The denominator problem: national hospital quality measures for acute myocardial infarction

Jeffrey Bruckel,[1] Xiu Liu,[2] Samuel F Hohmann,[3] Andrew S Karson,[2] Elizabeth Mort,[2] David M Shahian[2,4]

[1]Department of Cardiology, University of Rochester Medical Center, Rochester, New York, USA
[2]Center for Quality and Safety, Massachusetts General Hospital, Boston, Massachusetts, USA
[3]University HealthSystem Consortium, Chicago, Illinois, USA
[4]Department of Surgery, Center for Quality and Safety, Massachusetts General Hospital, Boston, Massachusetts, USA

Correspondence to
Dr Jeffrey Bruckel, Department of Cardiology, University of Rochester Medical Center, 601 Elmwood Ave, AC-G, Rochester, NY 14642, USA; jeffrey.bruckel@gmail.com

Linked

CrossMark

ABSTRACT

Background National Hospital Quality Measures (NHQM) should accurately reflect quality of care, as they increasingly impact reimbursement and reputation. However, similar to risk adjustment of outcomes measures, NHQM process measures pose unique methodological concerns, including lack of representativeness of the final denominator population after exclusions. This study determines population size and characteristics for each acute myocardial infarction (AMI) measure, reasons for exclusion from the measures, and variation in exclusion rates among hospitals.

Methods and results 163 144 discharges from 172 University HealthSystem Consortium hospitals between 2008-Q4 and 2013-Q3 were examined, including characteristics and propensity scores of included and excluded groups. Measure exclusions ranged from 17.8% (discharge aspirin) to 90.1% (percutaneous coronary intervention, PCI, within 90 min), with substantial variation across hospitals. Median annual denominator size (IQR) for PCI within 90 min was 28 (20, 44) at major teaching hospitals, versus 10 (0, 25) at non-teaching hospitals. Patients most likely to be excluded (in the 10th vs 1st propensity decile) were older (mean age (SD) of 78.1 (10.8) vs 50.3 (8.6) years), more likely to have Medicare (90.5% vs 0.9%), had more documented comorbidities (15.6 (4.6) vs 6.2 (2.5) hierarchical clinical condition categories) and higher admission mortality risk (Major or Extreme 80.9% vs 7.3%, respectively), and experienced higher inpatient mortality (10.0% vs 1.6%).

Conclusions Exclusion from AMI measures varied substantially among hospitals, sample sizes were very small for some measures (PCI and ACE inhibitor measures) and measures often excluded high-risk populations. This has implications for the representativeness and comparability of the measures and provides insight for future measure development.

BACKGROUND

The National Hospital Quality Measures (NHQM) are a set of performance measures, mostly evidence-based processes of care, derived from the results of the Cooperative Cardiovascular Project. This project found that many patients were not receiving indicated care for acute myocardial infarction (AMI).[1–5] In an attempt to incentivise hospitals to measure and improve care, these measures were ultimately adapted into reporting and pay-for-performance strategies, most significantly the Centers for Medicare and Medicaid Services (CMS) Value-Based Purchasing (VBP) programme.[6–10] For almost a decade, these measures have been the primary indicators of the quality of cardiovascular care provided in US hospitals. Consequently providers invested substantial resources towards data collection, analysis and reporting of these measures.

There is an ongoing debate about the merits of process measures as opposed to outcome measures for accountability and reimbursement.[10–12] Outcome measures are the ultimate validators of care quality, as noted by Professor Avedis Donabedian half a century ago. However, these measures require robust risk adjustment to fairly account for inherent differences in patient case mix among providers, and the adequacy of this adjustment has often been challenged.[13] For example, important risk predictors are sometimes not available, and coding of comorbidities may vary across institutions.

Process measures have been viewed as a potential solution to the challenges of outcomes measures. They theoretically avoid the need for risk adjustment, as all patients deserve to receive all indicated care. These measures are also directly

actionable, and results are immediately available. As process measures were more widely implemented, however, it became apparent that they also have troubling issues, some of which might limit their accuracy as measures of care quality (including, in some instances, a lack of correlation between process measure performance and outcomes).[14] From a methodological perspective, they appear to have a potential weakness that is somewhat analogous to that of risk adjustment in outcome measures. Not all generally desirable care processes are applicable to all patients, and virtually all process measures exclude certain populations.

Performance scores are based on the denominator population remaining after exclusions, but this creates two concerns. First, the final denominator population for a measure may be only a small fraction of the total number of patients with a given condition that the hospital treated. Second, because hospitals care for different types of patients (eg, tertiary teaching hospitals vs community hospitals), some may have more or less exclusions compared with others, which will potentially impact their performance scores. Although exclusions are necessary for clinical appropriateness and to ensure a homogeneous measurement population for accurate comparisons between hospitals, there is concern that variability in exclusions among hospitals may limit the comparability of measurements. Additionally, some large and potentially important groups of patients may be excluded from the measurement population.[15] Previous work has demonstrated that a significant percentage of Medicare beneficiaries are excluded from these measures, limiting their representativeness.[16] Exclusions from the measure population were reported to increase during the period from 1994 through 2001.[17] Most importantly, patients who were excluded from the measures had 1-year mortality rates double those of CMS-included patients.[18] Despite these important findings, previous studies have not assessed exclusion rates across a large population of inpatients, nor have they had access to specific reasons for exclusion from the measurement population.

The aims of this study were to (1) define the exclusion rates from the original populations for each of the measures, and the size of the final denominator populations; (2) determine the reasons for exclusion from the denominator measure population; (3) describe the variation in denominator exclusion rate among hospitals; and (4) investigate those patient and hospital factors that influence exclusion from the measures, and the resulting differences in composition of included and excluded populations.

