upper left quadrant improved between 2014 and 2015 when the alternative PDC method was used and declined in quality when the standard PDC was used. The reverse is again true for the lower right quadrant.

DISCUSSION

This article highlights how seemingly small changes in claims-based quality metric methodology can have a significant impact on quality performance, the association between quality performance and patient outcomes, and hospital quality rankings. Specifically, we examine how different methodologies for calculating quality performance, namely different exposure window durations for β-blocker use after discharge metrics and adherence methodologies for β-blocker use over time metrics impact the interpretation of quality for patients with HFrEF. We found that performance on both quality measures was sensitive to metric methodology and that the relationship between quality performance and readmissions was also sensitive to metric methodology, whereas the association with mortality was not. Finally, hospital quality rankings were sensitive

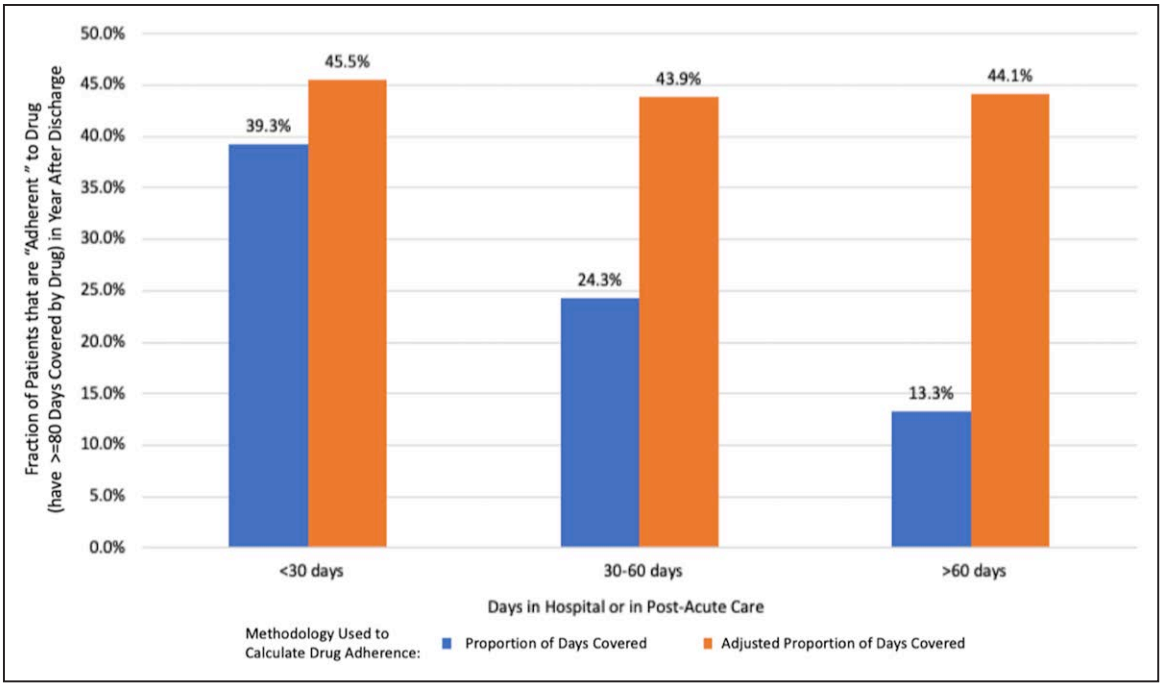

Figure 2. Variation in quality performance by adherence methodology and days spent in hospital/postacute care.
This figure shows measured performance on β-blocker use over time quality metrics, and how this varies based on the method used to calculate drug adherence over time. Adherence calculated using the standard proportion of days covered (PDC) method is orange. Adherence calculated using the adjusted PDC method is blue. Both methods determine the number of days the patient is covered by drug. The standard method does not consider days in hospital/postacute care. The adjusted method subtracts these days out of the number of days in the period (denominator).

to quality metric methodology and changing the methodology resulted in a substantial reordering of hospital quality ranks and a change in the interpretation of annual quality change for 18% to 20% of hospitals. These findings raise several important points: (1) the need for standard, universal methodologies for quality assessment and hospital ranking/reimbursement; (2) the importance of accounting for biases introduced with each quality metric methodology and the need for better risk adjustment; (3) the consistency of association between better quality performance (regardless of methodology) and lower mortality, in contrast to the unstable association between better quality and utilization-based outcomes, that is, readmissions.

