

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Diagnostic Knowledge, Diagnostic Error and Risk of Death, Hospitalization and Emergency Department Visits

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-041817
Article Type:	Original research
Date Submitted by the Author:	19-Jun-2020
Complete List of Authors:	Gray, Bradley Michael; American Board of Internal Medicine Vandergrift, Jonathan; American Board of Internal Medicine McCoy, Rozalina; Mayo Clinic, Division of Endocrinology, Department of Medicine Lipner, Rebecca; American Board of Internal Medicine Landon, Bruce; Harvard Medical School, Department of Health Care Policy
Keywords:	INTERNAL MEDICINE, MEDICAL EDUCATION & TRAINING, GENERAL MEDICINE (see Internal Medicine)

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1 Diagnostic Knowledge, Diagnostic Error and Risk of Death, Hospitalization and Emergency

2 Department Visits

3
4
5
6
7
8
9
10
11 4 Bradley M. Gray, PhD, corresponding author

12 5 Email: bgray@abim.org, Phone: 202-213 6646, FAX 202-213 6646

13 6 American Board of Internal Medicine

14 7 510 Walnut Street, Suite 1700, Philadelphia, Pennsylvania 19106, USA

15 8
16
17
18
19 9 Jonathan L. Vandergrift, MS

20 10 American Board of Internal Medicine Philadelphia, Pennsylvania, USA

21 11
22
23 12 Rozalina, G. McCoy, MD

24 13 Mayo Clinic, Rochester, Minnesota, USA

25 14
26 15 Rebecca S. Lipner, PhD

27 16 American Board of Internal Medicine Philadelphia, Pennsylvania, USA

28 17
29 18 Bruce E. Landon, MD

30 19 Harvard Medical School, Boston, Massachusetts, USA

31 20
32 21 Word count: 3,944,

33 22 Keywords: Internal Medicine, General Medicine, Medical Education & Training,

23 **Objective**

24 Diagnostic error is a key health care concern with large associations with morbidity and
25 mortality. Yet no study has quantified associations between outcomes whose cause was at risk
26 for diagnostic errors and one potentially large contributor to these errors: deficiencies in
27 diagnostic knowledge. Our objective was to measure that associations between diagnostic
28 knowledge and adverse outcomes at risk for diagnostic errors.

29 **Setting**

30 US primary care

31 **Participants**

32 1,410 general internists treating 42,407 Medicare beneficiaries during 48,632 outpatient visits.

33 **Outcome measures**

34 Using Medicare claims from general internists who recently took their American Board of
35 Internal Medicine Maintenance of Certification exam, we identified outpatient “index” visits for
36 new complaints at risk for diagnostic error because the presenting complaint was related to pre-
37 specified diagnostic error sensitive conditions.

38 **Design**

39 Using a cross-sectional design, we related performance on ABIM-MOC diagnostic exam
40 questions to 90-day risk of all-cause death, and, for outcome conditions related to the index visits
41 diagnosis, emergency department (ED) visits and hospitalizations.

42 **Results**

1
2
3 43 Rates of 90-day adverse outcomes per 1,000 index visits were 7 for death, 11 for
4
5 44 hospitalizations, and 14 for ED visits. Being seen by a physician in the top versus bottom third
6
7 45 of diagnostic knowledge during an index visit for a new compliant at risk for diagnostic error
8
9 46 was associated with 2.9 fewer all-cause deaths (95% confidence interval (CI) -5.0 to -0.7,
10
11 47 P=.008), and 4.1 fewer applicable hospitalizations (95% CI -6.9 to -1.2, P=0.006), and 4.9 fewer
12
13 48 applicable ED visits (95% CI -8.1% to -1.6%, P=0.003) per 1,000 visits.
14
15
16
17

18 49 **Conclusion**

19
20 50 Higher diagnostic knowledge was associated with lower risk of adverse outcomes at heightened
21
22 51 risk for diagnostic error.
23
24
25
26 52

1
2
3 53 **Strength Limitations**
4
5

6 54 • **Strengths**
7

- 8 ○ Unique diagnostic knowledge measure
9
10 ○ Linking diagnostic knowledge with adverse outcomes
11
12 ○ Scalable adverse outcome measures
13
14 ○ Extensive sensitivity analyses
15
16

17 58
18 59 • **Limitation**
19

- 20 60 ○ Omitted variable bias
21
22 61 ○ No direct diagnostic error measure
23
24

25 62
26
27

28 63
29
30

31 64
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57

65 Introduction

66
67 Diagnostic error has been identified as a key health care delivery concern and contributes to
68 significant potentially preventable morbidity and mortality.(1-3) Ambulatory care, and
69 especially primary care, is a practice setting with a particularly high risk for diagnostic error(4,
70 5) because of the wide variety of presentations encountered and the concomitant difficulty of
71 distinguishing harmful conditions from routine self-limited complaints, compounded by the well-
72 known time constraints faced by practitioners in that setting. Sing et. al., estimated that at least
73 5% of ambulatory visits are associated with diagnostic error, half of which may result in
74 considerable patient harm.(6) Similarly, Newman-Toker et al. reported that the cause of most
75 malpractice suits was diagnostic error and that the majority of these occurred in the ambulatory
76 care settings.(6, 7)

77
78 Deficiencies in diagnostic knowledge are likely to be an important contributor to these diagnostic
79 errors that could impact, for example, the breadth of diagnoses considered, appropriate ordering
80 and interpretation of tests, and/or synthesis of data more generally.(8-11) Because of this,
81 measuring physician diagnostic knowledge has become a major focus of organizations
82 throughout the developed world that are tasked with licensing and certifying physicians with the
83 underlying, although largely untested, hypothesis being that diagnostic knowledge will be a
84 measurable and strong predictor of diagnostic error.(12-15) Testing this hypothesis and
85 quantifying this relationship is therefore a critical public policy concern both in terms of the

1
2
3 86 importance of board certification and other programs designed to enhance lifelong learning for
4
5 87 physicians.
6
7

8 88
9
10
11 89 In the US, the American Board of Internal Medicine (ABIM) is a leading organization that
12
13 90 certifies primary care physicians, most notably general internists. In fact, most general internists
14
15 91 in the US are certified by the ABIM and these physicians represent about 45% of all adult
16
17 92 primary care physicians in the US.(16) Unlike medical licensure, board certification is not a legal
18
19 93 requirement to practice medicine in the US, though many hospitals require board certification as
20
21 94 one criterion to obtain privileges and insurers often require board certification to be included in
22
23 95 covered physician panels.(17, 18) To maintain their certification, general internists must pass an
24
25 96 initial certifying exam and, periodically, pass a recertification exam thereafter (referred to as
26
27 97 Maintenance of Certification (MOC) exams).(19, 20) Diagnostic knowledge is a major
28
29 98 component of these exams representing about half of all exam questions for the Internal
30
31 99 Medicine MOC (IM-MOC) exam.
32
33
34
35
36

37 100
38
39
40 101 One explanation for the lack of research on this topic is the difficulty in studying the relationship
41
42 102 between general diagnostic knowledge and diagnostic error because of the inability to quantify
43
44 103 diagnostic knowledge and identifying diagnostic errors at a population level, especially in the
45
46 104 outpatient setting.(21) We address this gap in the literature by applying a unique measure of
47
48 105 diagnostic knowledge, performance on diagnostic related questions on ABIM's IM-MOC exam,
49
50 106 and relating this measure to deaths, hospitalizations, and emergency department visits that
51
52 107 occurred after outpatient visits for new complaints at heightened risk for diagnostic error.
53
54
55
56
57

108 Methodology

109

110 *Physician and Index Visit Sample*

111 Our physician sample included general internists who were initially ABIM board certified in
112 2000 and took their IM-MOC exam between 2008 and 2011 (Figure 1). We identified Medicare
113 beneficiary outpatient Evaluation & Management (E&M) visits with these physicians using their
114 National Provider Identifier (NPI) during the calendar year following their exam (2009 to 2012).
115 These patients were age 65 or older and continuously enrolled in Medicare fee-for-service
116 (Medicare insures most of the US population over 65) during the physician's one year follow-up
117 period and the year prior. To ensure that any presenting complaints being evaluated were new
118 (i.e., not follow up), we restricted these visits to those that were the first visit for a new complaint
119 (the "*index visit*") because these visits were preceded by a 90-day clean period with no previous
120 inpatient or outpatient visit. The 90-day clean period is consistent with the US government
121 Centers for Medicare and Medicaid Services criteria used by its Bundled Payments for Care
122 Improvement Program for defining new episodes of care and with the patterns of visits we
123 observed (see Appendix Section 1 for related analysis).(22, 23)

124

125 We further restricted these index visits to those at heightened risk for diagnostic errors because
126 the recorded diagnosis in the Medicare claims (the "*index visit diagnosis*"), which includes
127 recording of symptom (e.g. loss of balance), could have been the initial presenting complaint for
128 one or more of 13 pre-specified diagnostic error sensitive conditions such as congestive heart
129 failure or bacteremia/sepsis (see Table 1). These 13 conditions (see Appendix Section 2 for a list

1
2
3 130 and applicable ICD-9 codes) were an acute non-cancerous subset of 20 conditions previously
4
5 131 noted by Schiff et al. to be at high risk for serious diagnostic error.(24) For instance, index visits
6
7
8 132 with diagnosis codes for chest pain, dyspepsia, shortness of breath, hypoxemia/hypoxia,
9
10 133 respiratory distress, weakness/fatigue, edema or ascites could all be the initial presentation of
11
12 134 congestive heart failure, which is one of the 13 diagnostic error sensitive conditions.
13
14

15 135
16
17
18 136 We used a three-step process to identify eligible index visit diagnoses. First, two physician
19
20 137 authors (RGM and BEL) identified all diagnoses that could be presenting complaints for the 13
21
22 138 diagnostic error sensitive conditions: what complaints/diagnoses might someone who ultimately
23
24 139 presented with a diagnostic-error sensitive condition have presented with initially? Second,
25
26 140 because the original list of identified index visit diagnoses was large (76), we reduced this list to
27
28 141 38 by applying a relative risk (RR) criteria. For a specific index visit diagnosis to meet this
29
30 142 criteria, all index visits with that diagnosis had to have a greater portion of later ED visits or
31
32 143 hospitalizations with the related outcome condition discharge diagnosis than index visits where
33
34 144 the specific at risk diagnosis was not present. For example, dizziness was chosen as an eligible
35
36 145 index visit diagnosis for stroke, one of the diagnostic error sensitive conditions, both because it
37
38 146 was identified as a potential presenting symptom of a stroke by physician authors and because
39
40 147 index visits with that diagnosis had a greater proportion of later hospitalization or ED visits for
41
42 148 stroke than visits without this diagnosis. Third, we also included index visits where the actual
43
44 149 diagnosis was one of the 13 diagnostic error sensitive conditions because we wanted to include
45
46 150 cases where diagnostic errors were and were not made. Therefore, we also included index visits
47
48 151 with a diagnosis of congestive heart failure itself as being at risk for the underlying condition
49
50 152 congestive heart failure.
51
52
53
54
55
56
57
58
59
60

1
2
3 153 There was no patient/public involvement in the design, conduct or reporting for this study.
4
5

6 154
7

8
9 155 ***Outcome Measures***
10

11
12 156 We examined the risk of three serious adverse outcomes within 90 days of the index visit that we
13
14 157 hypothesized would occur more frequently in cases of misdiagnosis: all-cause mortality,
15
16 158 hospitalizations, and ED visits. We did not count these events as adverse outcomes if they
17
18 159 occurred on the same day as the index visit because this may reflect a positive action (the
19
20 160 physician correctly diagnosed a patient with stroke and referred/admitted them to the hospital) or
21
22 161 be unavoidable regardless of the accuracy of the index visit diagnosis (the patient died despite
23
24 162 immediately admitting a patient to the hospital who exhibited stroke symptoms). Based on
25
26 163 Medicare billing codes, hospitalizations were limited to non-elective hospitalizations initiated
27
28 164 through the ED or trauma center. The ED and hospitalization outcomes were also limited to
29
30 165 cases where the discharge diagnosis was for one of the 13 diagnostic error sensitive conditions
31
32 166 following an index visit with the applicable diagnosis. We therefore presumed that these
33
34 167 discharge diagnoses were a reasonable representation of the underlying condition of the patient
35
36 168 at the time of the index visit. For example, we would count a hospitalization with a discharge
37
38 169 diagnosis of stroke as an adverse outcome if it occurred after an index visit for dizziness because
39
40 170 dizziness was identified as being a potential presenting complaint for stroke. However, we did
41
42 171 not count hospitalizations with a discharge diagnosis for acute coronary syndrome following an
43
44 172 index visit for dizziness because dizziness was not identified as a presenting complaint for acute
45
46 173 coronary syndrome. The rationale is that if there were no presenting complaints during the index
47
48 174 visit related to coronary syndrome, either because the underlying condition was not present or
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 175 could not be detected at the time of the index visit, then the index visit physician could not have
4
5 176 prevented the hospitalization regardless of their diagnostic knowledge.
6
7

8
9 177

10 11 178 *Measure of Diagnostic Knowledge*

12
13
14 179 Our measure of diagnostic knowledge was calculated as the percent of correct answers on the
15
16 180 IM-MOC exam for questions coded as “diagnosis related” by ABIM’s IM-MOC exam
17
18 181 committee. In our study, these questions comprised 53% of all IM-MOC exam questions, with
19
20 182 the remaining 42% addressing treatment and 5% related to other topics such as epidemiology or
21
22 183 pathophysiology. Exam questions are designed to replicate real world clinical scenarios and/or
23
24 184 patient encounters(25) and do not rely on rote memorization. Questions coded as “diagnosis
25
26 185 related” typically test knowledge and skills related to diagnostic inference, differential diagnosis,
27
28 186 and diagnostic testing and therefore are measuring diagnostic knowledge and decision-making.
29
30
31 187 Psychometric analysis indicates that scores on exam questions related to diagnosis were
32
33 188 meaningfully correlated, and thereby represent an independent underlying construct that could be
34
35 189 interpreted as diagnostic knowledge (see Appendix Section 3).
36
37
38
39

40 190 *Statistical methods*

41
42
43 191 Using Probit regression we estimated the associations with each adverse outcome, with standard
44
45 192 errors adjusted for correlations resulting from the nesting of visits within patients within
46
47 193 physicians.(26, 27) To measure associations with diagnostic knowledge we included categorical
48
49 194 regression explanatory variables for top and middle third of percent correct scores on diagnosis
50
51 195 related questions (bottom third was the reference category). Other exam level explanatory
52
53 196 variables included tertile indicators for performance on treatment-related questions and
54
55
56
57

1
2
3 197 performance on other question types. Since these variables measure knowledge unrelated to
4
5 198 diagnosis, they account for correlations between factors such as unmeasured practice or patient
6
7 199 characteristics that might be correlated with exam performance and our outcome measures (e.g.,
8
9 200 high scoring physicians may be more likely to practice in an academic setting or other such
10
11 201 settings that might be independently related to diagnostic error). Exam form indicators accounted
12
13 202 for differences in exam difficulty across exam administrations.
14
15
16
17
18 203

19
20 204 We also included physician, patient and visit level regression controls. Physician level controls
21
22 205 included: practice size (indicators for solo practice and practices larger than 50 physicians),
23
24 206 practice type (indicators for academic, group), demographic (gender), and training characteristics
25
26 207 (medical school location interacted with country of birth). Patient level controls included:
27
28 208 demographic characteristics (age and age squared, gender and race/ethnicity indicators) and a
29
30 209 Medicaid eligibility indicator. Lagged patient risk adjusters included 27 indicators for chronic
31
32 210 conditions and Medicare's Hierarchical Condition Category (HCC) risk adjustment score. Patient
33
34 211 index visit location level controls included: an indicator for residing in a rural ZIP code, ZIP
35
36 212 code median household income, and indicators for 10 US Health and Human Services regions.
37
38 213 Index visit level controls included: indicators of any outpatient visit, hospitalization or ED visits
39
40 214 within the prior year and number of days since the most recent of these events, visit year
41
42 215 indicators to control for secular changes in quality. We also included an indicator for whether or
43
44 216 not the patient had a previous contact with the index visit physician during the year prior to the
45
46 217 index visit to account for differences in physician-patient continuity (see Appendix Sections 4
47
48 218 and 6 for a full list of controls).
49
50
51
52
53
54
55
56 219

1
2
3 220 *Sensitivity Analysis*
4
5

6 221 We performed numerous sensitivity analyses to test the robustness of our results (detailed in
7
8 222 Appendix Section 5). First, we expanded the index visit sample to include all index visits with
9
10 223 the original 76 diagnoses identified by the physician authors regardless of whether they met the
11
12 224 relative risk criteria. Second, we expanded and contracted the index visit clean period by seven
13
14 225 days. Third, we excluded physician in academic medical centers to consider the possibility that
15
16 226 the unobserved physician characteristics related to where they worked or who they worked with
17
18 227 could be were independently both related to the underlying physician diagnostic skill and our
19
20 228 outcome measures. Fourth, to consider the possibility that these utilizations were only avoided
21
22 229 because the patient died, for the ED and hospitalization outcome, we also included instances
23
24 230 where the patient died. Fifth, as a falsification test we limited the index visits to those that were
25
26 231 unrelated to the 13 diagnostic error sensitive conditions. Under this sensitivity, we expected then
27
28 232 that the associations with diagnostic knowledge would decline. The index visit physician's
29
30 233 diagnostic knowledge cannot impact a future adverse outcome if the underlying condition that
31
32 234 caused that outcome was not present or detectible at the time of index visit. Therefore, this
33
34 235 reduction in association should be especially true for the hospitalization and ED measures where
35
36 236 adverse outcomes were limited to the 13 diagnostic error conditions and so were unrelated to the
37
38 237 index visit diagnoses in this sensitivity. Similarly, for the last sensitivity, we applied elective
39
40 238 hospitalizations as an outcome measure to consider the possibility that there could be a
41
42 239 correlation between the overall propensity to hospitalize in an area and physician knowledge.
43
44
45
46
47
48
49

50 240
51
52

53 241 The Advarra Institutional Review Board approved our study protocol and all analyses were
54
55 242 performed using Stata version 15 (College Station, TX). Patients and the public were not
56
57

1
2
3 243 involved in the design or execution of this study as the existing patient claims data used were de-
4
5 244 identified by the Center for Medicaid and Medicare Services prior to analysis.
6
7

8 245
9

10
11 246 **Results**
12

13
14 247
15

16
17 248 Of 2,492 general internists who initially certified in 2000 and who took an IM-MOC exam
18
19 249 between 2009 and 2012, 1,722 had outpatient visits with a fee-for-service Medicare beneficiary
20
21 250 during the study period. Those without visits generally practiced hospital medicine. Of these,
22
23 251 1,410 were included in the study because they had at least one outpatient index visit that met our
24
25 252 study inclusion criteria during the year after they took their IM-MOC exam. In total, 48,632
26
27 253 index visits with 42,407 patients treated by 1,410 physicians met study inclusion criteria (Figure
28
29 254 1). Table 1 lists frequency of index visits and subsequent outcomes for each diagnostic error
30
31 255 sensitivity condition.
32
33
34
35

36 256
37
38

39 257 The mean percent correct on diagnosis questions ranged from 84.3% among top third performers
40
41 258 to 65.5% among bottom third performers (Table 2). Patient and visit characteristics were similar
42
43 259 across tertiles of physician diagnostic knowledge. For example, there were no statistically
44
45 260 significant differences in the HCC risk adjuster across tertiles ($P=.19$). However, there were
46
47 261 differences in some physician and practice characteristics. When compared to physicians in the
48
49 262 bottom tertile of diagnostic knowledge, physicians in the top were significantly less likely to be
50
51 263 in solo practice (12.8% versus 24.4%, $P=0.009$), and more likely to be in academic practice
52
53 264 (9.7% versus 3.4%, $P<.001$). However, the proportion graduating from a US medical school was
54
55
56
57

1
2
3 265 similar across diagnostic knowledge tertiles (70.0% versus 63.3%, P=.30). Although
4
5 266 performance on diagnosis and treatment related questions were highly correlated, 36% of the
6
7 267 variation in diagnosis exam performance was not explained by performance on other parts of the
8
9
10 268 exam.

13 269 *Associations between diagnostic knowledge and patient adverse outcomes*

16 270 The overall rates of 90-day adverse outcomes per 1,000 index visits was 6.5 for death, 11.1 for
17
18 271 hospitalizations, and 13.6 for ED visits (with the latter two directly associated with one of the
19
20 272 diagnostic error sensitive conditions whose antecedent was present in the index applicable index
21
22 273 visit). Being seen by a physician scoring in the top versus bottom third of diagnostic knowledge
23
24 274 on the MOC exam was associated with 2.9 fewer deaths per 1,000 visits (95% confidence
25
26 275 interval (CI) -5.0 to -0.7, P=.008) which reflects a 35.3% lower risk of death (95% CI -52.8 to -
27
28 276 11.2, P=.008), (Table 3). Our finding also suggest that this difference in exam performance was
29
30 277 associated with a 4.1 fewer applicable hospitalizations (95% CI -6.9 to -1.2, P=0.006), and 4.9
31
32 278 fewer applicable ED visits (95% CI -8.1 to -1.6, P=0.003) per 1,000 visits (Table). These
33
34 279 reductions correspond with about a 30% lower risk for these utilization measures
35
36 280 (hospitalizations: -30.5%, 95% CI -46.1 to -10.4, P=.003, ED: -29.8%, 95% CI -44.4 to -11.4).

42 281

45 282 We also found a significant dose response relationship across all three regression adjusted
46
47 283 relative risk measures (P-trends <0.008). For example, the regression-adjusted 90-day risk of
48
49 284 death per 1,000 patients whose index visit physician scored in the top third of diagnostic
50
51 285 knowledge was 5.2 (95% CI 4.1 to 6.3), compared to 6.5 (95% CI 5.4 to 7.6) for the middle
52
53 286 third, and 8.1 (95% CI 6.5 to 9.7) for the bottom third (P-trend 0.008).

287 *Sensitivity Analyses*

288 Our sensitivity analyses (Appendix Section 5) confirmed that base case associations with
289 diagnostic knowledge were robust to different index visit clean periods, and diagnosis code
290 inclusion criteria. Suggesting that our results were not influenced by omitted variable bias, we
291 found that associations with diagnostic knowledge and our outcome measures became small and
292 statistically insignificant when we limited the sample to index visits with diagnoses unrelated to
293 any of the 13 diagnostic sensitive error conditions, and so were at lower risk for diagnostic error
294 ($P>0.50$ and associations were at most about a tenth of the base case percent difference between
295 top and bottom third of diagnostic knowledge). We also found no significant association between
296 lack of diagnostic knowledge and elective hospitalizations ($P=0.63$).

298 **Discussion**

299
300 We found that higher diagnostic knowledge among US outpatient internal medicine physicians
301 was associated with significant reductions in subsequent adverse outcomes whose cause was at
302 risk for diagnostic error. Indeed, for every 1,000 index visits for a new compliant at risk for
303 diagnostic error, being seen by a physician in the top versus bottom third of diagnostic
304 knowledge was associated with 2.9 fewer all-cause death and, for diagnostic error sensitive
305 conditions, 4.1 fewer hospitalizations and 4.9 fewer ED visits within 90 days. These figures
306 correspond to a reduction in risk for these adverse events by about a third. Although some prior
307 studies have demonstrated the high morbidity and mortality of diagnostic error(1-3), this is the
308 first study to demonstrate and quantify the direct association between serious adverse outcomes

1
2
3 309 and the diagnostic knowledge of their first contact primary care physician. These finding support
4
5 310 the notion that gaps in diagnostic knowledge between physicians is an important contributor to
6
7 311 the diagnostic error problem plaguing the healthcare system worldwide.
8
9

10
11 312
12
13
14 313 We measured the association between diagnostic knowledge and potential diagnostic error by
15
16 314 using Medicare claims data to identify patients who presented for outpatient visits with
17
18 315 complaints at heightened risk for serious diagnostic errors and examining the occurrence of
19
20 316 clinically relevant adverse outcomes soon thereafter. Although this approach lacks the precision
21
22 317 of individual chart audits(6), it is both clinically plausible and scalable in that it can be used to
23
24 318 monitor the care of large numbers of patients, making the method itself an important contribution
25
26 319 to the literature on diagnostic error. Although we did not directly measure diagnostic errors
27
28 320 through chart audits, the fact that we found associations with diagnostic knowledge and the
29
30 321 diagnostic error sensitive outcome conditions we studied coupled with the fact that we did not
31
32 322 find associations with treatment knowledge, nor did we find associations when the underlying
33
34 323 diagnostic error sensitive condition was likely not present during the outpatient index visit
35
36 324 because no antecedent diagnoses recorded indicates that the associations we report in this study
37
38 325 were likely driven by association with diagnostic errors that occurred during these visits.
39
40
41 326 Furthermore, our approach builds on prior studies that used claims data to infer diagnostic error
42
43 327 incidence for ED visits, in that we identified index visit diagnoses at risk for diagnostic error that
44
45 328 were clinically plausible and verified empirically, and we assured that we were studying new
46
47 329 problems by requiring that the patient not have had a visit over the previous 3 months
48
49 330 contacts.(28-30) We expanded on these studies by focusing on outpatient care and by examining
50
51 331 a much more comprehensive set of presenting complaints that may have been precursors to one
52
53
54
55
56
57
58
59
60

1
2
3 332 of 13 diagnostic error prone conditions that we studied. This approach was necessary in order to
4
5 333 study diagnostic error in the more low acuity setting of outpatient general internal medicine.
6
7

8 334
9
10
11 335 Our findings suggest an association between diagnostic knowledge and adverse outcomes. Yet,
12
13 336 there are important limitations to consider. We did not directly determine whether a diagnostic
14
15 337 error had occurred through chart review. Because our analyses were cross sectional, we cannot
16
17 338 rule out the possibility that observed associations were the result of omitted variable bias related
18
19 339 to either physician or patient characteristics, and do not reflect a causal relationship between
20
21 340 diagnostic knowledge and adverse outcomes. That said, there is no reason to believe that these
22
23 341 characteristics would be correlated with diagnostic knowledge independent of treatment
24
25 342 knowledge, which we were able to control for. Furthermore, had associations with diagnostic
26
27 343 knowledge been driven by omitted variable bias then we would have expected them to be similar
28
29 344 when estimated across index visits with lower or higher risk for diagnostic error, and they were
30
31 345 not. We also found that diagnosis exam performance was not associated with elective
32
33 346 hospitalizations, which are, presumably, unrelated to underlying diagnostic knowledge but may
34
35 347 be related to the overall propensity to hospitalize. Additional limitations include the fact that we
36
37 348 studied select conditions among older patients enrolled in the Medicare program so we cannot
38
39 349 extrapolate these findings to a younger population, other conditions we did not consider, or
40
41 350 populations with no or different health insurance coverage. Finally, diagnostic error may also
42
43 351 stem from factors outside of inadequate diagnostic knowledge, which are likely not represented
44
45 352 by test exam scores but could be correlated with diagnostic knowledge (e.g., poor
46
47 353 patient/physician communication skills and related system failures).(31, 32) That said, there is no
48
49 354 reason to believe that these other contributors to diagnostic error would not also be correlated
50
51
52
53
54
55
56
57
58
59
60

1
2
3 355 with the other aspects of the exam we do account for. Furthermore, based on an analysis of
4
5 356 malpractice claims, Newman-Toker et al. (7) reported that clinical judgement played an
6
7 357 important role in 86% of diagnostic errors, while poor patient/physician communication and
8
9 358 system failures played a role in far fewer diagnostic errors that resulted in malpractice suits (35%
10
11 359 and 22% respectively). Suggesting that improving communication will not reduce stroke related
12
13 360 diagnostic error, Kerber et al. (33) reported that frontline providers rarely ask the right questions
14
15 361 when patients present with dizziness. Communication ability is only valuable in terms of
16
17 362 reducing diagnostic error if the physician knows what questions to ask and what the answers
18
19 363 mean. Although we cannot say with certainty that our finding are driven by an underlying
20
21 364 association between diagnostic knowledge and diagnostic errors, at a minimum, our finding
22
23 365 suggest that diagnostic knowledge may be particularly important in terms avoiding these adverse
24
25 366 outcome at heightened risk for diagnostic errors.
26
27
28
29
30
31
32
33

34 368 We found that diagnostic knowledge at the point of primary care is a risk factor for outcomes at
35
36 369 heightened risk for diagnostic error. The fact that there exists a link between general diagnostic
37
38 370 knowledge and diagnostic error may not be surprising, the magnitude of the associations we
39
40 371 found suggests that interventions ignoring the role of physician knowledge maybe inadequate to
41
42 372 address the crisis of diagnostic error. Interventions targeted at improving diagnostic knowledge
43
44 373 could include such things as a greater focus on diagnostic training during graduate medical
45
46 374 education (i.e., medical school, residency, and fellowship). Knowledge-focused interventions
47
48 375 could also include incentivizing broad-based learning as well as targeted learning pursued
49
50 376 through continuing medical education (CME) activities (see Newman-Toker and McKay for a
51
52 377 similar observation(28)). During visits identified as being at risk for diagnostic errors,
53
54
55
56
57
58
59
60

1
2
3 378 physicians could be given related information at the point of care. ABIM and other certifying
4
5 379 organizations could also enhance awareness around deficiencies in diagnostic knowledge by
6
7 380 providing physicians with diagnosis specific exam performance feedback as well as resources to
8
9 381 improve their diagnostic knowledge. Physicians who have demonstrated heightened diagnostic
10
11 382 expertise might be utilized when patients present with symptoms at heightened risk for
12
13
14 383 diagnostic error.
15
16
17
18 384

19
20 385 In conclusion, gaps in diagnostic knowledge between first contact primary care physicians is an
21
22 386 important risk factor for serious diagnostic error sensitive outcomes and therefore should be a
23
24 387 target for interventions to reduce diagnostic errors.
25
26
27
28 388
29
30
31 389

390

Statements

391 A. Contribution statement: Design and conduct of the study; collection, management,
392 analysis, and interpretation of the data; and preparation, review, or approval of the
393 manuscript; and decision to submit the manuscript for publication were all conducted by
394 the authors independently of the American Board of Internal Medicine.