METHODS
Data source
The University HealthSystem Consortium (UHC), a national organisation of academic medical centre members and affiliated hospitals, collects data to facilitate quality measurement and comparative analysis. UHC has 117 participating academic medical centres with 338 affiliated hospitals. The UHC clinical database/resource manager includes patient-level data with detailed demographics and diagnosis and procedure codes (International Classification of Diseases 9, ICD-9-CM, codes). A subset of participating hospitals use UHC to collect data for submission to the NHQM, and these hospitals were the basis for our study (195 total hospitals in our source dataset). For patients meeting eligibility for inclusion in the NHQM, the database also includes patient-level information regarding the measure population, reasons for denominator exclusion and measure performance.

We matched data from UHC with the American Hospital Association Annual Survey (AHA survey) for 2011.[19] All but three hospitals in UHC were linked with the survey and these three hospitals were excluded from analysis. Teaching status was determined by whether the hospital had at least one training programme approved by the Accreditation Council for Graduate Medical Education, and whether the hospital was a member of the American Association of Medical Colleges Council of Teaching Hospitals (COTH). We classified hospitals into COTH teaching, non-COTH teaching and non-teaching hospitals.

Patient population and study period
Initial eligibility criteria for sampling in the AMI population required that patients were admitted for hospital inpatient care with an ICD-9-CM principal diagnosis code for AMI (410.x0/410.x1), were older than 18 years at admission and had length of stay less than 120 days.[6] From the patient population that meets these initial criteria, hospitals sample a portion of eligible AMI discharges based on sample size guidelines published in the measure specifications. Each hospital has some flexibility to determine their actual sample sizes as long as they comply with the sample size guidelines; consequently, sampling strategies and sizes differ among hospitals. For certain AMI measures (specifically (AMI-8a), primary percutaneous coronary intervention (PCI) within 90 min) many hospitals choose to include the entire eligible population.[6]

Our final study population consisted of all patients included by hospitals in their initial AMI population over 5 years between Quarter 4 of 2008 (2008-Q4) and Quarter 3 of 2013 (2013-Q3).

Definitions
The study period is covered by NHQM Measure Specification Manual Versions 2.5b through 4.2b. All measures except the AMI-10 statin at discharge measure were collected continuously during the study period; the AMI-10 measure was collected starting in

2010-Q4. This limitation is noted where applicable, specifically in the analysis of variation by hospital teaching status.

The three arrival measures (and their two derivatives) consist of aspirin at arrival (AMI-1) and time to primary PCI (median time, AMI-8; and therapy within 90 min of hospital arrival, AMI-8a). An exceedingly small group of patients were eligible for fibrinolysis and we do not report any data for this measure. The four discharge prescription measures include aspirin (AMI-2), ACE inhibitor (ACEi) or angiotensin receptor blocker (ARB) for patients with left ventricular systolic dysfunction (LVSD) (AMI-3), β-blocker (AMI-5) and statin (AMI-10).

Comorbidities and severity of illness

Comorbidities, severity of illness and risk of mortality were assessed. UHC generates severity and mortality scores at admission and discharge using the 3M All Patient Refined Diagnosis Related Group classification system for all inpatient encounters.[20–22] This system uses ICD-9 codes to group patients into four categories based on risk: 'Mild', 'Moderate', 'Major' and 'Extreme'.[22] We also used the patient's ICD-9 codes to determine specific comorbidities at the time of admission based on CMS risk-adjusted 30-day mortality model.[23–25] We used 'Present-on-Admission' flag to confirm whether a given comorbidity was present on admission and unlikely to be a complication developed during hospitalisation. We report data both for the number of total CMS Hierarchical Clinical Conditions (HCC) categories (out of 189 possible), as well as for a subset of 14 comorbidities used in the CMS AMI mortality model. We adjusted for the 14 comorbidities in the CMS AMI mortality model in our multivariable modelling (see below).

Analysis exclusions

For our analyses, we established a number of study exclusions (separate from the NHQM criteria) related to potential coding and data quality issues (see online supplementary eTable 1 for detailed information). For example, hospitals were completely excluded from the study if 100% of their total discharges were in the excluded group, if they had fewer than 20 discharges during the entire study period, or if they could not be matched with the AHA survey. Specific quarters were excluded from a hospital's data if there were fewer than five discharges in that quarter (which is consistent with the measure specification), or if the hospital reported 100% of their discharges in the excluded group for that quarter. Eight individual patients were excluded because of inaccurate diagnosis coding from the submitting hospitals (principal diagnosis was not AMI on the ICD-9 codes specification table). A small group of patients in 2013-Q3 had multiple reasons listed for exclusion from the measures. We took the first reason for exclusion in the measure specification

algorithm as the primary reason for exclusion for that patient.

Statistical methods

Overall hospital inclusion/exclusion rates and original and final denominator population sizes were determined for each process measure. Patient and hospital characteristics between included and excluded population were compared using χ^2 tests for categorical variables and t tests for continuous variables. Some analyses (exclusion variation by hospital status and detailed exclusion reasons) were limited to the period from 2010Q4 to 2013Q3 (after the addition of the AMI-10 measure in which the specification only had very minor change). We assessed unadjusted between-hospital variation in mean exclusion rate using coefficients of variation (CV, SD/Mean), where CV is the size of the SD relative to the mean, expressed as a decimal. Larger CV indicate having a larger SD compared with the mean. Adjusted between-hospital variation was assessed using the covariance parameter estimates from the generalised logistic mixed models, described below. We also determined whether there was a correlation between exclusion rate and performance on the measure using unadjusted Pearson correlation.

Propensity models were used to compare group differences on observed covariates, using multivariable logistic regression models. The goal of these analyses was to identify patients at high risk of exclusion from the measures. The propensity score analysis allowed us to summarise large numbers of covariates in a single score to determine whether characteristics differed significantly between the included and excluded groups, and to identify types of patients at low and high risk of exclusion. A propensity-score-stratified analysis comparing the 1st decile (least likely to be excluded) to the 10th decile (most likely to be excluded) of the propensity score distribution was performed. This allowed us to compare the association of patient and hospital characteristics with patient inclusion or exclusion.

