HSR Health Services Research

© Health Research and Educational Trust
DOI: 10.1111/1475-6773.12587
BEST OF THE 2016 ACADEMYHEALTH ANNUAL RESEARCH MEETING

Reliability of 30-Day Readmission Measures Used in the Hospital Readmission Reduction Program

Michael P. Thompson, Cameron M. Kaplan, Yu Cao, Gloria J. Bazzoli, and Teresa M. Waters

Objective. To assess the reliability of risk-standardized readmission rates (RSRRs) for medical conditions and surgical procedures used in the Hospital Readmission Reduction Program (HRRP).

Data Sources. State Inpatient Databases for six states from 2011 to 2013 were used to identify patient cohorts for the six conditions used in the HRRP, which was augmented with hospital characteristic and HRRP penalty data.

Study Design. Hierarchical logistic regression models estimated hospital-level RSRRs for each condition, the reliability of each RSRR, and the extent to which socioeconomic and hospital factors further explain RSRR variation. We used publicly available data to estimate payments for excess readmissions in hospitals with reliable and unreliable RSRRs.

Principal Findings. Only RSRRs for surgical procedures exceeded the reliability benchmark for most hospitals, whereas RSRRs for medical conditions were typically below the benchmark. Additional adjustment for socioeconomic and hospital factors modestly explained variation in RSRRs. Approximately 25 percent of payments for excess readmissions were tied to unreliable RSRRs.

Conclusions. Many of the RSRRs employed by the HRRP are unreliable, and one quarter of payments for excess readmissions are associated with unreliable RSRRs. Unreliable measures blur the connection between hospital performance and incentives, and threaten the success of the HRRP.

Key Words. Readmissions, reliability, risk-adjustment

Readmissions within 30 days of hospital discharge represent a substantial and potentially preventable burden on the health care system (Jencks, Williams, and Coleman 2009). To incentivize hospitals to reduce hospital readmissions, the Affordable Care Act (ACA) established the Hospital Readmission Reduction Program (HRRP), which levies financial penalties against hospitals with higher than expected readmission rates (Axon and Williams 2011; Centers for Medicare & Medicaid Services 2014). Initially, the HRRP focused on acute myocardial infarction (AMI), congestive heart failure (CHF), and pneumonia (PN); more recently, it has expanded to include readmissions for chronic obstructive pulmonary disease (COPD), total hip and/or knee arthroplasty (THKA), and coronary artery bypass graft surgery (CABG). In the first year of the HRRP, 2,217 hospitals were penalized more than $280 million, and through FY 2017, the HRRP has penalized hospitals almost $1.9 billion in total.

The 30-day readmission measures created by the Centers for Medicare and Medicaid Services (CMS) are risk adjusted to account for case mix differences between hospitals (Keenan et al. 2008; Bratzler et al. 2011; Krumholz et al. 2011). These measures assume that, after risk adjustment, any remaining hospital-level variation in 30-day readmission rates are due to underlying differences in hospital quality (Normand, Glickman, and Gatsonis 1997). One issue of particular concern is whether the CMS measures are able to discriminate systematic differences in hospital readmissions from statistical noise, referred to as measure reliability. Measure reliability is determined by the between-hospital variation in event rates (*the signal*) and the within-hospital variation in event rates (*the noise*) (Adams 2009; Adams et al. 2010). Reliability (*R*) is then estimated as $R = signal/(signal + noise)$, and it ranges from zero to one, with reliability increasing as this ratio approaches one. High reliability indicates that most of the observed variation in risk-standardized readmission rates (RSRRs) is due to systematic differences between hospitals, while low reliability indicates that most of the variation in RSRRs is due to random statistical variation.

Unreliable measures pose a potential threat to the financial incentives used in the HRRP. If readmission measures are unreliable, excess readmissions tied to hospitals may be the result of random chance, rather than truly higher or lower than expected readmission rates (i.e., increased false positives and false negatives). Because reliability and sample size are directly related, low volumes will be particularly vulnerable to statistical noise. This may elicit

Address correspondence to Michael P. Thompson, Ph.D., Department of Preventive Medicine, University of Tennessee Health Science Center, 66 N. Pauline, Suite 633, Memphis, TN 38163; e-mail: mthompson@uthsc.edu. Cameron M. Kaplan, Ph.D., and Teresa M. Waters, Ph.D., are with the Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, TN. Yu Cao, M.S., is with the Virginia Commonwealth University, Zion Crossroads, VA. Gloria J. Bazzoli, Ph.D., is with the Department of Health Administration, School of Allied Health Professions, Virginia Commonwealth University, Richmond, VA.

differential performance and responses to the HRRP across hospitals of different volumes. Noisy readmission measures may mask deficiencies and improvement in hospital readmission rates, distorting the link between hospital performance and financial incentives. Thus, unreliable measures may blur the relationship between HRRP financial penalties and hospital performance, and discourage quality improvement efforts. This may limit the extent to which HRRP achieves its primary objectives of reduced rates of hospital readmissions.

Many also argue that inadequate risk-adjustment unfairly penalizes hospitals treating disproportionately more minority or low-income individuals. (Joynt and Jha 2013; Gilman et al. 2014, 2015; Gu et al. 2014). Socioeconomic and hospital factors have been shown to predict hospital readmissions, above and beyond patient comorbidities (Joynt, Orav, and Jha 2011; Brown et al. 2014; Herrin et al. 2014). However, it is unclear how additional adjustment for these factors affects the reliability of RSRRs. The reliability of measures often declines when adding variables in risk adjustment, because the added factors account for more of the between-hospital variation in RSRRs previously assumed to be differences in quality. While many studies have examined the benefits of adjustment for these factors (Blum et al. 2014; Hu, Gonsahn, and Nerenz 2014; Bernheim et al. 2016; Glance et al. 2016), none have examined the potential loss that arises from additional adjustment in terms of reduced measure reliability.

