Appendix

The Association Between Primary Care Physician Diagnostic Knowledge and Death, Hospitalization and Emergency Department Visits Following an Outpatient Visit at Risk for Diagnostic Error: A Retrospective Cohort Study Using Medicare Claims

Bradley M. Gray, PhD, corresponding author
Email: bgray@abim.org 202-213 6646, FAX 202-213 6646
American Board of Internal Medicine
510 Walnut Street, Suite 1700, Philadelphia, Pennsylvania 19106, USA

Contents

Section 1: 90-day Index Visit Clean Period Derivation

Figure 1.1 displays the visit periodicity between each of the 921,416 visits to an internist in the sample and the most recent visit prior to that one.

To determine what the index visit clean period was we assumed that when two contacts happen "close" together they are more likely to be visits for the same acute episode of care. Therefore, if we exclude all but the first visit that happen "close" together then the remaining visits are highly likely to represent the first visit for a new episode of care (i.e., a new problem). However, the visit periodicity threshold that distinguishes visits that are "close" versus "not close" is unknown.

To help delineate this threshold, Figure 1.1 visit shows periodicity between each of the 921,416 visits to an internist in the sample and the most recent adjacent visit prior to that one. In Figure 1.1, you can see the slope of frequency curve is falling until about a 90 day gap between visits. This indicates that many of the visits prior to this point may be related to an existing episode of care. After 90 days, the periodicity slope begins to flatten out which indicates that the timing between visits is likely random and so it is less likely that the two visits are related to the same episode of care.

This flattening of the slope after 90 days is more clearly displayed in Figure 1.2 which displays the 15 day moving average of the change in visit counts per day (i.e., the changing slope). Here the slope stabilizes at about zero beginning around 90 days suggesting that a 90 day clean period for physician visits is likely to exclude most visits that are a follow-up to an ongoing episode of care from the index visit sample.

Figure 1.1. Visit Periodicity Plot for the 921,416 Outpatient Visits to Physicians in the Sample

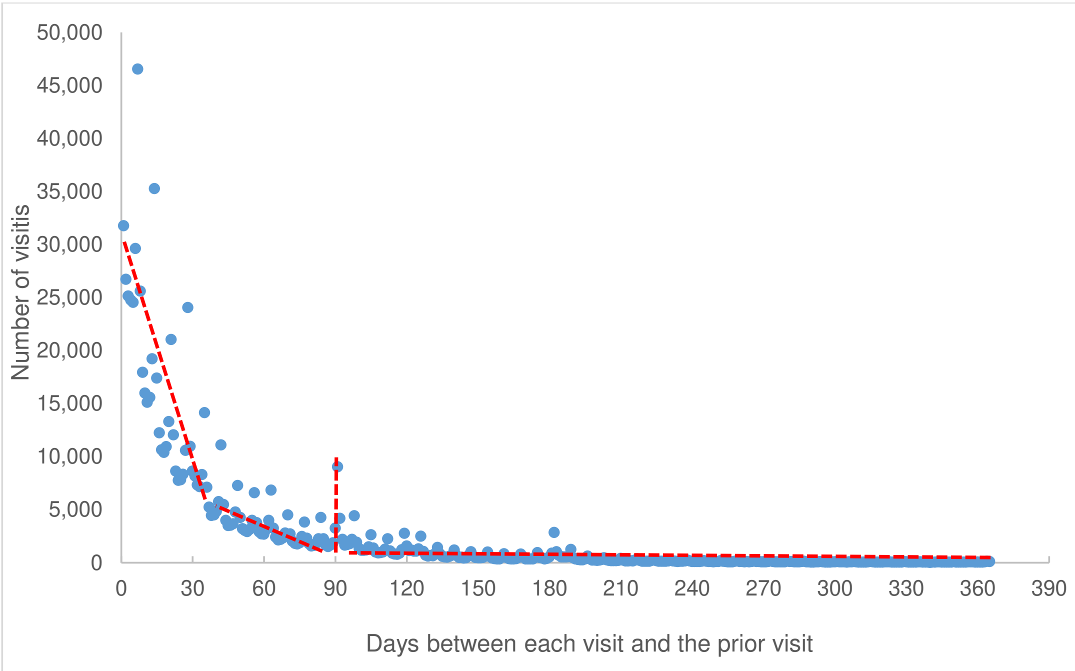
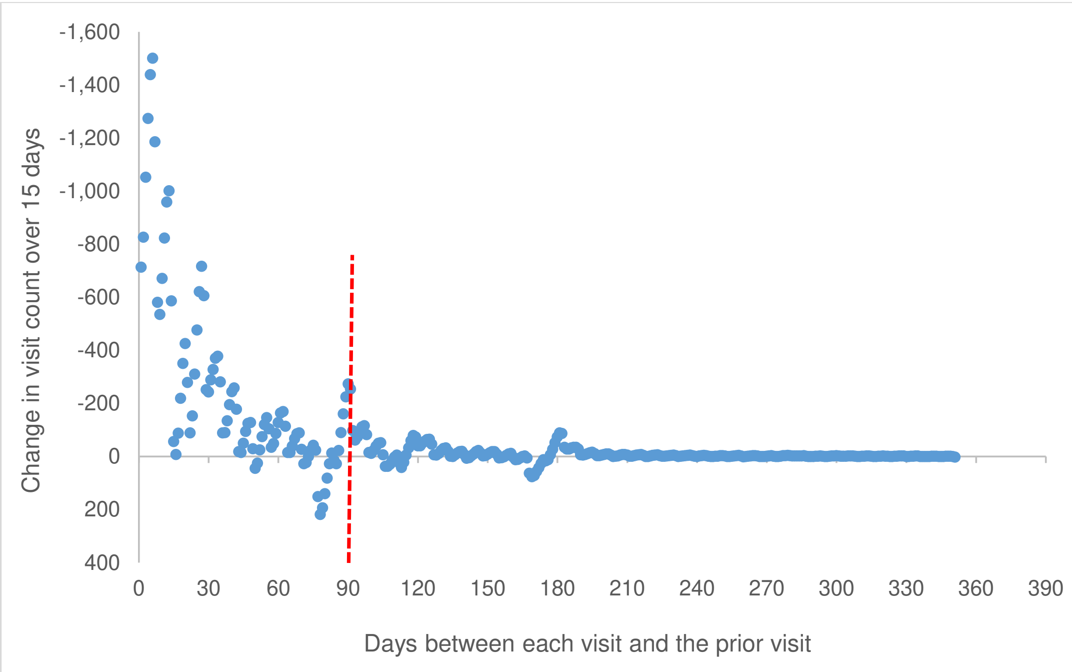