Multivariable logistic mixed models with a hospital-specific random effect were used to account for within-hospital clustering of patient demographic and clinical factors between included and excluded groups, and between-hospital variation. The outcome variable was patient exclusion status (inclusion or exclusion) rather than a clinical outcome such as mortality. SAS GLIMMIX was used to estimate the log-odds of exclusion from each measure. Propensity scores were calculated for each patient, which are the predicted probability of being in the excluded group. Patients with similar propensity scores have comparable characteristics.

Covariates were selected based on the CMS AMI mortality model and additional factors that we believed might impact exclusion rates (race, ethnicity,

insurance, severity of illness, weekend admission for arrival measures, weekend discharge and the use of intensive care unit for discharge measures). To maximise predictive accuracy, and given our large sample size, we employed a liberal approach to include as many available covariates as possible (an all-in approach).

All statistical analyses were performed using SAS software, V.9.4 (SAS Institute, Cary, North Carolina, USA). Selected graphics were produced with R (V.3.1.2), ggplot2 package (V.1.0.1) (R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2014).[26]

Ethical review

The study was approved by the institutional review board at Partners Healthcare, and given a waiver of informed consent.

RESULTS

The final eligible population for (AMI-1) through (AMI-8a) included 163 144 discharges (163 143 for AMI-8a) from 172 hospitals between 2008-Q4 and 2013-Q3. The study population for AMI-10 (statin at discharge) was 109 293 discharges from 163 hospitals between 2010-Q4 and 2013-Q3 (due to the later introduction of this measure). The majority of hospitals (61.6%) were COTH teaching hospitals, accounting for 83.3% of discharges. Among COTH hospitals, 96.1% were PCI-capable, compared with 50.1% and 53.7% of non-COTH teaching and non-teaching hospitals, respectively.

Final measure denominator size

Notably, the yearly measurement population denominator (eg, included patients) median size (IQR) for PCI within 90 min (AMI-8a) was only 28 (20, 44) at COTH hospitals, 12 (0, 29) at non-COTH teaching

hospitals and 10 (0, 25) at non-teaching hospitals. The final denominator population sizes for ACEi/ARB for LVSD (AMI-3) were likewise small (medians (IQR) 49 (33, 73), 13 (8, 30) and 9 (3, 20) at COTH, non-COTH teaching and non-teaching hospitals, respectively). The median and IQRs of the final denominator population sizes by hospital type are presented in table 1.

Measure exclusion rates

PCI within 90 min (AMI-8a) had the highest rate of exclusion, 90.1%. Aspirin at discharge (AMI-2) had the lowest rate of exclusion, 17.8%. There were some noticeable patterns in exclusion. Among the 109 293 patients eligible for all six measures, 1782 (1.6%) were included in all six and 7263 (6.6%) were excluded from all six measures. The most frequent measurement pattern observed was inclusion in all measures except AMI-8a and AMI-3, and this occurred in 58 429 (53.5%) of patients. The remaining 41 819 (38.3%) of patients were included in at least one measure. Exclusion rates and performance on each measure are provided in online supplementary eTable 2, and patterns in exclusion from more than one measure are shown in online supplementary eFigure 1.[27]

Correlation with performance

Exclusion rates were negatively correlated with performance on the quality measures; lower rates of measure exclusions were associated with higher rates of performance on the measures. Correlation coefficients ranged from −0.33 for AMI-3 to −0.69 for AMI-8a (p<0.001 for both) (see online supplementary eTable 3).

Variation in exclusion rates

Interhospital variation in exclusions was substantial. The measure with the greatest variation was aspirin at arrival (AMI-1), with CV as high as 0.709, 0.726 and

Table 1 Median eligible and included discharges per hospital per year

			All hospitals		COTH teaching hospitals		Non-COTH teaching hospitals		Non-teaching hospitals	
			Total	Included	Total	Included	Total	Included	Total	Included
AMI-1	Aspirin at arrival	Median (IQR)	242 (111, 362)	192 (92, 288)	305 (220, 435)	241 (186, 384)	123 (87, 255)	95 (51, 194)	92 (54, 189)	66 (38, 158)
AMI-8a	PCI within 90 min	Median (IQR)	242 (111, 362)	25 (11, 40)	305 (220, 435)	28 (20, 44)	123 (87, 255)	12 (0, 29)	92 (54, 189)	10 (0, 25)
AMI-2	Aspirin at discharge	Median (IQR)	242 (111, 362)	197 (72, 308)	305 (220, 435)	255 (184, 377)	123 (87, 255)	71 (35, 195)	92 (54, 189)	62 (22, 138)
AMI-3	ACEi/ARB for LVSD	Median (IQR)	242 (111, 362)	34 (14, 57)	305 (220, 435)	49 (33, 73)	123 (87, 255)	13 (8, 30)	92 (54, 189)	9 (3, 20)
AMI-5	β-blocker at discharge	Median (IQR)	242 (111, 362)	191 (71, 294)	305 (220, 435)	246 (173, 358)	123 (87, 255)	71 (34, 196)	92 (54, 189)	60 (23, 134)
AMI-10	Statin at discharge	Median (IQR)	251 (116, 374)	203 (80, 313)	308 (223, 418)	257 (183, 368)	130 (80, 231)	76 (34, 183)	90 (52, 177)	49 (19, 110)

ACEi, ACE inhibitor; AMI, acute myocardial infarction; ARB, angiotensin receptor blocker; COTH, Council of Teaching Hospitals; LVSD, left ventricular systolic dysfunction; PCI, percutaneous coronary intervention.

0.626, for COTH, non-COTH, and non-teaching hospitals, respectively. Measures AMI-2, AMI-5 and AMI-10 had similarly high CV. The measure with the lowest CV was ACEi/ARB for LVSD (AMI-3), with CV of 0.051, 0.063 and 0.044, for COTH, non-COTH and non-teaching hospitals, respectively. Figure 1 illustrates hospital-specific exclusion rates stratified by teaching status, using a subset population from 2010 Q4 to 2013. Each line represents the exclusion rate for a specific hospital for that specific measure. See online supplementary eTable 4a for mean, SD and CV of exclusion rates for each measure by hospital type. Based on multivariable models using the full population, covariance parameter estimates and SDs are presented in online supplementary eTable 4b. There is still significant between-hospital variation in exclusion rate for all measures after adjusting for patient characteristics. For example, the odds of a patient being excluded from aspirin at arrival (AMI-1) measure when admitted at a hospital with 1 SD above the average exclusion rate is 5.12 times that when admitted to a hospital with 1 SD below the average.

