Supplemental Appendix

Outcome-based Hospital Quality Measures

More formally, define as a binary indicator of whether patient treated at hospital experienced the outcome, and define $\mathbf{x}$ as a vector of patient-level characteristics. The current CMS model assumes the following:

(4)
$$Y_{ih}|\alpha_h, \beta, \mathbf{x_{ih}} \overset{ind}{\sim} \text{Logistic}(\alpha_h + \beta \mathbf{x_{ih}})$$

(5)
$$\alpha_h|\mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$$

where $\mathbf{x_{ih}}$ includes age and sex. Based on this model, the risk-standardized rate for hospital h (RSR_h) is estimated as

(6)
$$\widehat{\text{RSR}_h}(\mathbf{x_{ih}}) = \frac{\sum_{i=1}^{n_h} E(Y_{ih}|\hat{\alpha}_h, \hat{\mu}, \hat{\beta}, \mathbf{x_{ih}}, \hat{\sigma}^2)}{\sum_{i=1}^{n_h} E(Y_{ih}|\hat{\mu}, \hat{\beta}, \mathbf{x_{ih}}, \hat{\sigma}^2)} \bar{Y}$$

where $\bar{Y}$ is the national average outcome and n_h is the number of patients treated in hospital h (Ash et al. 2012).

In practice, equation (4) is estimated as a hierarchical Bayes logistic regression model that includes a hospital random effect (α_h). The underlying patient-level data and the estimated model parameters ($\hat{\alpha}_h, \hat{\mu}, \hat{\beta}, \hat{\sigma}^2$) are used to construct predicted values, which are then fed into equation (6) to obtain the risk-standardized rate for each hospital.

The numerator in equation (6) is simply the sum of predicted values for each patient in the hospital, where the predicted values are based on observed patient values, $\hat{\mu}, \hat{\beta}$, and the estimated hospital random effect ($\hat{\alpha}_h$). The denominator is similarly estimated as the sum of patient-level predictions, but only $\hat{\mu}$ – the estimated national mean of $\hat{\alpha}_h$ – is used. This measure is then multiplied by $\bar{Y}$ to place the risk standardized measure on the same scale as the outcome. In effect, outcome-based quality assessments are determined primarily by how much each hospital's estimated random effect ($\hat{\alpha}_h$) deviates from $\hat{\mu}$. The hypothetical reference hospital in the denominator of equation (6) has the same patient case-mix as the hospital in question, however it has average

quality, as reflected by the use of only $\hat{\mu}$ to summarize hospital-level factors. In this way each hospital is evaluated against a hypothetical reference hospital with an identical case-mix and with average quality.

Table A1: CMS Quality Measures Used to Construct Composite Domain Scores

Domain	Measure
Timely and Effective Care	
	Heart Failure Patients Given ACE Inhibitor or ARB for Left Ventricular Systolic Dysfunction (LVSD)
	AMI Patients Given Aspirin at Discharge
	Heart Failure Patients Given Assessment of Left Ventricular Function (LVF)
	Heart Failure Patients Given Discharge Instructions
	Pneumonia Patients Given the Most Appropriate Initial Antibiotic(s)
	Surgery Patients Who Received Preventative Antibiotic(s) One Hour Before Incision
	Surgery Patients Whose Preventative Antibiotic(s) are Stopped Within 24 hours After Surgery
Patient Experience of Care	
	Doctors always communicated well
	Nurses always communicated well
	Pain was always well controlled
	Patients always received help as soon as they wanted – Patients who gave a rating of 9 or 10 (high)
	Room was always clean
	Room always quiet at night
	Staff always explained
	Yes, patients would definitely recommend the hospital
	Yes, staff did give patients this information
30-Day Mortality Rates	
	AMI 30-Day Mortality Rate
	Pneumonia 30-Day Mortality Rate
	Heart Failure 30-Day Mortality Rate
30-Day Readmission Rates	
	AMI 30-Day Readmission Rate
	Pneumonia 30-Day Readmission Rate
	Heart Failure 30-Day Readmission Rate

Source: Archived Hospital Compare website data. The process and outcome measures are defined for specific patient conditions (e.g., AMI, pneumonia, heart failure) whereas the HCAHPS patient satisfaction surveys are distributed to a wider range of patients.