Our first key finding is that for claims-based quality assessment to be equitable, particularly for the purposes of reimbursement, standard, universal methods for calculating quality performance are necessary. To date, this has been challenging. In fact, one of the reasons quality reporting has been so burdensome is that, at present, different payers use different metrics and methodologies.[1,18] There have been efforts to simplify and standardize quality reporting across payers.[19,20] For example, the Core Quality Measures Collaborative, which included Center for Medicare and Medicaid Services and about 70% of private payers, agreed on a standard set of quality measures in 2015. Unfortunately, due to inadequate data systems and lack of electronic records, the initiative lost momentum. However, bypassing the need for provider reporting and/or electronic records, standardized claims-based quality assessment, across public and private payers, has the potential to succeed where prior initiatives have failed.

The second key finding is the importance of accounting for biases introduced by the methodology used to calculate quality performance and the need for better risk adjustment to avoid disadvantaging providers and hospitals that care for socioeconomically disadvantaged or medically complex patients who require large amounts of subsequent inpatient and/or postacute care. Because claims-based quality assessment relies on pharmacy claims, which in turn require an exposure window, immortal time bias may be introduced. Immortal time is the period of follow-up during which, by design, death cannot occur and so this bias will always favor a treatment that is measured over a time-interval (eg, drug exposure)[21] unless treatment group members can be sensibly matched at the end of the treatment group qualifying period. In this study, we highlight the trade-off in claims-data between the duration of the exposure window and the magnitude of the immortal time bias. Short exposure windows miss fills, lowering the observed quality performance, but capturing more outcomes, minimizing the immortal time bias—and vice versa. Therefore, one might conclude that shorter exposure windows are better; however, among patients discharged to postacute care, where pharmacy claims are silent, drug fills are delayed and so quality perfor-

Table 3. Association Between Quality Performance and 1-Year Patient Outcomes, by Adherence Calculation Methodology and Days in Hospital/Postacute Care (Long-Term Quality)

		Low CI	High CI	*P* Value
	Hazard ratio for death			
PDC	0.616	0.607	0.626	<0.001
<30 d in postacute care/hospital	0.611	0.601	0.621	<0.001
30–60 d in postacute care/hospital	0.676	0.646	0.707	<0.001
>60 d in postacute care/hospital	0.665	0.611	0.724	<0.001
Adjusted PDC	0.583	0.575	0.591	<0.001
<30 d in postacute care/hospital	0.545	0.536	0.554	<0.001
30–60 d in postacute care/hospital	0.741	0.715	0.769	<0.001
>60 d in postacute care/hospital	0.856	0.812	0.902	<0.001
	Hazard ratio for readmission			
PDC	0.778	0.771	0.786	<0.001
<30 d in postacute care/hospital	0.857	0.847	0.866	<0.001
30–60 d in postacute care/hospital	0.925	0.901	0.95	<0.001
>60 d in postacute care/hospital	0.955	0.912	1.000	0.045
Adjusted PDC	0.874	0.865	0.882	<0.001
<30 d in postacute care/hospital	0.862	0.853	0.872	<0.001
30–60 d in postacute care/hospital	0.949	0.927	0.971	<0.001
>60 d in postacute care/hospital	0.985	0.954	1.017	0.365

PDC indicates proportion of days covered by drug (β-blocker).

mance declines. Thus, while shorter exposure windows may minimize immortal time bias, they may also unfairly penalize providers and hospitals whose patients require postacute care.

This finding is confirmed at the hospital level. When we compare how often a hospital goes from improving in quality to declining in quality based solely on a change in how quality performance is calculated, we find that this occurs 18% to 20% of the time. Moreover, for hospitals in the top decile of subsequent inpatient and/or postacute care days (arguably those with the most socioeconomically disadvantaged or medically complex patients), this discordance occurs 28% to 30% of the time. This underscores the need for better risk adjustment algorithms to avoid unfairly penalizing critical access and/or tertiary referral centers who care for many of these more vulnerable patient populations. While standardization of quality metric methodologies across payers is a necessary first step, novel approaches such as including functional limitations or patient-reported outcomes in the assessment of quality may help level the playing field between institutions with different patient populations.[22–24]