In this study, we assessed the reliability of 30-day readmission measures used in the HRRP: AMI, CHF, PN, COPD, THKA, and CABG. We compared the reliability of hospital RSRRs to a commonly used benchmark for group-level comparisons (Scholle et al. 2008; Adams 2009; Krell et al. 2014), and estimate the hospital volumes required to meet this benchmark, and the proportion of sample hospitals that exceeded these volume thresholds. We also explored the extent to which socioeconomic and hospital factors accounted for the explained variation in RSRRs, beyond comorbidities already used in risk adjustment.

The HRRP requires a minimum 3-year hospital case volume of 25 cases for each condition to be eligible for inclusion in the financial penalty calculation. Since case volume varies by condition and hospital, each hospital may be eligible for a different set of measures. Therefore, we also examine how many hospitals met the eligibility criteria for all six or fewer than six measures, and the proportion of hospitals that met the benchmark for reliability on all measures for which they were assessed.

Finally, we sought to understand how measure reliability relates to the financial penalties allocated by the HRRP. Financial penalty rates are estimated as the sum of diagnosis-related group (DRG) payments attributed to excess readmissions for HRRP conditions, divided by DRG payments for all conditions, with penalty rates capped at certain values in each HRRP year. Excess readmissions, which form the numerator of this calculation, are determined using the RSRRs, and if the RSRRs are unreliable, this could be the result in the penalty rates being influenced by statistical noise rather than hospital differences in quality of care. Therefore, we estimated the amount and proportion of DRG payments attributed to excess readmissions for HRRP conditions that were tied to unreliable RSRRs. This analysis will provide insights on the extent to which the computation of HRRP penalties may be affected by random statistical variation.

METHODS

Data Sources

For this analysis, we pooled all discharges from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID) for six populous states (AR, FL, IA, MA, NY, WA) from 2011 to 2013 for each of the conditions covered by HRRP: AMI, CHF, PN, COPD, THKA, and CABG. We selected these states because the discharge records contain admission linkage variables, which can be used to calculate the days between successive hospital admissions for the same patient. We augmented patient-level data from HCUP with data from the 2011 American Hospital Association (AHA) Annual Survey to obtain hospital characteristics. We abstracted publically available data from CMS 2017 Inpatient Prospective Payment System (IPPS) Final Rule tables, including base operating and capital payments and HRRP Supplemental Data Files (Centers for Medicare & Medicaid Services 2016).

Patient Cohorts

We assembled condition-specific cohorts using published criteria developed by CMS using HCUP data based on published ICD-9 primary diagnosis codes to identify index admissions for patients with AMI, CHF, PN, COPD, THKA, and CABG (QualityNet 2016). Our analyses were limited to patients who were aged 65 and older, had Medicare listed as their primary payer, were discharged alive, were not transferred to another acute care hospital, and did not leave against medical advice. We linked HCUP data to AHA Annual Survey data through HCUP crosswalks linking patients to AHA hospital identifiers. Publicly available CMS data were linked using Medicare hospital provider IDs, which were also available in the AHA Annual Survey data. We then excluded patients if they were missing data on patient or hospital covariates. Lastly, per HRRP guidelines, we excluded patients treated in hospitals with fewer than 25 cases during the 2011–2013 period.

30-Day Readmissions

For each condition-specific cohort, we used the admission linkage variable provided by HCUP to calculate the days between successive admissions. Using the published CMS algorithm, we identified hospital admissions as readmissions if they occurred within 30 days of discharge for an identified index admission. Possible planned readmissions were not counted as readmissions.

Covariates

Patient characteristics were obtained from HCUP and included gender (female vs male), age, race/ethnicity (white, black, Hispanic, or other), dual eligible status, and median income quartile (as determined by HCUP). We also identified patient comorbidities used in CMS risk adjustment for each condition using publicly available ICD-9 coding (QualityNet 2016). Hospital characteristics from the AHA annual survey included teaching status (major, minor, or nonteaching), bed size (<100 beds, 100–300 beds, or +300 beds), member of hospital system, for-profit versus non-profit status, and rural versus urban geographic location. We also categorized hospitals as safety-net hospitals if they were in the highest quartile of disproportionate share index percentage (Gilman et al. 2015).

From the CMS FY 2017 IPPS Final Rule tables, we abstracted data on national adjusted operating standardized amounts and capital standard federal payment rates. We summed the average operating amount and capital payment rates to represent the base DRG cost amount used to standardize DRG payments, which amounted to $5,980.17. From the HRRP supplemental data, we abstracted the hospital-specific DRG weights for each condition, which was used to estimate the average DRG weight for each condition. By multiplying the sum of condition-specific DRGs to the base DRG cost amount, we can estimate the hospital payments associated with admissions for a specific condition.

Risk-Standardized Readmission Rates (RSRRs)

Following CMS methodology, we used hierarchical logistic regression models to estimate the predicted to expected number of readmissions for each hospital and condition, given their hospital case mix (i.e., the P/E ratio or excess readmission ratio); this ratio was then multiplied by the observed readmission rate for the entire condition-specific cohort. All models were adjusted for the comorbidities used in the CMS risk-adjustment models.

Reliability Analyses

We calculated reliability as the ratio of between-hospital variation in readmission rates (*the signal*) to the sum of between- (*the signal*) and within-hospital variation in event rates (*the noise*), that is, *signal*/(*signal* + *noise*). The between-hospital variation was the variance in hospital random intercepts generated from the hierarchical logistic regression models used to estimate RSRRs for each condition, and it is thus uniform for all hospitals. The within-hospital variation was calculated as the standard error of a proportion, or $\sqrt{[p(1-p)/n]}$, where p represents the average probability of a readmission within a given hospital (i.e., observed hospital readmission rate), and n is the hospital case volume (Adams 2009). Because both the observed readmission rate and case volume varies by hospital, the within-hospital variation varies by hospital. As such, each hospital has specific estimate of reliability. For each condition, we plotted the hospital-specific estimates of reliability against the hospital case volume and overlaid a fitted Loess curve to illustrate the overall volume-reliability relationship.