395 B. Conflicts: Bradley Gray, Jonathan Vandergrift, and Rebecca Lipner are paid employees of
396 the American Board of Internal Medicine. Bruce Landon is a paid consultants for the
397 American Board of Internal Medicine.

398 C. Funding: Financial and material support was provided by the American Board of Internal
399 Medicine.

400 D. Data Sharing: Administrative data describing physician characteristics and exam
401 performance can be obtain from the ABIM through a data sharing agreement that assures
402 physician confidentiality and its used for legitimate research purposes. Access to de-
403 identified Medicare claims data for this study were obtained through a special data use
404 agreement with the Centers for Medicare and Medicaid services which is a process
405 available to researchers in the US.

406 E. Patient and public statement: Patients and the public were not involved in the design or
407 execution of this study as the existing patient claims data used were de-identified by the
408 Center for Medicaid and Medicare Services prior to analysis. In terms of dissemination,
409 ABIM's communication department in collaboration with the authors of this study we write
410 a press release whose goal is to inform the public regarding the finding of the study.

411

412 **References**

- 413 1. National Academies of Sciences, Engineering, and Medicine. Improving Diagnosis in Health Care.
414 Washington, DC: The National Academies Press 2015 December 10, 2018.
- 415 2. Graber ML, Trowbridge R, Myers JS, et al. The next organizational challenge: finding and
416 addressing diagnostic error. *Jt Comm J Qual Patient Saf.* 2014;40(3):102-10.
- 417 3. Cresswell KM, Panesar SS, Salvilla SA, et al. Global research priorities to better understand the
418 burden of iatrogenic harm in primary care: an international Delphi exercise. *PLoS Med.*
419 2013;10(11):e1001554.
- 420 4. Kostopoulou O, Delaney BC, Munro CW. Diagnostic difficulty and error in primary care--a
421 systematic review. *Fam Pract.* 2008;25(6):400-13.
- 422 5. Goyder CR, Jones CH, Heneghan CJ, et al. Missed opportunities for diagnosis: lessons learned
423 from diagnostic errors in primary care. *Br J Gen Pract.* 2015;65(641):e838-44.
- 424 6. Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care:
425 estimations from three large observational studies involving US adult populations. *BMJ Qual Saf.*
426 2014;23(9):727-31.
- 427 7. Newman-Toker DE, Schaffer AC, Yu-Moe CW, et al. Serious misdiagnosis-related harms in
428 malpractice claims: The "Big Three" - vascular events, infections, and cancers. *Diagnosis (Berl).*
429 2019;6(3):227-40.
- 430 8. Gandhi TK, Kachalia A, Thomas EJ, et al. Missed and delayed diagnoses in the ambulatory
431 setting: a study of closed malpractice claims. *Ann Intern Med.* 2006;145(7):488-96.
- 432 9. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med.*
433 2005;165(13):1493-9.
- 434 10. Kachalia A, Gandhi TK, Puopolo AL, et al. Missed and delayed diagnoses in the emergency
435 department: a study of closed malpractice claims from 4 liability insurers. *Ann Emerg Med.*
436 2007;49(2):196-205.
- 437 11. Poon EG, Kachalia A, Puopolo AL, et al. Cognitive errors and logistical breakdowns contributing
438 to missed and delayed diagnoses of breast and colorectal cancers: a process analysis of closed
439 malpractice claims. *J Gen Intern Med.* 2012;27(11):1416-23.
- 440 12. Chisholm A, Askham J. A review of professional codes and standards for doctors in the UK, USA
441 and Canada: Picker Institute Europe Oxford; 2006.
- 442 13. Irvine D. Doctors in the UK: their new professionalism and its regulatory framework. *Lancet.*
443 2001;358(9295):1807-10.
- 444 14. Kovacs E, Schmidt AE, Szocska G, et al. Licensing procedures and registration of medical doctors
445 in the European Union. *Clinical Medicine.* 2014;14(3):229-38.
- 446 15. European Union of Medical Specialists. The european council for accreditation of medical
447 specialist qualifications (ECAMSQ) 2010 [Available from:
448 https://www.uems.eu/_data/assets/pdf_file/0009/1206/ECAMSQ_presentation.pdf.
- 449 16. Petterson SM, Liaw WR, Phillips RL, Jr., et al. Projecting US primary care physician workforce
450 needs: 2010-2025. *Ann Fam Med.* 2012;10(6):503-9.
- 451 17. Freed GL, Dunham KM, Singer D. Use of board certification and recertification in hospital
452 privileging: policies for general surgeons, surgical specialists, and nonsurgical subspecialists. *Arch Surg.*
453 2009;144(8):746-52.
- 454 18. Independence Blue Cross. Credentialing criteria 2019 [Available from:
455 https://www.ibx.com/pdfs/providers/interactive_tools/credentialing_criteria_ibx.pdf.
- 456 19. American Board of Medical Specialties. Certification Matters FAQs 2019 [Available from:
457 <https://www.certificationmatters.org/faqs/>.

- 1
2
3 458 20. Lipner RS, Bylsma WH, Arnold GK, et al. Who is maintaining certification in internal medicine--
4 459 and why? A national survey 10 years after initial certification. *Ann Intern Med.* 2006;144(1):29-36.
5 460 21. Balogh E, Miller BT, Ball J, et al. Improving diagnosis in health care. xxvii, 444 pages p.
6 461 22. Centers for Medicare & Medicaid Services. BPCI Advanced 2018 [Available from:
7 462 <https://innovation.cms.gov/initiatives/bpci-advanced>.
8 463 23. Centers for Medicare & Medicaid Services. Comprehensive Care for Joint Replacement Model
9 464 2018 [Available from: <https://innovation.cms.gov/initiatives/cjr>.
10 465 24. Schiff GD, Hasan O, Kim S, et al. Diagnostic error in medicine: analysis of 583 physician-reported
11 466 errors. *Arch Intern Med.* 2009;169(20):1881-7.
12 467 25. Lipner RS, Lucey CR. Putting the secure examination to the test. *JAMA.* 2010;304(12):1379-80.
13 468 26. Huber PJ. The behavior of maximum likelihood estimates under non-standard conditions. *Paper*
14 469 *presented at: Fifth Berkeley symposium on mathematical statistics and probability; Berkeley, CA.* 1967.
15 470 27. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for
16 471 heteroskedasticity. *Econometrica: Journal of the Econometric Society.* 1980:817-38.
17 472 28. Newman-Toker DE, Makary MA. Measuring Diagnostic Errors in Primary Care: The First Step on a
18 473 Path Forward Comment on "types and Origins of Diagnostic Errors in Primary Care Settings". *JAMA*
19 474 *internal medicine.* 2013;173(6):425-6.
20 475 29. Waxman DA, Kanzaria HK, Schriger DL. Unrecognized Cardiovascular Emergencies Among
21 476 Medicare Patients. *JAMA Intern Med.* 2018;178(4):477-84.
22 477 30. Liberman AL, Newman-Toker DE. Symptom-Disease Pair Analysis of Diagnostic Error (SPADE): a
23 478 conceptual framework and methodological approach for unearthing misdiagnosis-related harms using
24 479 big data. *BMJ Qual Saf.* 2018;27(7):557-66.
25 480 31. Davis Giardina T, King BJ, Ignaczak AP, et al. Root cause analysis reports help identify common
26 481 factors in delayed diagnosis and treatment of outpatients. *Health Aff (Millwood).* 2013;32(8):1368-75.
27 482 32. Zwaan L, Monteiro S, Sherbino J, et al. Is bias in the eye of the beholder? A vignette study to
28 483 assess recognition of cognitive biases in clinical case workups. *BMJ Qual Saf.* 2017;26(2):104-10.
29 484 33. Kerber KA, Newman-Toker DE. Misdiagnosing Dizzy Patients: Common Pitfalls in Clinical
30 485 Practice. *Neurol Clin.* 2015;33(3):565-75, viii.

486

487 **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)	Hospitalization ^{a,b}	Emergency department visit ^a	Death ^c
	Number (percent of index visits)	Number (percent of hospitalizations with a diagnostic error sensitive condition)	Number (percent of emergency department visits with a diagnostic error sensitive condition)	Number (percent 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 ^d	Not Reported ^d	121 (38.3)
Anemia	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	8,534 (17.5)	12 (2.2)	13 (2.0)	89 (28.2)
Spinal Cord Compression	6,386 (13.1)	Not Reported ^d	Not Reported ^d	36 (11.4)
Bacteremia / Sepsis	5,567 (11.4)	19 (3.5)	21 (3.2)	46 (14.6)
Appendicitis	2,584 (5.3)	Not Reported ^d	Not Reported ^d	17 (5.4)
Abscess	1,005 (2.1)	Not Reported ^d	13 (2.0)	Not Reported ^d

488 ^aCondition specific outcomes for one of the 13 diagnostic error sensitive conditions within 90 days of an
 489 outpatient index visit at risk for that condition

490 ^bHospitalizations include non-elective hospitalizations either initiated through the ED or a trauma center.

491 ^cAll cause mortality within 90 days of the index visit.

492 ^dNot reported because observations were less than 11.

493

494

495 **Table 2. Physician and Patient Characteristics by Diagnostic Exam Performance Tertile**

	Total	Diagnosis question percent correct			P-value ^a
		Top (78.5 to 95.8)	Middle (71.4 to 78.4)	Bottom (42.9 to 71.3)	
Exam performance, Mean (standard deviation)^a					
Diagnosis question percent correct	74.5 (0.4)	84.3 (0.3)	74.8 (0.1)	65.5 (0.3)	<.001
Other question percent correct	72.6 (0.7)	80.2 (1.0)	72.1 (1.1)	66.4 (1.5)	<.001
Treatment question percent correct	77.3 (0.3)	83.4 (0.4)	77.2 (0.4)	72.0 (0.5)	<.001
Physician Characteristics, count (%)					
Female Physician	19,428 (39.9)	6,546 (43.8)	6,357 (37.5)	6,525 (39.0)	0.37
US born physician	28,462 (58.5)	9,284 (62.1)	9,932 (58.6)	9,246 (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	9,452 (19.4)	1,914 (12.8)	3,462 (20.4)	4,076 (24.4)	0.009
Small group practice (2 to 10)	20,563 (42.3)	5,543 (37.1)	7,529 (44.4)	7,491 (44.8)	0.19
Medium physicians group practice (11 to 50)	7,442 (15.3)	2,899 (19.4)	2,402 (14.2)	2,141 (12.8)	0.25
Large physician group practice (>50 physicians)	5,391 (11.1)	2,150 (14.4)	1,655 (9.8)	1,586 (9.5)	0.14
Academic practice	2,708 (5.6)	1,447 (9.7)	697 (4.1)	564 (3.4)	<.001
Other practice	3,076 (6.3)	1,005 (6.7)	1,211 (7.1)	860 (5.1)	0.59
Beneficiary characteristics					
Beneficiary Race, count (percent)					
White	40,086 (82.4)	12,652 (84.6)	13,778 (81.3)	13,656 (81.7)	0.13
Black	3,958 (8.1)	926 (6.2)	1,609 (9.5)	1,423 (8.5)	0.03
Other	4,588 (9.4)	1,380 (9.2)	1,569 (9.3)	1,639 (9.8)	0.88
Beneficiary age (per year), Mean (SD) ^a	76.6 (0.1)	76.8 (0.1)	76.5 (0.1)	76.6 (0.1)	0.23
CCW chronic conditions, count (percent)					
Alzheimer's Disease and Related Disorders or Senile Dementia	5,151 (10.6)	1,497 (10.0)	1,793 (10.6)	1,861 (11.1)	0.16
Alzheimer's Disease	2,061 (4.2)	627 (4.2)	704 (4.2)	730 (4.4)	0.82
Acute Myocardial Infarction	1,408 (2.9)	394 (2.6)	494 (2.9)	520 (3.1)	0.13
Anemia	22,450 (46.2)	6,706 (44.8)	7,766 (45.8)	7,978 (47.7)	0.11
Asthma	4,424 (9.1)	1,313 (8.8)	1,548 (9.1)	1,563 (9.3)	0.39
Atrial Fibrillation	4,225 (8.7)	1,265 (8.5)	1,478 (8.7)	1,482 (8.9)	0.69
Breast Cancer	2,485 (5.1)	779 (5.2)	831 (4.9)	875 (5.2)	0.48
Colorectal Cancer	1,139 (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	1,662 (3.4)	507 (3.4)	600 (3.5)	555 (3.3)	0.66
Cataract	31,095 (63.9)	9,601 (64.2)	10,773 (63.5)	10,721 (64.1)	0.74
Heart Failure	9,207 (18.9)	2,786 (18.6)	3,155 (18.6)	3,266 (19.5)	0.54
Chronic Kidney Disease	6,904 (14.2)	2,083 (13.9)	2,392 (14.1)	2,429 (14.5)	0.62
Chronic Obstructive Pulmonary Disease	9,108 (18.7)	2,635 (17.6)	3,165 (18.7)	3,308 (19.8)	0.02
Depression	12,042 (24.8)	3,728 (24.9)	4,145 (24.4)	4,169 (24.9)	0.83
Diabetes	13,296 (27.3)	3,947 (26.4)	4,590 (27.1)	4,759 (28.5)	0.16
Glaucoma	10,030 (20.6)	3,086 (20.6)	3,501 (20.6)	3,443 (20.6)	0.99
Hip/Pelvic Fracture	1,531 (3.1)	430 (2.9)	535 (3.2)	566 (3.4)	0.15
Hyperlipidemia	37,132 (76.4)	11,266 (75.3)	12,898 (76.1)	12,968 (77.6)	0.11
Benign Prostatic Hyperplasia	5,815 (12.0)	1,792 (12.0)	1,987 (11.7)	2,036 (12.2)	0.76
Hypertension	37,607 (77.3)	11,345 (75.8)	13,011 (76.7)	13,251 (79.3)	<.001
Hypothyroidism	11,425 (23.5)	3,490 (23.3)	3,862 (22.8)	4,073 (24.4)	0.25

Ischemic Heart Disease	18,713 (38.5)	5,616 (37.5)	6,393 (37.7)	6,704 (40.1)	0.06
Osteoporosis	14,171 (29.1)	4,372 (29.2)	4,794 (28.3)	5,005 (29.9)	0.34
Rheumatoid Arthritis	23,352 (48.0)	6,879 (46.0)	8,275 (48.8)	8,198 (49.0)	0.02
Stroke	6,255 (12.9)	1,880 (12.6)	2,212 (13.0)	2,163 (12.9)	0.70
Number of chronic conditions, count (percent)					
<=4	5,066 (10.4)	1,459 (9.8)	1,744 (10.3)	1,863 (11.1)	0.08
5 to 7	16,861 (34.7)	5,392 (36.0)	5,981 (35.3)	5,488 (32.8)	0.006
8 to 10	16,230 (33.4)	4,907 (32.8)	5,664 (33.4)	5,659 (33.8)	0.35
>=11	10,475 (21.5)	3,200 (21.4)	3,567 (21.0)	3,708 (22.2)	0.28
Mental health visit, count (percent)	6,347 (13.1)	2,040 (13.6)	2,119 (12.5)	2,188 (13.1)	0.46
Hierarchical Condition Category (HCC) score, Mean (SD) ^a	0.98 (0.01)	0.96 (0.01)	0.98 (0.01)	0.99 (0.01)	0.19
Household medium income, mean \$ (SD) ^a	59,852 (643)	61,574 (1,106)	59,113 (1,144)	59,063 (1,075)	0.19
Medicaid dual eligible, count (percent)	6,392 (13.1)	1,793 (12.0)	2,411 (14.2)	2,188 (13.1)	0.28
Rural county residence, count (percent)	7,392 (15.2)	2,207 (14.8)	2,866 (16.9)	2,319 (13.9)	0.64
Visit characteristics					
Visit with same doctor in last year, Count (percent)	37,726 (77.6)	11,369 (76.0)	13,154 (77.6)	13,203 (79.0)	0.08
Visit with any physician in last year, count (percent)	44,852 (92.2)	13,711 (91.7)	15,647 (92.3)	15,494 (92.7)	0.08
Days since last visit with any physician (if any visit in last year), Mean (SD) ^a	144.2 (0.6)	147.1 (0.8)	144.4 (1.0)	141.4 (1.3)	<.001
ED visit in prior year, count (percent)	8,101 (16.7)	2,428 (16.2)	2,879 (17.0)	2,794 (16.7)	0.43
Days since last ED visits (if ED visit in last year), Mean (SD) ^a	222.8 (0.9)	221.2 (1.5)	223.5 (1.5)	223.4 (1.5)	0.47
Hospitalization in prior year, Count (percent)	4,227 (8.7)	1,280 (8.6)	1,489 (8.8)	1,458 (8.7)	0.85
Days since last hospitalization (if hospitalization in last year), Mean (SD) ^a	229.6 (1.2)	229.1 (2.1)	229.7 (2.1)	230.1 (1.9)	0.95
Index visit diagnosis groups, Count (percent)					
Abscess	1,005 (2.1)	268 (1.8)	394 (2.3)	343 (2.1)	0.21
Anemia	12,410 (25.5)	3,817 (25.5)	4,369 (25.8)	4,224 (25.3)	0.93
Aortic aneurysm	11,491 (23.6)	3,495 (23.4)	4,165 (24.6)	3,831 (22.9)	0.18
Appendicitis	2,584 (5.3)	845 (5.6)	949 (5.6)	790 (4.7)	0.01
Bacteremia	5,567 (11.4)	1,660 (11.1)	1,929 (11.4)	1,978 (11.8)	0.83
Congestive heart failure	12,137 (25.0)	3,633 (24.3)	4,221 (24.9)	4,283 (25.6)	0.67
Acute coronary syndrome	16,228 (33.4)	4,627 (30.9)	5,740 (33.9)	5,861 (35.1)	0.02
Depression	12,637 (26.0)	3,932 (26.3)	4,312 (25.4)	4,393 (26.3)	0.78
Fracture	13,409 (27.6)	4,324 (28.9)	4,364 (25.7)	4,721 (28.2)	0.11
Pulmonary embolism	8,534 (17.5)	2,683 (17.9)	2,984 (17.6)	2,867 (17.1)	0.71
Pneumonia	12,183 (25.1)	3,773 (25.2)	4,224 (24.9)	4,186 (25.0)	0.97
Spinal cord compression	6,386 (13.1)	1,985 (13.3)	2,218 (13.1)	2,183 (13.1)	0.94
Stroke	10,026 (20.6)	3,003 (20.1)	3,542 (20.9)	3,481 (20.8)	0.79

496 ^aP-values and standard deviation accounted for correlated errors within physicians

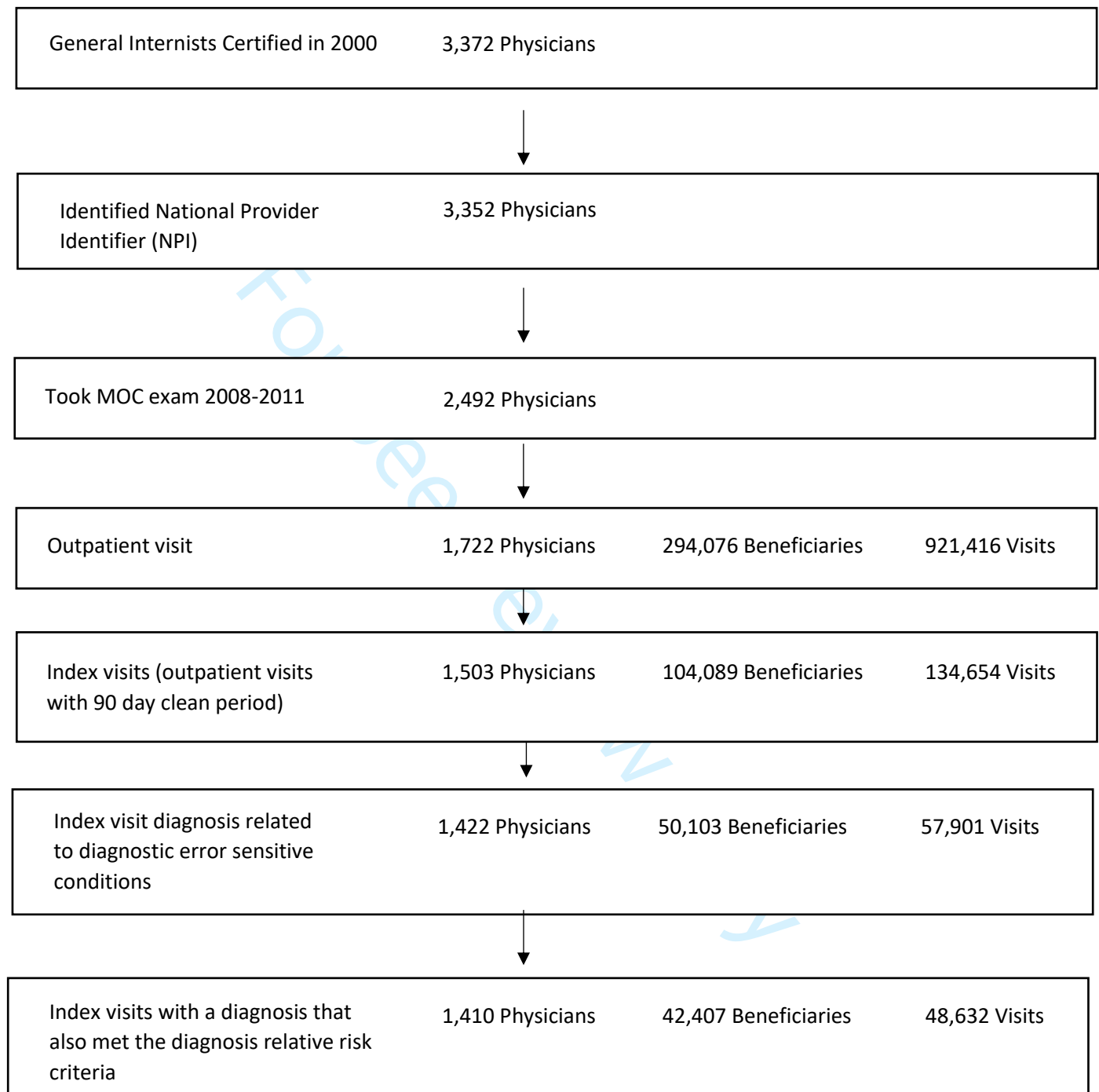
Table 3. Associations with diagnostic knowledge and adverse events per 1,000 index visits

Diagnostic knowledge tertile	Death ^a				Emergency department visit ^b				Hospitalization ^c			
	Unadjusted ^e	Regression adjusted ^{d,e}			Unadjusted ^e	Regression adjusted ^{d,e}			Unadjusted ^e	Regression adjusted ^{d,e}		
	Events per 1,000 visits (95% CI interval)	Events per 1,000 visits (95% CI interval)	Difference (95% CI)	P-value	Events per 1,000 visits (95%CI)	Events per 1,000 visits (95%CI)	Difference (95% CI)	P-value	Events per 1,000 visits (95% CI)	Events per 1,000 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	

^aAll cause mortality within 90 days of an outpatient index visit with a diagnosis at risk for one of 3 diagnostic error sensitive conditions.

^bEmergency 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.

^cHospitalizations were for non-elective hospitalizations either initiated through the ED or a trauma center 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.

Figure 1. Sample Selection

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Appendix

Diagnostic Knowledge, Diagnostic Error and Risk of Death, Hospitalization and Emergency Department Visits

Contents	
Section: Description	Page
Section 1: 90 day Clean Period Derivation	2
Section 2: Outcome Condition and Index Visit Eligibility Diagnoses Codes and Relative Risk.....	5
Section 3: Psychometric Analysis of Whether Diagnosis Related Questions Reflect an Underlying Construct	10
Section 4: Imputations for missing variables.....	11
Section 5: Regression Sensitivity Analyses.....	12
Section 6: Full Regression Coefficient Estimates and Explanatory Variable List	18
References	23

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

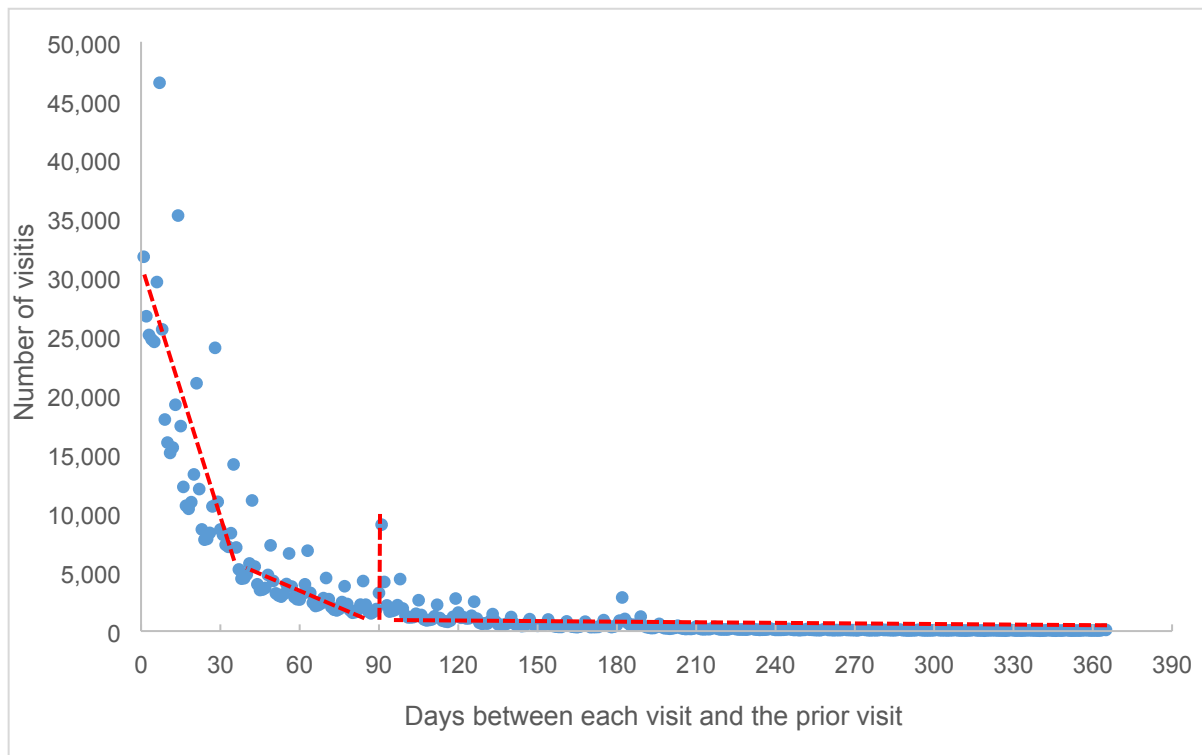
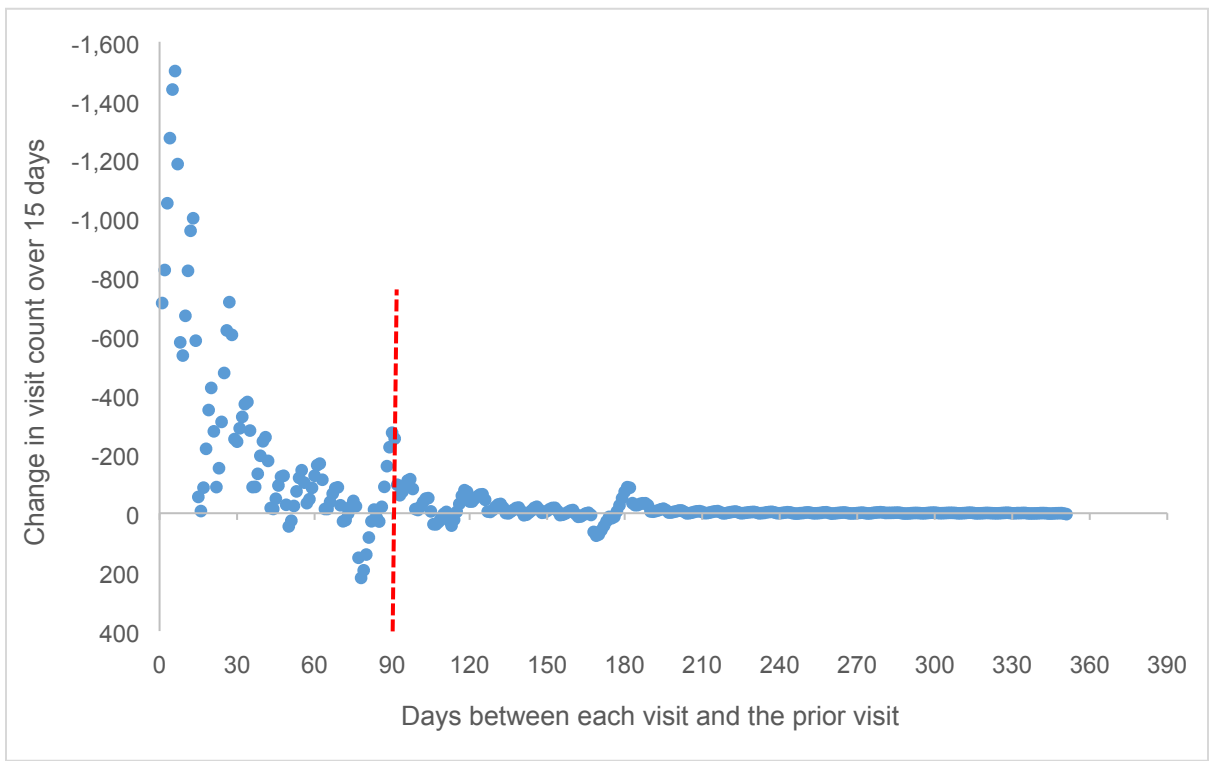


