BMJ Open

Association between primary care physician diagnostic knowledge and death, hospitalisation and emergency department visits following an outpatient visit at risk for diagnostic error: a retrospective cohort study using medicare claims

Bradley M Gray ,[1] Jonathan L Vandergrift,[1] Rozalina G McCoy ,[2] Rebecca S Lipner,[1] Bruce E Landon[3]

To cite: Gray BM, Vandergrift JL, McCoy RG, *et al*. Association between primary care physician diagnostic knowledge and death, hospitalisation and emergency department visits following an outpatient visit at risk for diagnostic error: a retrospective cohort study using medicare claims. *BMJ Open* 2021;**11**:e041817. doi:10.1136/bmjopen-2020-041817

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-041817).

Received 19 June 2020
Revised 04 March 2021
Accepted 04 March 2021

For numbered affiliations see end of article.

Correspondence to
Dr Bradley M Gray;
bgray@abim.org

ABSTRACT

Objective Diagnostic error is a key healthcare concern and can result in substantial morbidity and mortality. Yet no study has investigated the relationship between adverse outcomes resulting from diagnostic errors and one potentially large contributor to these errors: deficiencies in diagnostic knowledge. Our objective was to measure that associations between diagnostic knowledge and adverse outcomes after visits to primary care physicians that were at risk for diagnostic errors.

Setting/participants 1410 US general internists who recently took their American Board of Internal Medicine Maintenance of Certification (ABIM-IM-MOC) exam treating 42 407 Medicare beneficiaries who experienced 48 632 'index' outpatient visits for new problems at risk for diagnostic error because the presenting problem (eg, dizziness) was related to prespecified diagnostic error sensitive conditions (eg, stroke).

Outcome measures 90-day risk of all-cause death, and, for outcome conditions related to the index visits diagnosis, emergency department (ED) visits and hospitalisations.

Design Using retrospective cohort study design, we related physician performance on ABIM-IM-MOC diagnostic exam questions to patient outcomes during the 90-day period following an index visit at risk for diagnostic error after controlling for practice characteristics, patient sociodemographic and baseline clinical characteristics.

Results Rates of 90-day adverse outcomes per 1000 index visits were 7 for death, 11 for hospitalisations and 14 for ED visits. Being seen by a physician in the top versus bottom third of diagnostic knowledge during an index visit for a new problem at risk for diagnostic error was associated with 2.9 fewer all-cause deaths (95% CI −5.0 to −0.7, p=0.008), 4.1 fewer hospitalisations (95% CI −6.9 to −1.2, p=0.006) and 4.9 fewer ED visits (95% CI −8.1% to −1.6%, p=0.003) per 1000 visits.

Conclusion Higher diagnostic knowledge was associated with lower risk of adverse outcomes after visits for problems at heightened risk for diagnostic error.

Strengths and limitations of this study

- Unique diagnostic knowledge measure linking diagnostic knowledge with adverse outcomes.
- Scalable adverse outcome measures and extensive sensitivity analyses.
- Our assessment of diagnostic error is indirect (as indicated by adverse outcomes).
- Results are subject to selection bias if the mix of index visits or the severity of the patients or practice support differed for physicians with different levels of diagnostic knowledge.
- Results are only generalisable to physicians who elected to attempt American Board of Internal Medicine's certification exam and were about 10 years past initial certification and patients older than 65.

INTRODUCTION

Diagnostic error has been identified as a key healthcare delivery concern and contributes to significant potentially preventable morbidity and mortality.[1–3] Ambulatory care, and especially primary care, is a practice setting with a particularly high risk for diagnostic error[4 5] because of the wide variety of presentations encountered and the concomitant difficulty of distinguishing harmful conditions from routine self-limited problems, compounded by the well-known time constraints faced by practitioners in that setting. It has been estimated that at least 5% of ambulatory visits

are associated with diagnostic error, half of which may result in considerable patient harm. Diagnostic error is a common cause of malpractice suits and most frequently occurs in the ambulatory care settings.[6 7]

Deficiencies in diagnostic knowledge are likely to be an important contributor to these diagnostic errors that could impact, for example, the breadth of diagnoses considered, appropriate ordering and interpretation of tests and/or synthesis of data more generally.[8–11] Because of this, measuring physician diagnostic knowledge has become a major focus of organisations throughout the developed world that are tasked with licensing and certifying physicians with the underlying, although largely untested, hypothesis being that diagnostic knowledge will be a measurable and strong predictor of diagnostic error.[12–15] Testing this hypothesis and quantifying this relationship are therefore a critical public policy concern both in terms of the importance of board certification and other programmes designed to enhance lifelong learning for physicians.

In the USA, the American Board of Internal Medicine (ABIM) is a leading organisation that certifies primary care physicians, most notably general internists. In fact, most general internists in the USA are certified by the ABIM and these physicians represent about 45% of all adult primary care physicians in the USA.[16] Unlike medical licensure, board certification is not a legal requirement to practice medicine in the USA, though many hospitals require board certification as one criterion to obtain privileges and insurers often require board certification to be included in covered physician panels.[17 18] To maintain their certification, general internists must pass an initial certifying exam and, periodically, pass a recertification exam thereafter (referred to as Maintenance of Certification (MOC) exams).[19 20] Diagnostic knowledge is a major component of these exams representing about half of all exam questions for the Internal Medicine MOC (IM-MOC) exam.

One explanation for the lack of research on this topic is the difficulty in studying the relationship between general diagnostic knowledge and diagnostic error because of the inability to quantify diagnostic knowledge and identifying diagnostic errors at a population level, especially in the outpatient setting.[21] We address this gap in the literature by applying a unique measure of diagnostic knowledge, performance on diagnostic-related questions on ABIM's IM-MOC exam, and relating this measure to deaths, hospitalisations and emergency department (ED) visits that occurred after outpatient visits for new problems at heightened risk for diagnostic error.

METHODS

Physician and index visit sample

Our physician sample included general internists who were initially ABIM board certified in 2000 and took their IM-MOC exam between 2008 and 2011 (figure 1). We identified Medicare beneficiary outpatient Evaluation & Management visits with these physicians using their National Provider Identifier during the calendar year following their exam (2009–2012). These patients were age 65 or older and continuously enrolled in Medicare fee-for-service (Medicare insures most of the US population over 65) during the physician's 1-year follow-up period and the year prior. To ensure that any presenting problems being evaluated were new (ie, not follow-up), we restricted these visits to those that were the first visit for a new problem (the 'index visit') because these visits were preceded by a 90-day clean period with no previous inpatient or outpatient visit. The 90-day clean period is consistent with the US government Centers for Medicare and Medicaid Services criteria used by its Bundled Payments for Care Improvement Programme for defining new episodes of care and with the patterns of visits we observed (see online supplemental appendix section 1 for related analysis).[22 23]