Reasons for exclusion
Detailed reasons for exclusion from selected arrival and discharge measures are presented in figure 2. The single largest exclusion category consisted of 73 376 (67.1%) patients without LVSD who were not eligible for the ACEi/ARB at discharge for LVSD (AMI-3) measure. PCI within 90 min (AMI-8a) had two dominant reasons for exclusion: no documented ST-segment elevation (STEMI) or left bundle branch block (n=50 143 or 45.9%), and patient was transferred from another facility (n=41 610 or 38.1%).

Unadjusted measurement population differences
Most bivariate comparisons between included and excluded groups were statistically significant due to the large sample size (see online supplementary eTable 5 for complete unadjusted comparisons between the included and excluded groups for all measures). For the PCI within 90 min (AMI-8a) measure, the excluded patients were older (mean age (SD) 66.3 (14.2) vs 60.4 (12.5)), more likely to be discharged to another hospital (28.0% vs 12.4%) and had higher admission severity of illness and risk of mortality (Major or Extreme in 36.2% vs 21.5% for severity and 35.8% vs 17.6% for mortality).

Propensity score distributions
Model terms for each of the propensity models are presented in online supplementary eTable 6. Kernel plots of the log(PS) distributions are demonstrated in
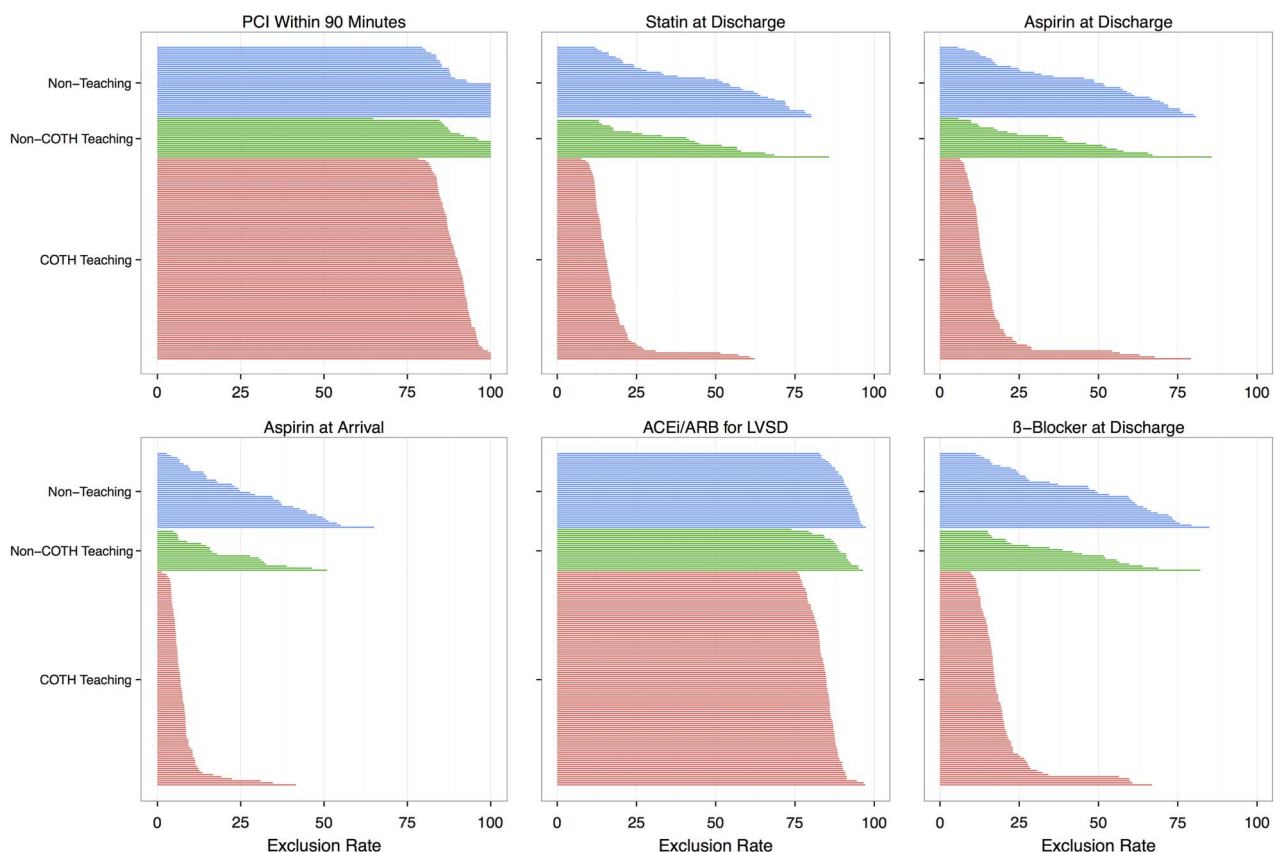