Figure 1.2. Average Change in Visit Count over the 15 days (15-day slope) Following each Data Point Listed in Figure 1



Review only

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	Outcome conditions ^a	Index visit eligibility diagnosis group	Relative Risk ^b	
Abscess	Fever	2.65	Acute Coronary Syndrome	Chest pain	8.38	
	Chills	0.00		Dyspnea	7.29	
Anemia	Gastrointestinal bleeding	25.20		Shortness of breath	3.65	
	Weight loss	4.09		Hypoxemia/hypoxia	2.01	
	Shortness of breath	3.51		Reflux disease	1.23	
	Weakness/Fatigue	2.35		Esophageal disease	1.22	
	Hypoxemia/Hypoxia	2.11		Weakness/Fatigue	1.14	
	Dyspnea	2.05		Nausea	1.05	
	Chest Pain	1.82		Other respiratory issue	0.86	
	Headache	1.29		Respiratory distress	0.00	
Aortic Aneurysm	Menorrhagia	0.00		Gastritis	0.00	
	Dyspnea	4.98		Heart Burn	0.00	
	Abdominal pain	4.93		Depression	Delirium	32.76
	Shortness of breath	3.80			Heart failure	6.16
	Chest pain	2.42	Anxiety		5.04	
	Other back pain	1.64	Dysthymia		4.99	
	Back pain	1.01	Weight loss		4.73	
Elevated blood pressure	0.00	Anemia	2.74			
		Fatigue	1.06			
Appendicitis	Vomiting	30.79	Alcohol	0.00		
	Diverticulitis	30.45	Amphetamines	0.00		
	Nausea	16.81	Cannabis	0.00		
	Abdominal pain	15.60	Cocaine	0.00		
	Abdominal tenderness	0.00	Drug Mental Disease	0.00		
Bacteremia/Sepsis	Fever	0.00	Hallucinogens	0.00		
	Vomiting	6.99	Opioids	0.00		
	Fever	5.10	Personality disorder	0.00		
	Nausea	3.82	related alcohol disease	0.00		
	Tachycardia	2.67	Sedatives	0.00		
Heart failure	Weakness/Fatigue	1.75	Stress	0.00		
	Hypoxemia/Hypoxia	9.99	Weight gain	0.00		
	Shortness of breath	5.09	Mood disorder	0.00		
	Dyspnea	3.33	Fracture	Gait instability	2.53	
	Edema	3.27		Edema	1.79	
	Chest Pain	2.46		Osteoporosis	1.66	
	Weakness/Fatigue	1.42		Hyperparathyroidism	1.09	
	Ascites	0.00		Vitamin D deficiency	1.08	
Respiratory Distress	0.00					

Outcome conditions ^a	Index visit eligibility diagnosis group	Relative Risk ^b	Outcome conditions ^a	Index visit eligibility diagnosis group	Relative Risk ^b
Fracture (con't)	Osteopenia	0.54	Spinal cord compression	Abdominal pain	31.20
	Celiac disease	0.00		Back pain	15.03
	Falls	0.00		Peripheral neuropathy	0.00
	Stress fracture	0.00		Weakness/Fatigue	0.00
Pulmonary embolism	Tachycardia	12.16	Stroke	Facial weakness	65.24
	Hypoxemia/hypoxia	10.98		Confusion	48.93
	Shortness of breath	6.75		Speech disturbance	19.60
	Dyspnea	6.54		Transient ischemic attack	7.82
	Abnormal respiration	6.35		Delirium	4.96
	Heart failure	4.51		Dizziness	3.20
	Chest pain	4.31		Lack coordination	2.92
	Cough	1.48		Gait instability	2.92
	Other respiratory issue	1.34		Vomiting	2.15
	Deep vein thrombosis	0.00		Weakness/Fatigue	1.54
	Respiratory distress	0.00		Headache	1.37
	Fever	0.00		Nausea	1.17
	Heart burn	0.00		Thunderclap headache	0.00
	Hemoptysis	0.00		Tension headache	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				

^aOutcomes include any hospitalization or emergency department visit for the diagnostic conditions within 90 days of the index visits including same day events

^bIndex 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.

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

For peer review only

Section 5. 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 5.1. 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 5.2.1 (for death), 5.2.2 (hospitalization) to 6.2.3 (for emergency department visit).

1
2
3 For the first sensitivity we expanding the eligible diagnoses code groups to all 76 identified by
4 physician authors versus 38 in the base case that also met the relative risk criteria.
5
6

7
8 For the third sensitivity we expand the index visit clean period to 97 days and contracted the
9 index visit clean period to 83 days.
10
11

12
13 For the fourth sensitivity, we excluded physician in academic medical centers to consider the
14 possibility that the unobserved physician characteristics related to where they worked or who
15 they worked with could be were independently both related to the underlying physician
16 diagnostic skill and our outcome measures.
17
18

19
20 For the fifth sensitivity we accounted for the possibility that adverse outcomes were avoided
21 because the patient died by altering the ED and hospitalization measures to include all-cause
22 mortality. For this sensitivity we added the following two outcome measures: base case
23 hospitalization or death and base case ED or death.
24
25

26
27
28 Overall results of robustness sensitivity analysis suggests that our base case results were not
29 highly sensitive to different underlying assumptions related to these factors (e.g., across all
30 robustness sensitivities percent change in the outcome measures between top versus bottom
31 diagnostic knowledge exam performers remained statistically significant ($P<0.05$)).
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 5.1. 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 5.2.1. 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
Exclusion of visits with physicians working in academic medical centers	45,924	5.1 (4.0 to 6.3)	6.5 (5.4 to 7.6)	8.0 (6.4 to 9.5)	-35.5 (-53.1 to -11.2)	-2.8 (-4.9 to -0.8)	0.008	-18.2 (-36.7 to 5.7)	-1.5 (-3.3 to 0.4)	0.13

Table 5.2.2. 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
Exclusion of visits with physicians working in academic medical centers	45,924	9.0 (7.4 to 10.7)	10.8 (9.2 to 12.4)	12.8 (10.8 to 14.9)	-29.4 (-45.9 to -7.9)	-3.8 (-6.7 to -0.9)	0.01	-15.9 (-32.7 to 5.1)	-2.0 (-4.7 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

Table 5.2.3. 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
Exclusion of visits with physicians working in academic medical centers	45,924	11.2 (9.4 to 13.0)	13.0 (11.2 to 14.8)	16.1 (13.7 to 18.4)	-30.5 (-45.4 to -11.4)	-4.9 (-8.2 to -1.6)	0.004	-19.3 (-34.3 to -0.7)	-3.1 (-6.1 to -0.1)	0.05
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

Section 6. Full Regression Coefficient Estimates and Explanatory Variables List

eTables 6.1 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 6.1. Probit Coefficient Associations and Regression Descriptive Statistics

Label	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	
	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

1							
2							
3	Asthma	-0.8 (1.1)	0.45	-0.2 (1.4)	0.88	0.1 (1.7)	0.95
4	Atrial Fibrillation	1.8 (1.1)	0.10	3.4 (1.5)	0.02	3.7 (1.7)	0.03
5	Breast Cancer	-2.0 (1.3)	0.13	-3.0 (1.8)	0.10	-2.0 (2.2)	0.35
6	Colorectal Cancer	-0.5 (1.7)	0.76	-3.4 (2.0)	0.10	-1.9 (2.5)	0.44
7	Endometrial Cancer	1.6 (4.2)	0.71	-1.1 (4.7)	0.82	-0.9 (5.5)	0.87
8	Lung Cancer	3.7 (3.4)	0.28	13.3 (6.2)	0.03	10.3 (6.2)	0.10
9	Prostate Cancer	-1.6 (1.4)	0.26	4.3 (2.5)	0.09	3.6 (2.8)	0.20
10	Cataract	-1.0 (0.9)	0.29	1.1 (1.0)	0.31	0.1 (1.2)	0.96
11	Heart Failure	1.8 (1.0)	0.08	2.9 (1.2)	0.02	3.3 (1.4)	0.02
12	Chronic Kidney Disease	-0.7 (0.8)	0.39	2.9 (1.2)	0.02	4.2 (1.4)	0.004
13	Chronic Obstructive Pulmonary Disease	1.1 (0.9)	0.26	1.3 (1.2)	0.27	0.9 (1.3)	0.50
14	Depression	0.2 (0.9)	0.79	0.9 (1.1)	0.43	0.7 (1.2)	0.59
15	Diabetes	0.0 (0.8)	0.99	3.3 (1.1)	0.003	2.4 (1.2)	0.04
16	Glaucoma	-0.9 (0.8)	0.29	0.0 (1.1)	1.00	-0.3 (1.2)	0.80
17	Hip/Pelvic Fracture	1.4 (1.6)	0.39	3.0 (2.4)	0.20	5.0 (2.8)	0.07
18	Hyperlipidemia	-1.8 (1.1)	0.09	-1.1 (1.3)	0.39	-0.7 (1.4)	0.63
19	Benign Prostatic Hyperplasia	-0.3 (1.1)	0.79	0.0 (1.5)	0.98	-0.8 (1.7)	0.63
20	Hypertension	0.8 (1.1)	0.46	2.4 (1.4)	0.09	1.2 (1.6)	0.44
21	Hypothyroidism	0.7 (0.8)	0.42	1.4 (1.1)	0.19	1.8 (1.2)	0.14
22	Ischemic Heart Disease	0.8 (0.9)	0.39	0.1 (1.1)	0.92	-1.3 (1.2)	0.29
23	Osteoporosis	-1.9 (0.8)	0.02	2.0 (1.2)	0.09	1.4 (1.3)	0.28
24	Rheumatoid Arthritis	-2.3 (0.8)	0.005	-0.2 (1.0)	0.87	0.3 (1.1)	0.81
25	Stroke	2.7 (1.1)	0.02	2.7 (1.3)	0.04	2.8 (1.5)	0.06
26	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
27	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
28	Hospitalization in prior year	8.8 (5.4)	0.10	0.9 (4.1)	0.84	0.4 (4.5)	0.93
29	ED visit in prior year	1.0 (2.5)	0.69	7.3 (4.0)	0.07	8.4 (4.6)	0.07
30	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
31	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
32	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
33	Index visit diagnosis group indicators						
34	Pulmonary embolism	-0.5 (1.3)	0.68	6.1 (1.4)	<.001	7.1 (1.6)	<.001
35	Acute coronary syndrome	-1.3 (1.1)	0.25	-3.0 (1.8)	0.11	-5.5 (2.0)	0.007
36	Stroke	-2.3 (1.4)	0.10	7.4 (1.5)	<.001	7.7 (1.7)	<.001
37	Congestive heart failure	0.2 (1.3)	0.88	10.1 (1.6)	<.001	12.1 (1.7)	<.001
38	Fracture	-1.3 (1.1)	0.22	3.2 (1.3)	0.02	5.1 (1.5)	<.001
39	Abscess	0.7 (2.3)	0.77	6.4 (3.3)	0.05	11.7 (3.3)	<.001
40	Pneumonia	2.3 (1.2)	0.05	5.6 (1.4)	<.001	6.7 (1.6)	<.001
41	Aortic aneurysm	1.0 (1.4)	0.50	-0.6 (2.0)	0.76	0.7 (2.2)	0.74
42	Appendicitis	2.0 (1.8)	0.28	5.9 (3.0)	0.05	9.6 (3.1)	0.002
43	Depression	0.0 (1.3)	0.99	3.0 (1.5)	0.05	2.4 (1.7)	0.15
44	Anemia	2.3 (1.1)	0.04	3.5 (1.8)	0.04	3.2 (2.0)	0.11
45	Bacteremia	0.5 (2.5)	0.85	-9.5 (3.0)	0.001	-8.3 (3.1)	0.008
46	Spinal cord compression	-0.5 (1.8)	0.79	-2.8 (2.8)	0.32	-7.0 (3.1)	0.02
47	Mental health visit	1.4 (1.2)	0.22	-0.9 (1.5)	0.53	0.1 (1.8)	0.97
48	HHS Region						
49	HHS Region 1	Reference		Reference		Reference	
50	HHS Region 2	1.6 (1.7)	0.35	-5.2 (2.2)	0.02	-6.7 (2.7)	0.01
51	HHS Region 3	2.7 (1.8)	0.12	2.1 (2.5)	0.40	1.3 (3.0)	0.66
52	HHS Region 4	0.4 (1.5)	0.77	-2.7 (2.2)	0.22	-4.9 (2.6)	0.07
53	HHS Region 5	0.3 (1.4)	0.81	0.8 (2.1)	0.69	-1.0 (2.6)	0.70
54	HHS Region 6	-0.9 (1.5)	0.53	-2.8 (2.2)	0.21	-4.4 (2.8)	0.11
55	HHS Region 7	0.0 (2.2)	0.99	3.2 (3.2)	0.31	0.9 (3.5)	0.79
56	HHS Region 8	-1.6 (2.2)	0.47	1.9 (3.8)	0.62	-2.0 (3.8)	0.61
57	HHS Region 9	0.0 (1.6)	0.99	-0.6 (2.5)	0.81	-3.2 (2.8)	0.26
58	HHS Region 10	-0.6 (2.2)	0.77	4.4 (3.5)	0.21	4.0 (4.3)	0.35
59	Study Year						
60	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:

^aMissing practice characteristics (1,432 or 2.94% of sample) were coded as “other unknown”.

^bMissing HCC (86 or .18% of sample) were replace by in sample mean HCC.

^cMissing rural indicator (22 or .05% of sample) were assumed to be non-rural

^bMissing ZIP code median income (708 or 1.46% of sample) were replace by in sample mean median income.

References

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

For peer review only

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract [Done]
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found [See abstract starting on page 2]
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported [See first paragraph page 6]
Objectives	3	State specific objectives, including any prespecified hypotheses [See last paragraph of first paragraph page 6]
Methods		
Study design	4	Present key elements of study design early in the paper [See Methodology section starting on page 7]
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection [See Physician and Index Visit sample subsections, (page 7)]
Participants	6	(a) <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants [See Methodology section and Figure 1 for physician, patient and visit sample stats] (b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed [NA] <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case [NA]
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable [See page 9 for outcome measures and page 9 for measures of knowledge]
Data sources/measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group [See page 10 for diagnostic knowledge measure]
Bias	9	Describe any efforts to address potential sources of bias [See Statistical Analyses page 11 first paragraph explanation for inclusion of other measures of knowledge and Sensitivity Analysis subsection page 12 and results of sensitivity analysis bottom page 13]
Study size	10	Explain how the study size was arrived at [See first and second paragraph page 6]
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why [See Statistical Methods subsection starting on page 9]
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding [See Statistical Methods first paragraph of page 12] (b) Describe any methods used to examine subgroups and interactions [NA] (c) Explain how missing data were addressed [See section 4 in the Supplement and reference to this in Table 3]

1
2 (d) *Cohort study*—If applicable, explain how loss to follow-up was addressed

3 *Case-control study*—If applicable, explain how matching of cases and controls was
4 addressed [NA]

5 *Cross-sectional study*—If applicable, describe analytical methods taking account of
6 sampling strategy [Sensitivity Analysis subsection starting on page 12 top]

7
8 (e) Describe any sensitivity analyses [Sensitivity Analysis subsection starting on top
9 of page 15]

10
11 Continued on next page
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60**Results**

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed [See Figure 1 and first paragraph page 12] (b) Give reasons for non-participation at each stage [See Figure 1, Table 1 and first paragraph page 11] (c) Consider use of a flow diagram [See Figure 1]
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders [See Page 13 second paragraph and Table 2] (b) Indicate number of participants with missing data for each variable of interest [NA] (c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount) [NA]
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time [NA] <i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure [NA] <i>Cross-sectional study</i> —Report numbers of outcome events or summary measures [See Table 1]
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included [See Table 3 and Supplementary Material for full set of controls and coefficients (Section 6)] (b) Report category boundaries when continuous variables were categorized [See Statistical Analysis Section and Supplementary Material Section 4] (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period [See Results section starting on top of page 13, Table 3]
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses [See Sensitivity Analysis subsection of Results section page top of Page 14 and Supplementary Material Section 5]

Discussion

Key results	18	Summarise key results with reference to study objectives [See Discussion subsection last paragraph of page 14]
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias [See first full paragraph starting on page 16]
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence [No other study has addressed our research question, however, in terms of methodology we compare our study to other in the Discussion section starting on bottom of page 15]
Generalisability	21	Discuss the generalisability (external validity) of the study results [See paragraph starting on last paragraph page 16, line 328]

Other information

Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based [See Funding section last paragraph on page 19]
---------	----	---

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

1
2
3 **Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and
4 published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely
5 available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at
6 <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is
7 available at www.strobe-statement.org.
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

BMJ Open

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

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-041817.R1
Article Type:	Original research
Date Submitted by the Author:	03-Feb-2021
Complete List of Authors:	Gray, Bradley; American Board of Internal Medicine Vandergrift, Jonathan; American Board of Internal Medicine McCoy, Rozalina; Mayo Clinic, Division of Endocrinology, Department of Medicine Lipner, Rebecca; American Board of Internal Medicine Landon, Bruce; Harvard Medical School, Department of Health Care Policy
Primary Subject Heading:	General practice / Family practice
Secondary Subject Heading:	Diagnostics, Medical education and training
Keywords:	INTERNAL MEDICINE, MEDICAL EDUCATION & TRAINING, GENERAL MEDICINE (see Internal Medicine)

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1
2
3 1 The Association Between Primary Care Physician Diagnostic Knowledge and Death,
4
5 2 Hospitalization and Emergency Department Visits Following an Outpatient Visit at Risk for
6
7
8 3 Diagnostic Error: A Retrospective Cohort Study Using Medicare Claims
9
10
11 4
12
13

14 5 Bradley M. Gray, PhD, corresponding author

15 6 Email: bgray@abim.org, Phone: 202-213 6646, FAX 202-213 6646

17 7 American Board of Internal Medicine

19 8 510 Walnut Street, Suite 1700, Philadelphia, Pennsylvania 19106, USA
20
21 9

23 10 Jonathan L. Vandergrift, MS

25 11 American Board of Internal Medicine Philadelphia, Pennsylvania, USA
26
27 12

29 13 Rozalina, G. McCoy, MD

31 14 Mayo Clinic, Rochester, Minnesota, USA
32
33 15

34 16 Rebecca S. Lipner, PhD

36 17 American Board of Internal Medicine Philadelphia, Pennsylvania, USA
37
38 18

40 19 Bruce E. Landon, MD

42 20 Harvard Medical School, Boston, Massachusetts, USA
43
44 21

45 22 Word count: 4,708,

47 23 Keywords: Internal Medicine, General Medicine, Medical Education & Training,
48
49
50
51
52
53
54
55
56
57
58
59
60

24 **Objective**

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

31 **Setting/Participants**

32 1,410 US general internists who recently took their American Board of Internal Medicine
33 Maintenance of Certification (ABIM-IM-MOC) exam treating 42,407 Medicare beneficiaries
34 who experienced 48,632 “index” outpatient visits for new complaints at risk for diagnostic error
35 because the presenting complaint (e.g., dizziness) was related to pre-specified diagnostic error
36 sensitive conditions (e.g. stroke).

37 **Outcome measures**

38 90-day risk of all-cause death, and, for outcome conditions related to the index visits diagnosis,
39 emergency department (ED) visits and hospitalizations.

40 **Design**

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

45 **Results**

46 Rates of 90-day adverse outcomes per 1,000 index visits were 7 for death, 11 for
47 hospitalizations, and 14 for ED visits. Being seen by a physician in the top versus bottom third
48 of diagnostic knowledge during an index visit for a new complaint at risk for diagnostic error
49 was associated with 2.9 fewer all-cause deaths (95% confidence interval (CI) -5.0 to -0.7,
50 P=.008), 4.1 fewer hospitalizations (95% CI -6.9 to -1.2, P=0.006), and 4.9 fewer ED visits (95%
51 CI -8.1% to -1.6%, P=0.003) per 1,000 visits.

52 **Conclusion**

53 Higher diagnostic knowledge was associated with lower risk of adverse outcomes after visits for
54 complaints at heightened risk for diagnostic error.

55

56 **Strengths and limitations of this study**

- 57 ○ Unique diagnostic knowledge measure linking diagnostic knowledge with adverse
58 outcomes
- 59 ○ Scalable adverse outcome measures and extensive sensitivity analyses
- 60 ○ Our assessment of diagnostic error is indirect (as indicated by adverse outcomes)
- 61 ○ Results are subject to selection bias if the mix of index visits or the severity of the
62 patients or practice support differed for physicians with different levels of
63 diagnostic knowledge.
- 64 ○ Results are only generalizable to physicians who elected to attempt ABIM's
65 certification exam and were about 10 years past initial certification and patients
66 older than 65.

Introduction

71
72
73 Diagnostic error has been identified as a key health care delivery concern and contributes to
74 significant potentially preventable morbidity and mortality.(1-3) Ambulatory care, and
75 especially primary care, is a practice setting with a particularly high risk for diagnostic error(4,
76 5) because of the wide variety of presentations encountered and the concomitant difficulty of
77 distinguishing harmful conditions from routine self-limited complaints, compounded by the well-
78 known time constraints faced by practitioners in that setting. It has been estimated that at least
79 5% of ambulatory visits are associated with diagnostic error, half of which may result in
80 considerable patient harm. Diagnostic error is a common cause of malpractice suits and most
81 frequently occurs in the ambulatory care settings.(6, 7)

82
83 Deficiencies in diagnostic knowledge are likely to be an important contributor to these diagnostic
84 errors that could impact, for example, the breadth of diagnoses considered, appropriate ordering
85 and interpretation of tests, and/or synthesis of data more generally.(8-11) Because of this,
86 measuring physician diagnostic knowledge has become a major focus of organizations
87 throughout the developed world that are tasked with licensing and certifying physicians with the
88 underlying, although largely untested, hypothesis being that diagnostic knowledge will be a
89 measurable and strong predictor of diagnostic error.(12-15) Testing this hypothesis and
90 quantifying this relationship is therefore a critical public policy concern both in terms of the
91 importance of board certification and other programs designed to enhance lifelong learning for
92 physicians.

1
2
3 93
4
5
6 94 In the US, the American Board of Internal Medicine (ABIM) is a leading organization that
7
8 95 certifies primary care physicians, most notably general internists. In fact, most general internists
9
10 96 in the US are certified by the ABIM and these physicians represent about 45% of all adult
11
12 97 primary care physicians in the US.(16) Unlike medical licensure, board certification is not a legal
13
14 98 requirement to practice medicine in the US, though many hospitals require board certification as
15
16 99 one criterion to obtain privileges and insurers often require board certification to be included in
17
18 100 covered physician panels.(17, 18) To maintain their certification, general internists must pass an
19
20 101 initial certifying exam and, periodically, pass a recertification exam thereafter (referred to as
21
22 102 Maintenance of Certification (MOC) exams).(19, 20) Diagnostic knowledge is a major
23
24 103 component of these exams representing about half of all exam questions for the Internal
25
26 104 Medicine MOC (IM-MOC) exam.
27
28
29
30
31

32 105
33
34
35 106 One explanation for the lack of research on this topic is the difficulty in studying the relationship
36
37 107 between general diagnostic knowledge and diagnostic error because of the inability to quantify
38
39 108 diagnostic knowledge and identifying diagnostic errors at a population level, especially in the
40
41 109 outpatient setting.(21) We address this gap in the literature by applying a unique measure of
42
43 110 diagnostic knowledge, performance on diagnostic related questions on ABIM's IM-MOC exam,
44
45 111 and relating this measure to deaths, hospitalizations, and emergency department visits that
46
47 112 occurred after outpatient visits for new complaints at heightened risk for diagnostic error.
48
49
50
51

52 113
53
54

55 114
56
57
58
59
60

115 **Methods**

116

117 ***Physician and Index Visit Sample***

118 Our physician sample included general internists who were initially ABIM board certified in
119 2000 and took their IM-MOC exam between 2008 and 2011 (Figure 1). We identified Medicare
120 beneficiary outpatient Evaluation & Management (E&M) visits with these physicians using their
121 National Provider Identifier (NPI) during the calendar year following their exam (2009 to 2012).
122 These patients were age 65 or older and continuously enrolled in Medicare fee-for-service
123 (Medicare insures most of the US population over 65) during the physician's one year follow-up
124 period and the year prior. To ensure that any presenting complaints being evaluated were new
125 (i.e., not follow up), we restricted these visits to those that were the first visit for a new complaint
126 (the "*index visit*") because these visits were preceded by a 90-day clean period with no previous
127 inpatient or outpatient visit. The 90-day clean period is consistent with the US government
128 Centers for Medicare and Medicaid Services criteria used by its Bundled Payments for Care
129 Improvement Program for defining new episodes of care and with the patterns of visits we
130 observed (see Appendix Section 1 for related analysis).(22, 23)

131

132 We further restricted these index visits to those at heightened risk for diagnostic errors because
133 the recorded diagnosis in the Medicare claims (the "*index visit diagnosis*"), which includes
134 recording of symptom (e.g. loss of balance), could have been the initial presenting complaint for
135 one or more of 13 pre-specified diagnostic error sensitive conditions such as congestive heart
136 failure or bacteremia/sepsis (see Table 1). These 13 conditions (see Appendix Section 2 for a list

1
2
3 137 and applicable ICD-9 codes) were an acute non-cancerous subset of 20 conditions previously
4
5 138 noted by Schiff et al. to be at high risk for serious diagnostic error.(24) For instance, index visits
6
7
8 139 with diagnosis codes for chest pain, dyspepsia, shortness of breath, hypoxemia/hypoxia,
9
10 140 respiratory distress, weakness/fatigue, edema or ascites could all be the initial presentation of
11
12 141 congestive heart failure, which is one of the 13 diagnostic error sensitive conditions.
13
14

15 142
16
17
18 143 We used a three-step process to identify eligible index visit diagnoses. First, two physician
19
20 144 authors (RGM and BEL) identified all diagnoses that could be presenting complaints for the 13
21
22 145 diagnostic error sensitive conditions: what complaints/diagnoses might someone who ultimately
23
24 146 presented with a diagnostic-error sensitive condition have presented with initially? Second,
25
26 147 because the original list of identified index visit diagnoses was large (76), we reduced this list to
27
28 148 38 by applying a relative risk (RR) criteria. For a specific index visit diagnosis to meet this
29
30 149 criteria, all index visits with that diagnosis had to have a greater portion of later ED visits or
31
32 150 hospitalizations with the related outcome condition discharge diagnosis than index visits where
33
34 151 the specific at risk diagnosis was not present. For example, dizziness was chosen as an eligible
35
36 152 index visit diagnosis for stroke, one of the diagnostic error sensitive conditions, both because it
37
38 153 was identified as a potential presenting symptom of a stroke by physician authors and because
39
40 154 index visits with that diagnosis had a greater proportion of later hospitalization or ED visits for
41
42 155 stroke than visits without this diagnosis. Third, we also included index visits where the actual
43
44 156 diagnosis was one of the 13 diagnostic error sensitive conditions because we wanted to include
45
46 157 cases where diagnostic errors were and were not made. Therefore, we also included index visits
47
48 158 with a diagnosis of congestive heart failure itself as being at risk for the underlying condition
49
50 159 congestive heart failure.
51
52
53
54
55
56
57
58
59
60

160 ***Outcome Measures***

161 We examined the risk of three serious adverse outcomes within 90 days of the index visit that we
162 hypothesized would occur more frequently in cases of misdiagnosis: all-cause mortality,
163 hospitalizations, and ED visits. We did not count these events as adverse outcomes if they
164 occurred on the same day as the index visit because this may reflect a positive action (the
165 physician correctly diagnosed a patient with stroke and referred/admitted them to the hospital) or
166 be unavoidable regardless of the accuracy of the index visit diagnosis (the patient died despite
167 immediately admitting the patient to the hospital who exhibited stroke symptoms). Based on
168 Medicare billing codes, hospitalizations were limited to non-elective hospitalizations initiated
169 through the ED or trauma center. The ED and hospitalization outcomes were also limited to
170 cases where the discharge diagnosis was for one of the 13 diagnostic error sensitive conditions
171 following an index visit with the applicable diagnosis. We therefore presumed that these
172 discharge diagnoses were a reasonable representation of the underlying condition of the patient
173 at the time of the index visit. For example, we would count a hospitalization with a discharge
174 diagnosis of stroke as an adverse outcome if it occurred after an index visit for dizziness because
175 dizziness was identified as being a potential presenting complaint for stroke. However, we did
176 not count hospitalizations with a discharge diagnosis for acute coronary syndrome following an
177 index visit for dizziness because dizziness was not identified as a presenting complaint for acute
178 coronary syndrome. The rationale is that if there were no presenting complaints during the index
179 visit related to coronary syndrome, either because the underlying condition was not present or
180 could not be detected at the time of the index visit, then the index visit physician could not have
181 prevented the hospitalization regardless of their diagnostic knowledge.