Median hospital-level reliability (R) was estimated for each condition, and it was compared to a commonly used benchmark for acceptable reliability, defined as $R = 0.70$ for group-level comparisons (Scholle et al. 2008; Adams 2009; Krell et al. 2014). The resulting estimate of reliability can be interpreted as the proportion of variation in readmission rates that is attributed to systematic differences between hospitals, which is assumed to be related to hospital quality. We estimated the average 3-year volume threshold required to meet acceptable reliability and the proportion of hospitals that exceeded this volume threshold. To assess how patient-level socioeconomic factors and hospital factors account for hospital-level event rate variation that is assumed to be differences in quality, we added socioeconomic factors (patient race, median income quartile, and dual eligible status) and hospital factors to the risk-adjustment models in a stepwise fashion. With each addition, we estimated the median reliability, the 3-year volume threshold required to meet acceptable reliability, and the proportion of hospitals exceeding the volume threshold. The arithmetic difference in reliability between two models represents the proportion of variation in readmission rates explained by the added variables.

Next, we compared measure eligibility and reliability within each hospital to describe how many hospitals meet the benchmark in all measures for which they are eligible. For each hospital, we summed the number of measures (AMI, CHF, PN, COPD, THKA, and CABG) that were eligible for inclusion into the HRRP penalty calculation (i.e., >25 cases in the 3-year period) and categorized them as eligible for all six measures or fewer than six measures. Within each of these categories, we calculated the number of hospitals that met the benchmark for reliability on all measures for which they were eligible, and the number of hospitals that did not achieve the benchmark on all measures.

Reliability and DRG Payments for Excess Readmissions

HRRP penalties are determined in part the by DRG payments attributed to excess readmissions for each condition. The algorithm used to estimate the HRRP penalty for a given hospital can be seen in the Appendix. For each HRRP condition, we estimated the sum of DRG payments by multiplying the number of admissions, the average DRG weight, and base operating/capital costs ($5,980.17) for each hospital. We estimated the DRG payments for excess readmissions by multiplying the sum of DRG admissions with the excess readmission ratio minus one (ERR—1) for each hospital. We then summed the total admissions, sum of DRG payments, and DRG payments for excess readmissions across hospitals with reliable RSRRs ($R \geq 0.7$) and unreliable RSRRs ($R < 0.7$). We then totaled these DRG payments for each condition and over all conditions, and estimated the proportion of DRG payments for excess readmissions tied to unreliable RSRRs. The resulting calculation estimates the proportion of DRG payments for excess readmissions that are tied to hospitals with unreliable RSRRs.

All analyses were conducted with *SAS version 9.4* (SAS Institute Inc., Cary, NC, USA). The University of Tennessee Health Science Center IRB deemed this project exempt from human subjects research review.

Results

The initial sample based on ICD-9 coding inclusion criteria and final sample following exclusions can be seen in Table S1. Sample patient and hospital characteristics for each condition-specific cohort can be seen in Table 1. Adjusting for patient comorbidities, the median RSRRs were 18.9 percent for AMI (range: 13.4–33.5 percent), 21.7 percent for CHF (16.1–36.2 percent), 18.4 percent for pneumonia (12.4–34.8 percent), 18.7 percent for COPD (13.4–29.8 percent), 7.1 percent for THKA (3.3–45.0 percent), and 18.8 percent for CABG (11.4–47.1 percent).

Figure 1 illustrates the volume-reliability relationship for each condition. In general, as volume increased, the reliability for all measures increased. Reliability of RSRRs was highest for THKA and lowest for CHF.

Adjusting for comorbidities outlined by CMS, the median reliability was below the benchmark *R*-value of 0.70 for AMI ($R = 0.58$), CHF ($R = 0.61$), PN ($R = 0.68$), and COPD ($R = 0.65$), and above the benchmark for THKA ($R = 0.94$) and CABG ($R = 0.78$) (Table 2). To exceed the reliability benchmark, hospitals needed to admit at least 605 AMI patients, 1,019 CHF patients, 442 pneumonia patients, 559 COPD patients, 9 THKA patients, and 107 CABG patients on average over a 3-year period. Only a small share of hospitals exceeded the R threshold for AMI (18.6 percent), CHF (16.6 percent), pneumonia (40.2 percent), and COPD (28.1 percent), whereas the vast majority exceeded the threshold for THKA (100 percent) and CABG (86.6 percent) (Table 2).

Patient socioeconomic factors account for 2–3 percent of the variation in RSRRs for AMI (i.e., 0.58–0.55 = 0.03 = 3 percent), CHF, and pneumonia and 0–1 percent of the variation for COPD, THKA, and CABG, after accounting for patient comorbidities. Hospital factors accounted for 2–4 percent of the variation in RSRRs for AMI, CHF, pneumonia, and CABG, and 0–1 percent of the variation in RSRRs for COPD and THKA, after adjustment for patient comorbidities and socioeconomic factors. After accounting for all factors, the percent of hospitals exceeding the volume threshold for acceptable reliability declined for AMI (from 18.6 to 4.7 percent), heart failure (from 16.6 to 5.2 percent), pneumonia (from 40.2 to 18.1 percent), COPD (from 28.1 to 21.2 percent), and CABG (from 86.6 to 78.4 percent), but not for THKA (remaining at 100 percent).