Figure 1.2. Average Change in Visit Count over the 15 days (15-day slope) Following each Data Point Listed in Figure 1



Section 2. ICD-9 Code Codes for Diagnostic Error sensitive Conditions and ICD-9 Code Groups for Index Visit Eligibility and Related Relative Risks

The Section includes a list the list of diagnostic error sensitive condition ICD-9 diagnoses (Table 2.1), index visit eligible ICD-9s diagnoses groups (Table 2.2), and relative risks for each index visit diagnosis group (Table 2.3).

eTable 2.1. ICD-9 Code Codes for Diagnostic Error Sensitive Conditions

Diagnostic Error Sensitive Conditions	ICD-9 groups
Abscess	681, 682
Acute Coronary Syndrome	410, 411.1
Anemia	280-284
Appendicitis	540-542, 543.0, 543.9
Aortic aneurysm	441
Bacteremia/Sepsis	038, 003.1, 020.2, 022.3, 036.2, 054.5, 449, 771.81, 790.7, 995.91, 995.92
Depression	296.2, 296.3
Fracture	800-829, 733.81
Congestive Heart failure	428
Pneumonia	480-486
Pulmonary embolism	415.1
Spinal cord compression	336.9
Stroke	430-437

eTable 2.2. ICD-9 Code Groups for Index Visit Eligibility

Index visit ICD-9 recorded diagnosis ICD-9 codes (76 different diagnoses)	ICD-9s	Met at least one diagnostic error sensitive relative risk criterion (38 met this criteria)
Abdominal pain	789.0	Yes
Abdominal tenderness	789.6	No
Abnormal respiration	786.0	Yes
Alcohol	291.0-291.5, 291.8, 291.9, 357.5, 425.5, 571.0-571.3, 303, 305.00-305.03, 535.30-535.31, E860.0	No
Amphetamines	304.4	No
Anxiety	300.0	Yes
Ascites	785.5	Yes
Back pain	724.5	Yes
Bronchitis	466.0, 466.1	Yes
Cannabis	304.3, 305.2	No
Celiac disease	579.0	No
Chest Pain	786.50, 786.51, 786.59	Yes
Chills	780.64	No
Cocaine	304.2, 305.6, E938.25	No
Confusion	298.2	Yes
Cough	786.2	Yes
Deep vein thrombosis	453.40	No
Delirium	293.0, 780.97	Yes
Diverticulitis	562.11	Yes
Dizziness	780.4	Yes
Drug Mental Disease	292	No
Dyspnea	786.09	Yes
Dysthymia	300.4	Yes
Edema	782.3	Yes
Elevated blood pressure	796.2	No
Esophageal disease	530.1, 530.3-530.9	Yes
Facial weakness	728.87	Yes
Falls	v15.88	No
Fatigue	780.7	Yes
Fever	780.60, 780.61	Yes
Gait instability	781.2	Yes
Gastritis	535	No
Gastrointestinal bleeding	578.9	Yes
Hallucinogens	304.5, 305.3, 969.6, E854.1, E939.6	No
Headache	339, 346, 784.0	Yes
Heart Burn	787.1	No
Hemoptysis	786.30, 786.39	Yes
Hyperparathyroidism	262.0	Yes
Hypoxemia/Hypoxia	799.02	Yes
Influenza	487.0, 487.1, 487.8, 488	No
Lack coordination	781.3	Yes
Lower respiratory disease	519.8	No
Lung cancer	162	Yes
Menorrhagia	626.2	No
Mood disorder	293.83, 293.84	No

Gray BM, *et al. BMJ Open* 2021; 11:e041817. doi: 10.1136/bmjopen-2020-041817

Nausea	787.01, 787.02	Yes
Opioids	304.0, 304.7, 305.5, 965.0, E850.0-E850.2, E935.0-E935.2	No
Osteopenia	733.90	No
Osteoporosis	733.0	Yes
Other back pain	721.2-721.9, 722.1, 722.2, 722.5, 722.6, 722.70, 722.72, 722.73, 722.80, 722.82, 722.83, 722.90 722.92, 722.93, 724.0, 724.1	Yes
Other respiratory issue	786.00,786.01, 786.06, 786.07, 786.52, 786.1, 786.2	Yes
Otitis media	381-383, 387, 055.2, 384.2, 384.8, 384.9, 385.0-385.2	No
Personality disorder	301	No
Pain respiration	786.52	Yes
Peripheral neuropathy	337.9, 337.1	No
Reflux disease	530.81	Yes
related alcohol disease	291.9, 292, 304.0-304.6	No
Respiratory Distress	518.81	No
Sedatives	304.1	No
Shortness of breath	786.05	Yes
Sinusitis	473	Yes
Speech disturbance	784.5	Yes
Stress	308	No
Stress fracture	733.94-733.98	No
Tachycardia	785.0	Yes
Tension headache	307.81	No
Thunderclap headache	339.43	No
Transient ischemic attack	435.0-435.3, 435.8, 435.9	Yes
Upper respiratory disease	472, 476, 477, 478.8	No
Viral illness	079.99	No
Vitamin D deficiency	268	Yes
Vomiting	787.01, 787.03	Yes
Weakness/Fatigue	728.87, 708.7	Yes
Weight gain	783.1	No
Weight loss	783.2	Yes

Gray BM, *et al. BMJ Open* 2021; 11:e041817. doi: 10.1136/bmjopen-2020-041817

eTable 2.3 Relative Risks for each Index Visit Diagnosis

Outcome conditions[a]	Index visit eligibility diagnosis group	Relative Risk[b]
Abscess	Fever	2.65
	Chills	0.00
Anemia	Gastrointestinal bleeding	25.20
	Weight loss	4.09
	Shortness of breath	3.51
	Weakness/Fatigue	2.35
	Hypoxemia/Hypoxia	2.11
	Dyspnea	2.05
	Chest Pain	1.82
	Headache	1.29
	Menorrhagia	0.00
Aortic Aneurysm	Dyspnea	4.98
	Abdominal pain	4.93
	Shortness of breath	3.80
	Chest pain	2.42
	Other back pain	1.64
	Back pain	1.01
	Elevated blood pressure	0.00
Appendicitis	Vomiting	30.79
	Diverticulitis	30.45
	Nausea	16.81
	Abdominal pain	15.60
	Abdominal tenderness	0.00
	Fever	0.00
Bacteremia/Sepsis	Vomiting	6.99
	Fever	5.10
	Nausea	3.82
	Tachycardia	2.67
	Weakness/Fatigue	1.75
Heart failure	Hypoxemia/Hypoxia	9.99
	Shortness of breath	5.09
	Dyspnea	3.33
	Edema	3.27
	Chest Pain	2.46
	Weakness/Fatigue	1.42
	Ascites	0.00
	Respiratory Distress	0.00