182 ***Measure of Diagnostic Knowledge***

1
2
3 183 Our measure of diagnostic knowledge was calculated as the percent of correct answers on the
4
5 184 IM-MOC exam for questions previously coded as “diagnosis-related” by ABIM’s IM-MOC
6
7
8 185 exam committee. In our study, these questions comprised 53% of all IM-MOC exam questions,
9
10 186 with the remaining 42% addressing treatment and 5% related to other topics such as
11
12 187 epidemiology or pathophysiology. More generally, exam questions are designed to replicate real
13
14
15 188 world clinical scenarios and/or patient encounters and without reliance on rote memorization.(25,
16
17 189 26)

18
19
20 190

21
22
23 191 The ABIM exam committee coded each question based on the primary function tested to assure
24
25 192 that the exam covers care typically rendered by outpatient primary care physicians. Questions
26
27 193 coded as “diagnosis related” typically test knowledge and skills related to diagnostic inference,
28
29
30 194 differential diagnosis, and diagnostic testing and therefore are measuring diagnostic knowledge
31
32 195 and related decision-making. Psychometric analysis indicates that scores on diagnosis related
33
34 196 exam questions were meaningfully correlated (i.e., Cronbach’s alpha score of 0.84), and thereby
35
36 197 represent an independent underlying construct that could be interpreted as diagnostic knowledge
37
38
39 198 (see Appendix Section 3 for more details).(27) Similarly, this analysis indicated that questions
40
41 199 coded as treatment related also represent an independent underlying construct (i.e., Cronbach’s
42
43 200 alpha score of 0.75). Although performance on diagnosis and treatment related questions were
44
45
46 201 correlated (Pearson Correlation=0.62), 59.5% of the variation in diagnosis exam performance for
47
48 202 the physician study sample was not explained by performance on other parts of the exam.
49

50
51 203 ***Statistical methods***
52
53
54
55
56
57
58
59
60

1
2
3 204 Using Probit regression we estimated the associations with each adverse outcome, with standard
4
5 205 errors adjusted for correlations resulting from the nesting of visits within patients within
6
7 206 physicians.(28, 29) To measure associations with diagnostic knowledge we included categorical
8
9 207 regression explanatory variables for top and middle third of percent correct scores on diagnosis
10
11 208 related questions (bottom third was the reference category). Other exam level explanatory
12
13 209 variables included tertile indicators for performance on treatment-related questions and
14
15 210 performance on other question types. Since these variables measure knowledge unrelated to
16
17 211 diagnosis, they account for correlations between factors such as unmeasured practice or patient
18
19 212 characteristics that might be correlated with exam performance and our outcome measures (e.g.,
20
21 213 high scoring physicians may be more likely to practice in an academic setting or other such
22
23 214 settings that might be independently related to diagnostic error). Exam form indicators accounted
24
25 215 for differences in exam difficulty across exam administrations.
26
27
28
29
30

31 216
32
33
34 217 We also included physician, patient and visit level regression controls. Physician level controls
35
36 218 included: practice size (indicators for solo practice and practices larger than 50 physicians),
37
38 219 practice type (indicators for academic, group), demographic (gender), and training characteristics
39
40 220 (medical school location interacted with country of birth). Patient level controls included:
41
42 221 demographic characteristics (age and age squared, gender and race/ethnicity indicators) and a
43
44 222 Medicaid eligibility indicator. Lagged patient risk adjusters included 27 indicators for chronic
45
46 223 conditions and Medicare's Hierarchal Condition Category (HCC) risk adjustment score. We
47
48 224 imputed values for a small number of missing values for controls (see Appendix Section 4).
49
50
51 225 Patient index visit location level controls included: an indicator for residing in a rural ZIP code,
52
53 226 ZIP code median household income, and indicators for 10 US Health and Human Services
54
55
56
57
58
59
60

1
2
3 227 regions. Index visit level controls included: indicators of any outpatient visit, hospitalization or
4
5 228 ED visits within the prior year and number of days since the most recent of these events, visit
6
7
8 229 year indicators to control for secular changes in quality. We also included an indicator for
9
10 230 whether or not the patient had a previous contact with the index visit physician during the year
11
12 231 prior to the index visit to account for differences in physician-patient continuity (see Appendix
13
14 232 Section 5 for a full list of controls).

17 233 *Sensitivity Analysis*

19
20 234 We performed numerous sensitivity analyses to test the robustness of our results (detailed in
21
22 235 Appendix Section 6). First, we expanded the index visit sample to include all index visits with
23
24 236 the original 76 diagnoses identified by the physician authors regardless of whether they met the
25
26 237 relative risk criteria. Second, we expanded and contracted the index visit clean period by seven
27
28 238 days. Third, excluded hospitalizations or ED events occurring the day after the index visit, in
29
30 239 addition to same day events, to consider the possibility that they might be triggered by a correct
31
32 240 diagnosis and therefore should not have been considered adverse outcomes. Fourth, we
33
34 241 considered the possibility that our results were biased due to omitted variables correlated with
35
36 242 practice size. For example, it could be that physicians in large practices have greater access to
37
38 243 specialists or other physicians for informal consultations than those in small practices and
39
40 244 therefore outcomes for these physicians may be less sensitive to their knowledge. To examine
41
42 245 this possibility, we estimated associations with knowledge and our two utilization measures
43
44 246 across a sample of physicians in either small (≤ 10 physicians, 54.5% (768/1,410) of physicians)
45
46 247 or large practices (> 50 or in academic medical centers, 23.7% (334/1,410) of physicians). We did
47
48 248 not conduct these sensitivities for death because there were too few deaths in the subgroups to
49
50 249 allow us to reliably estimate the associations (e.g., 39 deaths for physicians in large
51
52
53
54
55
56
57
58
59
60

1
2
3 250 practices).Fifth, to consider the possibility that these outcomes were only avoided because the
4
5 251 patient died, for the ED and hospitalization outcome, we also included instances where the
6
7 252 patient died. Sixth, as a falsification test we limited the index visits to those that were unrelated
8
9 253 to the 13 diagnostic error sensitive conditions. Under this sensitivity, we expected then that the
10
11 254 associations with diagnostic knowledge would decline. The index visit physician’s diagnostic
12
13 255 knowledge cannot impact a future adverse outcome if the underlying condition that caused that
14
15 256 outcome was not present or detectible at the time of index visit. Therefore, this reduction in
16
17 257 association should be especially true for the hospitalization and ED measures where adverse
18
19 258 outcomes were limited to the 13 diagnostic error conditions and so were unrelated to the index
20
21 259 visit diagnoses in this sensitivity. Similarly, for the last sensitivity, we applied elective
22
23 260 hospitalizations as an outcome measure to consider the possibility that there could be a
24
25 261 correlation between the overall propensity to hospitalize in an area and physician knowledge.
26
27
28
29
30
31
32
33

34 263 The Advarra Institutional Review Board approved our study protocol and all analyses were
35
36 264 performed using Stata version 15 (College Station, TX). Patients and the public were not
37
38 265 involved in the design or execution of this study as the existing patient claims data used were de-
39
40 266 identified by the Center for Medicaid and Medicare Services prior to analysis.
41
42
43

44 267 *Patient and Public Involvement*

45
46
47 268 Patients and/or the public were not involved in the design, or conduct, or reporting, or
48
49 269 dissemination plans of this research.
50
51
52
53
54

55 271 **Results**

1
2
3 272
4
5
6 273 Of 2,492 general internists who initially certified in 2000 and who took an IM-MOC exam
7
8 274 between 2009 and 2012, 1,722 had outpatient visits with a fee-for-service Medicare beneficiary
9
10 275 during the study period. Those without visits generally practiced hospital medicine. Of these,
11
12 276 1,410 were included in the study because they had at least one outpatient index visit that met our
13
14 277 study inclusion criteria during the year after they took their IM-MOC exam. In total, 48,632
15
16 278 index visits with 42,407 patients treated by 1,410 physicians met study inclusion criteria (Figure
17
18 279 1). Table 1 lists frequency of index visits and subsequent outcomes for each diagnostic error
19
20 280 sensitivity condition.
21
22
23
24
25 281

26
27
28 282 The mean percent correct on diagnosis questions ranged from 84.3% among top third performers
29
30 283 to 65.5% among bottom third performers (Table 2). Patient and visit characteristics were similar
31
32 284 across tertiles of physician diagnostic knowledge. For example, there were no statistically
33
34 285 significant differences in the HCC risk adjuster across tertiles ($P=.19$) However, there were
35
36 286 differences in some physician and practice characteristics. When compared to physicians in the
37
38 287 bottom tertile of diagnostic knowledge, physicians in the top were significantly less likely to be
39
40 288 in solo practice (12.8% versus 24.4%, $P=0.009$), and more likely to be in academic practice
41
42 289 (9.7% versus 3.4%, $P<.001$). However, the proportion graduating from a US medical school was
43
44 290 similar across diagnostic knowledge tertiles (70.0% versus 63.3%, $P=.30$).
45
46
47
48

49 291 *Associations between diagnostic knowledge and patient adverse outcomes*

50
51
52 292 The overall rates of 90-day adverse outcomes per 1,000 index visits was 6.5 for death, 11.1 for
53
54 293 hospitalizations, and 13.6 for ED visits (with the latter two directly associated with one of the
55
56
57

1
2
3 294 diagnostic error sensitive conditions whose antecedent was present in the applicable index visit).
4
5 295 Being seen by a physician scoring in the top versus bottom third of diagnostic knowledge on the
6
7
8 296 MOC exam was associated with 2.9 fewer deaths per 1,000 visits (95% confidence interval (CI) -
9
10 297 5.0 to -0.7, P=.008) which reflects a 35.3% lower risk of death (95% CI -52.8 to -11.2, P=.008),
11
12 298 (Table 3). Our finding also suggests that this difference in exam performance was associated
13
14
15 299 with 4.1 fewer applicable hospitalizations (95% CI -6.9 to -1.2, P=0.006), and 4.9 fewer
16
17 300 applicable ED visits (95% CI -8.1 to -1.6, P=0.003) per 1,000 visits (Table). These reductions
18
19 301 correspond with about a 30% lower risk for these utilization measures (hospitalizations: -30.5%,
20
21 302 95% CI -46.1 to -10.4, P=.003, ED: -29.8%, 95% CI -44.4 to -11.4).
22
23
24
25 303

26
27 304 We also found a significant dose response relationship across all three regression adjusted
28
29 305 relative risk measures (P-trends <0.008). For example, the regression-adjusted 90-day risk of
30
31 306 death per 1,000 patients whose index visit physician scored in the top third of diagnostic
32
33 307 knowledge was 5.2 (95% CI 4.1 to 6.3), compared to 6.5 (95% CI 5.4 to 7.6) for the middle
34
35 308 third, and 8.1 (95% CI 6.5 to 9.7) for the bottom third (P-trend 0.008).
36
37
38

39 309 *Sensitivity Analyses*

40
41
42 310 Our sensitivity analyses (Appendix Section 6) confirmed that base case associations with
43
44 311 diagnostic knowledge were robust to different index visit clean periods, and diagnosis code
45
46 312 inclusion criteria and next day coding of outcome measures. Associations with diagnostic
47
48 313 knowledge were also fairly robust to physician's practice size for both the ED and hospitalization
49
50 314 measures when we limited the sample to either small or large or academic practices.
51
52
53
54
55 315

1
2
3 316 Suggesting that our results were not influenced by omitted variable bias, we found that
4
5 317 associations with diagnostic knowledge and our outcome measures became small and
6
7 318 statistically insignificant when we limited the sample to index visits with diagnoses unrelated to
8
9 319 any of the 13 diagnostic sensitive error conditions, and so were at lower risk for diagnostic error
10
11 320 ($P>0.50$ and associations were at most about a tenth of the base case percent difference between
12
13 321 top and bottom third of diagnostic knowledge). We also found no significant association between
14
15 322 lack of diagnostic knowledge and elective hospitalizations ($P=0.63$).
16
17
18
19
20 323
21
22

23 324 **Discussion**

24
25 325
26
27
28 326 We found that higher diagnostic knowledge among US outpatient internal medicine physicians
29
30 327 was associated with significant reductions in subsequent adverse outcomes whose cause was at
31
32 328 risk for diagnostic error. Indeed, for every 1,000 index visits for a new complaint at risk for
33
34 329 diagnostic error, being seen by a physician in the top versus bottom third of diagnostic
35
36 330 knowledge was associated with 2.9 fewer all-cause death and, for diagnostic error sensitive
37
38 331 conditions, 4.1 fewer hospitalizations and 4.9 fewer ED visits within 90 days. These figures
39
40 332 correspond to a reduction in risk for these adverse events by about a third. Although some prior
41
42 333 studies have demonstrated the high morbidity and mortality of diagnostic error(1-3), this is the
43
44 334 first study to demonstrate and quantify the direct association between serious adverse outcomes
45
46 335 and the diagnostic knowledge of their first contact primary care physician. These finding support
47
48 336 the notion that gaps in diagnostic knowledge between physicians may be an important
49
50 337 contributor to the diagnostic error problem plaguing the healthcare system worldwide.
51
52
53
54
55
56
57
58
59
60

1
2
3 338
4
5
6 339 We measured the association between diagnostic knowledge and potential diagnostic error by
7
8 340 using Medicare claims data to identify patients who presented for outpatient visits with
9
10 341 complaints at heightened risk for serious diagnostic errors and examining the occurrence of
11
12 342 clinically relevant adverse outcomes soon thereafter. Although this approach lacks the precision
13
14 343 of individual chart audits(7), it is both clinically plausible and scalable in that it can be used to
15
16 344 monitor the care of large numbers of patients, making the method itself an important contribution
17
18 345 to the literature on diagnostic error. Although we did not directly measure diagnostic errors
19
20 346 through chart audits, the fact that we found associations with diagnostic knowledge and the
21
22 347 diagnostic error sensitive outcome conditions we studied coupled with the fact that we did not
23
24 348 find associations with treatment knowledge, nor did we find associations when the underlying
25
26 349 diagnostic error sensitive condition was likely not present during the outpatient index visit
27
28 350 because no antecedent diagnoses recorded indicates that the associations we report in this study
29
30 351 were likely driven by association with diagnostic errors that occurred during these visits.
31
32 352 Furthermore, our approach builds on prior studies that used claims data to infer diagnostic error
33
34 353 incidence for ED visits, in that we identified index visit diagnoses at risk for diagnostic error that
35
36 354 were clinically plausible and verified empirically, and we assured that we were studying new
37
38 355 problems by requiring that the patient not have had a visit over the previous 3 months
39
40 356 contacts.(30-32) We expanded on these studies by focusing on outpatient care and by examining
41
42 357 a much more comprehensive set of presenting complaints that may have been precursors to one
43
44 358 of 13 diagnostic error prone conditions that we studied. This approach was necessary in order to
45
46 359 study diagnostic error in the more low acuity setting of outpatient general internal medicine.
47
48
49
50
51
52
53
54
55 360
56
57
58
59
60

1
2
3 361 Our findings suggest an association between diagnostic knowledge and adverse outcomes. Yet,
4
5 362 there are important limitations to consider. We did not directly determine whether a diagnostic
6
7 363 error had occurred through such validated means as a chart review. Our findings cannot be
8
9
10 364 interpreted as causal given the cross-sectional nature of our study so we cannot rule out the
11
12 365 possibility that observed associations were the result of omitted variable bias related to either
13
14 366 physician or patient characteristics, and do not reflect a causal relationship between diagnostic
15
16 367 knowledge and adverse outcomes. That said, there is no reason to believe that these
17
18 368 characteristics would be correlated with diagnostic knowledge independent of treatment
19
20 369 knowledge which we were able to control for as both these knowledge measures should be
21
22 370 similarly correlated with unobserved factors such as ability of consulting colleagues.
23
24 371 Furthermore, had associations with diagnostic knowledge been driven by omitted variable bias
25
26 372 then we would have expected them to be similar when estimated across index visits with lower
27
28 373 or higher risk for diagnostic error, and they were not. We also found that diagnosis exam
29
30 374 performance was not associated with elective hospitalizations, which are, presumably, unrelated
31
32 375 to underlying diagnostic knowledge but may be related to the overall propensity to hospitalize.
33
34 376 That said, the fact that practice size was found to be correlated with diagnostic exam
35
36 377 performance is concerning. For example, as described above, practice size could be correlated
37
38 378 with access to specialists that intern might be related to our outcome measures. However,
39
40 379 sensitive analyses indicate that associations with knowledge and our utilization adverse outcome
41
42 380 measures were fairly similar across physicians practice size/type (small, and large or academic).
43
44 381 An additional limitation is that we studied select conditions among older patients enrolled in the
45
46 382 Medicare program so we cannot extrapolate these findings to a younger population, other
47
48 383 conditions we did not consider, or populations with no or different health insurance coverage.
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 384 Our findings might also not be applicable to older physicians who certified before 2000 or
4
5 385 younger physicians who certified after 2000 as well as physicians who choose not to attempt an
6
7 386 exam. While a physician's clinical knowledge might be related to their decision to not take the
8
9 387 MOC exam therefore not maintaining their certification, other factors certainly play a role in this
10
11 388 decision.
12
13
14

15 389
16
17
18 390 Another limitation of our study is that the IM-MOC exam was specifically designed to measure
19
20 391 clinical knowledge in general, it was not designed to measure diagnostic knowledge specifically.
21
22 392 That said, diagnostic knowledge is a major component of the exam and was found to meet the
23
24 393 criteria for measuring this underlying construct. Also diagnostic error may have stemmed from
25
26 394 factors outside of inadequate diagnostic knowledge, which are not covered by the exam but
27
28 395 could be correlated with our exam based diagnostic knowledge measure (e.g., poor
29
30 396 patient/physician communication skills and related system failures).(33, 34) That said, there is no
31
32 397 reason to believe that these other contributors to diagnostic error would not also be correlated
33
34 398 with the other aspects of the exam we do account for. Furthermore, based on an analysis of
35
36 399 malpractice claims, Newman-Toker et al. (6) reported that clinical judgement played an
37
38 400 important role in 86% of diagnostic errors, while poor patient/physician communication and
39
40 401 system failures played a role in far fewer diagnostic errors that resulted in malpractice suits (35%
41
42 402 and 22% respectively). Suggesting that improving communication will not reduce stroke related
43
44 403 diagnostic error, Kerber et al. (35) reported that frontline providers rarely ask the right questions
45
46 404 when patients present with dizziness. Communication ability is only valuable in terms of
47
48 405 reducing diagnostic error if the physician knows what questions to ask and what the answers
49
50
51
52
53
54
55 406 mean. Although we cannot say with certainty that our finding are driven by an underlying
56
57
58
59
60

1
2
3 407 association between diagnostic knowledge and diagnostic errors, at a minimum, our finding
4
5 408 suggest that patients treated by physicians who scored well on diagnostic exam questions may be
6
7 409 at lower risk for the adverse outcomes we studied. Finally, some might assert that a standardized
8
9 410 exam without access to medical reference material might be more a reflection of a physician's
10
11 411 rote memory and ability to recall medical facts than a test of their clinical knowledge and
12
13 412 judgement. Although this is a fundamental limitation of our study, it should be noted that the
14
15 413 exam is designed to mimic decision making in real life situations including have such things as
16
17 414 lab values and reference material embedded in questions and past research indicates that an
18
19 415 "open" book format that allows physicians access to reference material did not materially impact
20
21 416 exam performance.(36) It should also be noted that the necessary rapidity of decision making by
22
23 417 primary care physicians who have limited time per encounter might fairly be represented by an
24
25 418 exam with time constraints.
26
27
28
29
30

31 419
32
33
34 420 In this exploratory analysis, we found evidence that diagnostic knowledge of primary care
35
36 421 physicians seeing a patient for an index visit for a complaint that is at heightened risk of
37
38 422 diagnostic error is associated with adverse outcomes. The fact that there exists a link between
39
40 423 general diagnostic knowledge and diagnostic error may not be surprising, the magnitude of the
41
42 424 associations we found suggests that interventions ignoring the role of physician knowledge may
43
44 425 be inadequate to address the crisis of diagnostic error. Interventions targeted at improving
45
46 426 diagnostic knowledge could include such things as a greater focus on diagnostic training during
47
48 427 graduate medical education (i.e., medical school, residency, and fellowship). Knowledge-focused
49
50 428 interventions could also include incentivizing broad-based learning as well as targeted learning
51
52 429 pursued through continuing medical education (CME) activities.(30) During visits identified as
53
54
55
56
57
58
59
60

1
2
3 430 being at risk for diagnostic errors, physicians could be given related information at the point of
4
5 431 care including suggestions for specialty consultation.
6
7

8 432
9
10
11 433 Our results are important for two additional reasons. First, these results provide evidence that board
12
13 434 certification and maintenance of certification, which involves lifelong learning directed at maintaining
14
15 435 medical knowledge, might, in fact, be a valid approach to assuring the delivery of high quality care.
16
17 436 Many in the US complain about the time and expense of MOC and often point to the lack of rigorous
18
19 437 assessment between aspects of MOC and outcomes of interest to patients. These findings suggest that
20
21 438 processes such as MOC may translate into meaningful improvements in outcomes because they can
22
23 439 provide incentives for meaningful learning. This learning also could be enhanced through exam feedback
24
25 440 targeted at diagnostic knowledge. Second, the findings also suggest that interventions aimed at improving
26
27 441 diagnostic skills, whether knowledge-based or through, for instance, delivery of relevant information at
28
29 442 the point of care [this is in response to system changes] might be approaches that might be worthwhile if
30
31 443 the findings of this study are validated with additional research. Yet more research is needed to better
32
33 444 understand the link between diagnostic knowledge and diagnostic errors that are identified
34
35 445 through chart review or other methods of direct ascertainment and the extent to which such
36
37 446 errors result in adverse clinical outcomes.
38
39
40
41
42
43
44

45 448 In conclusion, gaps in diagnostic knowledge among first contact primary care physicians is
46
47 449 associated with serious diagnostic error sensitive outcomes. If this finding is confirmed in future
48
49 450 studies, diagnostic knowledge should be a target for interventions to reduce diagnostic errors.
50
51
52

53 451
54
55

56 452
57
58
59
60

1
2
3 453
4
5
6 454
7
8
9 455
10
11 456
12
13 457
14
15 458
16
17
18 459
19
20 460
21
22 461
23
24 462
25
26
27 463
28
29 464
30
31 465
32
33
34 466
35
36 467
37
38 468
39
40
41 469
42
43 470
44
45 471
46
47
48 472
49
50 473
51
52 474
53
54
55 475
56
57
58
59
60

Statements

- A. Contribution statement: Bradley Gray, Jonathan Vandergrift, Rebecca Lipner, Rozalina McCoy, and Bruce Landon met the ICMJE guidelines authorship criteria:
- a. Bradley Gray, Jonathan Vandergrift, Rebecca Lipner, Rozalina McCoy, and Bruce Landon substantially contributed to the conception and design of the work.
 - b. Bradley Gray, Jonathan Vandergrift, Rebecca Lipner contributed to the acquisition of the data.
 - c. Bradley Gray, Jonathan Vandergrift, Rebecca Lipner, Rozalina McCoy, and Bruce Landon substantially contributed to analysis or interpretation of data.
 - d. Bradley Gray, Jonathan Vandergrift, Rebecca Lipner, Rozalina McCoy, and Bruce Landon substantially contributed to the drafting the work and revising it critically for important intellectual content.
 - e. Bradley Gray, Jonathan Vandergrift, Rebecca Lipner, Rozalina McCoy, and Bruce Landon gave final approval of the version published.
 - f. Bradley Gray, Jonathan Vandergrift, Rebecca Lipner, Rozalina McCoy, and Bruce Landon 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.
- B. Competing Interests: Bradley Gray, Jonathan Vandergrift, and Rebecca Lipner are paid employees of the American Board of Internal Medicine. Bruce Landon is a paid consultants for the American Board of Internal Medicine.
- C. Funding Statement: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

1
2
3 476 D. Data Sharing: Administrative data describing physician characteristics and exam
4
5 477 performance can be obtained from the ABIM through a data sharing agreement that assures
6
7 478 physician confidentiality and its use for legitimate research purposes. Access to de-
8
9 identified Medicare claims data for this study were obtained through a special data use
10 479
11 agreement with the Centers for Medicare and Medicaid services which is a process
12 480
13 available to researchers in the US.
14 481

15
16
17 482 E. Dissemination to participants and related patient and public communities: As study data
18
19 483 were pseudonymised, it is not possible to send findings directly to the study participants.
20
21 484 ABIM's communication department in collaboration with the authors of this study will
22
23 write a press release whose goal is to inform the public regarding the findings of the study.
24 485
25

26 486
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

487 **References**

- 488 1. National Academies of Sciences, Engineering, and Medicine. Improving Diagnosis in Health
489 Care. Washington, DC: The National Academies Press 2015 December 10, 2018.
- 490 2. Graber ML, Trowbridge R, Myers JS, et al. The next organizational challenge: finding and
491 addressing diagnostic error. *Jt Comm J Qual Patient Saf.* 2014;40(3):102-10.
- 492 3. Cresswell KM, Panesar SS, Salvilla SA, et al. Global research priorities to better understand the
493 burden of iatrogenic harm in primary care: an international Delphi exercise. *PLoS Med.*
494 2013;10(11):e1001554.
- 495 4. Kostopoulou O, Delaney BC, Munro CW. Diagnostic difficulty and error in primary care--a
496 systematic review. *Fam Pract.* 2008;25(6):400-13.
- 497 5. Goyder CR, Jones CH, Heneghan CJ, et al. Missed opportunities for diagnosis: lessons learned
498 from diagnostic errors in primary care. *Br J Gen Pract.* 2015;65(641):e838-44.
- 499 6. Newman-Toker DE, Schaffer AC, Yu-Moe CW, et al. Serious misdiagnosis-related harms in
500 malpractice claims: The "Big Three" - vascular events, infections, and cancers. *Diagnosis (Berl).*
501 2019;6(3):227-40.
- 502 7. Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care:
503 estimations from three large observational studies involving US adult populations. *BMJ Qual Saf.*
504 2014;23(9):727-31.
- 505 8. Gandhi TK, Kachalia A, Thomas EJ, et al. Missed and delayed diagnoses in the ambulatory
506 setting: a study of closed malpractice claims. *Ann Intern Med.* 2006;145(7):488-96.
- 507 9. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med.*
508 2005;165(13):1493-9.
- 509 10. Kachalia A, Gandhi TK, Puopolo AL, et al. Missed and delayed diagnoses in the emergency
510 department: a study of closed malpractice claims from 4 liability insurers. *Ann Emerg Med.*
511 2007;49(2):196-205.
- 512 11. Poon EG, Kachalia A, Puopolo AL, et al. Cognitive errors and logistical breakdowns contributing
513 to missed and delayed diagnoses of breast and colorectal cancers: a process analysis of closed malpractice
514 claims. *J Gen Intern Med.* 2012;27(11):1416-23.
- 515 12. Chisholm A, Askham J. A review of professional codes and standards for doctors in the UK,
516 USA and Canada: Picker Institute Europe Oxford; 2006.
- 517 13. Irvine D. Doctors in the UK: their new professionalism and its regulatory framework. *Lancet.*
518 2001;358(9295):1807-10.
- 519 14. Kovacs E, Schmidt AE, Szocska G, et al. Licensing procedures and registration of medical
520 doctors in the European Union. *Clinical Medicine.* 2014;14(3):229-38.
- 521 15. European Union of Medical Specialists. The european council for accreditation of medical
522 specialist qualifications (ECAMSQ) 2010 [Available from:
523 https://www.uems.eu/_data/assets/pdf_file/0009/1206/ECAMSQ_presentation.pdf.
- 524 16. Petterson SM, Liaw WR, Phillips RL, Jr., et al. Projecting US primary care physician workforce
525 needs: 2010-2025. *Ann Fam Med.* 2012;10(6):503-9.
- 526 17. Freed GL, Dunham KM, Singer D. Use of board certification and recertification in hospital
527 privileging: policies for general surgeons, surgical specialists, and nonsurgical subspecialists. *Arch Surg.*
528 2009;144(8):746-52.
- 529 18. Independence Blue Cross. Credentialing criteria 2019 [Available from:
530 https://www.ibx.com/pdfs/providers/interactive_tools/credentialing_criteria_ibx.pdf.
- 531 19. American Board of Medical Specialties. Certification Matters FAQs 2019 [Available from:
532 <https://www.certificationmatters.org/faqs/>.
- 533 20. Lipner RS, Bylsma WH, Arnold GK, et al. Who is maintaining certification in internal medicine--
534 and why? A national survey 10 years after initial certification. *Ann Intern Med.* 2006;144(1):29-36.
- 535 21. Balogh E, Miller BT, Ball J, et al. Improving diagnosis in health care. xxvii, 444 pages p.