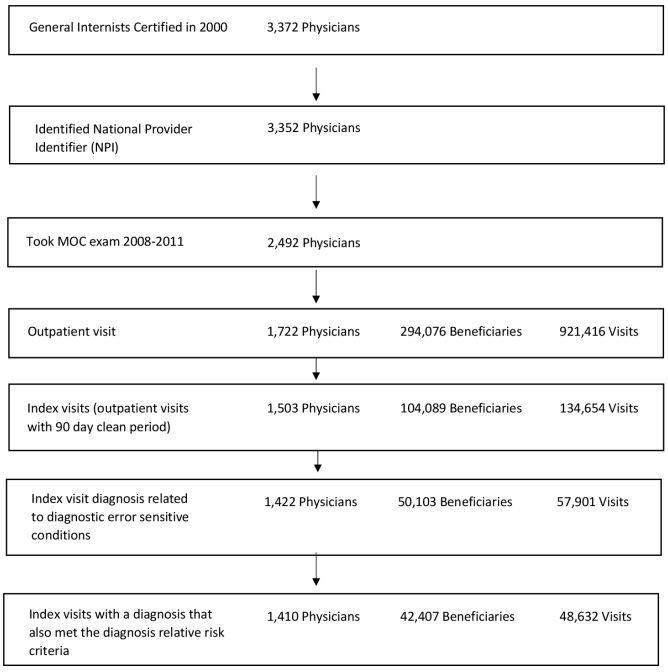


Figure 1 Sample selection. MOC, Maintenance of Certification.

We further restricted these index visits to those at heightened risk for diagnostic errors because the recorded diagnosis in the Medicare claims (the 'index visit diagnosis'), which includes recording of symptom (eg, loss of balance), could have been the initial presenting problem for one or more of 13 prespecified diagnostic error sensitive conditions such as congestive heart failure or bacteraemia/sepsis (see table 1). These 13 conditions (see online supplemental appendix section 2 for a list and applicable International Classification of Diseases, Ninth Revision codes (ICD-9 codes) were an acute non-cancerous subset of 20 conditions previously noted by Schiff *et al* to be at high risk for serious diagnostic error.[24] For instance, index visits with diagnosis codes for chest pain, dyspepsia, shortness of breath, hypoxaemia/hypoxia, respiratory distress, weakness/fatigue, oedema or ascites could all be

Table 1 Frequency of index visits related to each diagnostic error sensitive condition

Thirteen diagnostic error sensitive conditions	Index visits with a diagnosis code related to a diagnostic error sensitive condition (percentages can add to greater than 100% because of antecedent index visit diagnoses related to more than one diagnostic error sensitive condition) Number (per cent of index visits)	Hospitalisation*† Number (per cent of hospitalisations with a diagnostic error sensitive condition)	Emergency department (ED) visit* Number (per cent of ED visits with a diagnostic error sensitive condition)	Death‡ Number (per cent of deaths)
	48 632 (100.0)	541 (100)	663 (100)	316 (100)
Acute coronary syndrome	16 228 (33.4)	48 (8.9)	56 (8.4)	103 (32.6)
Fracture	13 409 (27.6)	60 (11.1)	100 (15.1)	60 (19.0)
Depression	12 637 (26.0)	Not reported§	Not reported§	121 (38.3)
Anaemia	12 410 (25.5)	54 (10.0)	59 (8.9)	110 (34.8)
Pneumonia	12 183 (25.1)	91 (16.8)	107 (16.1)	107 (33.9)
Congestive heart failure	12 137 (25.0)	227 (42.0)	254 (38.3)	120 (38.0)
Aortic aneurysm	11 491 (23.6)	17 (3.1)	23 (3.5)	79 (25.0)
Stroke	10 026 (20.6)	69 (12.8)	82 (12.4)	71 (22.5)
Pulmonary embolism	8534 (17.5)	12 (2.2)	13 (2.0)	89 (28.2)
Spinal cord compression	6386 (13.1)	Not reported§	Not reported§	36 (11.4)
Bacteraemia/sepsis	5567 (11.4)	19 (3.5)	21 (3.2)	46 (14.6)
Appendicitis	2584 (5.3)	Not reported§	Not reported§	17 (5.4)
Abscess	1005 (2.1)	Not reported§	13 (2.0)	Not reported

*Condition specific outcomes for one of the 13 diagnostic error sensitive conditions within 90 days of an outpatient index visit at risk for that condition.
†Hospitalisations include non-elective hospitalisations either initiated through the ED or a trauma centre.
‡All cause mortality within 90 days of the index visit.
§Not reported because observations were less than 11.

the initial presentation of congestive heart failure, which is one of the 13 diagnostic error sensitive conditions.

We used a three-step process to identify eligible index visit diagnoses. First, two physician authors (RGM and BL) identified all diagnoses that could be presenting problems for the 13 diagnostic error sensitive conditions: what problems/diagnoses might someone who ultimately presented with a diagnostic-error sensitive condition have presented with initially? Second, because the original list of identified index visit diagnoses was large (76), we reduced this list to 38 by applying a relative risk (RR) criteria. For a specific index visit diagnosis to meet this criteria, all index visits with that diagnosis had to have a greater portion of later ED visits or hospitalisations with the related outcome condition discharge diagnosis than index visits where the specific at risk diagnosis was not present. For example, dizziness was chosen as an eligible index visit diagnosis for stroke, one of the diagnostic error sensitive conditions, both because it was identified as a potential presenting symptom of a stroke by physician authors and because index visits with that diagnosis had a greater proportion of later hospitalisation or ED visits for stroke than visits without this diagnosis. Third, we also included index visits where the actual diagnosis was one of the 13 diagnostic error sensitive conditions because we wanted to include cases where diagnostic errors were and were not made. Therefore, we also included index visits with a diagnosis of congestive heart failure itself as being at risk for the underlying condition congestive heart failure.

Outcome measures

We examined the risk of three serious adverse outcomes within 90 days of the index visit that we hypothesised would occur more frequently in cases of misdiagnosis: all-cause mortality, hospitalisations and ED visits. We did not count these events as adverse outcomes if they occurred on the same day as the index visit because this may reflect

a positive action (the physician correctly diagnosed a patient with stroke and referred/admitted them to the hospital) or be unavoidable regardless of the accuracy of the index visit diagnosis (the patient died despite immediately admitting the patient to the hospital who exhibited stroke symptoms). Based on Medicare billing codes, hospitalisations were limited to non-elective hospitalisations initiated through the ED or trauma centre. The ED and hospitalisation outcomes were also limited to cases where the discharge diagnosis was for one of the 13 diagnostic error sensitive conditions following an index visit with the applicable diagnosis. We therefore presumed that these discharge diagnoses were a reasonable representation of the underlying condition of the patient at the time of the index visit. For example, we would count a hospitalisation with a discharge diagnosis of stroke as an adverse outcome if it occurred after an index visit for dizziness because dizziness was identified as being a potential presenting problem for stroke. However, we did not count hospitalisations with a discharge diagnosis for acute coronary syndrome following an index visit for dizziness because dizziness was not identified as a presenting problem for acute coronary syndrome. The rationale is that if there were no presenting problems during the index visit related to coronary syndrome, either because the underlying condition was not present or could not be detected at the time of the index visit, then the index visit physician could not have prevented the hospitalisation regardless of their diagnostic knowledge.