Outcome conditions[a]	Index visit eligibility diagnosis group	Relative Risk[b]
Acute Coronary Syndrome	Chest pain	8.38
	Dyspnea	7.29
	Shortness of breath	3.65
	Hypoxemia/hypoxia	2.01
	Reflux disease	1.23
	Esophageal disease	1.22
	Weakness/Fatigue	1.14
	Nausea	1.05
	Other respiratory issue	0.86
	Respiratory distress	0.00
	Gastritis	0.00
	Heart Burn	0.00
Depression	Delirium	32.76
	Heart failure	6.16
	Anxiety	5.04
	Dysthymia	4.99
	Weight loss	4.73
	Anemia	2.74
	Fatigue	1.06
	Alcohol	0.00
	Amphetamines	0.00
	Cannabis	0.00
	Cocaine	0.00
	Drug Mental Disease	0.00
	Hallucinogens	0.00
	Opioids	0.00
	Personality disorder	0.00
	related alcohol disease	0.00
	Sedatives	0.00
	Stress	0.00
	Weight gain	0.00
	Mood disorder	0.00
Fracture	Gait instability	2.53
	Edema	1.79
	Osteoporosis	1.66
	Hyperparathyroidism	1.09
	Vitamin D deficiency	1.08

Gray BM, *et al. BMJ Open* 2021; 11:e041817. doi: 10.1136/bmjopen-2020-041817

Outcome conditions[a]	Index visit eligibility diagnosis group	Relative Risk[b]
Fracture (con't)	Osteopenia	0.54
	Celiac disease	0.00
	Falls	0.00
	Stress fracture	0.00
Pulmonary embolism	Tachycardia	12.16
	Hypoxemia/hypoxia	10.98
	Shortness of breath	6.75
	Dyspnea	6.54
	Abnormal respiration	6.35
	Heart failure	4.51
	Chest pain	4.31
	Cough	1.48
	Other respiratory issue	1.34
	Deep vein thrombosis	0.00
	Respiratory distress	0.00
	Fever	0.00
	Heart burn	0.00
	Hemoptysis	0.00
Pneumonia	Hypoxemia/hypoxia	8.24
	Hemoptysis	7.57
	Lung cancer	7.53
	Fever	6.19
	Delirium	5.18
	Bronchitis	3.07
	Shortness of breath	2.99
	Cough	2.77
	Abnormal respiration	2.38
	Pain respiration	2.13
	Dyspnea	2.05
	Weakness/Fatigue	1.38
	Sinusitis	1.26
	Chest Pain	1.00
	Upper respiratory disease	0.71
	Otitis media	0.48
	Influenza	0.00
	Lower respiratory disease	0.00
	Viral illness	0.00

Outcome conditions[a]	Index visit eligibility diagnosis group	Relative Risk[b]
Spinal cord compression	Abdominal pain	31.20
	Back pain	15.03
	Peripheral neuropathy	0.00
	Weakness/Fatigue	0.00
Stroke	Facial weakness	65.24
	Confusion	48.93
	Speech disturbance	19.60
	Transient ischemic attack	7.82
	Delirium	4.96
	Dizziness	3.20
	Lack coordination	2.92
	Gait instability	2.92
	Vomiting	2.15
	Weakness/Fatigue	1.54
	Headache	1.37
	Nausea	1.17
	Thunderclap headache	0.00
	Tension headache	0.00

[a]Outcomes include any hospitalization or emergency department visit for the diagnostic conditions within 90 days of the index visits including same day events
[b]Index visit diagnoses groups applied in the analysis include those with a relative risk greater than one. Relative risks were computed as the probability of an outcome if the index visit diagnosis group was recorded in the index visit divided by the probability of an outcome if the diagnosis group was not recorded.

Gray BM, *et al. BMJ Open* 2021; 11:e041817. doi: 10.1136/bmjopen-2020-041817

Section 3. Psychometric Analysis of Whether Diagnosis Related Questions Reflect an Underlying Construct

To examine the degree to which treatment and diagnosis related questions represented an underlying construct, we calculated separate Cronbach's alpha indices to determine reliability of the subset of items for the 2010 IM-MOC examination for diagnosis related questions and treatment questions.[1] Overall, 170 of the questions were categorized as treatment or diagnostic related with 71 items classified as treatment and 99 items classified as diagnosis related questions. Overall, reliability for the diagnosis related questions was high, 0.84, suggesting that these questions hung together and were related to one underlying construct. The reliability for treatment related questions was also high, 0.75. This index, however, is partly a function of the number of items included in the calculation, where more items typically result in higher reliability. Consequently, it is not surprising that the diagnostic related questions have higher reliability given there were 28 more items than the treatment scale. To make these indices more equal, we computed the Spearman-Brown prophecy formula, which indicates expected reliability if the treatment scale was 99 items instead of 71. That formula resulted in a value of 0.81 for the treatment items which suggests that treatment related questions also measure one underlying construct.

Although performance on diagnosis and treatment related questions were correlated (Pearson Correlation=.62), 59.5% of the variation in diagnosis exam performance for the physician study sample was not explained by performance on other parts of the exam.

Gray BM, *et al. BMJ Open* 2021; 11:e041817. doi: 10.1136/bmjopen-2020-041817

Section 4. Imputations for missing variables

Missing practice characteristics (1,432 or 2.94% of sample) were coded as "other unknown".
Missing HCC (86 or .18% of sample) were replace by in sample mean HCC.
Missing rural indicator (22 or .05% of sample) were assumed to be non-rural
Missing ZIP code median income (708 or 1.46% of sample) were replace by in sample mean median income.

Section 5. Full Regression Coefficient Estimates and Explanatory Variables List

eTables 5.1 lists the probit coefficient associations with outcome measures across all explanatory variables as well as regression descriptive statistics. See Section 7 for percentage point associations with physician characteristics.

eTable 5.1. Probit Coefficient Associations and Regression Descriptive Statistics