- 1
2
3 536 22. Centers for Medicare & Medicaid Services. BPCI Advanced 2018 [Available from:
4 537 <https://innovation.cms.gov/initiatives/bpci-advanced>.
5 538 23. Centers for Medicare & Medicaid Services. Comprehensive Care for Joint Replacement Model
6 539 2018 [Available from: <https://innovation.cms.gov/initiatives/cjr>.
7 540 24. Schiff GD, Hasan O, Kim S, et al. Diagnostic error in medicine: analysis of 583 physician-
8 541 reported errors. *Arch Intern Med*. 2009;169(20):1881-7.
9 542 25. Gray B, Vandergrift J, Lipner RS, et al. Comparison of content on the American Board of Internal
10 543 Medicine Maintenance of Certification examination with conditions seen in practice by general internists.
11 544 *JAMA*. 2017;317(22):2317-24.
12 545 26. Samonte K, de la Cruz S, Garcia MJ. An Overview of the ABIM Cardiovascular Disease
13 546 Maintenance of Certification Examination. *J Am Coll Cardiol*. 2020;75(9):1083-6.
14 547 27. Bandalos DL. Measurement theory and applications for the social sciences: Guilford
15 548 Publications; 2018.
16 549 28. Huber PJ. The behavior of maximum likelihood estimates under non-standard conditions. *Paper*
17 550 *presented at: Fifth Berkeley symposium on mathematical statistics and probability; Berkeley, CA*. 1967.
18 551 29. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for
19 552 heteroskedasticity. *Econometrica: Journal of the Econometric Society*. 1980:817-38.
20 553 30. Newman-Toker DE, Makary MA. Measuring Diagnostic Errors in Primary Care: The First Step
21 554 on a Path Forward Comment on “types and Origins of Diagnostic Errors in Primary Care Settings”. *JAMA*
22 555 *internal medicine*. 2013;173(6):425-6.
23 556 31. Waxman DA, Kanzaria HK, Schriger DL. Unrecognized Cardiovascular Emergencies Among
24 557 Medicare Patients. *JAMA Intern Med*. 2018;178(4):477-84.
25 558 32. Liberman AL, Newman-Toker DE. Symptom-Disease Pair Analysis of Diagnostic Error
26 559 (SPADE): a conceptual framework and methodological approach for unearthing misdiagnosis-related
27 560 harms using big data. *BMJ Qual Saf*. 2018;27(7):557-66.
28 561 33. Davis Giardina T, King BJ, Ignaczak AP, et al. Root cause analysis reports help identify common
29 562 factors in delayed diagnosis and treatment of outpatients. *Health Aff (Millwood)*. 2013;32(8):1368-75.
30 563 34. Zwaan L, Monteiro S, Sherbino J, et al. Is bias in the eye of the beholder? A vignette study to
31 564 assess recognition of cognitive biases in clinical case workups. *BMJ Qual Saf*. 2017;26(2):104-10.
32 565 35. Kerber KA, Newman-Toker DE. Misdiagnosing Dizzy Patients: Common Pitfalls in Clinical
33 566 Practice. *Neurol Clin*. 2015;33(3):565-75, viii.
34 567 36. Lipner RS, Brossman BG, Samonte KM, et al. Effect of access to an electronic medical resource
35 568 on performance characteristics of a certification examination: A randomized controlled trial. *Ann Intern*
36 569 *Med*. 2017;167(5):302-10.

570

571 **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)	Hospitalization ^{a,b}	Emergency department visit ^a	Death ^c
	Number (percent of index visits)	Number (percent of hospitalizations with a diagnostic error sensitive condition)	Number (percent of emergency department visits with a diagnostic error sensitive condition)	Number (percent 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 ^d	Not Reported ^d	121 (38.3)
Anemia	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	8,534 (17.5)	12 (2.2)	13 (2.0)	89 (28.2)
Spinal Cord Compression	6,386 (13.1)	Not Reported ^d	Not Reported ^d	36 (11.4)
Bacteremia / Sepsis	5,567 (11.4)	19 (3.5)	21 (3.2)	46 (14.6)
Appendicitis	2,584 (5.3)	Not Reported ^d	Not Reported ^d	17 (5.4)
Abscess	1,005 (2.1)	Not Reported ^d	13 (2.0)	Not Reported ^d

572 ^aCondition specific outcomes for one of the 13 diagnostic error sensitive conditions within 90 days of an
573 outpatient index visit at risk for that condition

574 ^bHospitalizations include non-elective hospitalizations either initiated through the ED or a trauma center.

575 ^cAll cause mortality within 90 days of the index visit.

576 ^dNot reported because observations were less than 11.

577

578

579 Table 2. Physician and Patient Characteristics by Diagnostic Exam Performance Tertile

	Total	Diagnosis question percent correct			P-value ^a
		Top (78.5 to 95.8)	Middle (71.4 to 78.4)	Bottom (42.9 to 71.3)	
Exam performance, Mean (standard deviation)^a					
Diagnosis question percent correct	74.5 (0.4)	84.3 (0.3)	74.8 (0.1)	65.5 (0.3)	<.001
Other question percent correct	72.6 (0.7)	80.2 (1.0)	72.1 (1.1)	66.4 (1.5)	<.001
Treatment question percent correct	77.3 (0.3)	83.4 (0.4)	77.2 (0.4)	72.0 (0.5)	<.001
Physician Characteristics, count (%)					
Female Physician	19,428 (39.9)	6,546 (43.8)	6,357 (37.5)	6,525 (39.0)	0.37
US born physician	28,462 (58.5)	9,284 (62.1)	9,932 (58.6)	9,246 (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	9,452 (19.4)	1,914 (12.8)	3,462 (20.4)	4,076 (24.4)	0.009
Small group practice (2 to 10)	20,563 (42.3)	5,543 (37.1)	7,529 (44.4)	7,491 (44.8)	0.19
Medium physicians group practice (11 to 50)	7,442 (15.3)	2,899 (19.4)	2,402 (14.2)	2,141 (12.8)	0.25
Large physician group practice (>50 physicians)	5,391 (11.1)	2,150 (14.4)	1,655 (9.8)	1,586 (9.5)	0.14
Academic practice	2,708 (5.6)	1,447 (9.7)	697 (4.1)	564 (3.4)	<.001
Other practice	3,076 (6.3)	1,005 (6.7)	1,211 (7.1)	860 (5.1)	0.59
Beneficiary characteristics					
Beneficiary Race, count (percent)					
White	40,086 (82.4)	12,652 (84.6)	13,778 (81.3)	13,656 (81.7)	0.13
Black	3,958 (8.1)	926 (6.2)	1,609 (9.5)	1,423 (8.5)	0.03
Other	4,588 (9.4)	1,380 (9.2)	1,569 (9.3)	1,639 (9.8)	0.88
Beneficiary age (per year), Mean (SD) ^a	76.6 (0.1)	76.8 (0.1)	76.5 (0.1)	76.6 (0.1)	0.23
CCW chronic conditions, count (percent)					
Alzheimer's Disease and Related Disorders or Senile Dementia	5,151 (10.6)	1,497 (10.0)	1,793 (10.6)	1,861 (11.1)	0.16
Alzheimer's Disease	2,061 (4.2)	627 (4.2)	704 (4.2)	730 (4.4)	0.82
Acute Myocardial Infarction	1,408 (2.9)	394 (2.6)	494 (2.9)	520 (3.1)	0.13
Anemia	22,450 (46.2)	6,706 (44.8)	7,766 (45.8)	7,978 (47.7)	0.11
Asthma	4,424 (9.1)	1,313 (8.8)	1,548 (9.1)	1,563 (9.3)	0.39
Atrial Fibrillation	4,225 (8.7)	1,265 (8.5)	1,478 (8.7)	1,482 (8.9)	0.69
Breast Cancer	2,485 (5.1)	779 (5.2)	831 (4.9)	875 (5.2)	0.48
Colorectal Cancer	1,139 (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	1,662 (3.4)	507 (3.4)	600 (3.5)	555 (3.3)	0.66
Cataract	31,095 (63.9)	9,601 (64.2)	10,773 (63.5)	10,721 (64.1)	0.74
Heart Failure	9,207 (18.9)	2,786 (18.6)	3,155 (18.6)	3,266 (19.5)	0.54
Chronic Kidney Disease	6,904 (14.2)	2,083 (13.9)	2,392 (14.1)	2,429 (14.5)	0.62
Chronic Obstructive Pulmonary Disease	9,108 (18.7)	2,635 (17.6)	3,165 (18.7)	3,308 (19.8)	0.02
Depression	12,042 (24.8)	3,728 (24.9)	4,145 (24.4)	4,169 (24.9)	0.83
Diabetes	13,296 (27.3)	3,947 (26.4)	4,590 (27.1)	4,759 (28.5)	0.16
Glaucoma	10,030 (20.6)	3,086 (20.6)	3,501 (20.6)	3,443 (20.6)	0.99
Hip/Pelvic Fracture	1,531 (3.1)	430 (2.9)	535 (3.2)	566 (3.4)	0.15
Hyperlipidemia	37,132 (76.4)	11,266 (75.3)	12,898 (76.1)	12,968 (77.6)	0.11
Benign Prostatic Hyperplasia	5,815 (12.0)	1,792 (12.0)	1,987 (11.7)	2,036 (12.2)	0.76
Hypertension	37,607 (77.3)	11,345 (75.8)	13,011 (76.7)	13,251 (79.3)	<.001
Hypothyroidism	11,425 (23.5)	3,490 (23.3)	3,862 (22.8)	4,073 (24.4)	0.25

Ischemic Heart Disease	18,713 (38.5)	5,616 (37.5)	6,393 (37.7)	6,704 (40.1)	0.06
Osteoporosis	14,171 (29.1)	4,372 (29.2)	4,794 (28.3)	5,005 (29.9)	0.34
Rheumatoid Arthritis	23,352 (48.0)	6,879 (46.0)	8,275 (48.8)	8,198 (49.0)	0.02
Stroke	6,255 (12.9)	1,880 (12.6)	2,212 (13.0)	2,163 (12.9)	0.70
Number of chronic conditions, count (percent)					
<=4	5,066 (10.4)	1,459 (9.8)	1,744 (10.3)	1,863 (11.1)	0.08
5 to 7	16,861 (34.7)	5,392 (36.0)	5,981 (35.3)	5,488 (32.8)	0.006
8 to 10	16,230 (33.4)	4,907 (32.8)	5,664 (33.4)	5,659 (33.8)	0.35
>=11	10,475 (21.5)	3,200 (21.4)	3,567 (21.0)	3,708 (22.2)	0.28
Mental health visit, count (percent)	6,347 (13.1)	2,040 (13.6)	2,119 (12.5)	2,188 (13.1)	0.46
Hierarchical Condition Category (HCC) score, Mean (SD) ^a	0.98 (0.01)	0.96 (0.01)	0.98 (0.01)	0.99 (0.01)	0.19
Household medium income, mean \$ (SD) ^a	59,852 (643)	61,574 (1,106)	59,113 (1,144)	59,063 (1,075)	0.19
Medicaid dual eligible, count (percent)	6,392 (13.1)	1,793 (12.0)	2,411 (14.2)	2,188 (13.1)	0.28
Rural county residence, count (percent)	7,392 (15.2)	2,207 (14.8)	2,866 (16.9)	2,319 (13.9)	0.64
Visit characteristics					
Visit with same doctor in last year, Count (percent)	37,726 (77.6)	11,369 (76.0)	13,154 (77.6)	13,203 (79.0)	0.08
Visit with any physician in last year, count (percent)	44,852 (92.2)	13,711 (91.7)	15,647 (92.3)	15,494 (92.7)	0.08
Days since last visit with any physician (if any visit in last year), Mean (SD) ^a	144.2 (0.6)	147.1 (0.8)	144.4 (1.0)	141.4 (1.3)	<.001
ED visit in prior year, count (percent)	8,101 (16.7)	2,428 (16.2)	2,879 (17.0)	2,794 (16.7)	0.43
Days since last ED visits (if ED visit in last year), Mean (SD) ^a	222.8 (0.9)	221.2 (1.5)	223.5 (1.5)	223.4 (1.5)	0.47
Hospitalization in prior year, Count (percent)	4,227 (8.7)	1,280 (8.6)	1,489 (8.8)	1,458 (8.7)	0.85
Days since last hospitalization (if hospitalization in last year), Mean (SD) ^a	229.6 (1.2)	229.1 (2.1)	229.7 (2.1)	230.1 (1.9)	0.95
Index visit diagnosis groups, Count (percent)					
Abscess	1,005 (2.1)	268 (1.8)	394 (2.3)	343 (2.1)	0.21
Anemia	12,410 (25.5)	3,817 (25.5)	4,369 (25.8)	4,224 (25.3)	0.93
Aortic aneurysm	11,491 (23.6)	3,495 (23.4)	4,165 (24.6)	3,831 (22.9)	0.18
Appendicitis	2,584 (5.3)	845 (5.6)	949 (5.6)	790 (4.7)	0.01
Bacteremia	5,567 (11.4)	1,660 (11.1)	1,929 (11.4)	1,978 (11.8)	0.83
Congestive heart failure	12,137 (25.0)	3,633 (24.3)	4,221 (24.9)	4,283 (25.6)	0.67
Acute coronary syndrome	16,228 (33.4)	4,627 (30.9)	5,740 (33.9)	5,861 (35.1)	0.02
Depression	12,637 (26.0)	3,932 (26.3)	4,312 (25.4)	4,393 (26.3)	0.78
Fracture	13,409 (27.6)	4,324 (28.9)	4,364 (25.7)	4,721 (28.2)	0.11
Pulmonary embolism	8,534 (17.5)	2,683 (17.9)	2,984 (17.6)	2,867 (17.1)	0.71
Pneumonia	12,183 (25.1)	3,773 (25.2)	4,224 (24.9)	4,186 (25.0)	0.97
Spinal cord compression	6,386 (13.1)	1,985 (13.3)	2,218 (13.1)	2,183 (13.1)	0.94
Stroke	10,026 (20.6)	3,003 (20.1)	3,542 (20.9)	3,481 (20.8)	0.79

580 ^aP-values and standard deviation accounted for correlated errors within physicians

Table 3. Associations with diagnostic knowledge and adverse events per 1,000 index visits

Diagnostic knowledge tertile	Death ^a				Emergency department visit ^b				Hospitalization ^c			
	Unadjusted ^e	Regression adjusted ^{d,e}			Unadjusted ^e	Regression adjusted ^{d,e}			Unadjusted ^e	Regression adjusted ^{d,e}		
	Events per 1,000 visits (95% CI interval)	Events per 1,000 visits (95% CI interval)	Difference (95% CI)	P-value	Events per 1,000 visits (95%CI)	Events per 1,000 visits (95%CI)	Difference (95% CI)	P-value	Events per 1,000 visits (95% CI)	Events per 1,000 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	

^aAll cause mortality within 90 days of an outpatient index visit with a diagnosis at risk for one of 3 diagnostic error sensitive conditions.

^bEmergency 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.

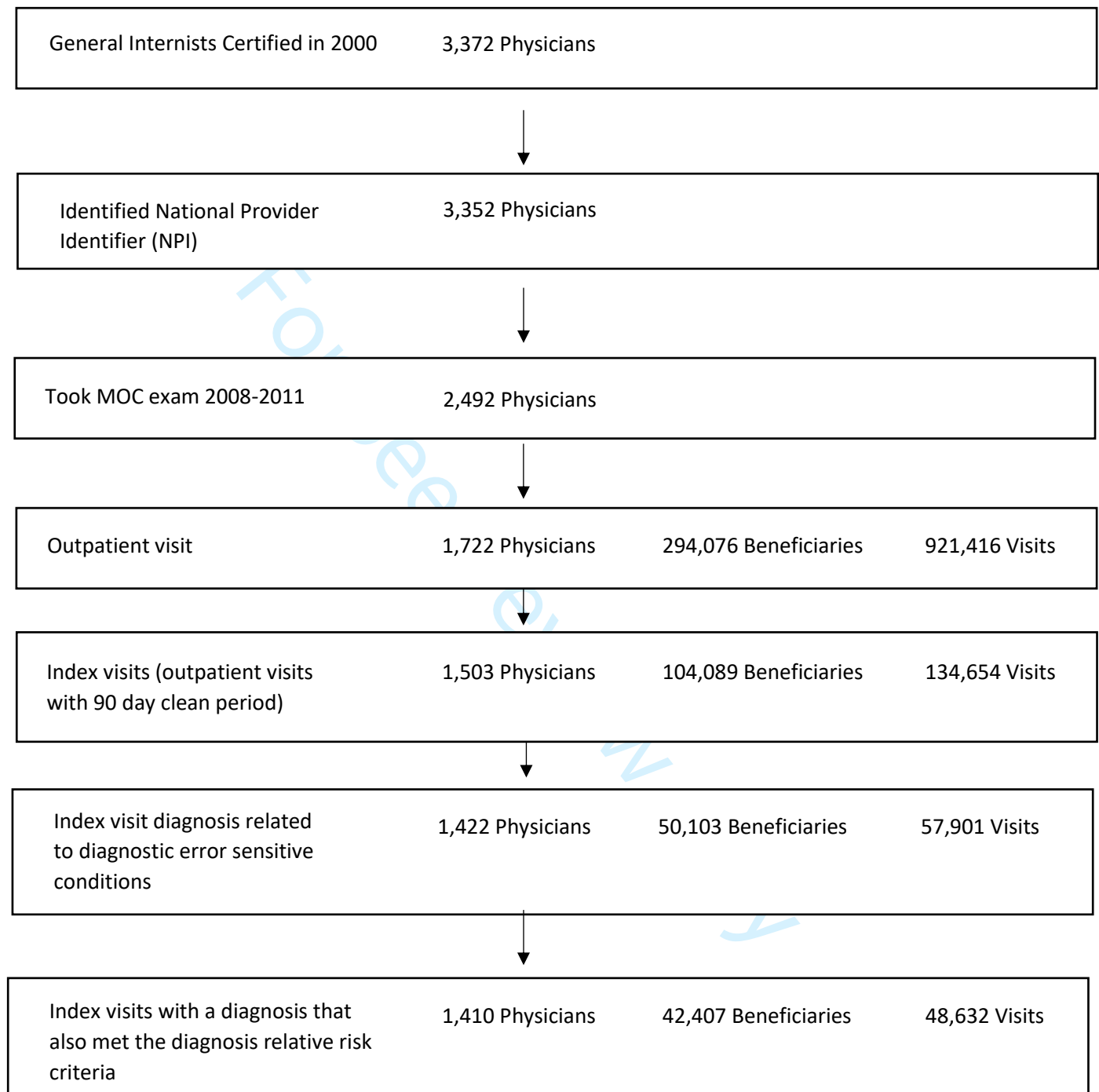
^cHospitalizations were for non-elective hospitalizations either initiated through the ED or a trauma center with a discharge diagnosis for one of 13 diagnostic error sensitive conditions within 90 days of the index visit with a diagnosis at risk for that condition.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

FIGURE LEGEND:

Figure 1. Sample Selection

For peer review only

Figure 1. Sample Selection

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: Description	Page
Section 1: 90 day Clean Period Derivation	2
Section 2: Outcome Condition and Index Visit Eligibility Diagnoses Codes and Relative Risk.....	5
Section 3: Psychometric Analysis of Whether Diagnosis Related Questions Reflect an Underlying Construct	10
Section 4: Imputations for missing variables.....	11
Section 5: Full Regression Coefficient Estimates and Explanatory Variable List.....	12
Section 6: Regression Sensitivity Analyses.....	15
References	21

Section 1: 90-day Index Visit Clean Period Derivation

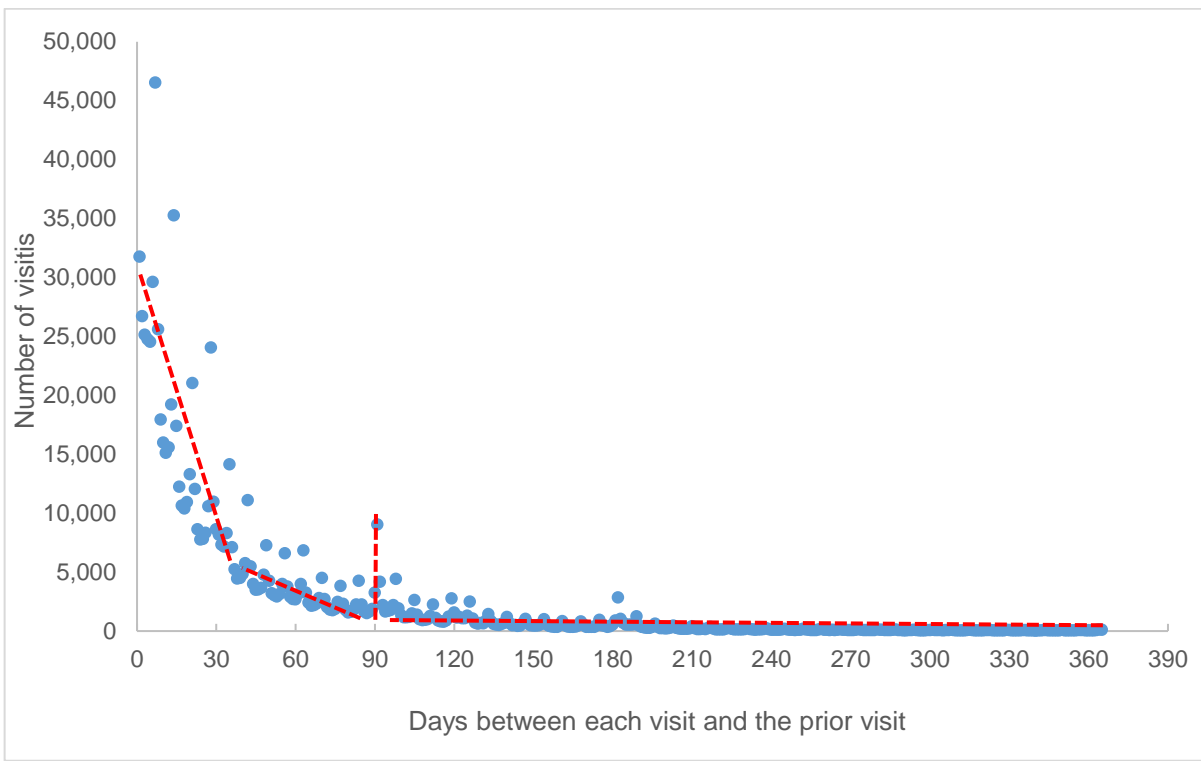
Figure 1.1 displays the visit periodicity between each of the 921,416 visits to an internist in the sample and the most recent visit prior to that one.

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.

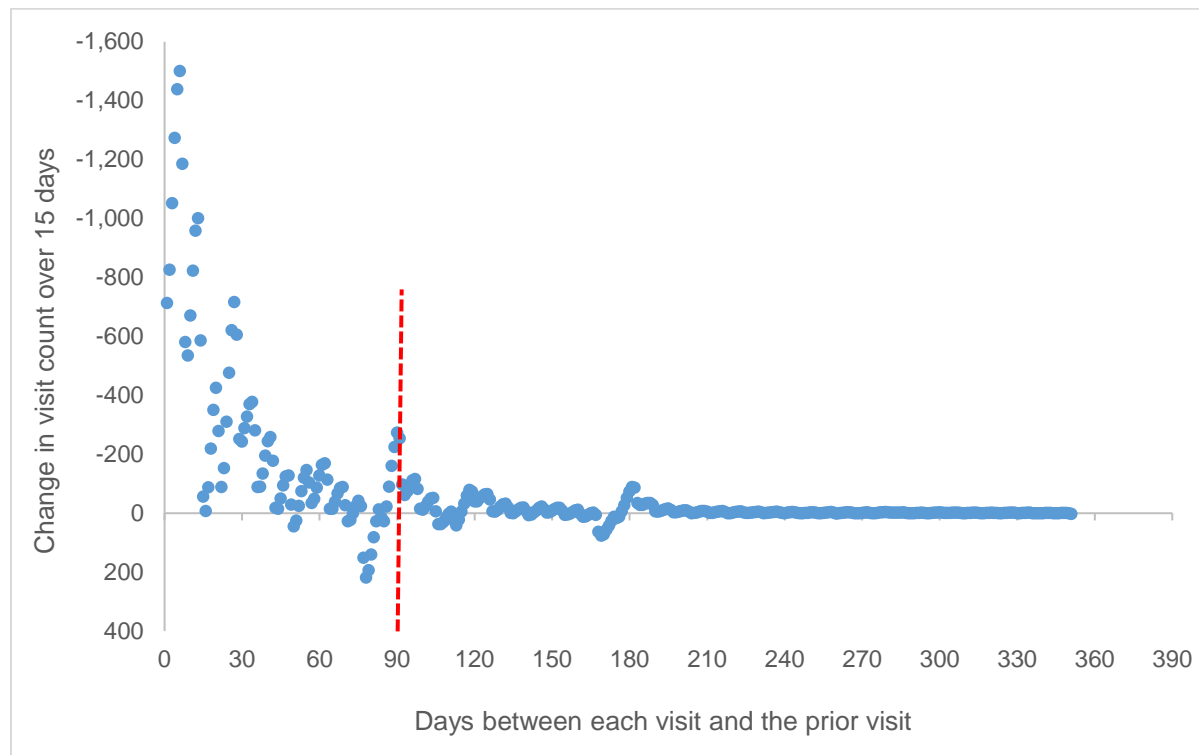
This flattening of the slope after 90 days is more clearly displayed in Figure 1.2 which displays the 15 day moving average of the change in visit counts per day (i.e., the changing slope). Here the slope stabilizes at about zero beginning around 90 days suggesting that a 90 day clean period for physician 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



Review only

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



Review only

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

Table 2.3 Relative Risks for each Index Visit Diagnosis

Outcome conditions ^a	Index visit eligibility diagnosis group	Relative Risk ^b	Outcome conditions ^a	Index visit eligibility diagnosis group	Relative Risk ^b	
Abscess	Fever	2.65	Acute Coronary Syndrome	Chest pain	8.38	
	Chills	0.00		Dyspnea	7.29	
Anemia	Gastrointestinal bleeding	25.20		Shortness of breath	3.65	
	Weight loss	4.09		Hypoxemia/hypoxia	2.01	
	Shortness of breath	3.51		Reflux disease	1.23	
	Weakness/Fatigue	2.35		Esophageal disease	1.22	
	Hypoxemia/Hypoxia	2.11		Weakness/Fatigue	1.14	
	Dyspnea	2.05		Nausea	1.05	
	Chest Pain	1.82		Other respiratory issue	0.86	
	Headache	1.29		Respiratory distress	0.00	
Aortic Aneurysm	Menorrhagia	0.00		Gastritis	0.00	
	Dyspnea	4.98		Heart Burn	0.00	
	Abdominal pain	4.93		Depression	Delirium	32.76
	Shortness of breath	3.80			Heart failure	6.16
	Chest pain	2.42	Anxiety		5.04	
	Other back pain	1.64	Dysthymia		4.99	
	Back pain	1.01	Weight loss		4.73	
Elevated blood pressure	0.00	Anemia	2.74			
		Fatigue	1.06			
Appendicitis	Vomiting	30.79	Alcohol		0.00	
	Diverticulitis	30.45	Amphetamines		0.00	
	Nausea	16.81	Cannabis		0.00	
	Abdominal pain	15.60	Cocaine		0.00	
	Abdominal tenderness	0.00	Drug Mental Disease		0.00	
	Fever	0.00	Hallucinogens		0.00	
Bacteremia/Sepsis	Vomiting	6.99	Opioids	0.00		
	Fever	5.10	Personality disorder	0.00		
	Nausea	3.82	related alcohol disease	0.00		
	Tachycardia	2.67	Sedatives	0.00		
	Weakness/Fatigue	1.75	Stress	0.00		
Heart failure	Hypoxemia/Hypoxia	9.99	Weight gain	0.00		
	Shortness of breath	5.09	Mood disorder	0.00		
	Dyspnea	3.33	Fracture	Gait instability	2.53	
	Edema	3.27		Edema	1.79	
	Chest Pain	2.46		Osteoporosis	1.66	
	Weakness/Fatigue	1.42		Hyperparathyroidism	1.09	
	Ascites	0.00		Vitamin D deficiency	1.08	
	Respiratory Distress	0.00				

Outcome conditions ^a	Index visit eligibility diagnosis group	Relative Risk ^b	Outcome conditions ^a	Index visit eligibility diagnosis group	Relative Risk ^b
Fracture (con't)	Osteopenia	0.54	Spinal cord compression	Abdominal pain	31.20
	Celiac disease	0.00		Back pain	15.03
	Falls	0.00		Peripheral neuropathy	0.00
	Stress fracture	0.00		Weakness/Fatigue	0.00
Pulmonary embolism	Tachycardia	12.16	Stroke	Facial weakness	65.24
	Hypoxemia/hypoxia	10.98		Confusion	48.93
	Shortness of breath	6.75		Speech disturbance	19.60
	Dyspnea	6.54		Transient ischemic attack	7.82
	Abnormal respiration	6.35		Delirium	4.96
	Heart failure	4.51		Dizziness	3.20
	Chest pain	4.31		Lack coordination	2.92
	Cough	1.48		Gait instability	2.92
	Other respiratory issue	1.34		Vomiting	2.15
	Deep vein thrombosis	0.00		Weakness/Fatigue	1.54
	Respiratory distress	0.00		Headache	1.37
	Fever	0.00		Nausea	1.17
	Heart burn	0.00		Thunderclap headache	0.00
	Hemoptysis	0.00		Tension headache	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				

^aOutcomes include any hospitalization or emergency department visit for the diagnostic conditions within 90 days of the index visits including same day events

^bIndex 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.