Measure of diagnostic knowledge

Our measure of diagnostic knowledge was calculated as the per cent of correct answers on the IM-MOC exam for questions previously coded as 'diagnosis-related' by ABIM's IM-MOC exam committee. In our study, these questions comprised 53% of all IM-MOC exam questions, with the remaining 42% addressing treatment and 5% related to other topics such as epidemiology or pathophysiology. More generally, exam questions are designed to replicate real world clinical scenarios and/or patient encounters and without reliance on rote memorisation.[25 26]

The ABIM exam committee coded each question based on the primary function tested to assure that the exam covers care typically rendered by outpatient primary care physicians. Questions coded as diagnosis related typically test knowledge and skills related to diagnostic inference, differential diagnosis and diagnostic testing and therefore are measuring diagnostic knowledge and related decision-making. Psychometric analysis indicates that scores on diagnosis related exam questions were meaningfully correlated (ie, Cronbach's alpha score of 0.84), and thereby represent an independent underlying construct that could be interpreted as diagnostic knowledge (see online supplemental appendix section 3 for more details).[27] Similarly, this analysis indicated that questions coded as treatment related also represent an independent underlying construct (ie, Cronbach's alpha score of 0.75). Although performance on diagnosis and treatment related questions were correlated (Pearson correlation=0.62), 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.

Statistical methods

Using Probit regression, we estimated the associations with each adverse outcome, with standard errors adjusted for correlations resulting from the nesting of visits within patients within physicians.[28 29] To measure associations with diagnostic knowledge, we included categorical regression explanatory variables for top and middle third of per cent correct scores on diagnosis-related questions (bottom third was the reference category). Other exam level explanatory variables included tertile indicators for performance on treatment-related questions and performance on other question types. Since these variables measure knowledge unrelated to diagnosis, they account for correlations between factors such as unmeasured practice or patient characteristics that might be correlated with exam performance and our outcome measures (eg, high scoring physicians may be more likely to practice in an academic setting or other such settings that might be independently related to diagnostic error). Exam form indicators accounted for differences in exam difficulty across exam administrations.

We also included physician, patient and visit level regression controls. Physician level controls included: practice size (indicators for solo practice and practices larger than 50 physicians), practice type (indicators for academic, group), demographic (gender) and training characteristics (medical school location interacted with country of birth). Patient level controls included: demographic characteristics (age and age squared, gender and race/ethnicity indicators) and a Medicaid eligibility indicator. Lagged patient risk adjusters included 27 indicators for chronic conditions and Medicare's Hierarchal Condition Category (HCC) risk adjustment score. We imputed values for a small number of missing values for controls (see online supplemental appendix section 4). Patient index visit location level controls included: an indicator for residing in a rural ZIP code, ZIP code median household income, and indicators for 10 US Health and Human Services regions. Index visit level controls included: indicators of any outpatient visit, hospitalisation or ED visits within the prior year and number of days since the most recent of these events, visit year indicators to control for secular changes in quality. We also included an indicator for whether or not the patient had a previous contact with the index visit physician during the year prior to the index visit to account for differences in physician–patient continuity (see online supplemental appendix section 5 for a full list of controls).

Sensitivity analysis

We performed numerous sensitivity analyses to test the robustness of our results (detailed in online supplemental

appendix section 6). First, we expanded the index visit sample to include all index visits with the original 76 diagnoses identified by the physician authors regardless of whether they met the RR criteria. Second, we expanded and contracted the index visit clean period by 7 days. Third, excluded hospitalisations or ED events occurring the day after the index visit, in addition to same day events, to consider the possibility that they might be triggered by a correct diagnosis and therefore should not have been considered adverse outcomes. Fourth, we considered the possibility that our results were biased due to omitted variables correlated with practice size. For example, it could be that physicians in large practices have greater access to specialists or other physicians for informal consultations than those is small practices and therefore outcomes for these physicians may be less sensitive to their knowledge. To examine this possibility, we estimated associations with knowledge and our two utilisation measures across a sample of physicians in either small (≤10 physicians, 54.5% (768/1410) of physicians) or large practices (>50 or in academic medical centres, 23.7% (334/1410) of physicians). We did not conduct these sensitivities for death because there were too few deaths in the subgroups to allow us to reliably estimate the associations (eg, 39 deaths for physicians in large practices). Fifth, to consider the possibility that these outcomes were only avoided because the patient died, for the ED and hospitalisation outcome, we also included instances where the patient died. Sixth, as a falsification test we limited the index visits to those that were unrelated to the 13 diagnostic error sensitive conditions. Under this sensitivity, we expected then that the associations with diagnostic knowledge would decline. The index visit physician's diagnostic knowledge cannot impact a future adverse outcome if the underlying condition that caused that outcome was not present or detectible at the time of index visit. Therefore, this reduction in association should be especially true for the hospitalisation and ED measures where adverse outcomes were limited to the 13 diagnostic error conditions and so were unrelated to the index visit diagnoses in this sensitivity. Similarly, for the last sensitivity, we applied elective hospitalisations as an outcome measure to consider the possibility that there could be a correlation between the overall propensity to hospitalise in an area and physician knowledge.

All analyses were performed using Stata V.15.

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

RESULTS

Of 2492 general internists who initially certified in 2000 and who took an IM-MOC exam between 2009 and 2012 and 1722 had outpatient visits with a fee-for-service Medicare beneficiary during the study period. Those without visits generally practised hospital medicine. Of these, 1410 were included in the study because they had at least one outpatient index visit that met our study inclusion criteria during the year after they took their IM-MOC exam. In total, 48 632 index visits with 42 407 patients treated by 1410 physicians met study inclusion criteria (figure 1). Table 1 lists frequency of index visits and subsequent outcomes for each diagnostic error sensitivity condition.

The mean per cent correct on diagnosis questions ranged from 84.3% among top third performers to 65.5% among bottom third performers (table 2). Patient and visit characteristics were similar across tertiles of physician diagnostic knowledge. For example, there were no statistically significant differences in the HCC risk adjuster across tertiles (p=0.19) However, there were differences in some physician and practice characteristics. When compared with physicians in the bottom tertile of diagnostic knowledge, physicians in the top were significantly less likely to be in solo practice (12.8% vs 24.4%, p=0.009), and more likely to be in academic practice (9.7% vs 3.4%, p<0.001). However, the proportion graduating from a US medical school was similar across diagnostic knowledge tertiles (70.0% vs 63.3%, p=0.30).