Of the 505 hospitals included in our sample, only 52 hospitals (10.3 percent) exceeded the reliability benchmark on all measure for which they

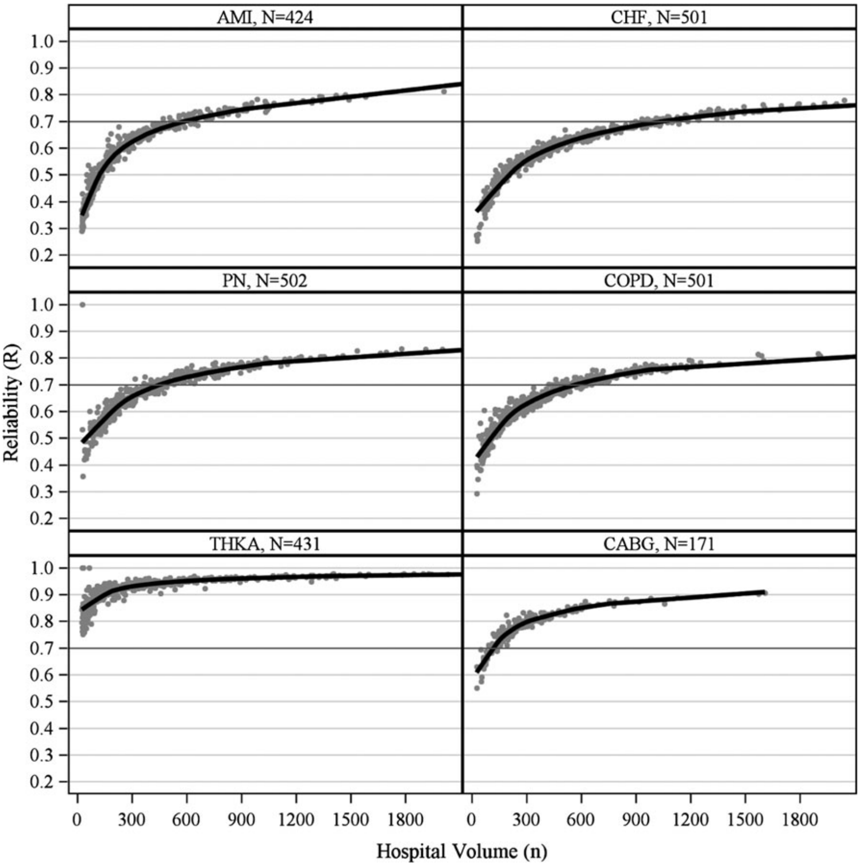
Table 1: Sample Size and Characteristics for Patients with Each HRRP Condition Pooled across the Study Period (2011–2013)

	AMI	*CHF*	*PN*	*COPD*	*THKA*	*CABG*
Hospitals, *N*	424	501	502	501	431	171
Patients, *n*	137,591	305,919	230,659	215,381	217,909	55,087
No. of readmissions	26,077	67,176	43,074	40,439	16,282	10,380
RSRRs (%)*, median (range)	18.9 (13.4–33.5)	21.7 (16.1–36.2)	18.4 (12.4–34.8)	18.7 (13.4–29.8)	7.1 (3.3–45.0)	18.8 (11.4–47.1)
Patient characteristics						
Male, %	52.7	46.6	45.9	40.8	36.7	70.0
Age, mean (SD)	78.6 (8.5)	81.1 (8.3)	80.3 (8.5)	77.2 (7.8)	74.0 (6.2)	74.4 (6.2)
Race, %						
White	80.1	76.4	81.6	79.4	87.5	83.8
Black	6.5	10.6	6.4	7.6	4.1	3.9
Hispanic	8.4	8.9	8.2	9.8	5.5	6.1
Other	5.0	4.1	3.9	3.2	3.0	6.1
Dual eligible, %	8.5	11.4	12.8	13.1	3.2	4.7
Median income, %						
4th Quartile—Highest	21.3	21.3	22.3	18.5	25.0	24.0
3rd Quartile	24.7	24.8	25.3	24.3	27.1	26.1
2nd Quartile	27.3	26.1	26.7	27.5	27.2	27.0
1st Quartile—Lowest	26.7	27.8	25.7	29.8	20.7	22.8
Hospital characteristics						
Teaching status, %						
Major	24.1	20.9	17.2	15.5	16.3	31.9
Minor	21.9	20.8	19.6	17.7	25.7	25.6
Nonteaching	54.1	58.3	63.2	66.9	58.0	42.5
Bed size, %						
300+	63.5	55.3	49.4	48.5	52.2	78.0
100–300	34.6	39.4	42.8	44.7	40.5	21.9
<100	2.0	5.2	7.9	6.8	7.3	0.2
Safety net, %	24.8	25.7	22.7	24.5	16.3	24.2
Ownership status, %						
For-profit	18.8	18.1	16.5	21.6	17.3	16.3
Not-for-profit	71.7	73.0	74.4	69.5	72.5	75.1
Public	9.5	9.0	9.1	8.9	10.2	8.6
Rural location, %	5.7	7.4	10.0	6.5	5.8	2.1

*Adjusted for patient comorbidities outlined by CMS.

were eligible (i.e., >25 cases in 3-year period). Almost a third of hospitals (33.1 percent) were eligible for assessment on all six measures, and 30 percent of those hospitals exceeded the reliability benchmark on all six

Figure 1: Relationship between Hospital Volume and Reliability of 30-Day Readmission Measures with Fitted Loess Curves for Each Condition



Note. Observations of volume >2,000 not shown.

(Table 3). For hospitals eligible for fewer than six measures, only two (0.6 percent) hospitals met the reliability benchmark for the measures on which they were assessed.

Table 4 shows the approximations for DRG payments tied to excess readmissions in hospitals with reliable and unreliable RSRRs. Total payments for excess readmissions for all conditions in this sample was estimated to be about $896 million. We estimate that 25 percent of these payments were tied to hospitals with unreliable condition-specific RSRRs. The proportion of payments tied to unreliable RSRRs is much higher for medical conditions (AMI,

Table 2: Median Reliability of Condition-Specific 30-Day Readmission Measures Cumulatively Adjusted for Comorbidities, Socioeconomic (SES) Factors, and Hospital Factors