Finally, regardless of the quality metric methodology used, we find a clear association between better quality performance and decreased mortality but not utilization-outcomes such as readmissions. There is a growing body of literature calling into question the use of utilization-based outcomes, in the assessment of quality, particularly for chronic diseases.[25,26] These results confirm and extend those findings, demonstrating with real-world data, how the association between quality performance and readmissions is sensitive to even small changes in quality metric methodology, whereas the association with mortality is not. Moreover, we find that this issue is particularly problematic in the assessment of early readmissions, and it is exacerbated among patients discharged to postacute care or among those who require many subsequent days inpatient or in postacute care during the following year. Thus, readmissions may be an even less reliable measure of quality among providers and hospitals that care for these patients.

Limitations

This study is limited by the nature of claims data and the absence of granular, clinical data. This limits our ability to further adjust for confounding in our models, though we make every effort to do so with as many measured variables as possible. However, it is possible that there are additional factors that affect adherence and clinicians' decision to not prescribe guideline-directed medical therapy that cannot be completely controlled for. As a result, this article can only comment on associations and makes no attempt to draw causal conclusions. Second, this study explores the limitations of claims-based quality assessment using Medi-

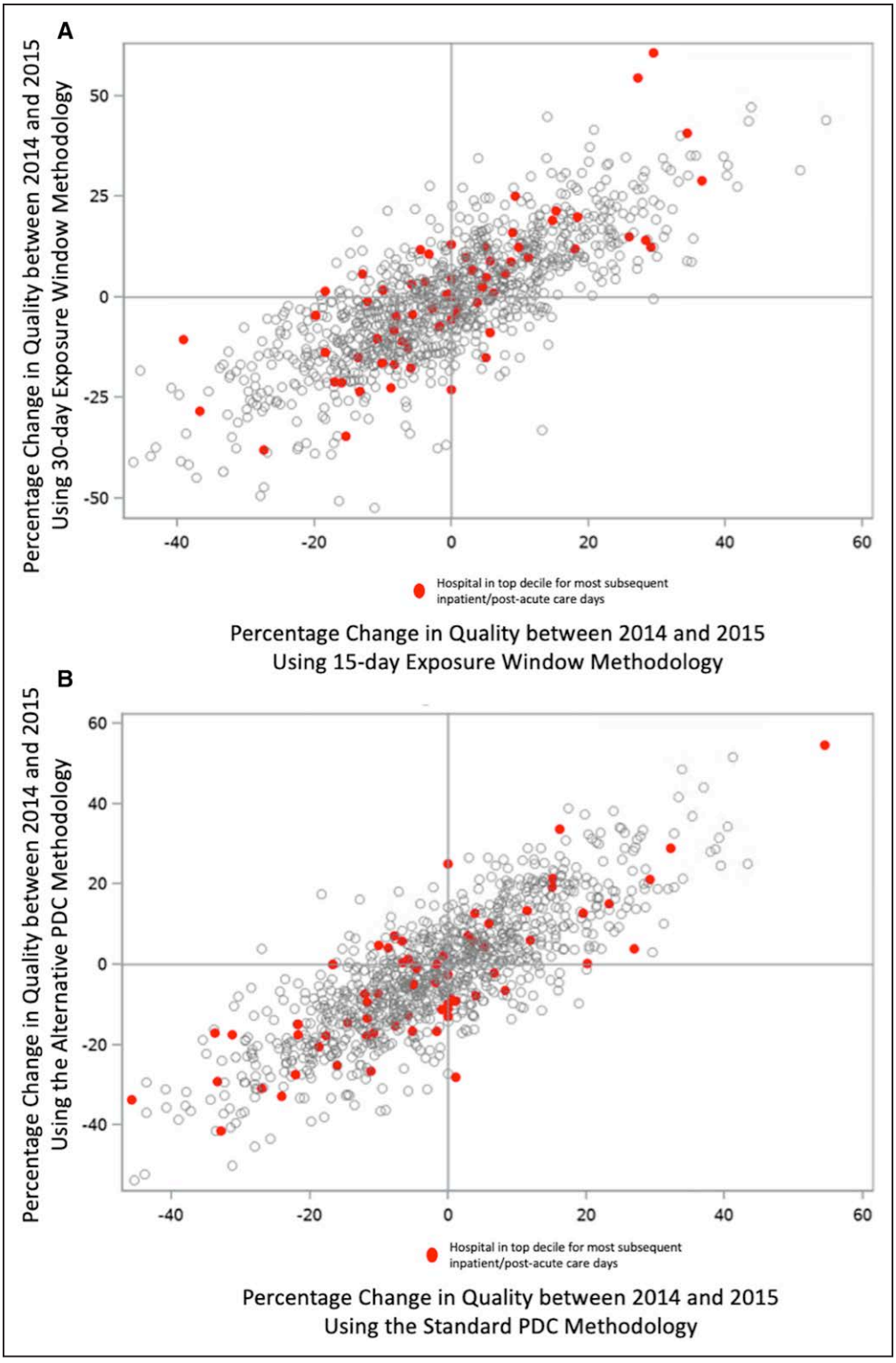

Figure 3. Change in annual hospital quality from 2014 to 2015 using different quality metric methodologies.
A, Comparison of the change from 2014 to 2015 in hospital performance on the β-blocker use after discharge quality metric calculated with a 15-d exposure window compared with the change from 2014 to 2015 in hospital performance on the β-blocker use after discharge quality metric calculated with a 30-d exposure window. **B**, Comparison of the change from 2014 to 2015 in hospital performance on the β-blocker use over time quality metric calculated with the standard proportion of days covered (PDC) method compared with the change from 2014 to 2015 in hospital performance on the β-blocker use over time quality metric calculated using the alternative PDC method.