Associations between diagnostic knowledge and patient adverse outcomes

The overall rates of 90-day adverse outcomes per 1000 index visits were 6.5 for death, 11.1 for hospitalisations and 13.6 for ED visits (with the latter two directly associated with one of the diagnostic error sensitive conditions whose antecedent was present in the applicable index visit). Being seen by a physician scoring in the top versus bottom third of diagnostic knowledge on the MOC exam was associated with 2.9 fewer deaths per 1000 visits (95% CI −5.0 to −0.7, p=0.008) which reflects a 35.3% lower risk of death (95% CI −52.8 to −11.2, p=0.008, table 3). Our finding also suggests that this difference in exam performance was associated with 4.1 fewer applicable hospitalisations (95% CI −6.9 to −1.2, p=0.006), and 4.9 fewer applicable ED visits (95% CI −8.1 to −1.6, p=0.003) per 1000 visits (table 3). These reductions correspond with about a 30% lower risk for these utilisation measures (hospitalisations: −30.5%, 95% CI −46.1 to −10.4, p=0.003, ED: −29.8%, 95% CI −44.4 to −11.4).

We also found a significant knowledge tertile dose response relationship across all three regression adjusted RR measures (p-trends<0.008). For example, the regression-adjusted 90-day risk of death per 1000 patients whose index visit physician scored in the top third of diagnostic knowledge was 5.2 (95% CI 4.1 to 6.3), compared with 6.5 (95% CI 5.4 to 7.6) for the middle third, and 8.1 (95% CI 6.5 to 9.7) for the bottom third (p-trend=0.008).

Sensitivity analyses

Our sensitivity analyses (online supplemental appendix section 6) confirmed that base case associations with diagnostic knowledge were robust to different index visit clean periods, and diagnosis code inclusion criteria and

Table 2 Physician and patient characteristics by diagnostic exam performance tertile

	Total	Diagnosis question per cent correct			P value*
		Top (78.5–95.8)	Middle (71.4–78.4)	Bottom (42.9–71.3)	
Exam performance, mean (SD)*					
Diagnosis question per cent correct	74.5 (0.4)	84.3 (0.3)	74.8 (0.1)	65.5 (0.3)	<0.001
Other question per cent correct	72.6 (0.7)	80.2 (1.0)	72.1 (1.1)	66.4 (1.5)	<0.001
Treatment question per cent correct	77.3 (0.3)	83.4 (0.4)	77.2 (0.4)	72.0 (0.5)	<0.001
Physician characteristics, count (%)					
Female physician	19 428 (39.9)	6546 (43.8)	6357 (37.5)	6525 (39.0)	0.37
US born physician	28 462 (58.5)	9284 (62.1)	9932 (58.6)	9246 (55.3)	0.37
US medical school	31 960 (65.7)	10 471 (70.0)	10 900 (64.3)	10 589 (63.3)	0.30
Practice type					
Solo physician practice	9452 (19.4)	1914 (12.8)	3462 (20.4)	4076 (24.4)	0.009
Small group practice (2–10)	20 563 (42.3)	5543 (37.1)	7529 (44.4)	7491 (44.8)	0.19
Medium physicians group practice (11–50)	7442 (15.3)	2899 (19.4)	2402 (14.2)	2141 (12.8)	0.25
Large physician group practice (>50 physicians)	5391 (11.1)	2150 (14.4)	1655 (9.8)	1586 (9.5)	0.14
Academic practice	2708 (5.6)	1447 (9.7)	697 (4.1)	564 (3.4)	<0.001
Other practice	3076 (6.3)	1005 (6.7)	1211 (7.1)	860 (5.1)	0.59
Beneficiary characteristics					
Beneficiary race, count (per cent)					
White	40 086 (82.4)	12 652 (84.6)	13 778 (81.3)	13 656 (81.7)	0.13
Black	3958 (8.1)	926 (6.2)	1609 (9.5)	1423 (8.5)	0.03
Other	4588 (9.4)	1380 (9.2)	1569 (9.3)	1639 (9.8)	0.88
Beneficiary age (per year), mean (SD)*	76.6 (0.1)	76.8 (0.1)	76.5 (0.1)	76.6 (0.1)	0.23
Chronic Condition Data Warehouse (CCW) conditions, count (per cent)					
Alzheimer's disease and related disorders or senile dementia	5151 (10.6)	1497 (10.0)	1793 (10.6)	1861 (11.1)	0.16
Alzheimer's disease	2061 (4.2)	627 (4.2)	704 (4.2)	730 (4.4)	0.82
Acute myocardial infarction	1408 (2.9)	394 (2.6)	494 (2.9)	520 (3.1)	0.13
Anaemia	22 450 (46.2)	6706 (44.8)	7766 (45.8)	7978 (47.7)	0.11
Asthma	4424 (9.1)	1313 (8.8)	1548 (9.1)	1563 (9.3)	0.39
Atrial fibrillation	4225 (8.7)	1265 (8.5)	1478 (8.7)	1482 (8.9)	0.69
Breast cancer	2485 (5.1)	779 (5.2)	831 (4.9)	875 (5.2)	0.48
Colorectal cancer	1139 (2.3)	357 (2.4)	406 (2.4)	376 (2.2)	0.68
Endometrial cancer	352 (0.7)	113 (0.8)	109 (0.6)	130 (0.8)	0.39
Lung cancer	435 (0.9)	151 (1.0)	152 (0.9)	132 (0.8)	0.19
Prostate cancer	1662 (3.4)	507 (3.4)	600 (3.5)	555 (3.3)	0.66
Cataract	31 095 (63.9)	9601 (64.2)	10 773 (63.5)	10 721 (64.1)	0.74
Heart failure	9207 (18.9)	2786 (18.6)	3155 (18.6)	3266 (19.5)	0.54
Chronic kidney disease	6904 (14.2)	2083 (13.9)	2392 (14.1)	2429 (14.5)	0.62
Chronic obstructive pulmonary disease	9108 (18.7)	2635 (17.6)	3165 (18.7)	3308 (19.8)	0.02
Depression	12 042 (24.8)	3728 (24.9)	4145 (24.4)	4169 (24.9)	0.83
Diabetes	13 296 (27.3)	3947 (26.4)	4590 (27.1)	4759 (28.5)	0.16
Glaucoma	10 030 (20.6)	3086 (20.6)	3501 (20.6)	3443 (20.6)	0.99
Hip/pelvic fracture	1531 (3.1)	430 (2.9)	535 (3.2)	566 (3.4)	0.15
Hyperlipidaemia	37 132 (76.4)	11 266 (75.3)	12 898 (76.1)	12 968 (77.6)	0.11