Figure 1 Plot of hospital-specific exclusion rates, stratified by teaching status (2010Q4–2013Q3). ACEi, ACE inhibitor; AMI, acute myocardial infarction; ARB, angiotensin receptor blocker; LVSD, left ventricular systolic dysfunction; PCI, percutaneous coronary intervention.
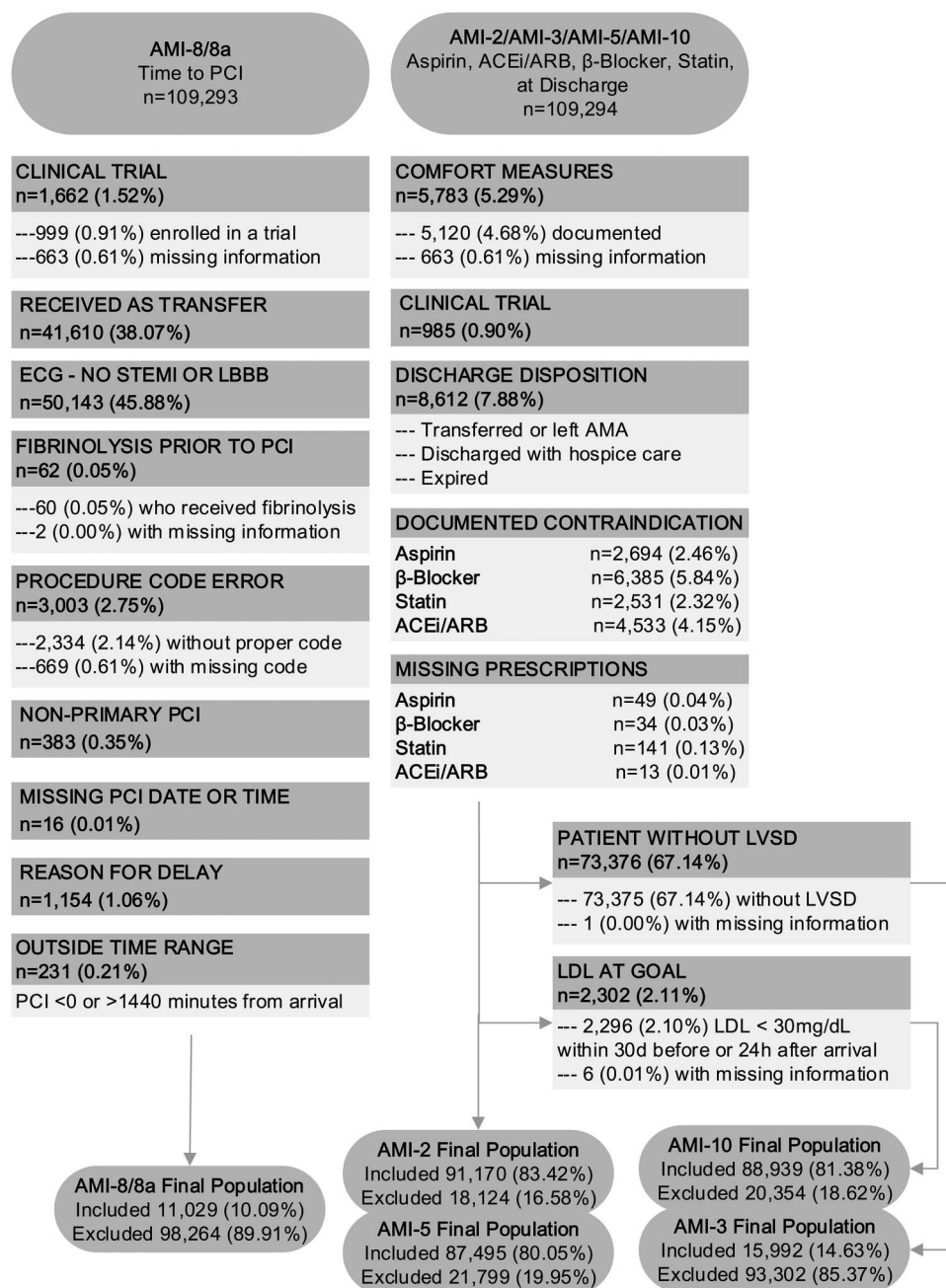
Figure 2 Measure-specific reasons for exclusion (2010-Q4–2013-Q3). Flow diagram demonstrating specific reasons for exclusion from arrival measures. Data are reported from 2010-Q4 to 2013-Q3 to allow comparability between measures (due to change in AMI-1 measure specification and addition of AMI-10 that occurred in 2010-Q4). ACEi, ACE inhibitor; AMA, against medical advice; AMI, acute myocardial infarction; ARB, angiotensin receptor blocker; LBBB, left bundle branch block; LDL, low-density lipoprotein; LVSD, left ventricular systolic dysfunction; PCI, percutaneous coronary intervention; STEMI, ST-segment elevation myocardial infarction.

figure 3, allowing a visual estimation of the population distributions of propensity score in the included and excluded populations. The population propensity score differences are most notable for the PCI within 90 min (AMI-8a) and the discharge measures. The included and excluded populations were most similar for aspirin at arrival (AMI-1).

Propensity-score-stratified analysis
Patient characteristics in Decile 1 (low probability of exclusion) and Decile 10 (high probability of