care data from older adults. While there is no reason the results should not apply to younger populations as well, care should be used in extrapolation. Third, it should be noted that hospital ranking (as opposed to a threshold) may be a poor measure of quality, particularly when the distribution of performance is tight.

Finally, the data used in this study is real-world Medicare claims. However, this study is illustrative. As such, the tables/figures are only meant to highlight challenges of different quality assessment methodologies, rather than measure current quality performance.

Conclusions

This article highlights how small changes in claims-based quality metric methodology can have a significant impact on measured quality performance, the association between quality performance, and patient outcomes and hospital rankings. While claims-based quality assessment has many advantages, failure to adequately understand its limitations risks unintended consequences, particularly among providers and hospitals who care for large numbers of socioeconomically disadvantaged and/or medically complex patients. These findings highlight the need for standardized quality metric methodologies, better methods for case-mix adjustment and raise further questions about the use of utilization-based outcomes, that is, readmissions, as a metric of quality in chronic diseases.

ARTICLE INFORMATION

Received September 18, 2019; accepted January 15, 2020.

The Data Supplement is available at https://www.ahajournals.org/doi/suppl/10.1161/CIRCOUTCOMES.119.006180.

Correspondence

Lauren Gilstrap, MD, MPH, Dartmouth-Hitchcock Medical Center, The Dartmouth Institute, Geisel School of Medicine, 1 Medical Center Rd, Lebanon, NH 03766. Email lauren.g.gilstrap@dartmouth.edu

Affiliations

Dartmouth-Hitchcock Medical Center, Heart and Vascular Center, Lebanon, NH (L.G., B.G.). The Dartmouth Institute for Health Policy and Clinical Practice (L.G., J.S.S., A.J.O., A.E.B., A.N.A.T., A.M.A.) and Department of Biomedical Data Science (A.J.O.), Geisel School of Medicine at Dartmouth, Lebanon, NH. Department of Medicine, Dartmouth-Hitchcock Medical Center, Lebanon, NH (L.G.).

Sources of Funding

Dr Gilstrap is supported by K23HL142835 from the National Heart Lung and Blood Institute. Drs. Austin, Skinner, Barnato, and O'Malley are supported, in part, by P01AG019783 from the National Institute on Aging. A.N.A. Tosteson is supported by UL1TR001086 from the National Center for Advancing Translational Sciences.

Disclosures

Dr Skinner is a consultant for Sutter Health, an investor in Dorsata, Inc, and was a consultant for the Government of Singapore. The other authors report no conflicts.