Continued

Table 2 Continued

	Total	Diagnosis question per cent correct			P value*
		Top (78.5–95.8)	Middle (71.4–78.4)	Bottom (42.9–71.3)	
Benign prostatic hyperplasia	5815 (12.0)	1792 (12.0)	1987 (11.7)	2036 (12.2)	0.76
Hypertension	37607 (77.3)	11345 (75.8)	13011 (76.7)	13251 (79.3)	<0.001
Hypothyroidism	11425 (23.5)	3490 (23.3)	3862 (22.8)	4073 (24.4)	0.25
Ischaemic heart disease	18713 (38.5)	5616 (37.5)	6393 (37.7)	6704 (40.1)	0.06
Osteoporosis	14171 (29.1)	4372 (29.2)	4794 (28.3)	5005 (29.9)	0.34
Rheumatoid arthritis	23352 (48.0)	6879 (46.0)	8275 (48.8)	8198 (49.0)	0.02
Stroke	6255 (12.9)	1880 (12.6)	2212 (13.0)	2163 (12.9)	0.70
Number of chronic conditions, count (per cent)					
≤4	5066 (10.4)	1459 (9.8)	1744 (10.3)	1863 (11.1)	0.08
5–7	16861 (34.7)	5392 (36.0)	5981 (35.3)	5488 (32.8)	0.006
8–10	16230 (33.4)	4907 (32.8)	5664 (33.4)	5659 (33.8)	0.35
≥11	10475 (21.5)	3200 (21.4)	3567 (21.0)	3708 (22.2)	0.28
Mental health visit, count (per cent)	6347 (13.1)	2040 (13.6)	2119 (12.5)	2188 (13.1)	0.46
Hierarchical Condition Category score, mean (SD)*	0.98 (0.01)	0.96 (0.01)	0.98 (0.01)	0.99 (0.01)	0.19
Household medium income, mean $ (SD)*	59852 (643)	61574 (1106)	59113 (1144)	59063 (1075)	0.19
Medicaid dual eligible, count (per cent)	6392 (13.1)	1793 (12.0)	2411 (14.2)	2188 (13.1)	0.28
Rural county residence, count (per cent)	7392 (15.2)	2207 (14.8)	2866 (16.9)	2319 (13.9)	0.64
Visit characteristics					
Visit with same doctor in last year, count (per cent)	37726 (77.6)	11369 (76.0)	13154 (77.6)	13203 (79.0)	0.08
Visit with any physician in last year, count (per cent)	44852 (92.2)	13711 (91.7)	15647 (92.3)	15494 (92.7)	0.08
Days since last visit with any physician (if any visit in last year), mean (SD)*	144.2 (0.6)	147.1 (0.8)	144.4 (1.0)	141.4 (1.3)	<0.001
ED visit in prior year, count (per cent)	8101 (16.7)	2428 (16.2)	2879 (17.0)	2794 (16.7)	0.43
Days since last ED visits (if ED visit in last year), mean (SD)*	222.8 (0.9)	221.2 (1.5)	223.5 (1.5)	223.4 (1.5)	0.47
Hospitalisation in prior year, count (per cent)	4227 (8.7)	1280 (8.6)	1489 (8.8)	1458 (8.7)	0.85
Days since last hospitalisation (if hospitalisation in last year), mean (SD)*	229.6 (1.2)	229.1 (2.1)	229.7 (2.1)	230.1 (1.9)	0.95
Index visit diagnosis groups, count (per cent)					
Abscess	1005 (2.1)	268 (1.8)	394 (2.3)	343 (2.1)	0.21
Anaemia	12410 (25.5)	3817 (25.5)	4369 (25.8)	4224 (25.3)	0.93
Aortic aneurysm	11491 (23.6)	3495 (23.4)	4165 (24.6)	3831 (22.9)	0.18
Appendicitis	2584 (5.3)	845 (5.6)	949 (5.6)	790 (4.7)	0.01
Bacteraemia	5567 (11.4)	1660 (11.1)	1929 (11.4)	1978 (11.8)	0.83
Congestive heart failure	12137 (25.0)	3633 (24.3)	4221 (24.9)	4283 (25.6)	0.67
Acute coronary syndrome	16228 (33.4)	4627 (30.9)	5740 (33.9)	5861 (35.1)	0.02
Depression	12637 (26.0)	3932 (26.3)	4312 (25.4)	4393 (26.3)	0.78
Fracture	13409 (27.6)	4324 (28.9)	4364 (25.7)	4721 (28.2)	0.11
Pulmonary embolism	8534 (17.5)	2683 (17.9)	2984 (17.6)	2867 (17.1)	0.71
Pneumonia	12183 (25.1)	3773 (25.2)	4224 (24.9)	4186 (25.0)	0.97
Spinal cord compression	6386 (13.1)	1985 (13.3)	2218 (13.1)	2183 (13.1)	0.94
Stroke	10026 (20.6)	3003 (20.1)	3542 (20.9)	3481 (20.8)	0.79

Continued

Table 2 Continued

	Total	Diagnosis question per cent correct: Top (78.5–95.8)	Middle (71.4–78.4)	Bottom (42.9–71.3)	P value*

*P values and SD accounted for correlated errors within physicians.
ED, emergency department.

next day coding of outcome measures. Associations with diagnostic knowledge were also fairly robust to physician's practice size for both the ED and hospitalisation measures when we limited the sample to either small or large or academic practices.

Suggesting that our results were not influenced by omitted variable bias, we found that associations with diagnostic knowledge and our outcome measures became small and statistically insignificant when we limited the sample to index visits with diagnoses unrelated to any of the 13 diagnostic sensitive error conditions, and so were at lower risk for diagnostic error (p>0.50 and associations were at most about a tenth of the base case per cent difference between top and bottom third of diagnostic knowledge). We also found no significant association between lack of diagnostic knowledge and elective hospitalisations (p=0.63).

DISCUSSION

We found that higher diagnostic knowledge among US outpatient internal medicine physicians was associated with significant reductions in subsequent adverse outcomes whose cause was at risk for diagnostic error. Indeed, for every 1000 index visits for a new problem at risk for diagnostic error, being seen by a physician in the top versus bottom third of diagnostic knowledge was associated with 2.9 fewer all-cause death and, for diagnostic error sensitive conditions, 4.1 fewer hospitalisations and 4.9 fewer ED visits within 90 days. These figures correspond to a reduction in risk for these adverse events by about a third. Although some prior studies have demonstrated the high morbidity and mortality of diagnostic error,[1–3] this is the first study to demonstrate and quantify the direct association between serious adverse outcomes and the diagnostic knowledge of their first contact primary care physician. These finding support the notion that gaps in diagnostic knowledge between physicians may be an important contributor to the diagnostic error problem plaguing the healthcare system worldwide.

We measured the association between diagnostic knowledge and potential diagnostic error by using Medicare claims data to identify patients who presented for outpatient visits with problems at heightened risk for serious diagnostic errors and examining the occurrence of clinically relevant adverse outcomes soon thereafter. Although this approach lacks the precision of individual chart audits,[7] it is both clinically plausible and scalable in that it can be used to monitor the care of large numbers of patients, making the method itself an important contribution to the literature on diagnostic error. Although we did not directly measure diagnostic errors through chart audits, the fact that we found associations with diagnostic knowledge and the diagnostic error sensitive outcome conditions we studied coupled with the fact that we did not find associations with treatment knowledge, nor did we find associations when the underlying diagnostic error sensitive condition was likely not present during the outpatient index visit because no antecedent diagnoses recorded indicates that the associations we report in this study were likely driven by association with diagnostic errors that occurred during these visits. Furthermore, our approach builds on prior studies that used claims data to infer diagnostic error incidence for ED visits, in that we identified index visit diagnoses at risk for diagnostic error that were clinically plausible and verified empirically, and we assured that we were studying new problems by requiring that the patient had not had an ED, hospital or outpatient visit over the previous 3 months.[30–32] We expanded on these studies by focusing on outpatient care and by examining a much more comprehensive set of presenting problems that may have been precursors to one of 13 diagnostic error prone conditions that we studied. This approach was necessary in order to study diagnostic error in the more low acuity setting of outpatient general internal medicine.