	Death		Hospitalization		Emergency Department Visit	
	Wald chi2(102): 815.36		Wald chi2(102): 1197.54		Wald chi2(102): 1201.10	
	Log pseudolikelihood -1588.8		Log pseudolikelihood = -2456.7		Log pseudolikelihood = -2989.0	
Label	Difference per 1,000 (SE)	P	Difference per 1,000 (SE)	P	Difference per 1,000 (SE)	P
Diagnosis question percent correct						
Diagnosis tertile 1	Reference		Reference		Reference	
Diagnosis tertile 2	-1.6 (1.0)	0.09	-2.3 (1.4)	0.09	-3.1 (1.5)	0.04
Diagnosis tertile 3	-2.9 (1.1)	0.008	-4.1 (1.5)	0.006	-4.9 (1.7)	0.003
Treatment question percent correct						
Treatment tertile 1	Reference		Reference		Reference	
Treatment tertile 2	0.7 (0.8)	0.41	-0.7 (1.2)	0.54	-0.3 (1.4)	0.82
Treatment tertile 3	1.6 (1.0)	0.13	1.6 (1.5)	0.29	1.6 (1.7)	0.33
Other question percent correct						
Other tertile 1	Reference		Reference		Reference	
Other tertile 2	1.3 (0.8)	0.12	0.3 (1.2)	0.78	0.1 (1.3)	0.95
Other tertile 3	2.5 (1.0)	0.01	-0.8 (1.3)	0.52	0.5 (1.5)	0.72
Female Physician	-1.2 (0.7)	0.08	-0.8 (1.0)	0.43	-0.7 (1.2)	0.54
Physician birth and medical school						
US born: US medical schools	Reference		Reference		Reference	
US born: Int'l medical schools	1.2 (1.8)	0.51	-1.9 (2.8)	0.50	-0.7 (2.8)	0.79
Int'l born: US medical schools	0.4 (1.1)	0.71	3.1 (1.5)	0.05	2.6 (1.9)	0.18
Int'l born: Int'l medical schools	0.6 (0.8)	0.43	0.2 (1.1)	0.86	0.5 (1.3)	0.70
Practice Type						
Academic practice	Reference		Reference		Reference	
Other practice, unknown[a]	3.5 (2.4)	0.14	-3.9 (2.7)	0.15	-3.9 (3.2)	0.22
Solo physician practice	-0.2 (1.8)	0.93	-5.0 (2.4)	0.04	-5.3 (2.7)	0.05
Small group practice (2 to 10)	-1.0 (1.7)	0.55	-5.6 (2.2)	0.01	-5.7 (2.5)	0.02
Medium physicians group practice (11 to 50)	1.7 (1.9)	0.37	-1.3 (2.4)	0.58	-3.3 (2.8)	0.25
Large physician group practice (>50 physicians)	-2.4 (1.8)	0.20	-1.4 (2.7)	0.62	-2.3 (3.1)	0.46
Female Beneficiaries	-3.6 (1.2)	0.002	-5.1 (1.5)	0.001	-6.2 (1.7)	<.001
Beneficiary Race						
White	Reference		Reference		Reference	
Black	0.5 (1.3)	0.73	4.9 (2.0)	0.01	3.6 (2.1)	0.09
Other	-3.1 (1.1)	0.004	-4.2 (1.5)	0.005	-5.5 (1.7)	0.001
Hierarchical Condition Category (HCC) score[b]	1.7 (0.4)	<.001	1.2 (0.5)	0.03	1.9 (0.6)	0.003
Medicaid Dual Eligible	0.1 (1.2)	0.91	2.3 (1.6)	0.16	1.9 (1.8)	0.27
Beneficiary age	0.7 (0.1)	<.001	0.4 (0.1)	<.001	0.5 (0.1)	<.001
Rural county residence[c]	-1.3 (0.9)	0.15	-1.3 (1.3)	0.31	0.5 (1.6)	0.76
Household medium income[d],	-3.1E-05 (1.6E-05)	0.05	8.7E-06 (2.2E-05)	0.69	-3.1E-06 (2.5E-05)	0.90
CCW chronic conditions						
Alzheimer's Disease and Related Disorders or Senile Dementia	1.7 (1.2)	0.18	3.0 (1.8)	0.09	3.7 (2.0)	0.07
Alzheimer's Disease	2.6 (1.7)	0.14	2.8 (2.5)	0.26	1.8 (2.6)	0.50
Acute Myocardial Infarction	2.7 (1.8)	0.14	2.3 (2.1)	0.27	2.5 (2.4)	0.29
Anemia	0.0 (0.8)	0.99	0.7 (1.1)	0.54	0.8 (1.2)	0.50