exclusion) of the propensity score distributions for PCI within 90 min and statin at discharge are presented in table 2. Additional patient-level characteristics in Decile 1 and Decile 10 of PS for the other measures are presented in online supplementary eTable 7. Compared with the bottom decile, patients in the top decile for exclusion from PCI within 90 min (AMI-8a) tended to be older (mean age of 78.1 (10.8) vs 50.3 (8.6) years), were more likely to have Medicare as a payer (90.5% vs 0.9%), were more likely to be transferred to another hospital
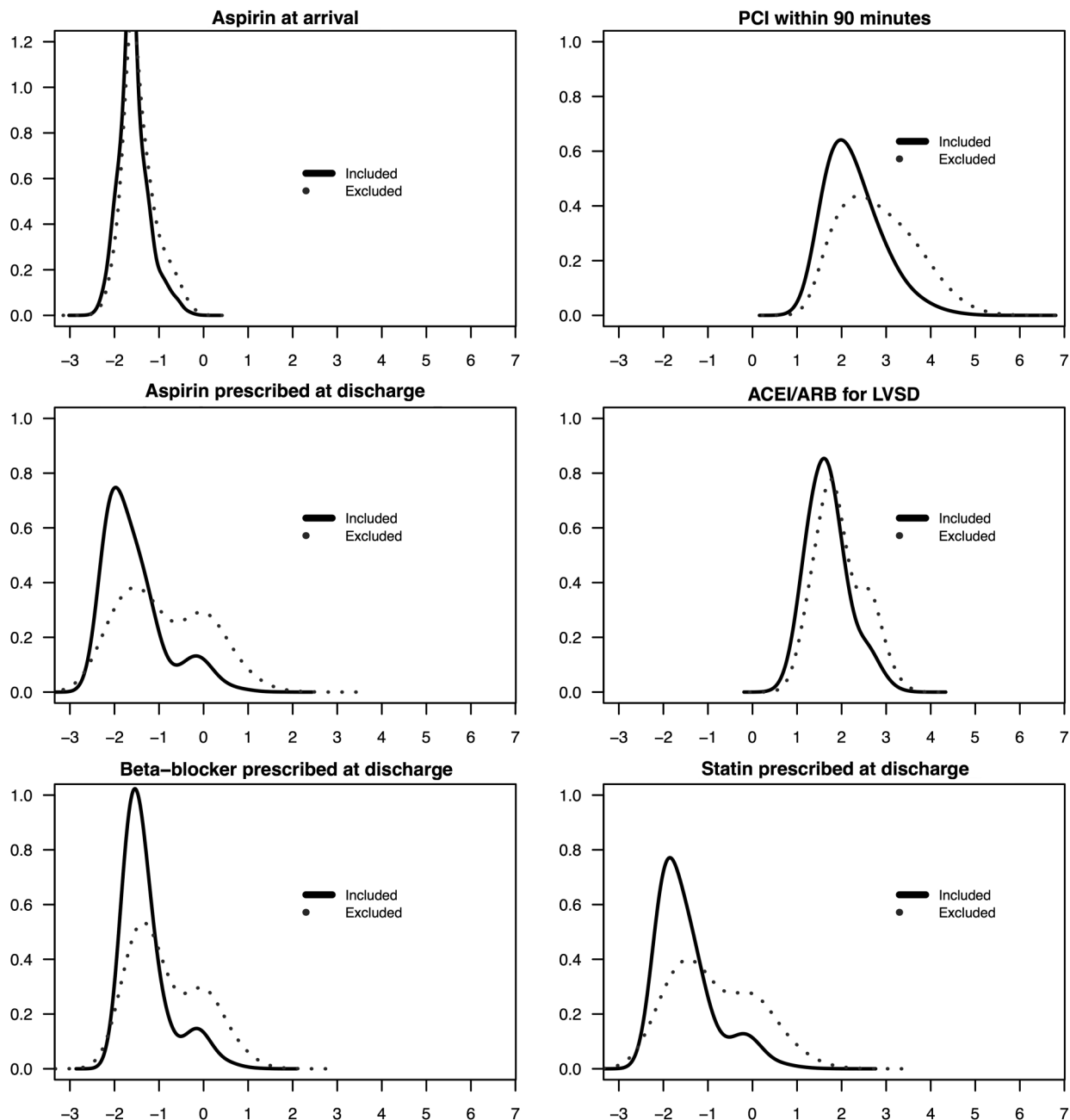
Figure 3 Kernel density plots of propensity score distributions in the included and excluded populations. Plots represent the estimated density of predicted probabilities of exclusion in the included and excluded groups. Regions of overlap identify populations with similar probabilities of exclusion. ACEi, ACE inhibitor; ARB, angiotensin receptor blocker; PCI, percutaneous coronary intervention; LVSD, left ventricular systolic dysfunction.

(53.8% vs 6.9%), had more comorbidities (mean 15.6 (4.6) vs 6.2 (2.5) HCC categories), had higher admission severity and mortality scores (Major or Extreme in 80.9% vs 7.9% and 80.9% vs 7.3%, respectively) and had a higher mortality rate during their hospitalisation (10.0% vs 1.6%).

DISCUSSION

Our study findings highlight a number of issues impacting the usefulness of NHQM process measures for AMI as well as other process measures that have population exclusion and inclusion criteria. Although

in some cases exclusions were for clinically appropriate reasons (ie, time to PCI and ACEi/ARB prescription), the rationale and appropriateness of other exclusion criteria (such as transfer exclusions) are more problematic. Excluded populations tended to include older patients on Medicare, with more comorbidities and more severe presentations.

Our data raise two primary concerns. First, high exclusion rates for some measures result in small measurement population denominators. Second, measures tend to exclude important high-risk populations (such as older patients and those with multiple

Table 2 Propensity-score-stratified patient characteristics

		PCI within 90 min AMI-8a		Statin at discharge AMI-10	
		Decile 1	Decile 10	Decile 1	Decile 10
Overall exclusion rate		90.06%		18.62%	
Stratified exclusion rate	n (%)	12 676 (78.71)	15 825 (98.26)	832 (7.71)	5273 (48.89)
Age (years)	x̄ (SD)	50.32 (8.58)	78.08 (10.80)	50.83 (8.79)	74.07 (12.34)
Male	%	96.54	40.46	80.57	57.06
Race					
White	%	72.44	75.23	61.89	76.38
Black	%	10.97	17.40	25.09	12.46
Asian/Pacific	%	3.51	1.65	4.34	2.09
Other	%	13.08	5.71	8.68	9.07
Hispanic	%	6.61	2.86	8.56	3.14
Primary payer					
Commercial	%	69.32	3.59	66.34	9.85
Medicaid	%	5.27	3.92	10.64	6.77
Medicare	%	0.85	90.52	12.93	73.73
Other	%	24.56	1.96	10.09	9.65
Discharge destination					
Home	%	90.51	31.74	90.73	20.04
Expired	%	1.61	10.02	0.11	30.35
Hospice	%	0.09	4.04	0.02	4.43
Discharge to another hospital	%	6.87	53.81	8.19	44.96
Left against medical advice	%	0.92	0.39	0.96	0.21
ICU stay	%	58.97	50.50	75.68	73.39
Number of HCC categories	x̄ (SD)	6.22 (2.53)	15.59 (4.58)	7.65 (2.76)	15.40 (5.23)
Number of HCC comorbidities	x̄ (SD)	0.41 (0.50)	4.08 (1.07)	1.46 (0.87)	2.90 (1.48)
Admission severity score					
Mild	%	59.87	0.74	53.59	0.76
Moderate	%	32.19	18.32	46.26	10.56
Major	%	2.38	65.43	0.15	33.89
Extreme	%	5.55	15.51	0	54.78
Admission mortality score					
Mild	%	65.57	0.53	64.90	1.17
Moderate	%	27.16	18.56	34.19	11.93
Major	%	2.27	62.70	0.88	37.47
Extreme	%	5.00	18.21	0.03	49.43
Hospitalisation severity score					
Mild	%	53.92	0.50	50.88	1.02
Moderate	%	33.04	12.65	48.90	0.32
Major	%	5.84	57.41	0.22	6.84
Extreme	%	7.20	29.44	0	92.81
Hospitalisation mortality score					
Mild	%	60.49	0.38	62.13	0.03
Moderate	%	27.75	11.56	36.17	0.44
Major	%	4.75	55.44	1.57	17.03
Extreme	%	7.01	32.61	0.13	82.51
Comorbidities					
Hypertension	%	38.57	79.52	84.64	62.16
Stroke	%	0.04	2.25	0	3.53
Cerebrovascular disease	%	0	22.00	3.27	5.42
Renal failure	%	0.01	82.93	1.66	60.91