For peer review only

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

Label	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	
	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

1							
2							
3	Asthma	-0.8 (1.1)	0.45	-0.2 (1.4)	0.88	0.1 (1.7)	0.95
4	Atrial Fibrillation	1.8 (1.1)	0.10	3.4 (1.5)	0.02	3.7 (1.7)	0.03
5	Breast Cancer	-2.0 (1.3)	0.13	-3.0 (1.8)	0.10	-2.0 (2.2)	0.35
6	Colorectal Cancer	-0.5 (1.7)	0.76	-3.4 (2.0)	0.10	-1.9 (2.5)	0.44
7	Endometrial Cancer	1.6 (4.2)	0.71	-1.1 (4.7)	0.82	-0.9 (5.5)	0.87
8	Lung Cancer	3.7 (3.4)	0.28	13.3 (6.2)	0.03	10.3 (6.2)	0.10
9	Prostate Cancer	-1.6 (1.4)	0.26	4.3 (2.5)	0.09	3.6 (2.8)	0.20
10	Cataract	-1.0 (0.9)	0.29	1.1 (1.0)	0.31	0.1 (1.2)	0.96
11	Heart Failure	1.8 (1.0)	0.08	2.9 (1.2)	0.02	3.3 (1.4)	0.02
12	Chronic Kidney Disease	-0.7 (0.8)	0.39	2.9 (1.2)	0.02	4.2 (1.4)	0.004
13	Chronic Obstructive Pulmonary Disease	1.1 (0.9)	0.26	1.3 (1.2)	0.27	0.9 (1.3)	0.50
14	Depression	0.2 (0.9)	0.79	0.9 (1.1)	0.43	0.7 (1.2)	0.59
15	Diabetes	0.0 (0.8)	0.99	3.3 (1.1)	0.003	2.4 (1.2)	0.04
16	Glaucoma	-0.9 (0.8)	0.29	0.0 (1.1)	1.00	-0.3 (1.2)	0.80
17	Hip/Pelvic Fracture	1.4 (1.6)	0.39	3.0 (2.4)	0.20	5.0 (2.8)	0.07
18	Hyperlipidemia	-1.8 (1.1)	0.09	-1.1 (1.3)	0.39	-0.7 (1.4)	0.63
19	Benign Prostatic Hyperplasia	-0.3 (1.1)	0.79	0.0 (1.5)	0.98	-0.8 (1.7)	0.63
20	Hypertension	0.8 (1.1)	0.46	2.4 (1.4)	0.09	1.2 (1.6)	0.44
21	Hypothyroidism	0.7 (0.8)	0.42	1.4 (1.1)	0.19	1.8 (1.2)	0.14
22	Ischemic Heart Disease	0.8 (0.9)	0.39	0.1 (1.1)	0.92	-1.3 (1.2)	0.29
23	Osteoporosis	-1.9 (0.8)	0.02	2.0 (1.2)	0.09	1.4 (1.3)	0.28
24	Rheumatoid Arthritis	-2.3 (0.8)	0.005	-0.2 (1.0)	0.87	0.3 (1.1)	0.81
25	Stroke	2.7 (1.1)	0.02	2.7 (1.3)	0.04	2.8 (1.5)	0.06
26	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
27	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
28	Hospitalization in prior year	8.8 (5.4)	0.10	0.9 (4.1)	0.84	0.4 (4.5)	0.93
29	ED visit in prior year	1.0 (2.5)	0.69	7.3 (4.0)	0.07	8.4 (4.6)	0.07
30	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
31	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
32	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
33	Index visit diagnosis group indicators						
34	Pulmonary embolism	-0.5 (1.3)	0.68	6.1 (1.4)	<.001	7.1 (1.6)	<.001
35	Acute coronary syndrome	-1.3 (1.1)	0.25	-3.0 (1.8)	0.11	-5.5 (2.0)	0.007
36	Stroke	-2.3 (1.4)	0.10	7.4 (1.5)	<.001	7.7 (1.7)	<.001
37	Congestive heart failure	0.2 (1.3)	0.88	10.1 (1.6)	<.001	12.1 (1.7)	<.001
38	Fracture	-1.3 (1.1)	0.22	3.2 (1.3)	0.02	5.1 (1.5)	<.001
39	Abscess	0.7 (2.3)	0.77	6.4 (3.3)	0.05	11.7 (3.3)	<.001
40	Pneumonia	2.3 (1.2)	0.05	5.6 (1.4)	<.001	6.7 (1.6)	<.001
41	Aortic aneurysm	1.0 (1.4)	0.50	-0.6 (2.0)	0.76	0.7 (2.2)	0.74
42	Appendicitis	2.0 (1.8)	0.28	5.9 (3.0)	0.05	9.6 (3.1)	0.002
43	Depression	0.0 (1.3)	0.99	3.0 (1.5)	0.05	2.4 (1.7)	0.15
44	Anemia	2.3 (1.1)	0.04	3.5 (1.8)	0.04	3.2 (2.0)	0.11
45	Bacteremia	0.5 (2.5)	0.85	-9.5 (3.0)	0.001	-8.3 (3.1)	0.008
46	Spinal cord compression	-0.5 (1.8)	0.79	-2.8 (2.8)	0.32	-7.0 (3.1)	0.02
47	Mental health visit	1.4 (1.2)	0.22	-0.9 (1.5)	0.53	0.1 (1.8)	0.97
48	HHS Region						
49	HHS Region 1	Reference		Reference		Reference	
50	HHS Region 2	1.6 (1.7)	0.35	-5.2 (2.2)	0.02	-6.7 (2.7)	0.01
51	HHS Region 3	2.7 (1.8)	0.12	2.1 (2.5)	0.40	1.3 (3.0)	0.66
52	HHS Region 4	0.4 (1.5)	0.77	-2.7 (2.2)	0.22	-4.9 (2.6)	0.07
53	HHS Region 5	0.3 (1.4)	0.81	0.8 (2.1)	0.69	-1.0 (2.6)	0.70
54	HHS Region 6	-0.9 (1.5)	0.53	-2.8 (2.2)	0.21	-4.4 (2.8)	0.11
55	HHS Region 7	0.0 (2.2)	0.99	3.2 (3.2)	0.31	0.9 (3.5)	0.79
56	HHS Region 8	-1.6 (2.2)	0.47	1.9 (3.8)	0.62	-2.0 (3.8)	0.61
57	HHS Region 9	0.0 (1.6)	0.99	-0.6 (2.5)	0.81	-3.2 (2.8)	0.26
58	HHS Region 10	-0.6 (2.2)	0.77	4.4 (3.5)	0.21	4.0 (4.3)	0.35
59	Study Year						
60	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:

^aMissing practice characteristics (1,432 or 2.94% of sample) were coded as “other unknown”.

^bMissing HCC (86 or .18% of sample) were replace by in sample mean HCC.

^cMissing rural indicator (22 or .05% of sample) were assumed to be non-rural

^bMissing 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).

1
2
3 For the first sensitivity we expanding the eligible diagnoses code groups to all 76 identified by
4 physician authors versus 38 in the base case that also met the relative risk criteria.
5
6
7

8 For the third sensitivity we expand the index visit clean period to 97 days and contracted the
9 index visit clean period to 83 days.
10
11
12

13 For the fourth sensitivity, we excluded physician in academic medical centers to consider the
14 possibility that the unobserved physician characteristics related to where they worked or who
15 they worked with could be were independently both related to the underlying physician
16 diagnostic skill and our outcome measures.
17
18
19
20

21 For the fifth sensitivity we accounted for the possibility that adverse outcomes were avoided
22 because the patient died by altering the ED and hospitalization measures to include all-cause
23 mortality. For this sensitivity we added the following two outcome measures: base case
24 hospitalization or death and base case ED or death.
25
26
27

28 Overall results of robustness sensitivity analysis suggests that our base case results were not
29 highly sensitive to different underlying assumptions related to these factors (e.g., across all
30 robustness sensitivities percent change in the outcome measures between top versus bottom
31 diagnostic knowledge exam performers remained statistically significant ($P < 0.05$)).
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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:	7,966 ^a	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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

References

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

For peer review only

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract [Done]
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found [See abstract starting on line 25]
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported [See Introduction section of the Manuscript first paragraph]
Objectives	3	State specific objectives, including any prespecified hypotheses [See last paragraph of first paragraph Introduction Section]
Methods		
Study design	4	Present key elements of study design early in the paper [See Methodology section]
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection [See Physician and Index Visit sample subsections, (starting on line 122)]
Participants	6	(a) <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants [See Methodology section and Figure 1 for physician, patient and visit sample stats]) (b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed [NA] <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case [NA]
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable [See Outcome Measures subsection of Mythology section starting on line 166 for Outcomes, starting on line 188 for diagnostic measure.
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group [See line 188 for diagnostic knowledge measure]
Bias	9	Describe any efforts to address potential sources of bias [See Statistical Methodes starting on line 208 first paragraph explanation for inclusion of other measures of knowledge and Sensitivity Analysis subsection start on line 237 and results of sensitivity analysis starting on line 308]
Study size	10	Explain how the study size was arrived at [See first and second paragraph page 6]
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why [See Statistical Methods section]
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding [See Statistical Methods first paragraph] (b) Describe any methods used to examine subgroups and interactions [NA] (c) Explain how missing data were addressed [See section 4 in the Supplement and reference to this in Table 3] (d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Case-control study—If applicable, explain how matching of cases and controls was addressed [NA]

Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy [Sensitivity Analysis subsection starting on line 238]

(e) Describe any sensitivity analyses [Sensitivity Analysis subsection starting on line 238]

Continued on next page

For peer review only

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed [See Figure 1 and first paragraph of the Results section] (b) Give reasons for non-participation at each stage [See Figure 1, Table 1 and first paragraph of the Results section] (c) Consider use of a flow diagram [See Figure 1]
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders [See Results section paragraph 2 and Table 2] (b) Indicate number of participants with missing data for each variable of interest [NA] (c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount) [NA]
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time [NA] <i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure [NA] <i>Cross-sectional study</i> —Report numbers of outcome events or summary measures [See Table 1]
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included [See Table 3 and Supplementary Material for full set of controls and coefficients (Section 6)] (b) Report category boundaries when continuous variables were categorized [See Statistical Analysis Section and Supplementary Material Section 4] (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period [See Results section and Table 3]
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses [See Sensitivity Analysis subsection of Results section and Supplementary Material Section 5]
Discussion		
Key results	18	Summarise key results with reference to study objectives [See Discussion subsection last paragraph of page 14]
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias [See several paragraphs in the Discussion section starting on line 363 to 420]
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence [No other study has addressed our research question, however, in terms of methodology we compare our study to other in the Discussion section starting on line 341]
Generalisability	21	Discuss the generalisability (external validity) of the study results [See study limitation bullet points after the abstract, lines 357 to 412 of the Discussion Section]
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based [See Funding section last paragraph]

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

1
2 **Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and
3 published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely
4 available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at
5 <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is
6 available at www.strobe-statement.org.
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

BMJ Open

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

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-041817.R2
Article Type:	Original research
Date Submitted by the Author:	04-Mar-2021
Complete List of Authors:	Gray, Bradley; American Board of Internal Medicine Vandergrift, Jonathan; American Board of Internal Medicine McCoy, Rozalina; Mayo Clinic, Division of Endocrinology, Department of Medicine Lipner, Rebecca; American Board of Internal Medicine Landon, Bruce; Harvard Medical School, Department of Health Care Policy
Primary Subject Heading:	General practice / Family practice
Secondary Subject Heading:	Diagnostics, Medical education and training
Keywords:	INTERNAL MEDICINE, MEDICAL EDUCATION & TRAINING, GENERAL MEDICINE (see Internal Medicine)

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1
2
3 1 The Association Between Primary Care Physician Diagnostic Knowledge and Death,
4
5 2 Hospitalization and Emergency Department Visits Following an Outpatient Visit at Risk for
6
7
8 3 Diagnostic Error: A Retrospective Cohort Study Using Medicare Claims
9
10
11 4
12
13

14 5 Bradley M. Gray, PhD, corresponding author

15 6 Email: bgray@abim.org, Phone: 202-213 6646, FAX 202-213 6646

17 7 American Board of Internal Medicine

19 8 510 Walnut Street, Suite 1700, Philadelphia, Pennsylvania 19106, USA
20
21 9

23 10 Jonathan L. Vandergrift, MS

25 11 American Board of Internal Medicine Philadelphia, Pennsylvania, USA
26
27 12

29 13 Rozalina, G. McCoy, MD

31 14 Mayo Clinic, Rochester, Minnesota, USA
32
33 15

34 16 Rebecca S. Lipner, PhD

36 17 American Board of Internal Medicine Philadelphia, Pennsylvania, USA
37
38 18

40 19 Bruce E. Landon, MD

42 20 Harvard Medical School, Boston, Massachusetts, USA
43
44 21

45 22 Word count: 4,678,

47 23 Keywords: Internal Medicine, General Medicine, Medical Education & Training,
48
49
50
51
52
53
54
55
56
57
58
59
60

24 **Objective**

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

31 **Setting/Participants**

32 1,410 US general internists who recently took their American Board of Internal Medicine
33 Maintenance of Certification (ABIM-IM-MOC) exam treating 42,407 Medicare beneficiaries
34 who experienced 48,632 “index” outpatient visits for new complaints at risk for diagnostic error
35 because the presenting complaint (e.g., dizziness) was related to pre-specified diagnostic error
36 sensitive conditions (e.g. stroke).

37 **Outcome measures**

38 90-day risk of all-cause death, and, for outcome conditions related to the index visits diagnosis,
39 emergency department (ED) visits and hospitalizations.

40 **Design**

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

45 **Results**

46 Rates of 90-day adverse outcomes per 1,000 index visits were 7 for death, 11 for
47 hospitalizations, and 14 for ED visits. Being seen by a physician in the top versus bottom third
48 of diagnostic knowledge during an index visit for a new complaint at risk for diagnostic error
49 was associated with 2.9 fewer all-cause deaths (95% confidence interval (CI) -5.0 to -0.7,
50 P=.008), 4.1 fewer hospitalizations (95% CI -6.9 to -1.2, P=0.006), and 4.9 fewer ED visits (95%
51 CI -8.1% to -1.6%, P=0.003) per 1,000 visits.

52 **Conclusion**

53 Higher diagnostic knowledge was associated with lower risk of adverse outcomes after visits for
54 complaints at heightened risk for diagnostic error.

55

56 **Strengths and limitations of this study**

- 57 ○ Unique diagnostic knowledge measure linking diagnostic knowledge with adverse
58 outcomes
- 59 ○ Scalable adverse outcome measures and extensive sensitivity analyses
- 60 ○ Our assessment of diagnostic error is indirect (as indicated by adverse outcomes)
- 61 ○ Results are subject to selection bias if the mix of index visits or the severity of the
62 patients or practice support differed for physicians with different levels of
63 diagnostic knowledge.
- 64 ○ Results are only generalizable to physicians who elected to attempt ABIM's
65 certification exam and were about 10 years past initial certification and patients
66 older than 65.

Introduction

71
72
73 Diagnostic error has been identified as a key health care delivery concern and contributes to
74 significant potentially preventable morbidity and mortality.(1-3) Ambulatory care, and
75 especially primary care, is a practice setting with a particularly high risk for diagnostic error(4,
76 5) because of the wide variety of presentations encountered and the concomitant difficulty of
77 distinguishing harmful conditions from routine self-limited complaints, compounded by the well-
78 known time constraints faced by practitioners in that setting. It has been estimated that at least
79 5% of ambulatory visits are associated with diagnostic error, half of which may result in
80 considerable patient harm. Diagnostic error is a common cause of malpractice suits and most
81 frequently occurs in the ambulatory care settings.(6, 7)

82
83 Deficiencies in diagnostic knowledge are likely to be an important contributor to these diagnostic
84 errors that could impact, for example, the breadth of diagnoses considered, appropriate ordering
85 and interpretation of tests, and/or synthesis of data more generally.(8-11) Because of this,
86 measuring physician diagnostic knowledge has become a major focus of organizations
87 throughout the developed world that are tasked with licensing and certifying physicians with the
88 underlying, although largely untested, hypothesis being that diagnostic knowledge will be a
89 measurable and strong predictor of diagnostic error.(12-15) Testing this hypothesis and
90 quantifying this relationship is therefore a critical public policy concern both in terms of the
91 importance of board certification and other programs designed to enhance lifelong learning for
92 physicians.

93

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

105

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

113

114

115 **Methods**

116

117 ***Physician and Index Visit Sample***

118 Our physician sample included general internists who were initially ABIM board certified in
119 2000 and took their IM-MOC exam between 2008 and 2011 (Figure 1). We identified Medicare
120 beneficiary outpatient Evaluation & Management (E&M) visits with these physicians using their
121 National Provider Identifier (NPI) during the calendar year following their exam (2009 to 2012).
122 These patients were age 65 or older and continuously enrolled in Medicare fee-for-service
123 (Medicare insures most of the US population over 65) during the physician's one year follow-up
124 period and the year prior. To ensure that any presenting complaints being evaluated were new
125 (i.e., not follow up), we restricted these visits to those that were the first visit for a new complaint
126 (the "*index visit*") because these visits were preceded by a 90-day clean period with no previous
127 inpatient or outpatient visit. The 90-day clean period is consistent with the US government
128 Centers for Medicare and Medicaid Services criteria used by its Bundled Payments for Care
129 Improvement Program for defining new episodes of care and with the patterns of visits we
130 observed (see Appendix Section 1 for related analysis).(22, 23)

131

132 We further restricted these index visits to those at heightened risk for diagnostic errors because
133 the recorded diagnosis in the Medicare claims (the "*index visit diagnosis*"), which includes
134 recording of symptom (e.g. loss of balance), could have been the initial presenting complaint for
135 one or more of 13 pre-specified diagnostic error sensitive conditions such as congestive heart
136 failure or bacteremia/sepsis (see Table 1). These 13 conditions (see Appendix Section 2 for a list

1
2
3 137 and applicable ICD-9 codes) were an acute non-cancerous subset of 20 conditions previously
4
5 138 noted by Schiff et al. to be at high risk for serious diagnostic error.(24) For instance, index visits
6
7
8 139 with diagnosis codes for chest pain, dyspepsia, shortness of breath, hypoxemia/hypoxia,
9
10 140 respiratory distress, weakness/fatigue, edema or ascites could all be the initial presentation of
11
12 141 congestive heart failure, which is one of the 13 diagnostic error sensitive conditions.
13
14

15 142
16
17
18 143 We used a three-step process to identify eligible index visit diagnoses. First, two physician
19
20 144 authors (RGM and BEL) identified all diagnoses that could be presenting complaints for the 13
21
22 145 diagnostic error sensitive conditions: what complaints/diagnoses might someone who ultimately
23
24 146 presented with a diagnostic-error sensitive condition have presented with initially? Second,
25
26 147 because the original list of identified index visit diagnoses was large (76), we reduced this list to
27
28 148 38 by applying a relative risk (RR) criteria. For a specific index visit diagnosis to meet this
29
30 149 criteria, all index visits with that diagnosis had to have a greater portion of later ED visits or
31
32 150 hospitalizations with the related outcome condition discharge diagnosis than index visits where
33
34 151 the specific at risk diagnosis was not present. For example, dizziness was chosen as an eligible
35
36 152 index visit diagnosis for stroke, one of the diagnostic error sensitive conditions, both because it
37
38 153 was identified as a potential presenting symptom of a stroke by physician authors and because
39
40 154 index visits with that diagnosis had a greater proportion of later hospitalization or ED visits for
41
42 155 stroke than visits without this diagnosis. Third, we also included index visits where the actual
43
44 156 diagnosis was one of the 13 diagnostic error sensitive conditions because we wanted to include
45
46 157 cases where diagnostic errors were and were not made. Therefore, we also included index visits
47
48 158 with a diagnosis of congestive heart failure itself as being at risk for the underlying condition
49
50 159 congestive heart failure.
51
52
53
54
55
56
57
58
59
60

160 ***Outcome Measures***

161 We examined the risk of three serious adverse outcomes within 90 days of the index visit that we
162 hypothesized would occur more frequently in cases of misdiagnosis: all-cause mortality,
163 hospitalizations, and ED visits. We did not count these events as adverse outcomes if they
164 occurred on the same day as the index visit because this may reflect a positive action (the
165 physician correctly diagnosed a patient with stroke and referred/admitted them to the hospital) or
166 be unavoidable regardless of the accuracy of the index visit diagnosis (the patient died despite
167 immediately admitting the patient to the hospital who exhibited stroke symptoms). Based on
168 Medicare billing codes, hospitalizations were limited to non-elective hospitalizations initiated
169 through the ED or trauma center. The ED and hospitalization outcomes were also limited to
170 cases where the discharge diagnosis was for one of the 13 diagnostic error sensitive conditions
171 following an index visit with the applicable diagnosis. We therefore presumed that these
172 discharge diagnoses were a reasonable representation of the underlying condition of the patient
173 at the time of the index visit. For example, we would count a hospitalization with a discharge
174 diagnosis of stroke as an adverse outcome if it occurred after an index visit for dizziness because
175 dizziness was identified as being a potential presenting complaint for stroke. However, we did
176 not count hospitalizations with a discharge diagnosis for acute coronary syndrome following an
177 index visit for dizziness because dizziness was not identified as a presenting complaint for acute
178 coronary syndrome. The rationale is that if there were no presenting complaints during the index
179 visit related to coronary syndrome, either because the underlying condition was not present or
180 could not be detected at the time of the index visit, then the index visit physician could not have
181 prevented the hospitalization regardless of their diagnostic knowledge.

182 ***Measure of Diagnostic Knowledge***

1
2
3 183 Our measure of diagnostic knowledge was calculated as the percent of correct answers on the
4
5 184 IM-MOC exam for questions previously coded as “diagnosis-related” by ABIM’s IM-MOC
6
7
8 185 exam committee. In our study, these questions comprised 53% of all IM-MOC exam questions,
9
10 186 with the remaining 42% addressing treatment and 5% related to other topics such as
11
12 187 epidemiology or pathophysiology. More generally, exam questions are designed to replicate real
13
14 188 world clinical scenarios and/or patient encounters and without reliance on rote memorization.(25,
15
16
17 189 26)

18
19
20 190
21
22
23 191 The ABIM exam committee coded each question based on the primary function tested to assure
24
25 192 that the exam covers care typically rendered by outpatient primary care physicians. Questions
26
27 193 coded as “diagnosis related” typically test knowledge and skills related to diagnostic inference,
28
29 194 differential diagnosis, and diagnostic testing and therefore are measuring diagnostic knowledge
30
31 195 and related decision-making. Psychometric analysis indicates that scores on diagnosis related
32
33 196 exam questions were meaningfully correlated (i.e., Cronbach’s alpha score of 0.84), and thereby
34
35 197 represent an independent underlying construct that could be interpreted as diagnostic knowledge
36
37 198 (see Appendix Section 3 for more details).(27) Similarly, this analysis indicated that questions
38
39 199 coded as treatment related also represent an independent underlying construct (i.e., Cronbach’s
40
41 200 alpha score of 0.75). Although performance on diagnosis and treatment related questions were
42
43 201 correlated (Pearson Correlation=0.62), 59.5% of the variation in diagnosis exam performance for
44
45 202 the physician study sample was not explained by performance on other parts of the exam.
46
47
48
49
50

51 203 ***Statistical methods***

1
2
3 204 Using Probit regression we estimated the associations with each adverse outcome, with standard
4
5 205 errors adjusted for correlations resulting from the nesting of visits within patients within
6
7 206 physicians.(28, 29) To measure associations with diagnostic knowledge we included categorical
8
9 207 regression explanatory variables for top and middle third of percent correct scores on diagnosis
10
11 208 related questions (bottom third was the reference category). Other exam level explanatory
12
13 209 variables included tertile indicators for performance on treatment-related questions and
14
15 210 performance on other question types. Since these variables measure knowledge unrelated to
16
17 211 diagnosis, they account for correlations between factors such as unmeasured practice or patient
18
19 212 characteristics that might be correlated with exam performance and our outcome measures (e.g.,
20
21 213 high scoring physicians may be more likely to practice in an academic setting or other such
22
23 214 settings that might be independently related to diagnostic error). Exam form indicators accounted
24
25 215 for differences in exam difficulty across exam administrations.
26
27
28
29
30

31 216
32
33
34 217 We also included physician, patient and visit level regression controls. Physician level controls
35
36 218 included: practice size (indicators for solo practice and practices larger than 50 physicians),
37
38 219 practice type (indicators for academic, group), demographic (gender), and training characteristics
39
40 220 (medical school location interacted with country of birth). Patient level controls included:
41
42 221 demographic characteristics (age and age squared, gender and race/ethnicity indicators) and a
43
44 222 Medicaid eligibility indicator. Lagged patient risk adjusters included 27 indicators for chronic
45
46 223 conditions and Medicare's Hierarchal Condition Category (HCC) risk adjustment score. We
47
48 224 imputed values for a small number of missing values for controls (see Appendix Section 4).
49
50
51 225 Patient index visit location level controls included: an indicator for residing in a rural ZIP code,
52
53 226 ZIP code median household income, and indicators for 10 US Health and Human Services
54
55
56
57
58
59
60

1
2
3 227 regions. Index visit level controls included: indicators of any outpatient visit, hospitalization or
4
5 228 ED visits within the prior year and number of days since the most recent of these events, visit
6
7
8 229 year indicators to control for secular changes in quality. We also included an indicator for
9
10 230 whether or not the patient had a previous contact with the index visit physician during the year
11
12 231 prior to the index visit to account for differences in physician-patient continuity (see Appendix
13
14 232 Section 5 for a full list of controls).

17 233 *Sensitivity Analysis*

19
20 234 We performed numerous sensitivity analyses to test the robustness of our results (detailed in
21
22 235 Appendix Section 6). First, we expanded the index visit sample to include all index visits with
23
24 236 the original 76 diagnoses identified by the physician authors regardless of whether they met the
25
26 237 relative risk criteria. Second, we expanded and contracted the index visit clean period by seven
27
28 238 days. Third, excluded hospitalizations or ED events occurring the day after the index visit, in
29
30 239 addition to same day events, to consider the possibility that they might be triggered by a correct
31
32 240 diagnosis and therefore should not have been considered adverse outcomes. Fourth, we
33
34 241 considered the possibility that our results were biased due to omitted variables correlated with
35
36 242 practice size. For example, it could be that physicians in large practices have greater access to
37
38 243 specialists or other physicians for informal consultations than those in small practices and
39
40 244 therefore outcomes for these physicians may be less sensitive to their knowledge. To examine
41
42 245 this possibility, we estimated associations with knowledge and our two utilization measures
43
44 246 across a sample of physicians in either small (≤ 10 physicians, 54.5% (768/1,410) of physicians)
45
46 247 or large practices (> 50 or in academic medical centers, 23.7% (334/1,410) of physicians). We did
47
48 248 not conduct these sensitivities for death because there were too few deaths in the subgroups to
49
50 249 allow us to reliably estimate the associations (e.g., 39 deaths for physicians in large
51
52
53
54
55
56
57
58
59
60

1
2
3 250 practices).Fifth, to consider the possibility that these outcomes were only avoided because the
4
5 251 patient died, for the ED and hospitalization outcome, we also included instances where the
6
7 252 patient died. Sixth, as a falsification test we limited the index visits to those that were unrelated
8
9 253 to the 13 diagnostic error sensitive conditions. Under this sensitivity, we expected then that the
10
11 254 associations with diagnostic knowledge would decline. The index visit physician’s diagnostic
12
13 255 knowledge cannot impact a future adverse outcome if the underlying condition that caused that
14
15 256 outcome was not present or detectible at the time of index visit. Therefore, this reduction in
16
17 257 association should be especially true for the hospitalization and ED measures where adverse
18
19 258 outcomes were limited to the 13 diagnostic error conditions and so were unrelated to the index
20
21 259 visit diagnoses in this sensitivity. Similarly, for the last sensitivity, we applied elective
22
23 260 hospitalizations as an outcome measure to consider the possibility that there could be a
24
25 261 correlation between the overall propensity to hospitalize in an area and physician knowledge.
26
27
28
29
30
31
32
33

34 263 The Advarra Institutional Review Board approved our study protocol and all analyses were
35
36 264 performed using Stata version 15 (College Station, TX).
37
38
39

40 265 *Patient and Public Involvement*

41
42 266 Patients and/or the public were not involved in the design, or conduct, or reporting, or
43
44 267 dissemination plans of this research.
45
46
47
48
49

50 51 269 **Results**

52
53
54 270
55
56
57
58
59
60

1
2
3 271 Of 2,492 general internists who initially certified in 2000 and who took an IM-MOC exam
4
5 272 between 2009 and 2012, 1,722 had outpatient visits with a fee-for-service Medicare beneficiary
6
7 273 during the study period. Those without visits generally practiced hospital medicine. Of these,
8
9
10 274 1,410 were included in the study because they had at least one outpatient index visit that met our
11
12 275 study inclusion criteria during the year after they took their IM-MOC exam. In total, 48,632
13
14 276 index visits with 42,407 patients treated by 1,410 physicians met study inclusion criteria (Figure
15
16 277 1). Table 1 lists frequency of index visits and subsequent outcomes for each diagnostic error
17
18
19 278 sensitivity condition.
20
21

22 279
23
24
25 280 The mean percent correct on diagnosis questions ranged from 84.3% among top third performers
26
27 281 to 65.5% among bottom third performers (Table 2). Patient and visit characteristics were similar
28
29
30 282 across tertiles of physician diagnostic knowledge. For example, there were no statistically
31
32 283 significant differences in the HCC risk adjuster across tertiles ($P=.19$) However, there were
33
34 284 differences in some physician and practice characteristics. When compared to physicians in the
35
36 285 bottom tertile of diagnostic knowledge, physicians in the top were significantly less likely to be
37
38 286 in solo practice (12.8% versus 24.4%, $P=0.009$), and more likely to be in academic practice
39
40 287 (9.7% versus 3.4%, $P<.001$). However, the proportion graduating from a US medical school was
41
42 288 similar across diagnostic knowledge tertiles (70.0% versus 63.3%, $P=.30$).
43
44
45

46 289 *Associations between diagnostic knowledge and patient adverse outcomes*

47
48

49 290 The overall rates of 90-day adverse outcomes per 1,000 index visits was 6.5 for death, 11.1 for
50
51 291 hospitalizations, and 13.6 for ED visits (with the latter two directly associated with one of the
52
53 292 diagnostic error sensitive conditions whose antecedent was present in the applicable index visit).
54
55
56
57

1
2
3 293 Being seen by a physician scoring in the top versus bottom third of diagnostic knowledge on the
4
5 294 MOC exam was associated with 2.9 fewer deaths per 1,000 visits (95% confidence interval (CI) -
6
7 295 5.0 to -0.7, P=.008) which reflects a 35.3% lower risk of death (95% CI -52.8 to -11.2, P=.008),
8
9 296 (Table 3). Our finding also suggests that this difference in exam performance was associated
10
11
12 297 with 4.1 fewer applicable hospitalizations (95% CI -6.9 to -1.2, P=0.006), and 4.9 fewer
13
14 298 applicable ED visits (95% CI -8.1 to -1.6, P=0.003) per 1,000 visits (Table). These reductions
15
16 299 correspond with about a 30% lower risk for these utilization measures (hospitalizations: -30.5%,
17
18 300 95% CI -46.1 to -10.4, P=.003, ED: -29.8%, 95% CI -44.4 to -11.4).