Gray BM, *et al. BMJ Open* 2021; 11:e041817. doi: 10.1136/bmjopen-2020-041817

Asthma	-0.8 (1.1)	0.45	-0.2 (1.4)	0.88	0.1 (1.7)	0.95
Atrial Fibrillation	1.8 (1.1)	0.10	3.4 (1.5)	0.02	3.7 (1.7)	0.03
Breast Cancer	-2.0 (1.3)	0.13	-3.0 (1.8)	0.10	-2.0 (2.2)	0.35
Colorectal Cancer	-0.5 (1.7)	0.76	-3.4 (2.0)	0.10	-1.9 (2.5)	0.44
Endometrial Cancer	1.6 (4.2)	0.71	-1.1 (4.7)	0.82	-0.9 (5.5)	0.87
Lung Cancer	3.7 (3.4)	0.28	13.3 (6.2)	0.03	10.3 (6.2)	0.10
Prostate Cancer	-1.6 (1.4)	0.26	4.3 (2.5)	0.09	3.6 (2.8)	0.20
Cataract	-1.0 (0.9)	0.29	1.1 (1.0)	0.31	0.1 (1.2)	0.96
Heart Failure	1.8 (1.0)	0.08	2.9 (1.2)	0.02	3.3 (1.4)	0.02
Chronic Kidney Disease	-0.7 (0.8)	0.39	2.9 (1.2)	0.02	4.2 (1.4)	0.004
Chronic Obstructive Pulmonary Disease	1.1 (0.9)	0.26	1.3 (1.2)	0.27	0.9 (1.3)	0.50
Depression	0.2 (0.9)	0.79	0.9 (1.1)	0.43	0.7 (1.2)	0.59
Diabetes	0.0 (0.8)	0.99	3.3 (1.1)	0.003	2.4 (1.2)	0.04
Glaucoma	-0.9 (0.8)	0.29	0.0 (1.1)	1.00	-0.3 (1.2)	0.80
Hip/Pelvic Fracture	1.4 (1.6)	0.39	3.0 (2.4)	0.20	5.0 (2.8)	0.07
Hyperlipidemia	-1.8 (1.1)	0.09	-1.1 (1.3)	0.39	-0.7 (1.4)	0.63
Benign Prostatic Hyperplasia	-0.3 (1.1)	0.79	0.0 (1.5)	0.98	-0.8 (1.7)	0.63
Hypertension	0.8 (1.1)	0.46	2.4 (1.4)	0.09	1.2 (1.6)	0.44
Hypothyroidism	0.7 (0.8)	0.42	1.4 (1.1)	0.19	1.8 (1.2)	0.14
Ischemic Heart Disease	0.8 (0.9)	0.39	0.1 (1.1)	0.92	-1.3 (1.2)	0.29
Osteoporosis	-1.9 (0.8)	0.02	2.0 (1.2)	0.09	1.4 (1.3)	0.28
Rheumatoid Arthritis	-2.3 (0.8)	0.005	-0.2 (1.0)	0.87	0.3 (1.1)	0.81
Stroke	2.7 (1.1)	0.02	2.7 (1.3)	0.04	2.8 (1.5)	0.06
Visit with same doctor in last year	-0.9 (1.1)	0.40	-1.3 (1.4)	0.34	-2.3 (1.5)	0.13
Visit with any physician in last year	-8.4 (3.5)	0.02	-3.5 (3.2)	0.28	-2.5 (3.3)	0.44
Hospitalization in prior year	8.8 (5.4)	0.10	0.9 (4.1)	0.84	0.4 (4.5)	0.93
ED visit in prior year	1.0 (2.5)	0.69	7.3 (4.0)	0.07	8.4 (4.6)	0.07
Days since last visit with any physician (per 30 d)	0.3 (0.2)	0.11	0.0 (0.3)	0.94	0.1 (0.3)	0.64
Days since last hospitalization (per 30 d)	-0.6 (0.4)	0.13	0.2 (0.5)	0.72	0.3 (0.5)	0.55
Days since last ED visits (per 30 d)	-0.2 (0.3)	0.60	-0.5 (0.4)	0.21	-0.7 (0.4)	0.13
Index visit diagnosis group indicators						
Pulmonary embolism	-0.5 (1.3)	0.68	6.1 (1.4)	<.001	7.1 (1.6)	<.001
Acute coronary syndrome	-1.3 (1.1)	0.25	-3.0 (1.8)	0.11	-5.5 (2.0)	0.007
Stroke	-2.3 (1.4)	0.10	7.4 (1.5)	<.001	7.7 (1.7)	<.001
Congestive heart failure	0.2 (1.3)	0.88	10.1 (1.6)	<.001	12.1 (1.7)	<.001
Fracture	-1.3 (1.1)	0.22	3.2 (1.3)	0.02	5.1 (1.5)	<.001
Abscess	0.7 (2.3)	0.77	6.4 (3.3)	0.05	11.7 (3.3)	<.001
Pneumonia	2.3 (1.2)	0.05	5.6 (1.4)	<.001	6.7 (1.6)	<.001
Aortic aneurysm	1.0 (1.4)	0.50	-0.6 (2.0)	0.76	0.7 (2.2)	0.74
Appendicitis	2.0 (1.8)	0.28	5.9 (3.0)	0.05	9.6 (3.1)	0.002
Depression	0.0 (1.3)	0.99	3.0 (1.5)	0.05	2.4 (1.7)	0.15
Anemia	2.3 (1.1)	0.04	3.5 (1.8)	0.04	3.2 (2.0)	0.11
Bacteremia	0.5 (2.5)	0.85	-9.5 (3.0)	0.001	-8.3 (3.1)	0.008
Spinal cord compression	-0.5 (1.8)	0.79	-2.8 (2.8)	0.32	-7.0 (3.1)	0.02
Mental health visit	1.4 (1.2)	0.22	-0.9 (1.5)	0.53	0.1 (1.8)	0.97
HHS Region						
HHS Region 1	Reference		Reference		Reference	
HHS Region 2	1.6 (1.7)	0.35	-5.2 (2.2)	0.02	-6.7 (2.7)	0.01
HHS Region 3	2.7 (1.8)	0.12	2.1 (2.5)	0.40	1.3 (3.0)	0.66
HHS Region 4	0.4 (1.5)	0.77	-2.7 (2.2)	0.22	-4.9 (2.6)	0.07
HHS Region 5	0.3 (1.4)	0.81	0.8 (2.1)	0.69	-1.0 (2.6)	0.70
HHS Region 6	-0.9 (1.5)	0.53	-2.8 (2.2)	0.21	-4.4 (2.8)	0.11
HHS Region 7	0.0 (2.2)	0.99	3.2 (3.2)	0.31	0.9 (3.5)	0.79
HHS Region 8	-1.6 (2.2)	0.47	1.9 (3.8)	0.62	-2.0 (3.8)	0.61
HHS Region 9	0.0 (1.6)	0.99	-0.6 (2.5)	0.81	-3.2 (2.8)	0.26
HHS Region 10	-0.6 (2.2)	0.77	4.4 (3.5)	0.21	4.0 (4.3)	0.35
Study Year						
2009	-3.1 (3.0)	0.30	-2.5 (4.3)	0.56	-1.9 (5.9)	0.75

Gray BM, *et al. BMJ Open* 2021; 11:e041817. doi: 10.1136/bmjopen-2020-041817

2010	-1.9 (2.9)	0.52	-2.1 (3.6)	0.55	-1.7 (5.2)	0.75
2011	1.1 (2.3)	0.63	1.4 (2.7)	0.60	-2.3 (4.2)	0.58
2012	Reference		Reference		Reference	
Yearly Quarter						
Q1	Reference		Reference		Reference	
Q2	-0.6 (0.9)	0.50	0.7 (1.2)	0.54	0.4 (1.4)	0.78
Q3	-0.7 (0.9)	0.43	-0.1 (1.2)	0.94	-0.8 (1.3)	0.55
Q4	0.6 (1.1)	0.55	1.9 (1.4)	0.18	1.2 (1.5)	0.42
Test Forms						
May 2008: A	3.7 (3.6)	0.30	-6.4 (4.2)	0.13	-5.5 (5.0)	0.27
May 2008: B	2.3 (4.3)	0.60	-4.1 (4.9)	0.41	-3.6 (6.1)	0.56
Nov. 2008: A	1.1 (3.1)	0.72	0.7 (4.2)	0.87	-2.7 (5.4)	0.62
Nov. 2008: B	Reference		Reference		Reference	
May 2009: A	5.1 (3.7)	0.16	3.6 (3.9)	0.36	-2.2 (5.0)	0.66
May 2009: B	0.2 (5.0)	0.96	-0.1 (5.0)	0.98	-6.0 (6.0)	0.31
Nov 2009: A	1.5 (3.1)	0.64	2.7 (3.4)	0.43	-1.5 (4.9)	0.75
Nov 2009: B	6.7 (3.9)	0.08	8.2 (3.9)	0.04	5.2 (5.2)	0.32
Nov 2009: C	Reference		Reference		Reference	
May 2010: A	0.9 (2.4)	0.69	-0.8 (2.6)	0.75	-4.6 (4.1)	0.26
May 2010: B	1.7 (2.4)	0.48	-0.2 (2.7)	0.94	-3.6 (4.3)	0.41
Nov. 2010: A	1.3 (2.4)	0.59	-1.2 (2.5)	0.63	-4.8 (4.1)	0.24
Nov. 2010: B	1.4 (2.3)	0.55	-0.5 (2.6)	0.85	-3.6 (4.1)	0.38
Nov. 2010: C	Reference		Reference		Reference	
May 2011: A	3.3 (3.2)	0.30	1.5 (3.7)	0.69	-1.3 (5.2)	0.80
May 2011: B	1.9 (3.4)	0.59	-1.4 (4.0)	0.74	-5.8 (5.3)	0.27
Nov. 2011: A	-1.8 (3.2)	0.57	-3.6 (4.5)	0.42	-3.2 (7.2)	0.65
Nov. 2011: B	0.2 (2.7)	0.94	-4.6 (3.7)	0.21	-8.0 (5.3)	0.13
Nov. 2011: C	Reference		Reference		Reference	