Table 2 Continued

		PCI within 90 min AMI-8a		Statin at discharge AMI-10	
		Decile 1	Decile 10	Decile 1	Decile 10
COPD	%	0.04	40.25	8.14	22.26
Pneumonia	%	0.43	26.75	0.39	29.85
Diabetes	%	0.83	65.62	41.49	40.21
Protein/calorie malnutrition	%	0.06	9.89	0.02	15.84
Dementia	%	0.01	18.81	0.03	8.17
Functional disorder	%	0	4.09	0.71	1.55
Peripheral vascular disease	%	0.07	44.73	3.64	24.00
Metastatic cancer	%	0.02	4.09	0	6.45
Psychiatric disease	%	1.16	3.65	2.18	4.12
Liver disease	%	0.04	2.97	0	5.42

AMI, acute myocardial infarction; COPD, chronic obstructive pulmonary disease; HCC, hierarchical clinical conditions; ICU, intensive care unit; PCI, percutaneous coronary intervention.

comorbidities). With regard to the first concern, measures with small denominators may be unduly influenced at the margin by single patient encounters, and they may lack sufficient power to differentiate performance between hospitals. Assuming the median sample size of 28 patients at COTH hospitals for the AMI-8a measure, measure non-performance for a single patient will change the performance rate by 3.6%. Performance failure on more than two patients at this sample size would place the hospital below the 2013 VBP achievement threshold for the AMI-8a measure.[28] The effect is even more pronounced at smaller hospitals with smaller measurement population sizes (despite the presence of strict sample size guidelines that relieve hospitals with small population sizes from the reporting requirement).

A second concern is that the final measure denominators excluded some important subpopulations of interest, including Medicare patients and those with high medical complexity and severity of illness. Consequently, while performance measures may reflect a hospital's ability to provide care for a typical subset of patients, they may not accurately characterise their effectiveness in caring for more complex subpopulations. For example, the AMI measures may not accurately reflect the care provided at academic medical centres, where treatment of the most complex patients, who require appropriate and timely care, is an important distinguishing factor.[29] More frequent exclusion of Medicare patients from the denominator populations compounds the difficulty in accurately assessing and (through VBP programmes) improving the quality of care provided to this population.

Our study had several limitations. First, the data were derived from the UHC database, which includes primarily academic medical centres and their affiliates. Additionally, we were only able to obtain data from hospitals that use UHC to assist in submission of NHQM data to CMS. Although this limits the generalisability of the data to other groups of hospitals, UHC represents the single largest source of data available that allows determination of the reason for exclusion from the measures. Our analyses did not allow for accurate determination of patient presentation (ST-segment myocardial infarction vs non-ST segment myocardial infarction), which limited some analyses (particularly the (AMI-8a) measure).

Our study is the first to examine specific reasons for exclusion from the measurement populations in a large, nationwide sample of patients, and among contemporary measures included in VBP. These findings impact the interpretation of the AMI measures for single hospitals and weaken the ability to make meaningful comparisons between hospitals, and have important implications for design of future measures for clinical conditions and pay-for-performance strategies. Our data highlight the need for transition to risk-adjusted outcomes measures with broad inclusion criteria that can more fully account for care provided to complex patients. Although AMI process measures have been retired from the VBP programme and replaced with risk-adjusted outcome measures for mortality and readmissions, newer process measure sets for stroke and venous thromboembolism have been introduced.[30] These newer measure sets are largely similar in structure to earlier AMI process measures and rely on a similar framework.

A number of potential refinements to process-measure-based programmes could help mitigate the problems with small sample sizes and differences in representativeness between the included and excluded groups. Small sample sizes could be improved by increasing the minimum sample size requirement (currently 311 patients per quarter for hospitals with ≥1551 initial cases), or using a longer measurement period. Multicomponent composite measures might also increase effective sample size and provide more comprehensive assessment of quality.

Measures might be developed that stratify performance for lower-risk patients versus more complex patients, rather than excluding the latter, as is currently done. Finally, requiring that hospitals publish the denominator exclusions and demographics of the initial measurement population and the final denominator population would provide external stakeholders a much better understanding of these measures, including their limitations.

CONCLUSION

Improving the quality of care provided to patients with AMI is a noble and important goal. To make real gains in quality, it is crucial that measurement is accurate and responsive to change, and that measures are representative of the populations of patients treated at various hospitals. Progressive refinement of measurement systems is critical, with the goal of ensuring relative homogeneity of the study cohorts while also achieving adequate sample size and representativeness. These considerations will further contribute to advancing the quality of care for patients with AMI, and the lessons learned from these measures are applicable to many other quality measurement programmes.