Our findings suggest an association between diagnostic knowledge and adverse outcomes. Yet, there are important limitations to consider. We did not directly determine whether a diagnostic error had occurred through such validated means as a chart review. Our findings cannot be interpreted as causal given the cross-sectional nature of our study so we cannot rule out the possibility that observed associations were the result of omitted variable bias related to either physician or patient characteristics, and do not reflect a causal relationship between diagnostic knowledge and adverse outcomes. That said, there is no reason to believe that these characteristics would be correlated with diagnostic knowledge independent of treatment knowledge which we were able to control for as both these knowledge measures should be similarly correlated with unobserved factors such as ability of consulting colleagues. Furthermore, had associations with diagnostic knowledge been driven by omitted variable bias then we would have expected them to be similar when estimated across index visits with lower or higher risk for diagnostic error, and they were not. We also

Table 3 Associations with diagnostic knowledge and adverse events per 1000 index visits

Diagnostic knowledge tertile	Death*				Emergency department visit†				Hospitalisation‡			
	Unadjusted	Regression adjusted§¶			Unadjusted¶	Regression adjusted§¶			Unadjusted¶	Regression adjusted§¶		
	Events per 1000 visits (95% CI)	Events per 1000 visits (95% CI)	Difference (95% CI)	P value	Events per 1000 visits (95% CI)	Events per 1000 visits (95% CI)	Difference (95% CI)	P value	Events per 1000 visits (95% CI)	Events per 1000 visits (95% CI)	Difference (95% CI)	P value
Top	6.2 (5.0 to 7.4)	5.2 (4.1 to 6.3)	−2.9 (−5.0 to −0.7)	0.008	13.0 (11.2 to 14.8)	11.5 (9.8 to 13.2)	−4.9 (−8.1 to −1.6)	0.003	10.4 (8.8 to 12.1)	9.2 (7.7 to 10.8)	−4.1 (−6.9 to −1.2)	0.006
Middle	6.6 (5.4 to 7.8)	6.5 (5.4 to 7.6)	−1.6 (−3.6 to 0.3)	0.09	13.0 (11.2 to 14.7)	13.2 (11.5 to 15.0)	−3.1 (−6.1 to −0.1)	0.04	10.8 (9.2 to 12.4)	11.0 (9.4 to 12.6)	−2.3 (−4.9 to 0.4)	0.09
Bottom	6.6 (5.5 to 7.8)	8.1 (6.5 to 9.7)	Reference		14.9 (13.0 to 16.8)	16.4 (14.0 to 18.7)	Reference		12.1 (10.4 to 13.8)	13.3 (11.2 to 15.4)	Reference	

*All cause mortality within 90 days of an outpatient index visit with a diagnosis at risk for one of 3 diagnostic error sensitive conditions.
†Emergency department visit for one of the 13 diagnostic error sensitive conditions within 90 days of an outpatient index visit with a visit at risk for that condition.
‡Hospitalisations were for non-elective hospitalisations either initiated through the ED or a trauma centre with a discharge diagnosis for one of 13 diagnostic error sensitive conditions within 90 days of the index visit with a diagnosis at risk for that condition.

found that diagnosis exam performance was not associated with elective hospitalisations, which are, presumably, unrelated to underlying diagnostic knowledge but may be related to the overall propensity to hospitalise. That said, the fact that practice size was found to be correlated with diagnostic exam performance is concerning. For example, as described above, practice size could be correlated with access to specialists that in turn might be related to our outcome measures. However, sensitive analyses indicate that associations with knowledge and our utilisation adverse outcome measures were fairly similar across physicians practice size/type (small, and large or academic). An additional limitation is that we studied select conditions among older patients enrolled in the Medicare programme so we cannot extrapolate these findings to a younger population, other conditions we did not consider, or populations with no or different health insurance coverage. Our findings might also not be applicable to older physicians who certified before 2000 or younger physicians who certified after 2000 as well as physicians who choose not to attempt an exam. While a physician's clinical knowledge might be related to their decision to not take the MOC exam therefore not maintaining their certification, other factors certainly play a role in this decision.

Another limitation of our study is that the IM-MOC exam was specifically designed to measure clinical knowledge in general, it was not designed to measure diagnostic knowledge specifically. That said, diagnostic knowledge is a major component of the exam and was found to meet the criteria for measuring this underlying construct. Also diagnostic error may have stemmed from factors outside of inadequate diagnostic knowledge, which are not covered by the exam but could be correlated with our exam based diagnostic knowledge measure (eg, poor patient/physician communication skills and related system failures).[33 34] That said, there is no reason to believe that these other contributors to diagnostic error would not also be correlated with the other aspects of the exam we do account for. Furthermore, based on an analysis of malpractice claims, Newman-Toker *et al*[6] reported that clinical judgement played an important role in 86% of diagnostic errors, while poor patient/physician communication and system failures played a role in far fewer diagnostic errors that resulted in malpractice suits (35% and 22%, respectively). Suggesting that improving communication will not reduce stroke related diagnostic error, Kerber and Newman-Toker[35] reported that frontline providers rarely ask the right questions when patients present with dizziness. Communication ability is only valuable in terms of reducing diagnostic error if the physician knows what questions to ask and what the answers mean. Although we cannot say with certainty that our finding is driven by an underlying association between diagnostic knowledge and diagnostic errors, at a minimum, our finding suggests that patients treated by physicians who scored well on diagnostic

exam questions may be at lower risk for the adverse outcomes we studied. Finally, some might assert that a standardised exam without access to medical reference material might be more a reflection of a physician's rote memory and ability to recall medical facts than a test of their clinical knowledge and judgement. Although this is a fundamental limitation of our study, it should be noted that the exam is designed to mimic decision making in real life situations including such things as patient's laboratory results and reference material impeded in the exam and past research indicates that an 'open' book format that allows physicians access to reference material did not materially impact exam performance.[36] It should also be noted that the necessary rapidity of decision-making by primary care physicians who have limited time per encounter might fairly be represented by an exam with time constraints.

In this exploratory analysis, we found evidence that diagnostic knowledge of primary care physicians seeing a patient for an index visit for a problem that is at heightened risk of diagnostic error is associated with adverse outcomes. The fact that there exists a link between general diagnostic knowledge and diagnostic error may not be surprising, the magnitude of the associations we found suggests that interventions ignoring the role of physician knowledge may be inadequate to address the crisis of diagnostic error. Interventions targeted at improving diagnostic knowledge could include such things as a greater focus on diagnostic training during graduate medical education (ie, medical school, residency and fellowship). Knowledge-focused interventions could also include incentivising broad-based learning as well as targeted learning pursued through continuing medical education activities.[30] During visits identified as being at risk for diagnostic errors, physicians could be given related information at the point of care including suggestions for specialty consultation.