Note:
[a]Missing practice characteristics (1,432 or 2.94% of sample) were coded as "other unknown".
[b]Missing HCC (86 or .18% of sample) were replace by in sample mean HCC.
[c]Missing rural indicator (22 or .05% of sample) were assumed to be non-rural
[b]Missing ZIP code median income (708 or 1.46% of sample) were replace by in sample mean median income.

Gray BM, *et al. BMJ Open* 2021; 11:e041817. doi: 10.1136/bmjopen-2020-041817

Section 6. Regression Sensitivity Analyses

In this section we describe the results of falsification and robustness sensitivities. Falsification sensitivities examine associations with diagnostic knowledge under scenarios where we expect the underlying associations to be weaker than in the base case. Robustness sensitivities examine the degree to which base case associations with diagnostic knowledge were robust to assumptions regarding index visit diagnoses eligibility, outcome variable construction, and regression control variables.

Falsification sensitivities

Results of falsification sensitivities are exhibited in eTable 6.1. These sensitivities include applying the index visit sample that did not meet any diagnoses eligibility criteria and applying elective hospitalizations as an outcome measure. Presumably diagnostic knowledge would not impact outcomes with the diagnostic error sensitive conditions after index visits where related diagnoses codes for these conditions were not present. That is, either because the underlying condition was not present or not detectable at the time of the index visit and therefore was not preventable. However, outcomes after these index visits could be associated with omitted variables that were both correlated with our outcome measures and exam performance. For example, it could be that physicians with low diagnostic knowledge also have less healthy patients in ways we do not control for and therefore would be more likely to experience adverse events more generally. We also assume that elective hospitalizations would be related to the overall propensity to hospitalize but would not be related to underlying diagnostic skill.

Overall the results of falsification sensitivities support the validity of our base case finding. For example, although the overall risk of each adverse outcome was comparable to the base case, all associations with diagnostic knowledge were very small in absolute terms and none were statistically significant (P>0.05). For example, applying for the sample of index visits without eligible diagnoses codes, scoring in the top versus bottom tertile of diagnostic knowledge was associated with a 0.0 (95% CI -1.3 to 1.3, p=0.99) difference in the risk of death within 90 days of the index visit or under one tenth of the statistically significant 2.9 (95% CI: -5.0 to -0.7, p=0.008) fewer death per 1,000 observed in the base case. Yet, the mean risk of death in the base case and this sensitivity was comparable (0.7% in the base case versus 0.4% in this sensitivity). This sensitivity also addressed another limitation of our study, that we did not have a direct measure of cause of death since if the associations we found were driven by reductions in death due to the 13 diagnostic error prone conditions applied in our study we would expect that the associations with death and diagnostic exam performance would be much smaller when estimated using the index sample without eligible diagnoses codes for these conditions. Similarly we found that the associations between diagnostic knowledge and risk of an elective hospitalization were statistically insignificant, top compared to bottom tertile association P was 0.63, and was wrong signed.

Robustness sensitivities

Results of robustness sensitivities are exhibited in eTables 6.2.1 (for death), 6.2.2 (hospitalization) to 6.2.3 (for emergency department visit).

For the first sensitivity we expanding the eligible diagnoses code groups to all 76 identified by physician authors versus 38 in the base case that also met the relative risk criteria.

For the third sensitivity we expand the index visit clean period to 97 days and contracted the index visit clean period to 83 days.

For the fourth sensitivity, we excluded physician in academic medical centers to consider the possibility that the unobserved physician characteristics related to where they worked or who they worked with could be were independently both related to the underlying physician diagnostic skill and our outcome measures.

For the fifth sensitivity we accounted for the possibility that adverse outcomes were avoided because the patient died by altering the ED and hospitalization measures to include all-cause mortality. For this sensitivity we added the following two outcome measures: base case hospitalization or death and base case ED or death.

Overall results of robustness sensitivity analysis suggests that our base case results were not highly sensitive to different underlying assumptions related to these factors (e.g., across all robustness sensitivities percent change in the outcome measures between top versus bottom diagnostic knowledge exam performers remained statistically significant ($P<0.05$)).

Gray BM, *et al. BMJ Open* 2021; 11:e041817. doi: 10.1136/bmjopen-2020-041817

Table 6.1. Results of Falsification Sensitivity Analyses for All Adverse Outcomes

Adverse outcome measure / Sensitivity	Number of index visits	Regression adjusted outcomes per 1,000 index visits, (95% CI)			Top versus bottom tertile of diagnostic knowledge			Middle versus bottom tertile of diagnostic knowledge		
		Top	Middle	Bottom	Percent difference (95% CI)	Difference per 1,000 index visits (95% CI)	P-value	Percent difference (95% CI)	Difference per 1,000 index visits (95% CI)	P-value
Death										
Base	48,632	5.2 (4.1 to 6.3)	6.5 (5.4 to 7.6)	8.1 (6.5 to 9.7)	-35.3 (-52.8 to -11.2)	-2.9 (-5.0 to -0.7)	0.008	-20.2 (-38.3 to 3.2)	-1.6 (-3.6 to 0.3)	0.09
Falsification sensitivity										
Index visits sample that did not meet the diagnoses code eligibility criteria.	84,497	3.9 (3.0 to 4.7)	4.3 (3.5 to 5.0)	3.9 (3.1 to 4.7)	0.2 (-27.9 to 39.4)	0.0 (-1.3 to 1.3)	0.99	10.1 (-17.2 to 46.5)	0.4 (-0.8 to 1.5)	0.51
Hospitalization										
Base	48,632	9.2 (7.7 to 10.8)	11.0 (9.4 to 12.6)	13.3 (11.2 to 15.4)	-30.5 (-46.1 to -10.4)	-4.1 (-6.9 to -1.2)	0.006	-17.1 (-33.2 to 3.0)	-2.3 (-4.9 to 0.4)	0.09
Falsification sensitivities										
Index visits sample that did not meet the diagnoses code eligibility criteria.	84,497	13.8 (12.0 to 15.5)	13.1 (11.8 to 14.4)	14.0 (12.5 to 15.5)	-1.5 (-18.2 to 18.6)	-0.2 (-2.8 to 2.4)	0.87	-6.0 (-19.1 to 9.1)	-0.8 (-2.9 to 1.2)	0.42
Elective hospitalization	48,264	9.6 (7.7 to 11.5)	9.0 (7.6 to 10.3)	8.9 (7.4 to 10.5)	7.6 (-19.6 to 43.9)	0.7 (-2.0 to 3.4)	0.63	0.4 (-20.1 to 26.3)	0.0 (-2.0 to 2.1)	0.97
Emergency Department Visit										
Base	48,632	11.5 (9.8 to 13.2)	13.2 (11.5 to 15.0)	16.4 (14.0 to 18.7)	-29.8 (-44.4 to -11.4)	-4.9 (-8.1 to -1.6)	0.003	-19.0 (-33.8 to -1.0)	-3.1 (-6.1 to -0.1)	0.04
Falsification sensitivity										
Index visits sample that did not meet the diagnoses code eligibility criteria.	84,497	18.0 (16.0 to 20.0)	17.7 (16.1 to 19.2)	18.7 (17.0 to 20.4)	-3.8 (-17.8 to 12.6)	-0.7 (-3.6 to 2.2)	0.63	-5.5 (-16.9 to 7.3)	-1.0 (-3.4 to 1.3)	0.38