Condition	*Risk-Adjustment Model*	*Reliability (R), Median (IQR)*	*Volume Threshold for R = 0.70, Median (IQR)*	*Hospitals with R ≥ 0.70, N (%)*
AMI	Comorbidities*	0.58 (0.45–0.68)	605 (527–685)	79 (18.6)
(N = 424)	+ SES Factors†	0.55 (0.42–0.65)	785 (683–888)	50 (11.8)
	+ Hospital Factors‡	0.51 (0.38–0.61)	1,103 (960–1,248)	20 (4.7)
CHF	Comorbidities*	0.61 (0.53–0.68)	1,019 (938–1,098)	83 (16.6)
(N = 501)	+ SES Factors†	0.58 (0.50–0.65)	1,318 (1,215–1,421)	46 (9.2)
	+ Hospital Factors‡	0.56 (0.47–0.62)	1,616 (1,489–1,742)	26 (5.2)
PN	Comorbidities*	0.68 (0.62–0.73)	442 (402–488)	202 (40.2)
(N = 502)	+ SES Factors†	0.66 (0.60–0.71)	532 (484–587)	148 (29.5)
	+ Hospital Factors‡	0.62 (0.57–0.68)	710 (645–783)	91 (18.1)
COPD	Comorbidities*	0.65 (0.58–0.70)	559 (504–613)	141 (28.1)
(N = 501)	+ SES Factors†	0.64 (0.57–0.70)	587 (529–644)	122 (24.4)
	+ Hospital Factors‡	0.63 (0.56–0.69)	638 (575–699)	106 (21.2)
THKA	Comorbidities*	0.94 (0.91–0.96)	9 (7–11)	431 (100)
(N = 431)	+ SES Factors†	0.94 (0.91–0.96)	9 (7–11)	431 (100)
	+ Hospital Factors‡	0.94 (0.91–0.96)	9 (7–12)	431 (100)
CABG	Comorbidities*	0.78 (0.74–0.82)	107 (93–123)	148 (86.6)
(N = 171)	+ SES Factors†	0.78 (0.73–0.82)	112 (97–129)	148 (86.6)
	+ Hospital Factors‡	0.75 (0.71–0.80)	144 (125–165)	134 (78.4)

*Adjusted for comorbidities outlined by CMS condition-specific measures.

†Adjusted for race (white, black, Hispanic, or other), median income quartile, and dual eligible status.

‡Adjusted for hospital teaching status, bed size category, safety-net status, hospital ownership, and rural location.

Table 3: Number of Hospitals Achieving Acceptable Reliability ($R \geq 0.70$) on All Measures (AMI, CHF, Pneumonia, COPD, THKA, and CABG) for Which They Were Eligible (>25 Cases) (N = 505 Hospitals)

		*Hospitals Meeting Benchmark (R ≥ 0.70), N (%)**	
Measure Eligibility	*Hospitals, N (%)*	*All Measures*	*Not All Measures*
All six measures	167 (33.1)	50 (29.9)	123 (74.1)
Fewer than six measures	338 (66.9)	2 (0.6)	338 (99.4)
Total	505 (100)	52 (10.3)	453 (89.7)

*Percents reflect row percentages.

Table 4: Approximation of DRG Payments Attributable to Excess Readmissions for Hospitals with Reliable and Unreliable RSRRs for Each Condition, and the Total and Percent of DRG Payments Tied to Unreliable RSRRs

		Hospitals with Reliable RSRRs (R ≥ 0.70)			*Hospitals with Unreliable RSRRs (R < 0.70)*			*DRG Payments for Excess Readmissions*	
Condition	*Average DRG Weight*	*Total Admits*	*Sum of DRG Payments**	*DRG Payments for Excess Readmits*†	*Total Admits*	*Sum of DRG Payments**	*DRG Payments for Excess Readmits*†	*Total Payments*	*% Tied to Unreliable RSRRs*
AMI	2.358	69,804	$984,323,017	$29,558,832	67,787	$955,880,814	$49,895,732	$79,454,564	63
CHF	1.340	128,080	$1,026,359,833	$31,468,163	177,839	$1,425,099,987	$73,177,899	$104,646,062	70
PN	1.518	155,853	$1,414,817,646	$67,288,716	74,806	$679,081,424	$47,237,616	$114,526,332	41
COPD	1.196	119,625	$855,591,892	$35,954,623	95,756	$684,874,042	$44,826,502	$80,781,125	55
THKA	2.173	217,909	$2,831,707,715	$396,677,618	0	$0	$0	$396,677,618	0
CABG	5.325	53,273	$1,696,447,001	$111,761,966	1,814	$57,765,751	$8,500,593	$120,262,559	7
Total	–	744,544	$8,809,247,104	$672,709,918	418,002	$3,802,702,018	$223,638,342	$896,348,260	25

Notes. If a hospital had an ERR less than or equal to one (i.e., no excess readmissions), they had no DRG payments for excess readmissions.
*Sum of DRG Payments = Total Admits*Average DRG Weight*Base Operating Cost ($5,980.17).
†DRG Payments for Excess Readmits = Sum of DRG Payments*[Excess Readmission Ratio (ERR) − 1].

CHF, PN, and COPD), but contributed less payments overall. For instance, 63 percent of payments for excess AMI readmissions were tied to hospitals with unreliable AMI RSRRs, yet they contributed to just under 9 percent of the total DRG payments for excess readmissions. Conversely, RSRRs for surgical procedures (THKA and CABG) accounted for almost 58 percent of total payments for excess readmissions, but lower proportions were tied to unreliable hospitals because the reliability of these measures was better when contrasted to HRRP medical conditions.

DISCUSSION

Under the ACA, the HRRP was mandated to levy financial penalties on hospitals with higher than expected readmissions for selected health conditions. The purpose of these penalties was to incentivize hospitals to reduce costly and preventable readmissions. If readmission measures are unreliable, excess readmissions tied to hospitals may be the result of random chance, rather than truly higher than expected readmission rates. Our study found that the reliability of RSRRs did not meet the benchmark for medical conditions (AMI, CHF, PN, and COPD) for the vast majority of hospitals, but it was above the benchmark for surgical procedures (THKA and CABG) for the vast majority of hospitals providing these services. Furthermore, most hospitals did not have sufficient reliability on the measures for which they were evaluated to determine HRRP financial penalties. Lastly, we found that a quarter of payments for excess readmissions were tied to unreliable RSRRs, suggesting that a sizeable proportion of HRRP penalties, particularly those associated with medical conditions, may be affected by statistical noise rather than actual differences in quality of care.