21
22 301
23
24
25 302 We also found a significant dose response relationship across all three regression adjusted
26
27 303 relative risk measures (P-trends <0.008). For example, the regression-adjusted 90-day risk of
28
29 304 death per 1,000 patients whose index visit physician scored in the top third of diagnostic
30
31 305 knowledge was 5.2 (95% CI 4.1 to 6.3), compared to 6.5 (95% CI 5.4 to 7.6) for the middle
32
33 306 third, and 8.1 (95% CI 6.5 to 9.7) for the bottom third (P-trend 0.008).

37 307 *Sensitivity Analyses*

38
39
40 308 Our sensitivity analyses (Appendix Section 6) confirmed that base case associations with
41
42 309 diagnostic knowledge were robust to different index visit clean periods, and diagnosis code
43
44 310 inclusion criteria and next day coding of outcome measures. Associations with diagnostic
45
46 311 knowledge were also fairly robust to physician's practice size for both the ED and hospitalization
47
48 312 measures when we limited the sample to either small or large or academic practices.
49
50

51
52 313
53
54
55
56
57
58
59
60

1
2
3 314 Suggesting that our results were not influenced by omitted variable bias, we found that
4
5 315 associations with diagnostic knowledge and our outcome measures became small and
6
7
8 316 statistically insignificant when we limited the sample to index visits with diagnoses unrelated to
9
10 317 any of the 13 diagnostic sensitive error conditions, and so were at lower risk for diagnostic error
11
12 318 ($P>0.50$ and associations were at most about a tenth of the base case percent difference between
13
14 319 top and bottom third of diagnostic knowledge). We also found no significant association between
15
16
17 320 lack of diagnostic knowledge and elective hospitalizations ($P=0.63$).
18
19
20 321

22 **Discussion**

23 322
24
25
26 323
27
28
29 324 We found that higher diagnostic knowledge among US outpatient internal medicine physicians
30
31 325 was associated with significant reductions in subsequent adverse outcomes whose cause was at
32
33 326 risk for diagnostic error. Indeed, for every 1,000 index visits for a new complaint at risk for
34
35
36 327 diagnostic error, being seen by a physician in the top versus bottom third of diagnostic
37
38 328 knowledge was associated with 2.9 fewer all-cause death and, for diagnostic error sensitive
39
40 329 conditions, 4.1 fewer hospitalizations and 4.9 fewer ED visits within 90 days. These figures
41
42
43 330 correspond to a reduction in risk for these adverse events by about a third. Although some prior
44
45 331 studies have demonstrated the high morbidity and mortality of diagnostic error(1-3), this is the
46
47 332 first study to demonstrate and quantify the direct association between serious adverse outcomes
48
49 333 and the diagnostic knowledge of their first contact primary care physician. These finding support
50
51 334 the notion that gaps in diagnostic knowledge between physicians may be an important
52
53
54 335 contributor to the diagnostic error problem plaguing the healthcare system worldwide.
55
56
57
58
59
60

1
2
3 336
4
5
6 337 We measured the association between diagnostic knowledge and potential diagnostic error by
7
8 338 using Medicare claims data to identify patients who presented for outpatient visits with
9
10 339 complaints at heightened risk for serious diagnostic errors and examining the occurrence of
11
12 340 clinically relevant adverse outcomes soon thereafter. Although this approach lacks the precision
13
14 341 of individual chart audits(7), it is both clinically plausible and scalable in that it can be used to
15
16 342 monitor the care of large numbers of patients, making the method itself an important contribution
17
18 343 to the literature on diagnostic error. Although we did not directly measure diagnostic errors
19
20 344 through chart audits, the fact that we found associations with diagnostic knowledge and the
21
22 345 diagnostic error sensitive outcome conditions we studied coupled with the fact that we did not
23
24 346 find associations with treatment knowledge, nor did we find associations when the underlying
25
26 347 diagnostic error sensitive condition was likely not present during the outpatient index visit
27
28 348 because no antecedent diagnoses recorded indicates that the associations we report in this study
29
30 349 were likely driven by association with diagnostic errors that occurred during these visits.
31
32 350 Furthermore, our approach builds on prior studies that used claims data to infer diagnostic error
33
34 351 incidence for ED visits, in that we identified index visit diagnoses at risk for diagnostic error that
35
36 352 were clinically plausible and verified empirically, and we assured that we were studying new
37
38 353 problems by requiring that the patient had not had an ED, hospital or outpatient visit over the
39
40 354 previous 3 months.(30-32) We expanded on these studies by focusing on outpatient care and by
41
42 355 examining a much more comprehensive set of presenting complaints that may have been
43
44 356 precursors to one of 13 diagnostic error prone conditions that we studied. This approach was
45
46 357 necessary in order to study diagnostic error in the more low acuity setting of outpatient general
47
48 358 internal medicine.
49
50
51
52
53
54
55
56
57
58
59
60

359

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 found that diagnosis exam performance was not associated with elective hospitalizations, which are, presumably, unrelated to underlying diagnostic knowledge but may be related to the overall propensity to hospitalize. 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 utilization 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 program so we cannot extrapolate these findings to a younger population, other

1
2
3 382 conditions we did not consider, or populations with no or different health insurance coverage.
4

5 383 Our findings might also not be applicable to older physicians who certified before 2000 or
6

7 384 younger physicians who certified after 2000 as well as physicians who choose not to attempt an
8

9 385 exam. While a physician's clinical knowledge might be related to their decision to not take the
10

11 386 MOC exam therefore not maintaining their certification, other factors certainly play a role in this
12

13 387 decision.
14
15

16
17 388

18
19
20 389 Another limitation of our study is that the IM-MOC exam was specifically designed to measure
21

22 390 clinical knowledge in general, it was not designed to measure diagnostic knowledge specifically.
23

24 391 That said, diagnostic knowledge is a major component of the exam and was found to meet the
25

26 392 criteria for measuring this underlying construct. Also diagnostic error may have stemmed from
27

28 393 factors outside of inadequate diagnostic knowledge, which are not covered by the exam but
29

30 394 could be correlated with our exam based diagnostic knowledge measure (e.g., poor
31

32 395 patient/physician communication skills and related system failures).(33, 34) That said, there is no
33

34 396 reason to believe that these other contributors to diagnostic error would not also be correlated
35

36 397 with the other aspects of the exam we do account for. Furthermore, based on an analysis of
37

38 398 malpractice claims, Newman-Toker et al. (6) reported that clinical judgement played an
39

40 399 important role in 86% of diagnostic errors, while poor patient/physician communication and
41

42 400 system failures played a role in far fewer diagnostic errors that resulted in malpractice suits (35%
43

44 401 and 22% respectively). Suggesting that improving communication will not reduce stroke related
45

46 402 diagnostic error, Kerber et al. (35) reported that frontline providers rarely ask the right questions
47

48 403 when patients present with dizziness. Communication ability is only valuable in terms of
49

50 404 reducing diagnostic error if the physician knows what questions to ask and what the answers
51
52
53
54
55
56
57
58
59
60

1
2
3 405 mean. Although we cannot say with certainty that our findings are driven by an underlying
4
5 406 association between diagnostic knowledge and diagnostic errors, at a minimum, our findings
6
7 407 suggest that patients treated by physicians who scored well on diagnostic exam questions may be
8
9 408 at lower risk for the adverse outcomes we studied. Finally, some might assert that a standardized
10
11 409 exam without access to medical reference material might be more a reflection of a physician's
12
13 410 rote memory and ability to recall medical facts than a test of their clinical knowledge and
14
15 411 judgement. Although this is a fundamental limitation of our study, it should be noted that the
16
17 412 exam is designed to mimic decision making in real life situations including such things as patient's
18
19 413 laboratory results and reference material impeded in the exam and past research indicates that an
20
21 414 "open" book format that allows physicians access to reference material did not materially impact
22
23 415 exam performance.⁽³⁶⁾ It should also be noted that the necessary rapidity of decision making by
24
25 416 primary care physicians who have limited time per encounter might fairly be represented by an
26
27 417 exam with time constraints.
28
29
30
31
32
33
34 418
35
36

37 419 In this exploratory analysis, we found evidence that diagnostic knowledge of primary care
38
39 420 physicians seeing a patient for an index visit for a complaint that is at heightened risk of
40
41 421 diagnostic error is associated with adverse outcomes. The fact that there exists a link between
42
43 422 general diagnostic knowledge and diagnostic error may not be surprising, the magnitude of the
44
45 423 associations we found suggests that interventions ignoring the role of physician knowledge may
46
47 424 be inadequate to address the crisis of diagnostic error. Interventions targeted at improving
48
49 425 diagnostic knowledge could include such things as a greater focus on diagnostic training during
50
51 426 graduate medical education (i.e., medical school, residency, and fellowship). Knowledge-focused
52
53 427 interventions could also include incentivizing broad-based learning as well as targeted learning
54
55
56
57
58
59
60

1
2
3 428 pursued through continuing medical education (CME) activities.(30) During visits identified as
4
5 429 being at risk for diagnostic errors, physicians could be given related information at the point of
6
7
8 430 care including suggestions for specialty consultation.
9
10

11 431
12
13
14 432 Our results are important for two additional reasons. First, these results provide evidence that board
15
16 433 certification and maintenance of certification, which involves lifelong learning directed at maintaining
17
18 434 medical knowledge, might, in fact, be a valid approach to assuring the delivery of high quality care.
19
20 435 Many in the US complain about the time and expense of MOC and often point to the lack of rigorous
21
22 436 assessment between aspects of MOC and outcomes of interest to patients. These findings suggest that
23
24 437 processes such as MOC may translate into meaningful improvements in outcomes because they can
25
26 438 provide incentives for meaningful learning. This learning also could be enhanced through exam feedback
27
28 439 targeted at diagnostic knowledge. Second, the findings also suggest that interventions aimed at improving
29
30 440 diagnostic skills, whether knowledge-based or through, for instance, delivery of relevant information at
31
32 441 the point of care [this is in response to system changes] might be approaches that might be worthwhile if
33
34 442 the findings of this study are validated with additional research. Yet more research is needed to better
35
36 443 understand the link between diagnostic knowledge and diagnostic errors that are identified
37
38 444 through chart review or other methods of direct ascertainment and the extent to which such
39
40 445 errors result in adverse clinical outcomes.
41
42
43
44
45 446

46
47
48 447 In conclusion, gaps in diagnostic knowledge among first contact primary care physicians is
49
50 448 associated with serious diagnostic error sensitive outcomes. If this finding is confirmed in future
51
52 449 studies, diagnostic knowledge should be a target for interventions to reduce diagnostic errors.
53
54

55 450
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

451

For peer review only

1
2
3 452
4
5
6 453
7
8
9 454
10
11 455
12
13 456
14
15 457
16
17
18 458
19
20 459
21
22 460
23
24 461
25
26 462
27
28
29 463
30
31 464
32
33 465
34
35 466
36
37 467
38
39 468
40
41 469
42
43 470
44
45 471
46
47 472
48
49 473
50
51 474
52
53
54
55
56
57
58
59
60

Statements

- A. Contribution statement: Bradley Gray, Jonathan Vandergrift, Rebecca Lipner, Rozalina McCoy, and Bruce Landon met the ICMJE guidelines authorship criteria:
- a. Bradley Gray, Jonathan Vandergrift, Rebecca Lipner, Rozalina McCoy, and Bruce Landon substantially contributed to the conception and design of the work.
 - b. Bradley Gray, Jonathan Vandergrift, Rebecca Lipner contributed to the acquisition of the data.
 - c. Bradley Gray, Jonathan Vandergrift, Rebecca Lipner, Rozalina McCoy, and Bruce Landon substantially contributed to analysis or interpretation of data.
 - d. Bradley Gray, Jonathan Vandergrift, Rebecca Lipner, Rozalina McCoy, and Bruce Landon substantially contributed to the drafting the work and revising it critically for important intellectual content.
 - e. Bradley Gray, Jonathan Vandergrift, Rebecca Lipner, Rozalina McCoy, and Bruce Landon gave final approval of the version published.
 - f. Bradley Gray, Jonathan Vandergrift, Rebecca Lipner, Rozalina McCoy, and Bruce Landon 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.
- B. Competing Interests: Bradley Gray, Jonathan Vandergrift, and Rebecca Lipner are paid employees of the American Board of Internal Medicine. Bruce Landon is a paid consultants for the American Board of Internal Medicine.
- C. Funding Statement: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

1
2
3 475 D. Data Sharing: Administrative data describing physician characteristics and exam
4
5 476 performance can be obtained from the ABIM through a data sharing agreement that assures
6
7 477 physician confidentiality and its use for legitimate research purposes. Access to de-
8
9 identified Medicare claims data for this study were obtained through a special data use
10 478
11 agreement with the Centers for Medicare and Medicaid services which is a process
12 479
13 available to researchers in the US.
14 480

15
16
17 481 E. Dissemination to participants and related patient and public communities: As study data
18
19 482 were pseudonymised, it is not possible to send findings directly to the study participants.
20
21 483 ABIM's communication department in collaboration with the authors of this study will
22
23 write a press release whose goal is to inform the public regarding the findings of the study.
24 484
25

26 485
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

486 **References**

- 487 1. National Academies of Sciences, Engineering, and Medicine. Improving Diagnosis in Health
488 Care. Washington, DC: The National Academies Press 2015 December 10, 2018.
- 489 2. Graber ML, Trowbridge R, Myers JS, et al. The next organizational challenge: finding and
490 addressing diagnostic error. *Jt Comm J Qual Patient Saf.* 2014;40(3):102-10.
- 491 3. Cresswell KM, Panesar SS, Salvilla SA, et al. Global research priorities to better understand the
492 burden of iatrogenic harm in primary care: an international Delphi exercise. *PLoS Med.*
493 2013;10(11):e1001554.
- 494 4. Kostopoulou O, Delaney BC, Munro CW. Diagnostic difficulty and error in primary care--a
495 systematic review. *Fam Pract.* 2008;25(6):400-13.
- 496 5. Goyder CR, Jones CH, Heneghan CJ, et al. Missed opportunities for diagnosis: lessons learned
497 from diagnostic errors in primary care. *Br J Gen Pract.* 2015;65(641):e838-44.
- 498 6. Newman-Toker DE, Schaffer AC, Yu-Moe CW, et al. Serious misdiagnosis-related harms in
499 malpractice claims: The "Big Three" - vascular events, infections, and cancers. *Diagnosis (Berl).*
500 2019;6(3):227-40.
- 501 7. Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care:
502 estimations from three large observational studies involving US adult populations. *BMJ Qual Saf.*
503 2014;23(9):727-31.
- 504 8. Gandhi TK, Kachalia A, Thomas EJ, et al. Missed and delayed diagnoses in the ambulatory
505 setting: a study of closed malpractice claims. *Ann Intern Med.* 2006;145(7):488-96.
- 506 9. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med.*
507 2005;165(13):1493-9.
- 508 10. Kachalia A, Gandhi TK, Puopolo AL, et al. Missed and delayed diagnoses in the emergency
509 department: a study of closed malpractice claims from 4 liability insurers. *Ann Emerg Med.*
510 2007;49(2):196-205.
- 511 11. Poon EG, Kachalia A, Puopolo AL, et al. Cognitive errors and logistical breakdowns contributing
512 to missed and delayed diagnoses of breast and colorectal cancers: a process analysis of closed malpractice
513 claims. *J Gen Intern Med.* 2012;27(11):1416-23.
- 514 12. Chisholm A, Askham J. A review of professional codes and standards for doctors in the UK,
515 USA and Canada: Picker Institute Europe Oxford; 2006.
- 516 13. Irvine D. Doctors in the UK: their new professionalism and its regulatory framework. *Lancet.*
517 2001;358(9295):1807-10.
- 518 14. Kovacs E, Schmidt AE, Szocska G, et al. Licensing procedures and registration of medical
519 doctors in the European Union. *Clinical Medicine.* 2014;14(3):229-38.
- 520 15. European Union of Medical Specialists. The european council for accreditation of medical
521 specialist qualifications (ECAMSQ) 2010 [Available from:
522 https://www.uems.eu/_data/assets/pdf_file/0009/1206/ECAMSQ_presentation.pdf.
- 523 16. Petterson SM, Liaw WR, Phillips RL, Jr., et al. Projecting US primary care physician workforce
524 needs: 2010-2025. *Ann Fam Med.* 2012;10(6):503-9.
- 525 17. Freed GL, Dunham KM, Singer D. Use of board certification and recertification in hospital
526 privileging: policies for general surgeons, surgical specialists, and nonsurgical subspecialists. *Arch Surg.*
527 2009;144(8):746-52.
- 528 18. Independence Blue Cross. Credentialing criteria 2019 [Available from:
529 https://www.ibx.com/pdfs/providers/interactive_tools/credentialing_criteria_ibx.pdf.
- 530 19. American Board of Medical Specialties. Certification Matters FAQs 2019 [Available from:
531 <https://www.certificationmatters.org/faqs/>.
- 532 20. Lipner RS, Bylsma WH, Arnold GK, et al. Who is maintaining certification in internal medicine--
533 and why? A national survey 10 years after initial certification. *Ann Intern Med.* 2006;144(1):29-36.
- 534 21. Balogh E, Miller BT, Ball J, et al. Improving diagnosis in health care. xxvii, 444 pages p.

- 1
2
3 535 22. Centers for Medicare & Medicaid Services. BPCI Advanced 2018 [Available from:
4 536 <https://innovation.cms.gov/initiatives/bpci-advanced>.
5 537 23. Centers for Medicare & Medicaid Services. Comprehensive Care for Joint Replacement Model
6 538 2018 [Available from: <https://innovation.cms.gov/initiatives/cjr>.
7 539 24. Schiff GD, Hasan O, Kim S, et al. Diagnostic error in medicine: analysis of 583 physician-
8 540 reported errors. *Arch Intern Med*. 2009;169(20):1881-7.
9 541 25. Gray B, Vandergrift J, Lipner RS, et al. Comparison of content on the American Board of Internal
10 542 Medicine Maintenance of Certification examination with conditions seen in practice by general internists.
11 543 *JAMA*. 2017;317(22):2317-24.
12 544 26. Samonte K, de la Cruz S, Garcia MJ. An Overview of the ABIM Cardiovascular Disease
13 545 Maintenance of Certification Examination. *J Am Coll Cardiol*. 2020;75(9):1083-6.
14 546 27. Bandalos DL. Measurement theory and applications for the social sciences: Guilford
15 547 Publications; 2018.
16 548 28. Huber PJ. The behavior of maximum likelihood estimates under non-standard conditions. *Paper*
17 549 *presented at: Fifth Berkeley symposium on mathematical statistics and probability; Berkeley, CA*. 1967.
18 550 29. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for
19 551 heteroskedasticity. *Econometrica: Journal of the Econometric Society*. 1980:817-38.
20 552 30. Newman-Toker DE, Makary MA. Measuring Diagnostic Errors in Primary Care: The First Step
21 553 on a Path Forward Comment on “types and Origins of Diagnostic Errors in Primary Care Settings”. *JAMA*
22 554 *internal medicine*. 2013;173(6):425-6.
23 555 31. Waxman DA, Kanzaria HK, Schriger DL. Unrecognized Cardiovascular Emergencies Among
24 556 Medicare Patients. *JAMA Intern Med*. 2018;178(4):477-84.
25 557 32. Liberman AL, Newman-Toker DE. Symptom-Disease Pair Analysis of Diagnostic Error
26 558 (SPADE): a conceptual framework and methodological approach for unearthing misdiagnosis-related
27 559 harms using big data. *BMJ Qual Saf*. 2018;27(7):557-66.
28 560 33. Davis Giardina T, King BJ, Ignaczak AP, et al. Root cause analysis reports help identify common
29 561 factors in delayed diagnosis and treatment of outpatients. *Health Aff (Millwood)*. 2013;32(8):1368-75.
30 562 34. Zwaan L, Monteiro S, Sherbino J, et al. Is bias in the eye of the beholder? A vignette study to
31 563 assess recognition of cognitive biases in clinical case workups. *BMJ Qual Saf*. 2017;26(2):104-10.
32 564 35. Kerber KA, Newman-Toker DE. Misdiagnosing Dizzy Patients: Common Pitfalls in Clinical
33 565 Practice. *Neurol Clin*. 2015;33(3):565-75, viii.
34 566 36. Lipner RS, Brossman BG, Samonte KM, et al. Effect of access to an electronic medical resource
35 567 on performance characteristics of a certification examination: A randomized controlled trial. *Ann Intern*
36 568 *Med*. 2017;167(5):302-10.
37
38
39
40 569

570 **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)	Hospitalization ^{a,b}	Emergency department visit ^a	Death ^c
	Number (percent of index visits)	Number (percent of hospitalizations with a diagnostic error sensitive condition)	Number (percent of emergency department visits with a diagnostic error sensitive condition)	Number (percent 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 ^d	Not Reported ^d	121 (38.3)
Anemia	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	8,534 (17.5)	12 (2.2)	13 (2.0)	89 (28.2)
Spinal Cord Compression	6,386 (13.1)	Not Reported ^d	Not Reported ^d	36 (11.4)
Bacteremia / Sepsis	5,567 (11.4)	19 (3.5)	21 (3.2)	46 (14.6)
Appendicitis	2,584 (5.3)	Not Reported ^d	Not Reported ^d	17 (5.4)
Abscess	1,005 (2.1)	Not Reported ^d	13 (2.0)	Not Reported ^d

571 ^aCondition specific outcomes for one of the 13 diagnostic error sensitive conditions within 90 days of an
572 outpatient index visit at risk for that condition

573 ^bHospitalizations include non-elective hospitalizations either initiated through the ED or a trauma center.

574 ^cAll cause mortality within 90 days of the index visit.

575 ^dNot reported because observations were less than 11.

576

577

578 Table 2. Physician and Patient Characteristics by Diagnostic Exam Performance Tertile

	Total	Diagnosis question percent correct			P-value ^a
		Top (78.5 to 95.8)	Middle (71.4 to 78.4)	Bottom (42.9 to 71.3)	
Exam performance, Mean (standard deviation)^a					
Diagnosis question percent correct	74.5 (0.4)	84.3 (0.3)	74.8 (0.1)	65.5 (0.3)	<.001
Other question percent correct	72.6 (0.7)	80.2 (1.0)	72.1 (1.1)	66.4 (1.5)	<.001
Treatment question percent correct	77.3 (0.3)	83.4 (0.4)	77.2 (0.4)	72.0 (0.5)	<.001
Physician Characteristics, count (%)					
Female Physician	19,428 (39.9)	6,546 (43.8)	6,357 (37.5)	6,525 (39.0)	0.37
US born physician	28,462 (58.5)	9,284 (62.1)	9,932 (58.6)	9,246 (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	9,452 (19.4)	1,914 (12.8)	3,462 (20.4)	4,076 (24.4)	0.009
Small group practice (2 to 10)	20,563 (42.3)	5,543 (37.1)	7,529 (44.4)	7,491 (44.8)	0.19
Medium physicians group practice (11 to 50)	7,442 (15.3)	2,899 (19.4)	2,402 (14.2)	2,141 (12.8)	0.25
Large physician group practice (>50 physicians)	5,391 (11.1)	2,150 (14.4)	1,655 (9.8)	1,586 (9.5)	0.14
Academic practice	2,708 (5.6)	1,447 (9.7)	697 (4.1)	564 (3.4)	<.001
Other practice	3,076 (6.3)	1,005 (6.7)	1,211 (7.1)	860 (5.1)	0.59
Beneficiary characteristics					
Beneficiary Race, count (percent)					
White	40,086 (82.4)	12,652 (84.6)	13,778 (81.3)	13,656 (81.7)	0.13
Black	3,958 (8.1)	926 (6.2)	1,609 (9.5)	1,423 (8.5)	0.03
Other	4,588 (9.4)	1,380 (9.2)	1,569 (9.3)	1,639 (9.8)	0.88
Beneficiary age (per year), Mean (SD) ^a	76.6 (0.1)	76.8 (0.1)	76.5 (0.1)	76.6 (0.1)	0.23
CCW chronic conditions, count (percent)					
Alzheimer's Disease and Related Disorders or Senile Dementia	5,151 (10.6)	1,497 (10.0)	1,793 (10.6)	1,861 (11.1)	0.16
Alzheimer's Disease	2,061 (4.2)	627 (4.2)	704 (4.2)	730 (4.4)	0.82
Acute Myocardial Infarction	1,408 (2.9)	394 (2.6)	494 (2.9)	520 (3.1)	0.13
Anemia	22,450 (46.2)	6,706 (44.8)	7,766 (45.8)	7,978 (47.7)	0.11
Asthma	4,424 (9.1)	1,313 (8.8)	1,548 (9.1)	1,563 (9.3)	0.39
Atrial Fibrillation	4,225 (8.7)	1,265 (8.5)	1,478 (8.7)	1,482 (8.9)	0.69
Breast Cancer	2,485 (5.1)	779 (5.2)	831 (4.9)	875 (5.2)	0.48
Colorectal Cancer	1,139 (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	1,662 (3.4)	507 (3.4)	600 (3.5)	555 (3.3)	0.66
Cataract	31,095 (63.9)	9,601 (64.2)	10,773 (63.5)	10,721 (64.1)	0.74
Heart Failure	9,207 (18.9)	2,786 (18.6)	3,155 (18.6)	3,266 (19.5)	0.54
Chronic Kidney Disease	6,904 (14.2)	2,083 (13.9)	2,392 (14.1)	2,429 (14.5)	0.62
Chronic Obstructive Pulmonary Disease	9,108 (18.7)	2,635 (17.6)	3,165 (18.7)	3,308 (19.8)	0.02
Depression	12,042 (24.8)	3,728 (24.9)	4,145 (24.4)	4,169 (24.9)	0.83
Diabetes	13,296 (27.3)	3,947 (26.4)	4,590 (27.1)	4,759 (28.5)	0.16
Glaucoma	10,030 (20.6)	3,086 (20.6)	3,501 (20.6)	3,443 (20.6)	0.99
Hip/Pelvic Fracture	1,531 (3.1)	430 (2.9)	535 (3.2)	566 (3.4)	0.15
Hyperlipidemia	37,132 (76.4)	11,266 (75.3)	12,898 (76.1)	12,968 (77.6)	0.11
Benign Prostatic Hyperplasia	5,815 (12.0)	1,792 (12.0)	1,987 (11.7)	2,036 (12.2)	0.76
Hypertension	37,607 (77.3)	11,345 (75.8)	13,011 (76.7)	13,251 (79.3)	<.001
Hypothyroidism	11,425 (23.5)	3,490 (23.3)	3,862 (22.8)	4,073 (24.4)	0.25