REFERENCES

1. Casalino LP, Gans D, Weber R, Cea M, Tuchovsky A, Bishop TF, Miranda Y, Frankel BA, Ziehler KB, Wong MM, Evenson TB. US physician practices spend more than $15.4 billion annually to report quality measures. *Health Aff (Millwood)*. 2016;35:401–406. doi: 10.1377/hlthaff.2015.1258
2. MedPAC. Report to Congress: Medicare Payment Policy. Medicare Payment Advisory Comission. 2019. http://www.medpac.gov/docs/default-source/reports/mar19_medpac_entirereport_sec.pdf. Accessed January 16, 2020.
3. Ross JS, Siu AL. The importance of population-based performance measures. *Health Serv Res*. 2007;42(1 pt 1):1–6. doi: 10.1111/j.1475-6773.2006.00693.x
4. Birkhäuer J, Gaab J, Kossowsky J, Hasler S, Krummenacher P, Werner C, Gerger H. Trust in the health care professional and health outcome: a meta-analysis. *PLoS One*. 2017;12:e0170988. doi: 10.1371/journal.pone.0170988
5. Gupta A, Allen LA, Bhatt DL, Cox M, DeVore AD, Heidenreich PA, Hernandez AF, Peterson ED, Matsouaka RA, Yancy CW, Fonarow GC. Association of the hospital readmissions reduction program implementation with readmission and mortality outcomes in heart failure. *JAMA Cardiol*. 2018;3:44–53. doi: 10.1001/jamacardio.2017.4265
6. Wadhera RK, Joynt Maddox KE, Wasfy JH, Haneuse S, Shen C, Yeh RW. Association of the hospital readmissions reduction program with mortality among medicare beneficiaries hospitalized for heart failure, acute myocardial infarction, and pneumonia. *JAMA*. 2018;320:2542–2552. doi: 10.1001/jama.2018.19232
7. Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, Ikonomidis JS, Khavjou O, Konstam MA, Maddox TM, Nichol G, Pham M, Piña IL, Trogdon JG; American Heart Association Advocacy Coordinating Committee; Council on Arteriosclerosis, Thrombosis and Vascular Biology; Council on Cardiovascular Radiology and Intervention; Council on Clinical Cardiology; Council on Epidemiology and Prevention; Stroke Council. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. *Circ Heart Fail*. 2013;6:606–619. doi: 10.1161/HHF.0b013e318291329a
8. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr, Drazner MH, Fonarow GC, Geraci SA, Horwich T, Januzzi JL, Johnson MR, Kasper EK, Levy WC, Masoudi FA, McBride PE, McMurray JJ, Mitchell JE, Peterson PN, Riegel B, Sam F, Stevenson LW, Tang WH, Tsai EJ, Wilkoff BL; WRITING COMMITTEE MEMBERS; American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. *Circulation*. 2013;128:e240–e327. doi: 10.1161/CIR.0b013e31829e8776
9. Bonow RO, Ganiats TG, Beam CT, Blake K, Casey DE Jr, Goodlin SJ, Grady KL, Hundley RF, Jessup M, Lynn TE, Masoudi FA, Nilasena D, Pina IL, Rockswold PD, Sadwin LB, Sikkema JD, Sincak CA, Spertus J, Torcson PJ, Torres E, Williams MV, Wong JB, ACCF/AHA Task Force on Performance Measures, Peterson ED, Masoudi FA, DeLong E, Erwin JP 3rd, Fonarow GC, Goff DC Jr, Grady KL, Green LA, Heidenreich PA, Jenkins KJ, Loth A, Shahian DM; American Academy of Family Physicians; American Academy of Hospice and Palliative Medicine; American Nurses Association; American Society of Health-System Pharmacists; Heart Rhythm Society; Society of Hospital Medicine. ACCF/AHA/AMA-PCPI 2011 performance measures for adults with heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Performance Measures and the American Medical Association-Physician Consortium for Performance Improvement. *J Am Coll Cardiol*. 2012;59:1812–1832. doi: 10.1016/j.jacc.2012.03.013
10. Quality Measures. MACRA/MIPS: Quality Payment Program. 2017. https://qpp.cms.gov/mips/explore-measures/quality-measures?py=2019#measures. Accessed January 16, 2020.
11. Li Q, Glynn RJ, Dreyer NA, Liu J, Mogun H, Setoguchi S. Validity of claims-based definitions of left ventricular systolic dysfunction in Medicare patients. *Pharmacoepidemiol Drug Saf*. 2011;20:700–708. doi: 10.1002/pds.2146
12. Loop MS, Van Dyke MK, Chen L, Brown TM, Durant RW, Safford MM, Levitan EB. Comparison of length of stay, 30-day mortality, and 30-day readmission rates in medicare patients with heart failure and with reduced versus preserved ejection fraction. *Am J Cardiol*. 2016;118:79–85. doi: 10.1016/j.amjcard.2016.04.015
13. Center for Medicare and Medicaid Services. Quality ID #8 (NQF 0083): Heart Failure (HF): Beta-Blocker Therapy for Left Ventricular Systolic Dysfunction (LVSD). 2019. https://qpp.cms.gov/docs/QPP_quality_mea-