Our findings have important implications for the HRRP. The purpose of the HRRP is to use negative incentives (i.e., financial penalties) to motivate hospitals to reduce readmission rates. To the extent that RSRRs are subject to random noise, financial penalties may be assessed due to random chance, thereby obscuring the link between hospital performance and HRRP financial penalties. This could weaken overall program incentives and thus may limit HRRP from fully meeting its policy objectives. Since we found that many hospitals did not have acceptable reliability for any of their RSRR measures covered under the HRRP, their penalties are likely the result of statistical noise and unlikely to provide constructive information about areas needing improvement. Employing more reliable measures in the HRRP would ensure penalties are due to systematic differences between hospitals, rather than statistical noise.

Furthermore, we found that the excess readmissions for surgical procedures, which are much more reliable, account for almost just under half of the DRG payments for excess readmissions. Therefore, the good news is that hospitals should focus efforts on reducing readmissions for these measures in particular, as they are much less subject to statistical noise; performance improvements for these areas are likely to be reflected in RSRRs. Conversely, hospitals may want to consider placing less emphasis on reducing readmissions for targeted medical conditions (AMI, CHF, PN, and COPD), as they are more subject to statistical noise, and performance improvements for these patients may be less reflected in subsequent RSRRs. However, it is important to note that the number of hospitals that are eligible to receive penalties on THKA and CABG measures is relatively small compared to those eligible to receive penalties for medical conditions. This leaves many hospitals to be evaluated solely on medical condition RSRRs, which are much less reliable.

Measures such as RSRRs are also of importance to hospitals because they provide feedback on their performance. If RSRRs employed by the HRRP are unreliable, hospital investments into quality improvement may be inefficient, as they may mask deficiencies and improvement. Reducing readmissions often requires resource investment into multifaceted interventions (Kripalani et al. 2007; Hansen et al. 2011). Intervening on natural variation in a measure is referred to as “tampering” in the field of quality improvement and control, and it has been shown to increase, rather than reduce, observed variation in quality measures (The W. Edwards Deming Institute 2016). Waste in health care is of critical importance (Berwick and Hackbarth 2012), and government policies intended to improve outcomes and reduce costs should be cognizant of hospital responses to pay-for-performance programs built on unreliable measures.

Inefficient investment of resources could also be especially problematic for resource-scarce hospitals, such as hospitals serving vulnerable and low-income populations. Safety net, urban, and minority-serving hospitals have already been shown to disproportionately receive HRRP penalties (Joynt and Jha 2013; Gilman et al. 2014, 2015; Gu et al. 2014). Larger financial penalties, coupled with the inability of financially strapped hospitals to effectively direct scarce resources to improve quality, could ultimately widen disparities in care, which echoes previous concerns regarding the HRRP (Bhalla and Kalkut 2010) and in other pay-for-performance initiatives (Casalino et al. 2007; Ryan 2013). As these hospitals serve an important function in serving vulnerable and low-income populations, policy makers should be careful to design policies sensitive to these effects.

Adjusting readmission rates for socioeconomic or hospital factors related to treating low-income or underserved populations could ease the penalty burden on safety-net hospitals. While certain hospital and socioeconomic factors are associated with higher 30-day readmission rates, how these factors explained the variation in readmission rates between hospitals has not been well documented (Joynt, Orav, and Jha 2011; Brown et al. 2014; Hu, Gonsahn, and Nerenz 2014; Sheingold, Zuckerman, and Shartzer 2016). A previous study found that hospital characteristics explained a modest amount of the variation in readmissions between hospitals, but this study examined all-cause readmissions, and not the condition-specific readmissions employed by the HRRP (Singh et al. 2014). When new risk-adjustment factors are under consideration, researchers should balance the relative gains of adjustment to the loss in measure reliability. Our findings suggest that the loss in reliability of adding socioeconomic or hospital factors is likely small.

Our findings offer lessons for policy makers when exploring pay-for-performance initiatives or measures that tie financial incentives to hospital performance. For measures to be reliable for group-level comparisons, they must have sufficient between-group variation (e.g., physicians, hospitals, accountable care organizations) and adequate sample size. The absence of either of these elements limits the measure's usefulness in hospital performance profiling. For instance, the condition with the largest sample size, CHF, has among the worst reliability because hospital-level RSRRs do not vary substantially (range: 16.1–36.2 percent). Conversely, RSRRs for CABG have much higher reliability, despite its small sample size, because the variation in RSRRs is much larger (range: 11.4–47.1 percent). Figure 1 illustrates how reliability declines markedly as hospital volume declines. When pay-for-performance initiatives are assessing current or exploring new measures to tie payment to quality, it is essential to consider these attributes.

A number of approaches could be taken to improve the reliability of HRRP measures. First, the minimum case volume could be increased from the current standard of 25 cases over 3 years. Second, a hospital's eligibility for the HRRP could be determined by the estimated reliability of their RSRRs. While both of these approaches would improve the reliability of RSRRs overall, it would also exclude many low volume hospitals from the HRRP. Excluding these hospitals may not be desirable, since they tend to have worse outcomes across a spectrum of procedures and conditions, although findings for readmissions have been mixed (Halm, Lee, and Chassin 2002; Horwitz et al. 2015).

To avoid excluding low-volume hospitals, the HRRP could also choose to employ a composite readmission measure, such as the hospital-wide all-cause (HWAC) readmissions measure. The HWAC readmission measure has the benefit of greater hospital case volume and a more comprehensive view of hospital readmissions (Horwitz et al. 2014). An annual measure may also allow hospitals to demonstrate year-to-year improvements in readmissions, rather than relying on improvements averaged over a 3-year period. However, the HWAC may not provide meaningful opportunities for intervention and would require a clean start to the HRRP, as a recent study showed little agreement between the HWAC readmission rates and condition-specific rates currently employed by the HRRP (Rosen et al. 2016).