Table A2: Balance: Non-Deferrable Condition Sample

Measure	Timely and Effective Care	Patient Satisfaction
162 Malignant neoplasm of trachea, bronchus, and lung	-0.00019	-0.00030
197 Secondary malignant neoplasm of respiratory and digestive systems	-0.00083	0.0013
410 Acute myocardial infarction	0.0023	-0.0035
431 Intracerebral hemorrhage	0.0017	0.0018
433 Occlusion and stenosis of precerebral arteries	0.00034	-0.00075
434 Occlusion of cerebral arteries	0.0038	-0.0039
435 Transient cerebral ischemia	0.000091	-0.0014
482 Other bacterial pneumonia	-0.00071	-0.00026
486 Pneumonia, organism unspecified	-0.011**	0.00027
507 Pneumonitis due to solids and liquids	0.00028	0.00055
518 Other diseases of lung	0.00091	0.0046
530 Diseases of esophagus	-0.0018	-0.0016
531 Gastric ulcer	0.0019	0.00063
532 Duodenal ulcer	0.00014	0.00031
557 Vascular insuciency of intestine	-0.00019	-0.00011
558 Other and unspecified noninfectious gastroenteritis and colitis	-0.0024*	0.000077
560 Intestinal obstruction without mention of hernia	-0.0047*	-0.0055**
599 Other disorders of urethra and urinary tract	-0.0089**	-0.0027
728 Disorders of muscle, ligament, and fascia	0.00073	-0.00083
780 General symptoms	0.0045	-0.0076*
807 Fracture of rib(s), sternum, larynx, and trachea	0.0016	0.0016
808 Fracture of pelvis	-0.00087	0.0055***
820 Fracture of neck of femur	0.010*	0.021***
823 Fracture of tibia and fibula	0.00014	0.00012
824 Fracture of ankle	0.0013	0.0025*
959 Injury, other and unspecified	0.000023	0.00040
965 Poisoning by analgesics, antipyretics, and antirheumatics	0.00076	0.0010
969 Poisoning by psychotropic agents	0.000010	-0.00031

Table shows balance across principal diagnosis controls used in regressions. Balance is assessed using pairwise regressions of each characteristic on each instrument based on the specification in Equation 5. The reported estimates (rows) show the coefficient on the instrument from each of these regressions fit separately for each instrument (columns). Sample size = 546,700

Table A3: Balance: Non-Deferrable Condition Sample

Measure	30-Day Mortality Rate	30-Day Readmission Rate
162 Malignant neoplasm of trachea, bronchus, and lung	0.00060	-0.0016
197 Secondary malignant neoplasm of respiratory and digestive systems	-0.0011	0.00010
410 Acute myocardial infarction	-0.013***	0.0023
431 Intracerebral hemorrhage	0.00027	0.0028*
433 Occlusion and stenosis of precerebral arteries	0.00077	0.00041
434 Occlusion of cerebral arteries	0.0013	0.00057
435 Transient cerebral ischemia	-0.00034	0.00037
482 Other bacterial pneumonia	-0.00088	0.00013
486 Pneumonia, organism unspecified	0.013***	0.00084
507 Pneumonitis due to solids and liquids	-0.0061**	0.0036
518 Other diseases of lung	-0.0028	-0.00044
530 Diseases of esophagus	-0.0019	0.00061
531 Gastric ulcer	0.0018	-0.00049
532 Duodenal ulcer	-0.0014	0.00065
557 Vascular insuciency of intestine	0.00060	-0.0015
558 Other and unspecified noninfectious gastroenteritis and colitis	0.0014	0.0012
560 Intestinal obstruction without mention of hernia	-0.0011	-0.0021
599 Other disorders of urethra and urinary tract	0.0063*	0.0058
728 Disorders of muscle, ligament, and fascia	0.0029**	-0.00087
780 General symptoms	-0.0042	0.011***
807 Fracture of rib(s), sternum, larynx, and trachea	-0.000027	-0.0012
808 Fracture of pelvis	0.0014	-0.0034*
820 Fracture of neck of femur	0.013***	-0.017***
823 Fracture of tibia and fibula	-0.00082	-0.000040
824 Fracture of ankle	0.00012	-0.0013
959 Injury, other and unspecified	-0.000014	0.00019
965 Poisoning by analgesics, antipyretics, and antirheumatics	-0.0015**	0.00071
969 Poisoning by psychotropic agents	0.00047	0.00038

Table shows balance across principal diagnosis controls used in regressions. Balance is assessed using pairwise regressions of each characteristic on each instrument based on the specification in Equation 5. The reported estimates (rows) show the coefficient on the instrument from each of these regressions fit separately for each instrument (columns). Sample size = 546,700