REFERENCES

1 Mehta RH, Montoye CK, Faul J, et al. Enhancing quality of care for acute myocardial infarction: shifting the focus of improvement from key indicators to process of care and tool use: the American College of Cardiology Acute Myocardial Infarction Guidelines Applied in Practice Project in Michigan: Flint and Saginaw Expansion. *J Am Coll Cardiol* 2004;43:2166–73.

2 Centers for Medicare and Medicaid Services. Hospital Compare. http://www.medicare.gov/hospitalcompare/Data/About.html (accessed 18 Dec 2014).

3 Davies HT, Marshall MN. Public disclosure of performance data: does the public get what the public wants? *Lancet* 1999;353:1639–40.

4 Schneider EC, Epstein AM. Use of public performance reports: a survey of patients undergoing cardiac surgery. *JAMA* 1998;279:1638–42.

5 Ellerbeck EF, Jencks SF, Radford MJ, et al. Quality of care for Medicare patients with acute myocardial infarction. A four-state pilot study from the Cooperative Cardiovascular Project. *JAMA* 1995;273:1509–14.

6 The Joint Commission and Centers for Medicare and Medicaid Services. Specifications Manual for National Inpatient Hospital Quality Measures (Versions 2.5b through 4.3b). http://www.

qualitynet.org/dcs/ContentServer?pagename=QnetPublic%2FPage%2FQnetTier2&cid=1141662756099 (accessed 15 Oct 2014).

7 Richardson WC (chair). *Crossing the quality chasm: a new health system for the 21st century.* Institute of Medicine, Committee on Quality of Health Care in America. Washington DC: National Academies Press, 2001.

8 Centers for Medicare and Medicaid Services. Hospital Inpatient Quality Reporting Program. 2013. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalRHQDAPU.html (accessed 18 Dec 2014).

9 The Joint Commission. Core Measure Sets. http://www.jointcommission.org/core_measure_sets.aspx (accessed 18 Dec 2014).

10 Blumenthal D and Jena AB. Hospital value-based purchasing. *J Hosp Med* 2013;8:271–7.

11 Bilimoria KY. Facilitating quality improvement: pushing the pendulum back toward process measures. *JAMA* 2015;314:1333–4.

12 Gross PA. Editorial process versus outcome measures: the end of the debate. *Med Care* 2012;50:200–2.

13 Lilford RJ, Brown CA, Nicholl J. Use of process measures to monitor the quality of clinical practice. *BMJ* 2007;335:648–50.

14 Shahian DM, Meyer GS, Mort E, et al. Association of National Hospital Quality Measure adherence with long-term mortality and readmissions. *BMJ Qual Saf* 2012;21:325–36.

15 Gurm HS, Valle JA, Smith DE, et al. Eroding the denominator: the incomplete story of door-to-balloon time reporting. *J Am Coll Cardiol* 2012;60:789–90.

16 Rathore SS, Wang Y, Radford MJ, et al. Quality of care of Medicare beneficiaries with acute myocardial infarction: who is included in quality improvement measurement? *J Am Geriatr Soc* 2003;51:466–75.

17 Bernheim SM, Wang Y, Bradley EH, et al. Who is missing from the measures? Trends in the proportion and treatment of patients potentially excluded from publicly reported quality measures. *Am Heart J* 2010;160:943–50.e1-5.

18 McCabe JM, Kennedy KF, Eisenhauer AC, et al. Reporting trends and outcomes in ST-segment-elevation myocardial infarction National Hospital Quality Assessment Programs. *Circulation* 2014;129:194–202.

19 American Hospital Association. AHA Annual Survey Database™ Fiscal Year 2013. http://www.ahadataviewer.com/book-cd-products/AHA-Survey/ (accessed 1/2/2014).

20 Baram D, Daroowalla F, Garcia R, et al. Use of the All Patient Refined-Diagnosis Related Group (APR-DRG) risk of mortality score as a severity adjustor in the medical ICU. *Clin Med Circ Respir Pulm Med* 2008;2:19–25.

21 Iezzoni LI, Ash AS, Shwartz M, et al. Predicting who dies depends on how severity is measured: implications for evaluating patient outcomes. *Ann Intern Med* 1995;123:763–70.

22 3M. Overview of the 3M™ APR DRG Classification System. http://multimedia.3m.com/mws/media/910941O/ebook-overview-of-the-3m-apr-drgs-10-13.pdf?fn=3M_APR_DRG_eBook.pdf (accessed 02/11/2015).18.

23 Li P, Kim MM, Doshi JA. Comparison of the performance of the CMS Hierarchical Condition Category (CMS-HCC) risk adjuster with the Charlson and Elixhauser comorbidity measures in predicting mortality. *BMC Health Serv Res* 2010;10:245.

24 Pope GC, Kautter J, Ellis RP, *et al*. Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financ Rev* 2004;25:119–41.

25 Kautter J, Pope GC, Ingber M, *et al*. The HHS-HCC risk adjustment model for individual and small group markets under the Affordable Care Act. *Medicare Medicaid Res Rev* 2014;4:pii: mmrr2014-004-03-a03.

26 Wickham H, Chang W. *ggplot2: elegant graphics for data analysis*. New York: Springer, 2009.

27 Reynolds A, Gerrard D. Venn/Euler Diagram of Four or More Sets: Euler Grid. BioStar: An Online Question & Answer Resource for the Bioinformatics Community 2013. https://www.biostars.org/p/77362/#77377, https://www.biostars.org/p/77362/#80850 (accessed 1/2015).

28 Centers for Medicare and Medicaid Services. *Frequently asked questions, hospital value-based purchasing program*. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/HVBP/HVBP-FAQs.pdf. Washington, DC, 2012.

29 Hochman JS, Sleeper LA, Webb JG, *et al*. Early revascularization in acute myocardial infarction complicated by cardiogenic shock. SHOCK Investigators. Should We Emergently Revascularize Occluded Coronaries for Cardiogenic Shock. *N Engl J Med* 1999;341:625–34.

30 Medicare Program; Hospital Inpatient Prospective Payment Systems for Acute Care Hospitals Final Rule. Department of Health and Human Services. CMS-1607-F. Published 8/22/2014.