There are limitations to our study that should be considered. First, since we did not use Medicare administrative claims, we could not strictly employ the risk-adjustment model used by CMS. We attempted to replicate this model by identifying and adjusting for the same comorbidities used by CMS, but we are likely to miss diagnoses for comorbidities that are typically coded in outpatient settings. Second, while we limited our analyses to individuals aged 65 and older with Medicare listed as primary payer in the discharge records (or secondary payer only if primary payer was Medicaid), this does not technically ensure all individuals are Medicare beneficiaries. Third, we did not have 100 percent of US hospitals in our sample, so repeating our analyses in Medicare claims data would be important to confirm our findings in all hospitals subject to the HRRP. Fourth, we made several assumptions in our approximations for DRG payments attributable to excess readmissions. This includes using an average DRG weight for each condition, the average base DRG cost, and excess readmissions, all of which may vary when using Medicare administrative claims data.

Conclusion

The HRRP levies financial penalties on hospitals with excess readmissions relative to their peers for targeted conditions. Our study found that the RSRRs for medical conditions used by the HRRP have limited reliability, while RSRRs for surgical procedures were shown to be reliable for group-level comparisons. Ultimately, few hospitals have acceptable reliability on all measures for which they are assessed by HRRP, and a sizeable portion of DRG payments attributable to excess readmissions for these conditions were tied to unreliable measures. Unreliable measures potentially blur the connection between hospital performance and incentives, which could limit the ability of HRRP to achieve its objectives. Steps should be taken to address the reliability of measures used by the HRRP to compare relative performance on hospital readmissions.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This research was supported by funding from the Agency for Healthcare Research and Quality (R01HS023783; TM Waters, PI). Preliminary findings of this research were presented at the AcademyHealth 2016 Annual Research Meeting on June 27, 2016.

Disclosures: None.

Disclaimers: None.

REFERENCES

Adams, J. L. 2009. *The Reliability of Provider Profiling: A Tutorial.* Santa Monica, CA: RAND Corporation.

Adams, J. L., A. Mehrotra, J. W. Thomas, and E. A. McGlynn. 2010. "Physician Cost Profiling—Reliability and Risk of Misclassification." *New England Journal of Medicine* 362 (11): 1014–21.

Axon, R., and M. V. Williams. 2011. "Hospital Readmission as an Accountability Measure." *Journal of the American Medical Association* 305 (5): 504–5.

Bernheim, S. M., C. S. Parzynski, L. Horwitz, Z. Lin, M. J. Araas, J. S. Ross, E. E. Drye, L. G. Suter, S.-L. T. Normand, and H. M. Krumholz. 2016. "Accounting For Patients' Socioeconomic Status Does Not Change Hospital Readmission Rates." *Health Affairs* 35 (8): 1461–70.

Berwick, D. M., and A. D. Hackbarth. 2012. "Eliminating Waste in US Health Care." *Journal of the American Medical Association* 307 (14): 1513–6.

Bhalla, R., and G. Kalkut. 2010. "Could Medicare Readmission Policy Exacerbate Health Care System Inequity?" *Annals of Internal Medicine* 152 (2): 114–7.

Blum, A. B., N. N. Egorova, E. A. Sosunov, A. C. Gelijns, E. DuPree, A. J. Moskowitz, A. D. Federman, D. D. Ascheim, and S. Keyhani. 2014. "Impact of Socioeconomic Status Measures on Hospital Profiling in New York City." *Circulation: Cardiovascular Quality and Outcomes* 7 (3): 391–7.

Bratzler, D. W., S.-L. T. Normand, Y. Wang, W. J. O'Donnell, M. Metersky, L. F. Han, M. T. Rapp, and H. M. Krumholz. 2011. "An Administrative Claims Model for

Profiling Hospital 30-Day Mortality Rates for Pneumonia Patients." *PLoS ONE* 6 (4): e17401.

Brown, J. R., C.-H. Chang, W. Zhou, T. A. MacKenzie, D. J. Malenka, and D. C. Goodman. 2014. "Health System Characteristics and Rates of Readmission after Acute Myocardial Infarction in the United States." *Journal of the American Heart Association* 3 (e000714): 1–16.

Casalino, L. P., A. Elster, A. Eisenberg, E. Lewis, J. Montgomery, and D. Ramos. 2007. "Will Pay-For-Performance and Quality Reporting Affect Health Care Disparities?" *Health Affairs* 26 (3): w405–14.

Centers for Medicare & Medicaid Services. 2014. "Readmissions Reduction Program" [accessed on December 12, 2014, 2014]. Available at http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html

Centers for Medicare & Medicaid Services. 2016. "Acute Inpatient PPS" [accessed on August 3, 2016]. Available at https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/index.html

Gilman, M., E. K. Adams, J. M. Hockenberry, I. B. Wilson, A. S. Milstein, and E. R. Becker. 2014. "California Safety-Net Hospitals Likely to Be Penalized by ACA Value, Readmission, and Meaningful-Use Programs." *Health Affairs* 33 (8): 1314–22.

Gilman, M., E. K. Adams, J. M. Hockenberry, A. S. Milstein, I. B. Wilson, and E. R. Becker. 2015. "Safety-Net Hospitals More Likely Than Other Hospitals to Fare Poorly under Medicare's Value-Based Purchasing." *Health Affairs* 34 (3): 398–405.

Glance, L. G., A. L. Kellermann, T. M. Osler, Y. Li, W. Li, and A. W. Dick. 2016. "Impact of Risk Adjustment for Socioeconomic Status on Risk-Adjusted Surgical Readmission Rates." *Annals of Surgery* 263 (4): 698–704.

Gu, Q., L. Koenig, J. Faerberg, C. R. Steinberg, C. Vaz, and M. P. Wheatley. 2014. "The Medicare Hospital Readmissions Reduction Program: Potential Unintended Consequences for Hospitals Serving Vulnerable Populations." *Health Services Research* 49 (3): 818–37.

Halm, E. A., C. Lee, and M. R. Chassin. 2002. "Is Volume Related to Outcome in Health Care? A Systematic Review and Methodologic Critique of the Literature." *Annals of Internal Medicine* 137 (6): 511–20.