Ischemic Heart Disease	18,713 (38.5)	5,616 (37.5)	6,393 (37.7)	6,704 (40.1)	0.06
Osteoporosis	14,171 (29.1)	4,372 (29.2)	4,794 (28.3)	5,005 (29.9)	0.34
Rheumatoid Arthritis	23,352 (48.0)	6,879 (46.0)	8,275 (48.8)	8,198 (49.0)	0.02
Stroke	6,255 (12.9)	1,880 (12.6)	2,212 (13.0)	2,163 (12.9)	0.70
Number of chronic conditions, count (percent)					
<=4	5,066 (10.4)	1,459 (9.8)	1,744 (10.3)	1,863 (11.1)	0.08
5 to 7	16,861 (34.7)	5,392 (36.0)	5,981 (35.3)	5,488 (32.8)	0.006
8 to 10	16,230 (33.4)	4,907 (32.8)	5,664 (33.4)	5,659 (33.8)	0.35
>=11	10,475 (21.5)	3,200 (21.4)	3,567 (21.0)	3,708 (22.2)	0.28
Mental health visit, count (percent)	6,347 (13.1)	2,040 (13.6)	2,119 (12.5)	2,188 (13.1)	0.46
Hierarchical Condition Category (HCC) score, Mean (SD) ^a	0.98 (0.01)	0.96 (0.01)	0.98 (0.01)	0.99 (0.01)	0.19
Household medium income, mean \$ (SD) ^a	59,852 (643)	61,574 (1,106)	59,113 (1,144)	59,063 (1,075)	0.19
Medicaid dual eligible, count (percent)	6,392 (13.1)	1,793 (12.0)	2,411 (14.2)	2,188 (13.1)	0.28
Rural county residence, count (percent)	7,392 (15.2)	2,207 (14.8)	2,866 (16.9)	2,319 (13.9)	0.64
Visit characteristics					
Visit with same doctor in last year, Count (percent)	37,726 (77.6)	11,369 (76.0)	13,154 (77.6)	13,203 (79.0)	0.08
Visit with any physician in last year, count (percent)	44,852 (92.2)	13,711 (91.7)	15,647 (92.3)	15,494 (92.7)	0.08
Days since last visit with any physician (if any visit in last year), Mean (SD) ^a	144.2 (0.6)	147.1 (0.8)	144.4 (1.0)	141.4 (1.3)	<.001
ED visit in prior year, count (percent)	8,101 (16.7)	2,428 (16.2)	2,879 (17.0)	2,794 (16.7)	0.43
Days since last ED visits (if ED visit in last year), Mean (SD) ^a	222.8 (0.9)	221.2 (1.5)	223.5 (1.5)	223.4 (1.5)	0.47
Hospitalization in prior year, Count (percent)	4,227 (8.7)	1,280 (8.6)	1,489 (8.8)	1,458 (8.7)	0.85
Days since last hospitalization (if hospitalization in last year), Mean (SD) ^a	229.6 (1.2)	229.1 (2.1)	229.7 (2.1)	230.1 (1.9)	0.95
Index visit diagnosis groups, Count (percent)					
Abscess	1,005 (2.1)	268 (1.8)	394 (2.3)	343 (2.1)	0.21
Anemia	12,410 (25.5)	3,817 (25.5)	4,369 (25.8)	4,224 (25.3)	0.93
Aortic aneurysm	11,491 (23.6)	3,495 (23.4)	4,165 (24.6)	3,831 (22.9)	0.18
Appendicitis	2,584 (5.3)	845 (5.6)	949 (5.6)	790 (4.7)	0.01
Bacteremia	5,567 (11.4)	1,660 (11.1)	1,929 (11.4)	1,978 (11.8)	0.83
Congestive heart failure	12,137 (25.0)	3,633 (24.3)	4,221 (24.9)	4,283 (25.6)	0.67
Acute coronary syndrome	16,228 (33.4)	4,627 (30.9)	5,740 (33.9)	5,861 (35.1)	0.02
Depression	12,637 (26.0)	3,932 (26.3)	4,312 (25.4)	4,393 (26.3)	0.78
Fracture	13,409 (27.6)	4,324 (28.9)	4,364 (25.7)	4,721 (28.2)	0.11
Pulmonary embolism	8,534 (17.5)	2,683 (17.9)	2,984 (17.6)	2,867 (17.1)	0.71
Pneumonia	12,183 (25.1)	3,773 (25.2)	4,224 (24.9)	4,186 (25.0)	0.97
Spinal cord compression	6,386 (13.1)	1,985 (13.3)	2,218 (13.1)	2,183 (13.1)	0.94
Stroke	10,026 (20.6)	3,003 (20.1)	3,542 (20.9)	3,481 (20.8)	0.79

579 ^aP-values and standard deviation accounted for correlated errors within physicians

Table 3. Associations with diagnostic knowledge and adverse events per 1,000 index visits

Diagnostic knowledge tertile	Death ^a				Emergency department visit ^b				Hospitalization ^c			
	Unadjusted ^e	Regression adjusted ^{d,e}			Unadjusted ^e	Regression adjusted ^{d,e}			Unadjusted ^e	Regression adjusted ^{d,e}		
	Events per 1,000 visits (95% CI interval)	Events per 1,000 visits (95% CI interval)	Difference (95% CI)	P-value	Events per 1,000 visits (95%CI)	Events per 1,000 visits (95%CI)	Difference (95% CI)	P-value	Events per 1,000 visits (95% CI)	Events per 1,000 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	

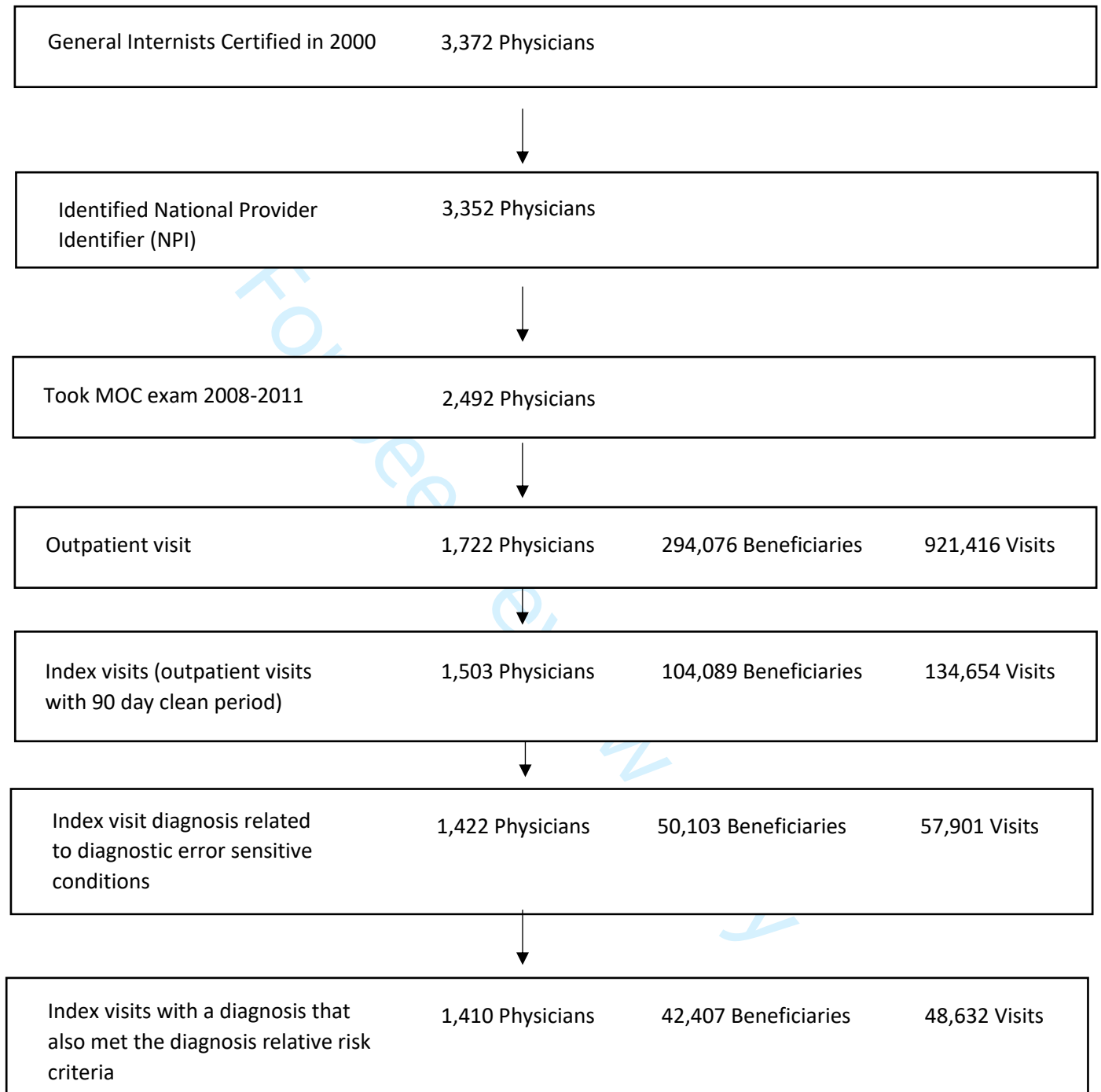
^aAll cause mortality within 90 days of an outpatient index visit with a diagnosis at risk for one of 3 diagnostic error sensitive conditions.

^bEmergency 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.

^cHospitalizations were for non-elective hospitalizations either initiated through the ED or a trauma center 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.

FIGURE LEGEND:**Figure 1. Sample Selection**

For peer review only

Figure 1. Sample Selection

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: Description	Page
Section 1: 90 day Clean Period Derivation	2
Section 2: Outcome Condition and Index Visit Eligibility Diagnoses Codes and Relative Risk.....	5
Section 3: Psychometric Analysis of Whether Diagnosis Related Questions Reflect an Underlying Construct	10
Section 4: Imputations for missing variables.....	11
Section 5: Full Regression Coefficient Estimates and Explanatory Variable List.....	12
Section 6: Regression Sensitivity Analyses.....	15
References	21

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

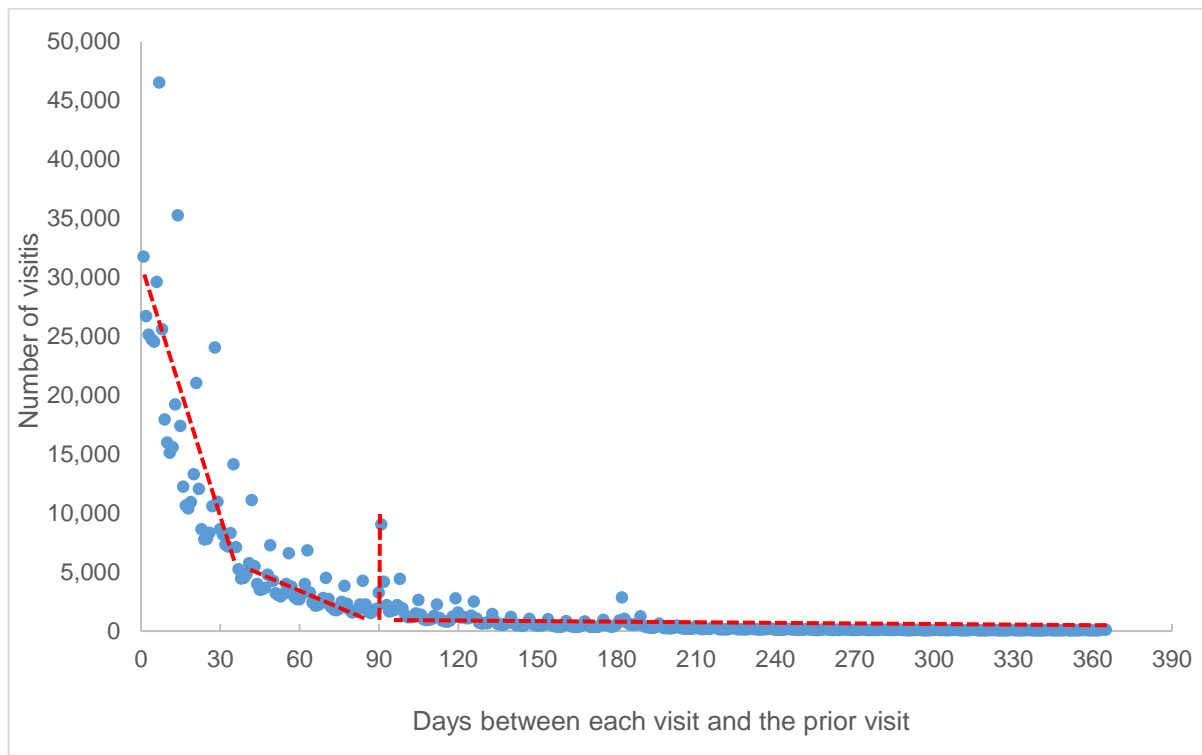
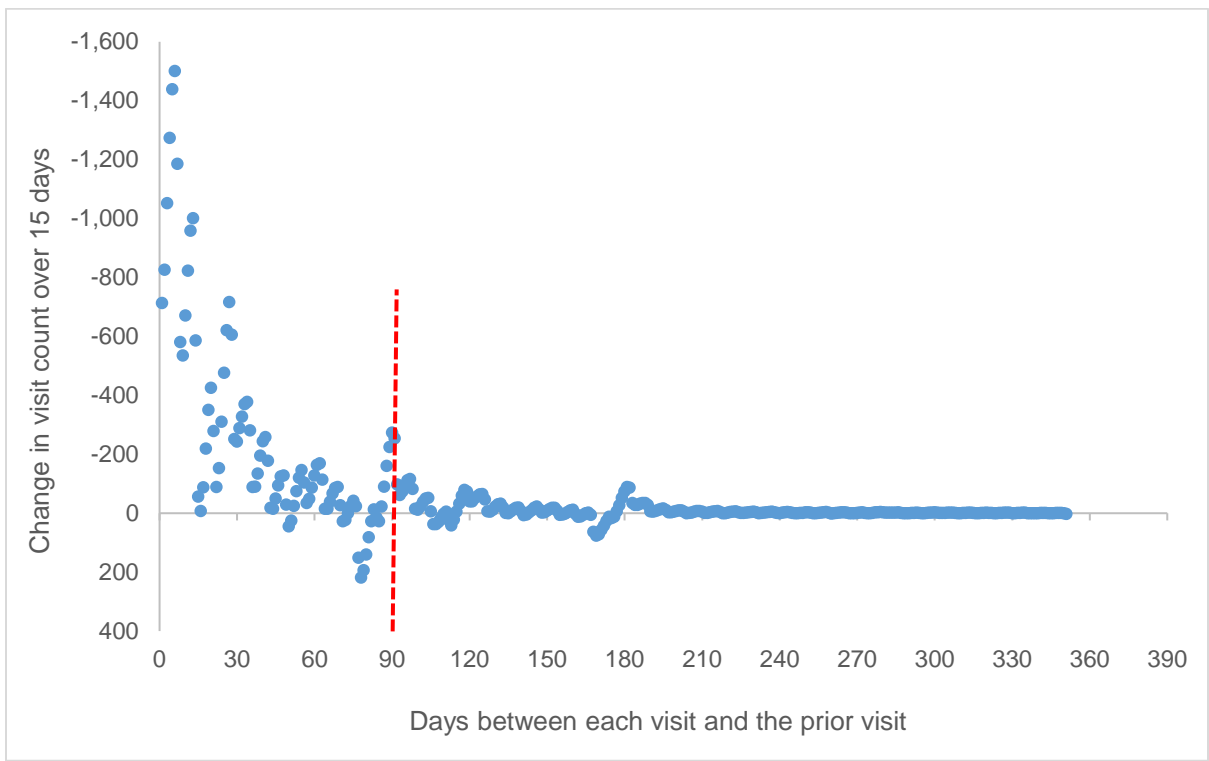


Figure 1.2. Average Change in Visit Count over the 15 days (15-day slope) Following each Data Point Listed in Figure 1



Review only

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	Outcome conditions ^a	Index visit eligibility diagnosis group	Relative Risk ^b	
Abscess	Fever	2.65	Acute Coronary Syndrome	Chest pain	8.38	
	Chills	0.00		Dyspnea	7.29	
Anemia	Gastrointestinal bleeding	25.20		Shortness of breath	3.65	
	Weight loss	4.09		Hypoxemia/hypoxia	2.01	
	Shortness of breath	3.51		Reflux disease	1.23	
	Weakness/Fatigue	2.35		Esophageal disease	1.22	
	Hypoxemia/Hypoxia	2.11		Weakness/Fatigue	1.14	
	Dyspnea	2.05		Nausea	1.05	
	Chest Pain	1.82		Other respiratory issue	0.86	
	Headache	1.29		Respiratory distress	0.00	
Aortic Aneurysm	Menorrhagia	0.00		Gastritis	0.00	
	Dyspnea	4.98		Heart Burn	0.00	
	Abdominal pain	4.93		Depression	Delirium	32.76
	Shortness of breath	3.80			Heart failure	6.16
	Chest pain	2.42	Anxiety		5.04	
	Other back pain	1.64	Dysthymia		4.99	
	Back pain	1.01	Weight loss		4.73	
Elevated blood pressure	0.00	Anemia	2.74			
		Fatigue	1.06			
Appendicitis	Vomiting	30.79	Alcohol		0.00	
	Diverticulitis	30.45	Amphetamines		0.00	
	Nausea	16.81	Cannabis		0.00	
	Abdominal pain	15.60	Cocaine		0.00	
	Abdominal tenderness	0.00	Drug Mental Disease		0.00	
Bacteremia/Sepsis	Fever	0.00	Hallucinogens		0.00	
	Vomiting	6.99	Opioids	0.00		
	Fever	5.10	Personality disorder	0.00		
	Nausea	3.82	related alcohol disease	0.00		
	Tachycardia	2.67	Sedatives	0.00		
Heart failure	Weakness/Fatigue	1.75	Stress	0.00		
	Hypoxemia/Hypoxia	9.99	Weight gain	0.00		
	Shortness of breath	5.09	Mood disorder	0.00		
	Dyspnea	3.33	Fracture	Gait instability	2.53	
	Edema	3.27		Edema	1.79	
	Chest Pain	2.46		Osteoporosis	1.66	
	Weakness/Fatigue	1.42		Hyperparathyroidism	1.09	
	Ascites	0.00		Vitamin D deficiency	1.08	
Respiratory Distress	0.00					

Outcome conditions ^a	Index visit eligibility diagnosis group	Relative Risk ^b	Outcome conditions ^a	Index visit eligibility diagnosis group	Relative Risk ^b
Fracture (con't)	Osteopenia	0.54	Spinal cord compression	Abdominal pain	31.20
	Celiac disease	0.00		Back pain	15.03
	Falls	0.00		Peripheral neuropathy	0.00
	Stress fracture	0.00		Weakness/Fatigue	0.00
Pulmonary embolism	Tachycardia	12.16	Stroke	Facial weakness	65.24
	Hypoxemia/hypoxia	10.98		Confusion	48.93
	Shortness of breath	6.75		Speech disturbance	19.60
	Dyspnea	6.54		Transient ischemic attack	7.82
	Abnormal respiration	6.35		Delirium	4.96
	Heart failure	4.51		Dizziness	3.20
	Chest pain	4.31		Lack coordination	2.92
	Cough	1.48		Gait instability	2.92
	Other respiratory issue	1.34		Vomiting	2.15
	Deep vein thrombosis	0.00		Weakness/Fatigue	1.54
	Respiratory distress	0.00		Headache	1.37
	Fever	0.00		Nausea	1.17
	Heart burn	0.00		Thunderclap headache	0.00
	Hemoptysis	0.00		Tension headache	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				

^aOutcomes include any hospitalization or emergency department visit for the diagnostic conditions within 90 days of the index visits including same day events

^bIndex 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.

For peer review only

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

Label	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	
	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

1							
2							
3	Asthma	-0.8 (1.1)	0.45	-0.2 (1.4)	0.88	0.1 (1.7)	0.95
4	Atrial Fibrillation	1.8 (1.1)	0.10	3.4 (1.5)	0.02	3.7 (1.7)	0.03
5	Breast Cancer	-2.0 (1.3)	0.13	-3.0 (1.8)	0.10	-2.0 (2.2)	0.35
6	Colorectal Cancer	-0.5 (1.7)	0.76	-3.4 (2.0)	0.10	-1.9 (2.5)	0.44
7	Endometrial Cancer	1.6 (4.2)	0.71	-1.1 (4.7)	0.82	-0.9 (5.5)	0.87
8	Lung Cancer	3.7 (3.4)	0.28	13.3 (6.2)	0.03	10.3 (6.2)	0.10
9	Prostate Cancer	-1.6 (1.4)	0.26	4.3 (2.5)	0.09	3.6 (2.8)	0.20
10	Cataract	-1.0 (0.9)	0.29	1.1 (1.0)	0.31	0.1 (1.2)	0.96
11	Heart Failure	1.8 (1.0)	0.08	2.9 (1.2)	0.02	3.3 (1.4)	0.02
12	Chronic Kidney Disease	-0.7 (0.8)	0.39	2.9 (1.2)	0.02	4.2 (1.4)	0.004
13	Chronic Obstructive Pulmonary Disease	1.1 (0.9)	0.26	1.3 (1.2)	0.27	0.9 (1.3)	0.50
14	Depression	0.2 (0.9)	0.79	0.9 (1.1)	0.43	0.7 (1.2)	0.59
15	Diabetes	0.0 (0.8)	0.99	3.3 (1.1)	0.003	2.4 (1.2)	0.04
16	Glaucoma	-0.9 (0.8)	0.29	0.0 (1.1)	1.00	-0.3 (1.2)	0.80
17	Hip/Pelvic Fracture	1.4 (1.6)	0.39	3.0 (2.4)	0.20	5.0 (2.8)	0.07
18	Hyperlipidemia	-1.8 (1.1)	0.09	-1.1 (1.3)	0.39	-0.7 (1.4)	0.63
19	Benign Prostatic Hyperplasia	-0.3 (1.1)	0.79	0.0 (1.5)	0.98	-0.8 (1.7)	0.63
20	Hypertension	0.8 (1.1)	0.46	2.4 (1.4)	0.09	1.2 (1.6)	0.44
21	Hypothyroidism	0.7 (0.8)	0.42	1.4 (1.1)	0.19	1.8 (1.2)	0.14
22	Ischemic Heart Disease	0.8 (0.9)	0.39	0.1 (1.1)	0.92	-1.3 (1.2)	0.29
23	Osteoporosis	-1.9 (0.8)	0.02	2.0 (1.2)	0.09	1.4 (1.3)	0.28
24	Rheumatoid Arthritis	-2.3 (0.8)	0.005	-0.2 (1.0)	0.87	0.3 (1.1)	0.81
25	Stroke	2.7 (1.1)	0.02	2.7 (1.3)	0.04	2.8 (1.5)	0.06
26	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
27	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
28	Hospitalization in prior year	8.8 (5.4)	0.10	0.9 (4.1)	0.84	0.4 (4.5)	0.93
29	ED visit in prior year	1.0 (2.5)	0.69	7.3 (4.0)	0.07	8.4 (4.6)	0.07
30	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
31	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
32	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
33	Index visit diagnosis group indicators						
34	Pulmonary embolism	-0.5 (1.3)	0.68	6.1 (1.4)	<.001	7.1 (1.6)	<.001
35	Acute coronary syndrome	-1.3 (1.1)	0.25	-3.0 (1.8)	0.11	-5.5 (2.0)	0.007
36	Stroke	-2.3 (1.4)	0.10	7.4 (1.5)	<.001	7.7 (1.7)	<.001
37	Congestive heart failure	0.2 (1.3)	0.88	10.1 (1.6)	<.001	12.1 (1.7)	<.001
38	Fracture	-1.3 (1.1)	0.22	3.2 (1.3)	0.02	5.1 (1.5)	<.001
39	Abscess	0.7 (2.3)	0.77	6.4 (3.3)	0.05	11.7 (3.3)	<.001
40	Pneumonia	2.3 (1.2)	0.05	5.6 (1.4)	<.001	6.7 (1.6)	<.001
41	Aortic aneurysm	1.0 (1.4)	0.50	-0.6 (2.0)	0.76	0.7 (2.2)	0.74
42	Appendicitis	2.0 (1.8)	0.28	5.9 (3.0)	0.05	9.6 (3.1)	0.002
43	Depression	0.0 (1.3)	0.99	3.0 (1.5)	0.05	2.4 (1.7)	0.15
44	Anemia	2.3 (1.1)	0.04	3.5 (1.8)	0.04	3.2 (2.0)	0.11
45	Bacteremia	0.5 (2.5)	0.85	-9.5 (3.0)	0.001	-8.3 (3.1)	0.008
46	Spinal cord compression	-0.5 (1.8)	0.79	-2.8 (2.8)	0.32	-7.0 (3.1)	0.02
47	Mental health visit	1.4 (1.2)	0.22	-0.9 (1.5)	0.53	0.1 (1.8)	0.97
48	HHS Region						
49	HHS Region 1	Reference		Reference		Reference	
50	HHS Region 2	1.6 (1.7)	0.35	-5.2 (2.2)	0.02	-6.7 (2.7)	0.01
51	HHS Region 3	2.7 (1.8)	0.12	2.1 (2.5)	0.40	1.3 (3.0)	0.66
52	HHS Region 4	0.4 (1.5)	0.77	-2.7 (2.2)	0.22	-4.9 (2.6)	0.07
53	HHS Region 5	0.3 (1.4)	0.81	0.8 (2.1)	0.69	-1.0 (2.6)	0.70
54	HHS Region 6	-0.9 (1.5)	0.53	-2.8 (2.2)	0.21	-4.4 (2.8)	0.11
55	HHS Region 7	0.0 (2.2)	0.99	3.2 (3.2)	0.31	0.9 (3.5)	0.79
56	HHS Region 8	-1.6 (2.2)	0.47	1.9 (3.8)	0.62	-2.0 (3.8)	0.61
57	HHS Region 9	0.0 (1.6)	0.99	-0.6 (2.5)	0.81	-3.2 (2.8)	0.26
58	HHS Region 10	-0.6 (2.2)	0.77	4.4 (3.5)	0.21	4.0 (4.3)	0.35
59	Study Year						
60	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:

^aMissing practice characteristics (1,432 or 2.94% of sample) were coded as “other unknown”.

^bMissing HCC (86 or .18% of sample) were replace by in sample mean HCC.

^cMissing rural indicator (22 or .05% of sample) were assumed to be non-rural

^bMissing 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).

1
2
3 For the first sensitivity we expanding the eligible diagnoses code groups to all 76 identified by
4 physician authors versus 38 in the base case that also met the relative risk criteria.
5
6
7

8 For the third sensitivity we expand the index visit clean period to 97 days and contracted the
9 index visit clean period to 83 days.
10
11
12

13 For the fourth sensitivity, we excluded physician in academic medical centers to consider the
14 possibility that the unobserved physician characteristics related to where they worked or who
15 they worked with could be were independently both related to the underlying physician
16 diagnostic skill and our outcome measures.
17
18
19
20

21 For the fifth sensitivity we accounted for the possibility that adverse outcomes were avoided
22 because the patient died by altering the ED and hospitalization measures to include all-cause
23 mortality. For this sensitivity we added the following two outcome measures: base case
24 hospitalization or death and base case ED or death.
25
26
27

28 Overall results of robustness sensitivity analysis suggests that our base case results were not
29 highly sensitive to different underlying assumptions related to these factors (e.g., across all
30 robustness sensitivities percent change in the outcome measures between top versus bottom
31 diagnostic knowledge exam performers remained statistically significant ($P<0.05$)).
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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:	7,966 ^a	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.

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract [See page 3, last paragraph] <hr/> (b) Provide in the abstract an informative and balanced summary of what was done and what was found [See Abstract page 3 Design section]
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported [See page 6, Introduction section of the Manuscript first and second paragraph]
Objectives	3	State specific objectives, including any prespecified hypotheses [See page 7, last paragraph]
Methods		
Study design	4	Present key elements of study design early in the paper [See last sentence on page 7 and Methods section starting on page 8]
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection [See Physician and Index Visit sample subsections starting on page 7]
Participants	6	(a) <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants [See Methods section page 8 and Figure 1 for physician, patient and visit sample stats, page 33] <hr/> (b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed [NA] <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case [NA]
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable [See Outcome Measures subsection of Methods section, first paragraph of page 10 for outcomes. See paragraph starting on page 11 for diagnostic knowledge measure.]
Data sources/measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group [See paragraph starting on page 11 for diagnostic knowledge measure].
Bias	9	Describe any efforts to address potential sources of bias [See Statistical Methodes starting first paragraph of page 11, explanation for inclusion of other measures of knowledge and Sensitivity Analysis subsection on page 12]
Study size	10	Explain how the study size was arrived at [See first and second paragraph page 7]
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why [See Statistical Methods an page 12 for description for specification of regression explanatory variables]
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding [See Statistical Methods page 12 first paragraph] <hr/> (b) Describe any methods used to examine subgroups and interactions [NA] <hr/> (c) Explain how missing data were addressed [See second to last sentence page 12]

1
2
3 [and Appendix page 44](#)

4 (d) *Cohort study*—If applicable, explain how loss to follow-up was addressed

5 *Case-control study*—If applicable, explain how matching of cases and controls was
6 addressed [\[NA\]](#)

7 *Cross-sectional study*—If applicable, describe analytical methods taking account of
8 sampling strategy [\[Sensitivity Analysis subsection page 13 \]](#)

9 (e) Describe any sensitivity analyses [\[Sensitivity Analysis subsection page 13\]](#)

10
11 Continued on next page
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed [See first paragraph of the Results section page 15] (b) Give reasons for non-participation at each stage [See page 15 first paragraph as well as Figure 1 (page 33) and Table 1 (page 28)] (c) Consider use of a flow diagram [See Figure 1 (page 33)]
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders [See Results section paragraph 2 (page 14) and Table 2 (page 28)] (b) Indicate number of participants with missing data for each variable of interest [See second to last sentence page 12 and Appendix Section 4, page 44] (c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount) [NA]
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time [NA] <i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure [NA] <i>Cross-sectional study</i> —Report numbers of outcome events or summary measures [See Table 1 page 28]
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included [See Table 3 page 30 for adjusted and unadjusted estimates and page 12 first a second paragraph controls and Appendix Section 5, page 45, for coefficient estimates listed as absolute differences] (b) Report category boundaries when continuous variables were categorized [See Statistical Analysis Section page 16, Results page 16 and Table 3, page 30] (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period [See Results starting on the last paragraph of page 15]
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses [See Sensitivity Analysis subsection of Results section last paragraph page 15 and Appendix Section 5, page 45]
Discussion		
Key results	18	Summarise key results with reference to study objectives [See first paragraph of Discussion section page subsection last paragraph of page 16]
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias [See page Discussion section page 18 to page 20]
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence [No other study has addressed our research question, however, in terms of methodology we compare our study to other in the Discussion section page 17]
Generalisability	21	Discuss the generalisability (external validity) of the study results [See Discussion section starting on the bottom of page 18 and continuing on page 19]
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based [See Funding bullet on page 23]

1
2 *Give information separately for cases and controls in case-control studies and, if applicable, for exposed and
3 unexposed groups in cohort and cross-sectional studies.
4

5
6 **Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and
7 published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely
8 available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at
9 <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is
10 available at www.strobe-statement.org.
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60