Gray BM, *et al. BMJ Open* 2021; 11:e041817. doi: 10.1136/bmjopen-2020-041817

Table 6.2.1. Results of Robustness Sensitivity Analyses for the Death Adverse Outcome

	Number of index visits	Regression adjusted deaths per 1,000 index visits (95% CI)			Top versus bottom tertile of diagnostic knowledge			Middle versus bottom tertile of diagnostic knowledge		
		Top	Middle	Bottom	Percent difference (95% CI)	Difference per 1,000 index visits (95% CI)	P-value	Percent difference (95% CI)	Difference per 1,000 index visits (95% CI)	P-value
Base	**48,632**	**5.2 (4.1 to 6.3)**	**6.5 (5.4 to 7.6)**	**8.1 (6.5 to 9.7)**	**-35.3 (-52.8 to -11.2)**	**-2.9 (-5.0 to -0.7)**	**0.008**	**-20.2 (-38.3 to 3.2)**	**-1.6 (-3.6 to 0.3)**	**0.09**
Sensitivities										
Applying larger list of index visit diagnoses eligibility (all 76 diagnoses identified by physician authors)	57,749	4.9 (3.9 to 5.9)	5.7 (4.8 to 6.7)	7.0 (5.7 to 8.4)	-30.2 (-48.9 to -4.5)	-2.1 (-4.0 to -0.3)	0.03	-18.4 (-36.7 to 5.2)	-1.3 (-2.9 to 0.4)	0.13
97 day index visit clean period	40,417	7.5 (5.8 to 9.1)	6.8 (5.6 to 8.1)	4.9 (3.8 to 6.0)	-34.7 (-53.9 to -7.6)	-2.6 (-4.8 to -0.4)	0.02	-8.5 (-31.0 to 21.2)	-0.6 (-2.7 to 1.4)	0.54
83 day index visit clean period	54,169	5.5 (4.4 to 6.6)	6.8 (5.7 to 7.8)	8.4 (6.8 to 10.0)	-34.7 (-51.7 to -11.7)	-2.9 (-5.0 to -0.8)	0.007	-19.5 (-37.0 to 3.0)	-1.6 (-3.6 to 0.3)	0.09
Small practices (visits with physicians with practices of 10 or less physicians)	29,242	4.5 (3.2 to 5.9)	5.9 (4.6 to 7.2)	8.2 (6.3 to 10.2)	-44.9 (-63.6 to -16.7)	-3.7 (-6.3 to -1.1)	.0047	-28.6 (-48.9 to .1)	-2.4 (-4.8 to 0.1)	.058
Large (>50 physicians)/academic medical center practices:	6,308[a]	6.4 (3.6 to 9.1)	6.4 (3.4 to 9.4)	5.7 (2.1 to 9.2)	12.9 (-50.8 to 159.0)	0.7 (-4.2 to 5.6)	.7714	13.3 (-43.0 to -125.1)	0.8 (-3.3 to 4.8)	0.72
Not counting next day death as an adverse outcome	48,632	5.2 (4.1 to 6.3)	6.4 (5.3 to 7.5)	8.1 (6.5 to 9.7)	-35.7 (-53.1 to -11.8)	-2.9 (-5.0 to -0.8)	.000729	-21.0 (-38.9 to 2.1)	-1.7 (-3.6 to 0.2)	.081

[a] 1,791 observations excluded due to lack of variation in outcomes within control test administrations or other controls

Table 6.2.2. Results of robustness sensitivity analyses for the hospitalization adverse outcome