Hansen, L. O., R. S. Young, K. Hinami, A. Leung, and M. V. Williams. 2011. "Interventions to Reduce 30-Day Rehospitalization: A Systematic Review." *Annals of Internal Medicine* 155 (8): 520–8.

Herrin, J., J. St. Andre, K. Kenward, M. S. Joshi, A.-M. J. Audet, and S. C. Hines. 2014. "Community Factors and Hospital Readmission Rates." *Health Services Research* 50(1): 20–39.

Horwitz, L. I., C. Partovian, Z. Lin, J. N. Grady, J. Herrin, M. Conover, J. Montague, C. Dillaway, K. Bartczak, L. G. Suter, J. S. Ross, S. M. Bernheim, H. M. Krumholz, and E. E. Drye. 2014. "Development and Use of an Administrative Claims Measure for Profiling Hospital-wide Performance on 30-Day Unplanned ReadmissionAdministrative Claims Measure for Profiling

Hospital-Wide Performance on Readmission." *Annals of Internal Medicine* 161 (10 Suppl.): S66–75.

Horwitz, L. I., Z. Lin, J. Herrin, S. Bernheim, E. E. Drye, H. M. Krumholz, H. J. Hines Jr, and J. S. Ross. 2015. "Association of Hospital Volume with Readmission Rates: A Retrospective Cross-Sectional Study." *British Medical Journal (Clinical Research Ed.)* 350: h447.

Hu, J., M. D. Gonsahn, and D. R. Nerenz. 2014. "Socioeconomic Status and Readmissions: Evidence from an Urban Teaching Hospital." *Health Affairs* 33 (5): 778–85.

Jencks, S. F., M. V. Williams, and E. A. Coleman. 2009. "Rehospitalizations among Patients in the Medicare Fee-for-Service Program." *New England Journal of Medicine* 360 (14): 1418–28.

Joynt, K. E., and A. K. Jha. 2013. "Characteristics of Hospitals Receiving Penalties under the Hospital Readmissions Reduction Program." *Journal of the American Medical Association* 309 (4): 342–3.

Joynt, K. E., E. Orav, and A. K. Jha. 2011. "Thirty-Day Readmission Rates for Medicare Beneficiaries by Race and Site of Care." *Journal of the American Medical Association* 305 (7): 675–81.

Keenan, P. S., S.-L. T. Normand, Z. Lin, E. E. Drye, K. R. Bhat, J. S. Ross, J. D. Schuur, B. D. Stauffer, S. M. Bernheim, A. J. Epstein, Y. Wang, J. Herrin, J. Chen, J. J. Federer, J. A. Mattera, Y. Wang, and H. M. Krumholz. 2008. "An Administrative Claims Measure Suitable for Profiling Hospital Performance on the Basis of 30-Day All-Cause Readmission Rates among Patients with Heart Failure." *Circulation: Cardiovascular Quality and Outcomes* 1: 29–37.

Krell, R. W., A. Hozain, L. S. Kao, and J. B. Dimick. 2014. "Reliability of Risk-Adjusted Outcomes for Profiling Hospital Surgical Quality." *Journal of the American Medical Association Surgery* 149 (5): 467–74.

Kripalani, S., F. LeFevre, C. O. Phillips, M. V. Williams, P. Basaviah, and D. W. Baker. 2007. "Deficits in Communication and Information Transfer between Hospital-Based and Primary Care Physicians: Implications for Patient Safety and Continuity of Care." *Journal of the American Medical Association* 297 (8): 831–41.

Krumholz, H. M., Z. Lin, E. E. Drye, M. Desai, L. F. Han, M. T. Rapp, J. A. Mattera, and S.-L. T. Normand. 2011. "An Administrative Claims Measure Suitable for Profiling Hospital Performance Based on 30-Day All-Cause Readmission Rates among Patients with AMI." *Circulation: Cardiovascular Quality and Outcomes* 4: 243–52.

Normand, S.-L. T., M. E. Glickman, and C. A. Gatsonis. 1997. "Statistical Methods for Profiling Providers of Medical Care: Issues and Applications." *Journal of the American Statistical Association* 92: 803–14.

QualityNet. 2016. "Measure Methodology Reports—Readmission Measures" [accessed on August 17, 2016]. Available at https://www.qualitynet.org/dcs/ContentServer?cid=1219069855841&pagename=QnetPublic%2FPage%2FQnetTier4&c=Page

Rosen, A. K., Q. Chen, M. Shwartz, C. Pilver, H. J. Mull, K. F. M. Itani, and A. Borzecki. 2016. "Does Use of a Hospital-wide Readmission Measure versus

Condition-Specific Readmission Measures Make a Difference for Hospital Profiling and Payment Penalties?" *Medical Care* 54 (2): 155–61.

Ryan, A. M. 2013. "Will Value-Based Purchasing Increase Disparities in Care?" *New England Journal of Medicine* 369 (26): 2472–4.

Scholle, S. H., J. Roski, J. L. Adams, D. L. Dunn, E. A. Kerr, D. P. Dugan, and R. E. Jensen. 2008. "Benchmarking Physician Performance: Reliability of Individual and Composite Measures." *The American Journal of Managed Care* 14 (12): 833–8.

Sheingold, S. H., R. Zuckerman, and A. Shartzer. 2016. "Understanding Medicare Hospital Readmission Rates and Differing Penalties between Safety-Net and Other Hospitals." *Health Affairs* 35 (1): 124–31.

Singh, S., Y.-L. Lin, Y.-F. Kuo, A. B. Nattinger, and J. S. Goodwin. 2014. "Variation in the Risk of Readmission among Hospitals: The Relative Contribution of Patient, Hospital and Inpatient Provider Characteristics." *Journal of General Internal Medicine* 29 (4): 572–8.

The W. Edwards Deming Institute. 2016. "The Funnel Experiment" [accessed on April 12, 2016]. Available at https://www.deming.org/theman/theories/funnelexperiment

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix SA2: Calculation of Readmission Penalties.

Table S1. Sample Derivation Based on Inclusion and Exclusion Criteria for Each Condition.