Our results are important for two additional reasons. First, these results provide evidence that board certification and maintenance of certification, which involves lifelong learning directed at maintaining medical knowledge, might, in fact, be a valid approach to assuring the delivery of high-quality care. Many in the USA report problem about the time and expense of MOC and often point to the lack of rigorous assessment between aspects of MOC and outcomes of interest to patients. These findings suggest that processes such as MOC may translate into meaningful improvements in outcomes because they can provide incentives for meaningful learning. This learning also could be enhanced through exam feedback targeted at diagnostic knowledge. Second, the findings also suggest that interventions aimed at improving diagnostic skills, whether knowledge-based or through, for instance, delivery of relevant information at the point of care (this is in response to system changes) might be approaches that might be worthwhile if the findings of this study are validated with additional research. Yet more research is needed to better understand the link between diagnostic knowledge and diagnostic errors that are identified through chart review or other methods of direct ascertainment and the extent to which such errors result in adverse clinical outcomes.

In conclusion, gaps in diagnostic knowledge among first contact primary care physicians are associated with serious diagnostic error sensitive outcomes. If this finding is confirmed in future studies, diagnostic knowledge should be a target for interventions to reduce diagnostic errors.

Author affiliations
[1]Assessment and Research, American Board of Internal Medicine, Philadelphia, Pennsylvania, USA
[2]Division of Endocrinology, Department of Medicine, Mayo Clinic, Rochester, Minnesota, USA
[3]Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts, USA

Contributors All authors substantially contributed to the conception and design of the work. BG, JLV and RSL contributed to the acquisition of the data. All authors substantially contributed to analysis or interpretation of data. All authors substantially contributed to the drafting the work and revising it critically for important intellectual content. All authors gave final approval of the version published. All agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work were appropriately investigated and resolved.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests BG, JV and RL are paid employees of the American Board of Internal Medicine. BL is a paid consultant for the American Board of Internal Medicine.

Patient consent for publication Not required.

Ethics approval Advarra Institutional Review Board approved our study protocol. Continuing review number (CR00144650) and the protocol number (Pro00026550).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. Administrative data describing physician characteristics and exam performance can be obtained from the ABIM through a data sharing agreement that assures physician confidentiality and its use for legitimate research purposes. Access to deidentified Medicare claims data for this study were obtained through a special data use agreement with the Centers for Medicare and Medicaid services which is a process available to researchers in the US.

ORCID iDs
Bradley M Gray http://orcid.org/0000-0001-8979-019X
Rozalina G McCoy http://orcid.org/0000-0002-2289-3183

REFERENCES

1 National Academies of Sciences, Engineering, and Medicine. *Improving diagnosis in health care*. Washington, DC: The National Academies Press, 2018.
2 Graber ML, Trowbridge R, Myers JS, *et al*. The next organizational challenge: finding and addressing diagnostic error. *Jt Comm J Qual Patient Saf* 2014;40:102–10.
3 Cresswell KM, Panesar SS, Salvilla SA, *et al*. Global research priorities to better understand the burden of iatrogenic harm in primary care: an international Delphi exercise. *PLoS Med* 2013;10:e1001554.
4 Kostopoulou O, Delaney BC, Munro CW. Diagnostic difficulty and error in primary care--a systematic review. *Fam Pract* 2008;25:400–13.
5 Goyder CR, Jones CHD, Heneghan CJ, *et al*. Missed opportunities for diagnosis: lessons learned from diagnostic errors in primary care. *Br J Gen Pract* 2015;65:e838–44.
6 Newman-Toker DE, Schaffer AC, Yu-Moe CW, *et al*. Serious misdiagnosis-related harms in malpractice claims: The "Big Three" - vascular events, infections, and cancers. *Diagnosis* 2019;6:227–40.
7 Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving us adult populations. *BMJ Qual Saf* 2014;23:727–31.
8 Gandhi TK, Kachalia A, Thomas EJ, *et al*. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. *Ann Intern Med* 2006;145:488–96.
9 Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165:1493–9.
10 Kachalia A, Gandhi TK, Puopolo AL, *et al*. Missed and delayed diagnoses in the emergency department: a study of closed malpractice claims from 4 liability insurers. *Ann Emerg Med* 2007;49:196–205.
11 Poon EG, Kachalia A, Puopolo AL, *et al*. Cognitive errors and logistical breakdowns contributing to missed and delayed diagnoses of breast and colorectal cancers: a process analysis of closed malpractice claims. *J Gen Intern Med* 2012;27:1416–23.
12 Chisholm A, Askham J. *A review of professional codes and standards for doctors in the UK*. USA and Canada: Picker Institute Europe Oxford, 2006.
13 Irvine D. Doctors in the UK: their new professionalism and its regulatory framework. *Lancet* 2001;358:1807–10.
14 Kovacs E, Schmidt AE, Szocska G, *et al*. Licensing procedures and registration of medical doctors in the European Union. *Clin Med* 2014;14:229–38.
15 European Union of Medical Specialists. The European Council for accreditation of medical specialist qualifications (ECAMSQ), 2010. Available: https://www.uems.eu/__data/assets/pdf_file/0009/1206/ECAMSQ_presentation.pdf
16 Petterson SM, Liaw WR, Phillips RL, *et al*. Projecting US primary care physician workforce needs: 2010-2025. *Ann Fam Med* 2012;10:503–9.
17 Freed GL, Dunham KM, Singer D. Use of board certification and recertification in hospital privileging: policies for general surgeons, surgical specialists, and nonsurgical subspecialists. *Arch Surg* 2009;144:746–52.
18 Independence Blue Cross. Credentialing criteria, 2019. Available: https://www.ibx.com/pdfs/providers/interactive_tools/credentialing_criteria_ibx.pdf
19 American Board of Medical Specialties. Certification matters FAQs, 2019. Available: https://www.certificationmatters.org/faqs/
20 Lipner RS, Bylsma WH, Arnold GK, *et al*. Who is maintaining certification in internal medicine--and why? A national survey 10 years after initial certification. *Ann Intern Med* 2006;144:29–36.
21 Balogh E, Miller BT, Ball J. *Improving diagnosis in health care*: xxvii, 444 pages p.
22 Centers for Medicare & Medicaid Services. BPCI advanced, 2018. Available: https://innovation.cms.gov/initiatives/bpci-advanced
23 Centers for Medicare & Medicaid Services. Comprehensive care for joint replacement model, 2018. Available: https://innovation.cms.gov/initiatives/cjr
24 Schiff GD, Hasan O, Kim S, *et al*. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med* 2009;169:1881–7.
25 Gray B, Vandergrift J, Lipner RS, *et al*. Comparison of content on the American Board of internal medicine maintenance of certification examination with conditions seen in practice by general internists. *JAMA* 2017;317:2317–24.
26 Samonte K, de la Cruz S, Garcia MJ. An Overview of the ABIM Cardiovascular Disease Maintenance of Certification Examination. *J Am Coll Cardiol* 2020;75:1083–6.
27 Bandalos DL. *Measurement theory and applications for the social sciences*. Guilford Publications, 2018.
28 Huber PJ. The behavior of maximum likelihood estimates under non-standard conditions. *Paper presented at: Fifth Berkeley symposium on mathematical statistics and probability*, Berkeley, CA, 1967.
29 White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 1980;48:817–38.
30 Newman-Toker DE, Makary MA. Measuring diagnostic errors in primary care: the first step on a path forward. Comment on "Types and origins of diagnostic errors in primary care settings". *JAMA Intern Med* 2013;173:425–6.
31 Waxman DA, Kanzaria HK, Schriger DL. Unrecognized cardiovascular emergencies among Medicare patients. *JAMA Intern Med* 2018;178:477–84.
32 Liberman AL, Newman-Toker DE. Symptom-Disease pair analysis of diagnostic error (spade): a conceptual framework and methodological approach for unearthing misdiagnosis-related harms using big data. *BMJ Qual Saf* 2018;27:557–66.
33 Giardina TD, King BJ, Ignaczak AP, *et al*. Root cause analysis reports help identify common factors in delayed diagnosis and treatment of outpatients. *Health Aff* 2013;32:1368–75.
34 Zwaan L, Monteiro S, Sherbino J, *et al*. Is bias in the eye of the beholder? A vignette study to assess recognition of cognitive biases in clinical case workups. *BMJ Qual Saf* 2017;26:104–10.
35 Kerber KA, Newman-Toker DE. Misdiagnosing dizzy patients: common pitfalls in clinical practice. *Neurol Clin* 2015;33:565–75. viii.
36 Lipner RS, Brossman BG, Samonte KM, *et al*. Effect of access to an electronic medical resource on performance characteristics of a certification examination: a randomized controlled trial. *Ann Intern Med* 2017;167:302–10.