	Number of index visits	Regression adjusted risk of emergency department hospitalization per 1,000 index visits, (95% CI)			Top versus bottom tertile of diagnostic knowledge			Middle versus bottom tertile of diagnostic knowledge		
		Top	Middle	Bottom	Percent difference (95% CI)	Difference per 1,000 index visits (95% CI)	P-value	Percent difference (95% CI)	Difference per 1,000 index visits (95% CI)	P-value
Base	**48,632**	**9.2 (7.7 to 10.8)**	**11.0 (9.4 to 12.6)**	**13.3 (11.2 to 15.4)**	**-30.5 (-46.1 to -10.4)**	**-4.1 (-6.9 to -1.2)**	**0.006**	**-17.1 (-33.2 to 3.0)**	**-2.3 (-4.9 to 0.4)**	**0.09**
Sensitivities										
Applying larger list of index visit diagnoses eligibility (all 76 diagnoses identified by physician authors)	57,749	11.3 (9.6 to 13.0)	9.7 (8.3 to 11.0)	8.3 (6.9 to 9.7)	-26.6 (-43.0 to -5.4)	-3.0 (-5.5 to -0.5)	0.02	-14.6 (-31.0 to 5.6)	-1.7 (-3.9 to 0.6)	0.15
97 day index visit clean period	40,417	8.3 (6.7 to 9.9)	10.4 (8.7 to 12.1)	13.4 (11.0 to 15.9)	-38.4 (-54.2 to -17.3)	-5.2 (-8.4 to -1.9)	0.002	-22.8 (-39.6 to -1.3)	-3.1 (-6.0 to -0.1)	0.04
83 day index visit clean period	54,169	9.3 (7.8 to 10.8)	11.2 (9.7 to 12.8)	13.2 (11.2 to 15.3)	-29.7 (-45.3 to -9.7)	-3.9 (-6.8 to -1.1)	0.007	-15.1 (-31.2 to 4.8)	-2.0 (-4.6 to 0.6)	0.13
Hospitalization visit or death (hospitalization base case measure or death base case measure)	48,632	13.7 (11.9 to 15.4)	16.4 (14.5 to 18.2)	19.8 (17.4 to 22.2)	-30.9 (-43.3 to -15.8)	-6.1 (-9.4 to -2.8)	<.001	-17.4 (-30.5 to -1.9)	-3.4 (-6.6 to -0.3)	0.03
Shortening the outcome period from 90 day to 14 days	48,632	2.0 (1.3 to 2.7)	3.2 (2.4 to 4.1)	3.3 (2.4 to 4.3)	-40.3 (-63.3 to -3.0)	-1.4 (-2.6 to -0.1)	0.04	-3.7 (-35.2 to 43.2)	-0.1 (-1.4 to 1.2)	0.85
Small practices (visits with physicians with practices of 10 or less physicians)	29,242	7.8 (5.8 to 9.8)	12.1 (10.0 to 14.2)	11.8 (9.5 to 14.0)	-33.4 (-53.0 to -5.6)	-3.9 (-7.2 to -0.6)	0.02	-18.8 (-39.3 to 8.5)	-2.2 (-5.3 to 0.9)	0.16
Large (>50 physicians)/academic medical center practices:	7,966a	10.4 (7.3 to 13.5)	12.0 (7.8 to 16.2)	22.5 (13.5 to 31.5)	-53.7 (-73.2 to -20.2)	-12.1 (-22.2 to -2.0)	0.02	-46.7 (-68.0 to -8.7)	-10.5 (-20.5 to -0.5)	0.04
Not counting next day hospitalizations as an adverse outcome	48,632	8.7 (7.2 to 10.2)	9.9 (8.4 to 11.5)	12.5 (10.4 to 14.5)	-30.0 (-46.1 to -9.0)	-3.7 (-6.5 to -0.9)	0.0087	-20.2 (36.3 to 0.0)	-2.5 (-5.1 to 0)	.054604

[a] 133 observations excluded due to lack of variation in outcomes within control test administrations or other controls

Gray BM, *et al. BMJ Open* 2021; 11:e041817. doi: 10.1136/bmjopen-2020-041817

Table 6.2.3. Results of robustness sensitivity analyses for the emergency department visit adverse outcome

	Number of index visits	Regression adjusted risk of emergency department visit per 1,000 index visits, (95% CI)			Top versus bottom tertile of diagnostic knowledge			Middle versus bottom tertile of diagnostic knowledge		
		Top	Middle	Bottom	Percent difference (95% CI)	Difference per 1,000 index visits (95% CI)	P-value	Percent difference (95% CI)	Difference per 1,000 index visits (95% CI)	P-value
Base	**48,632**	**11.5 (9.8 to 13.2)**	**13.2 (11.5 to 15.0)**	**16.4 (14.0 to 18.7)**	**-29.8 (-44.4 to -11.4)**	**-4.9 (-8.1 to -1.6)**	**0.003**	**-19.0 (-33.8 to -1.0)**	**-3.1 (-6.1 to -0.1)**	**0.04**
Sensitivities										
Applying larger list of index visit diagnoses eligibility (all 76 diagnoses identified by physician authors)	57,740	10.4 (8.8 to 12.0)	11.7 (10.2 to 13.2)	13.9 (12.0 to 15.8)	-25.2 (-40.5 to -6.0)	-3.5 (-6.3 to -0.7)	0.01	-16.3 (-31.1 to 1.7)	-2.3 (-4.8 to 0.2)	0.08
97 day index visit clean period	40,417	10.5 (8.7 to 12.3)	12.6 (10.7 to 14.5)	16.7 (13.9 to 19.5)	-37.2 (-51.9 to -18.0)	-6.2 (-9.9 to -2.5)	<.001	-24.5 (-39.9 to -5.3)	-4.1 (-7.5 to -0.7)	0.02
83 day index visit clean period	54,169	11.6 (9.9 to 13.2)	13.4 (11.7 to 15.1)	16.4 (14.1 to 18.8)	-29.5 (-43.8 to -11.6)	-4.8 (-8.0 to -1.7)	0.003	-18.4 (-32.6 to -1.1)	-3.0 (-5.9 to -0.1)	0.04
Emergency department visit or death (hospitalization base case measure or death base case measure)	48,632	15.7 (13.7 to 17.7)	18.5 (16.5 to 20.5)	22.6 (20.0 to 25.2)	-30.6 (-42.6 to -16.0)	-6.9 (-10.6 to -3.3)	<.001	-18.0 (-30.4 to -3.4)	-4.1 (-7.5 to -0.7)	0.02
Shortening the outcome period from 90 day to 14 days	48,632	2.7 (1.9 to 3.4)	3.7 (2.8 to 4.7)	4.0 (2.9 to 5.1)	-34.4 (-57.8 to 2.1)	-1.4 (-2.9 to 0.1)	0.07	-7.5 (-36.2 to 34.2)	-0.3 (-1.8 to 1.1)	0.68
Small practices (visits with physicians with practices of 10 or less physicians)	29,242	10.3 (8.0 to 12.5)	12.1 (10.0 to 14.2)	14.7 (12.3 to 17.1)	-30.1 (-48.2 to -5.8)	-4.4 (-8.0 to -0.8)	.016	-17.7 (-36.2 to 6.3)	-2.6 (-6.0 to 0.8)	.138
Large (>50 physicians)/academic medical center practices:	7,966a	13.3 (9.3 to 17.2)	12.6 (8.4 to 16.8)	24.2 (15.2 to 33.2)	-45.3 (-67.8 to -6.9)	-11.0 (-21.7 to -0.3)	0.045	-48.1 (-68.3 to -14.8)	-11.6 (-21.5 to -1.8)	0.021
Not counting next day emergency department visits as an adverse outcome	48,632	10.6 (9.0 to 12.3)	12.0 (10.3 to 13.7)	15.0 (23.7 to 17.3)	-29.2 (44.2 to 10.2)	-4.4 (-7.5 to -1.3)	.0055	-20.1 (35.2 to 1.3)	-3.0 (-5.9 to -0.1)	.040

[a] **133 observations excluded due to lack of variation in outcomes within control test administrations or other controls**

References

1. Bandalos DL. *Measurement theory and applications for the social sciences*: Guilford Publications; 2018.

Gray BM, *et al. BMJ Open* 2021; 11:e041817. doi: 10.1136/bmjopen-2020-041817