Table A4: 2SLS Results - Unstandardized Measures

Outcome:		30D Readmission	30D Mortality	365D Mortality
Quality Measure	Mean [SD]	(1)	(2)	(3)
Timely and Effective Care Composite	93.40	-0.00035	-0.0014	-0.0038
	[5.01]	(0.0007)	(0.0008)	(0.001)
Patient Experience Composite	66.89	-0.0021	-0.00057	-0.0028
	[5.02]	(0.0007)	(0.0008)	(0.001)
30-Day Mortality Rate Composite	12.69	-0.0048	0.008	0.010
	[1.29]	(0.003)	(0.003)	(0.004)
30-Day Readmission Rate Composite	20.98	0.009	-0.0026	0.002
	[1.48]	(0.002)	(0.003)	(0.003)

Each cell reports two-stage least squares (2SLS) coefficient estimates for a separate regression. Outcome means: 30D Readmission = 15.0%, 30D Mortality = 17.0%, 365D Mortality = 37.2%. Sample sizes: 546,700 (30-Day outcomes), 451,503 (1-Year Mortality). All models include patient demographic and ambulance controls as listed in Table 1, as well as the diagnosis controls as listed in Table A2. Models also include ZIP code-patient origin fixed effects. Standard errors, clustered at Health Service Area (HSA) level, are reported in parentheses.

Table A5: 2SLS Results - Concurrent Quality Measures

Outcome:		30D Readmission	30D Mortality	365D Mortality
Quality Measure	Mean [SD]	(1)	(2)	(3)
Timely and Effective Care Composite	91.09	-0.0030	-0.013	-0.036
	[6.10]	(0.007)	(0.008)	(0.014)
Patient Experience Composite	66.12	-0.024	-0.0034	-0.026
	[5.17]	(0.007)	(0.008)	(0.010)
30-Day Mortality Rate Composite	12.67	-0.0071	0.021	0.022
	[1.26]	(0.007)	(0.007)	(0.010)
30-Day Readmission Rate Composite	20.93	0.026	-0.0050	0.004
	[1.49]	(0.007)	(0.007)	(0.010)

This table replicates the results from Table 4 but using measures of quality with measurment that runs concurrent to the patient data. For example, the process quality measure used for patients admitted in 2008 reflects the values published in 2009, since those values were measured based on 2008 admissions. Each cell reports two-stage least squares (2SLS) coefficient estimates for a separate regression. Outcome means: 30D Readmission = 15.0%, 30D Mortality = 17.0%, 365D Mortality = 37.2%. Sample sizes: 546,700 (30-Day outcomes), 451,503 (1-Year Mortality). All models include patient demographic and ambulance controls as listed in Table 1, as well as the diagnosis controls as listed in Table A2. Models also include ZIP code-patient origin fixed effects. Standard errors, clustered at Health Service Area (HSA) level, are reported in parentheses.

Table A6: 2SLS Results - Excluding AMI, and PN Patients

Outcome:		30D Readmission	30D Mortality	365D Mortality
Quality Measure	Mean [SD]	(1)	(2)	(3)
Timely and Effective Care Composite	93.43	-0.0053	-0.015	-0.023
	[4.98]	(0.008)	(0.009)	(0.015)
Patient Experience Composite	66.87	-0.022	0.003	-0.019
	[5.02]	(0.008)	(0.008)	(0.012)
30-Day Mortality Rate Composite	12.69	-0.015	0.016	0.021
	[1.29]	(0.008)	(0.008)	(0.011)
30-Day Readmission Rate Composite	20.98	0.029	-0.010	-0.0026
	[1.49]	(0.008)	(0.008)	(0.011)

This table replicates the results from Table 4 but using nondeferrable conditions that do not include AMI, HF, and PN patients. Each cell reports two-stage least squares (2SLS) coefficient estimates for a separate regression. Outcome means: 30D Readmission = 14.5%, 30D Mortality = 17.2%, 365D Mortality = 37.1%. Sample sizes: 438,911 (30-Day outcomes), 361,958 (1-Year Mortality). All models include patient demographic and ambulance controls as listed in Table 1, as well as the diagnosis controls as listed in Table A2. Models also include ZIP code-patient origin fixed effects. Standard errors, clustered at Health Service Area (HSA) level, are reported in parentheses.