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

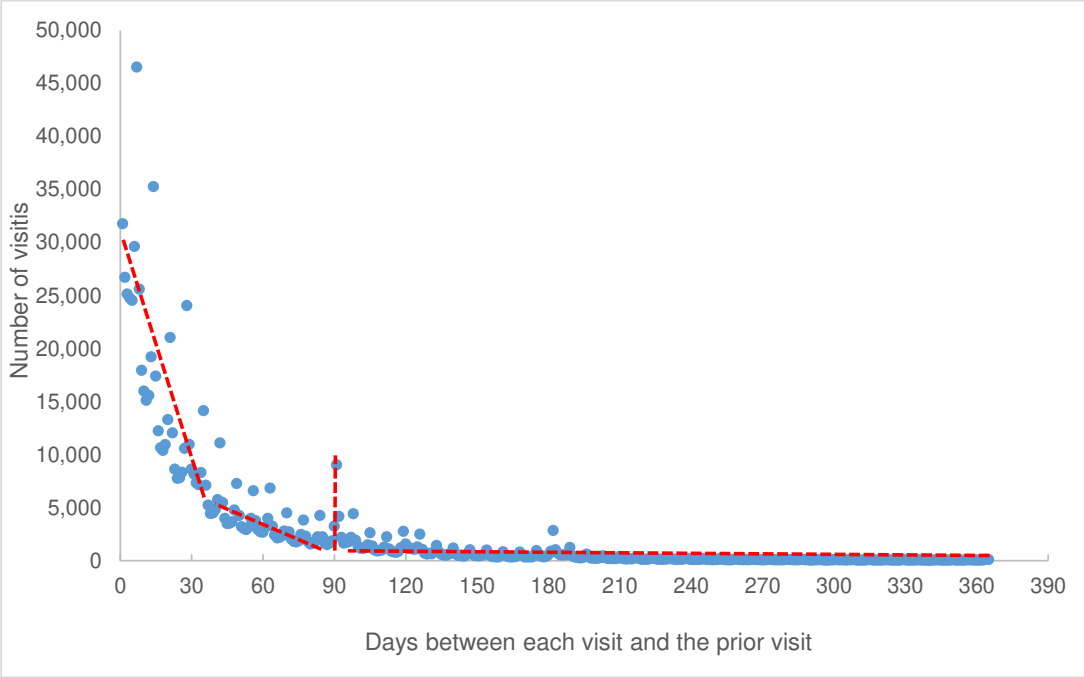
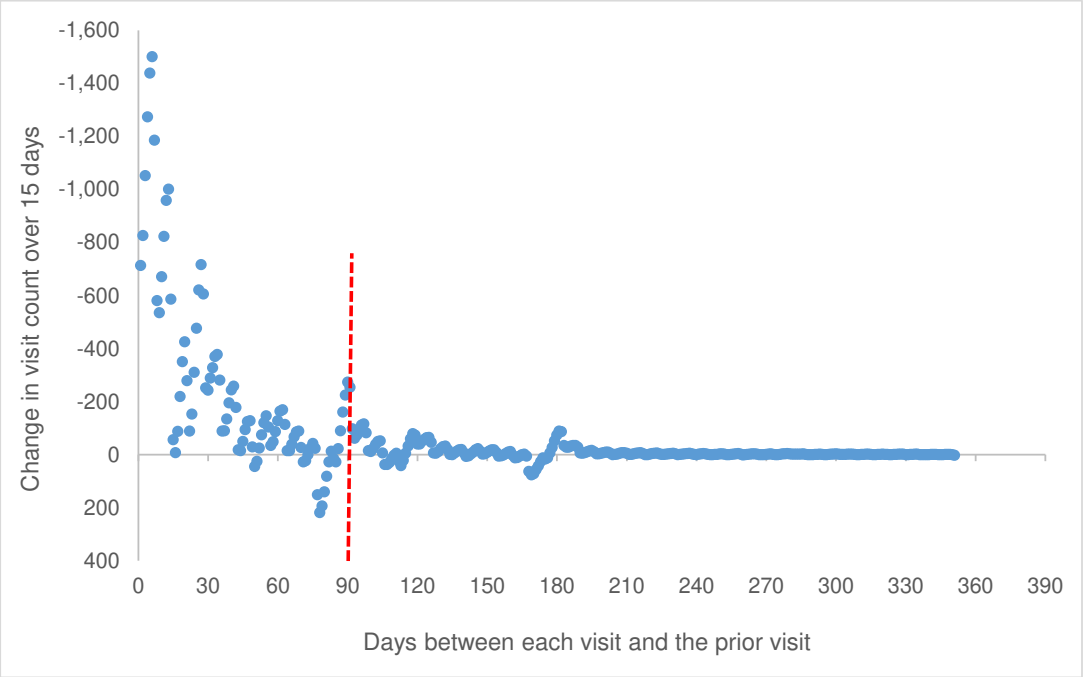


Figure 1.2. Average Change in Visit Count over the 15 days (15-day slope) Following each Data Point Listed in Figure 1



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

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

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

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
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.

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.

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

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

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.

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 ($P<0.05$)).

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

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

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.