sure_specifications/CQM-Measures/2019_Measure_008_MIPSCQM.pdf. Accessed January 16, 2020.
14. Dayoub EJ, Seigerman M, Tuteja S, Kobayashi T, Kolansky DM, Giri J, Groeneveld PW. Trends in platelet adenosine diphosphate P2Y12 receptor inhibitor use and adherence among antiplatelet-naive patients after percutaneous coronary intervention, 2008-2016. *JAMA Intern Med*. 2018;178:943–950. doi: 10.1001/jamainternmed.2018.0783
15. National Quality Foundation. Proportion of Days Covered (PDC). 2019. http://www.qualityforum.org/Measure_Details.aspx?id=9993. Accessed January 16, 2020.
16. Centers for Disease Control. Calculating Proportion of Days Covered (PDC) for Antihyperive and Antidiabetic Medications. 2015. https://www.cdc.gov/dhdsp/docs/med-adherence-evaluation-tool.pdf. Accessed January 16, 2020.
17. Haynes RB. A critical review of the "determinants" of patient compliance with therapeutic regimens. In: Sackett DL and Haynes RB, eds. *Compliance With Therapeutic Regimens*. Baltimore, MD: Johns Hopkins Univeristy Press; 1976: 26–39.
18. National Quality Forum. MAP 2018 Considerations for Implementing Measures in Federal Programs - PAC-LTC. 2018. https://www.qualityforum.org/Publications/2018/02/MAP_2018_Considerations_for_Implementing_Measures_in_Federal_Programs_-_PAC-LTC.aspx. Accessed January 16, 2020.
19. Verma S. Remarks by Administrator Seema Verma at the Health Care Payment Learning and Action Network (LAN) Fall Summit. 2017. https://www.cms.gov/newsroom/fact-sheets/speech-remarks-administrator-seema-verma-health-care-payment-learning-and-action-network-lan-fall. Accessed January 16, 2020.
20. CMS. Core Measures. 2017. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/Core-Measures. Accessed January 16, 2020.
21. Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ*. 2010;340:b5087. doi: 10.1136/bmj.b5087
22. Governemt Accountability Office. Medicare Advantage: Benefits and Challenges of Payment Adjustments Based on Beneficiaries' Ability to Perform Daily Tasks. 2018. https://www.gao.gov/assets/700/694425.pdf. Accessed January 16, 2020.
23. Heidenreich PA. Patient-reported outcomes: the future of heart failure care. *JACC Heart Fail*. 2019;7:875–877. doi: 10.1016/j.jchf.2019.06.006
24. Chandra A, Vaduganathan M, Lewis EF, Claggett BL, Rizkala AR, Wang W, Lefkowitz MP, Shi VC, Anand IS, Ge J, Lam CSP, Maggioni AP, Martinez F, Packer M, Pfeffer MA, Pieske B, Redfield MM, Rouleau JL, Van Veldhuisen DJ, Zannad F, Zile MR, McMurray JJV, Solomon SD; PARAGON-HF Investigators. Health-related quality of life in heart failure with preserved ejection fraction: the PARAGON-HF Trial. *JACC Heart Fail*. 2019;7:862–874. doi: 10.1016/j.jchf.2019.05.015
25. McWilliams JM, Chernew ME, Landon BE. Medicare ACO program savings not tied to preventable hospitalizations or concentrated among high-risk patients. *Health Aff (Millwood)*. 2017;36:2085–2093. doi: 10.1377/hlthaff.2017.0814
26. Gilstrap LG, Chernew ME, Nguyen CA, Alam S, Bai B, McWilliams JM, Landon BE, Landrum MB. Association between clinical practice group adherence to quality measures and adverse outcomes among adult patients with diabetes. *JAMA Netw Open*. 2019;2:e199139. doi: 10.1001/jamanetworkopen.2019.9139