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

Incidence, duration and risk factors associated with delayed and missed diagnostic opportunities associated with Tuberculosis: A population-based longitudinal study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-045605
Article Type:	Original research
Date Submitted by the Author:	08-Oct-2020
Complete List of Authors:	Miller, Aaron; University of Iowa, Epidemiology Arakkal, Alan; The University of Iowa, Epidemiology Koeneman, Scott; The University of Iowa, Biostatistics Cavanaugh, Joe; The University of Iowa, Biostatistics Gerke, Alicia; The University of Iowa, Internal Medicine Hornick, Douglas; University of Iowa, Internal Medicine Polgreen, Philip; University of Iowa Roy J and Lucille A Carver College of Medicine, Internal Medicine
Keywords:	Tuberculosis < INFECTIOUS DISEASES, Epidemiology < INFECTIOUS DISEASES, GENERAL MEDICINE (see Internal Medicine), Respiratory infections < THORACIC 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.

For peer review only

Incidence, duration and risk factors associated with delayed and missed diagnostic opportunities associated with Tuberculosis: A population-based longitudinal study

Aaron C Miller[1]*, Alan Arakkal[1], Scott Koeneman[2], Joseph E Cavanaugh[2], Alicia K Gerke[3], Douglas B Hornick[3], Philip M Polgreen[3,1]

[1]Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, Iowa; [2]Department of Biostatistics, College of Public Health, University of Iowa, Iowa City, Iowa; [3]Department of Internal Medicine, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, Iowa.

*Corresponding Author: Aaron C Miller, Department of Epidemiology, University of Iowa College of Public Health, 145 N. Riverside Drive, Iowa City, IA 52242. Email: aaron-miller@uiowa.edu. Phone: (319) 384-1584.

Word Count: 299 (abstract), 3,997 (manuscript)

Funding: This work was supported by the *Agency for Healthcare Research and Quality* grant number 5R01HS027375.

Competing Interests: The authors have no competing interests to declare.

For peer review only

ABSTRACT

Objectives: Missed opportunities to diagnose tuberculosis are costly to patients and society. This study aims to estimate (1) the frequency and duration of diagnostic delays among patients with active pulmonary tuberculosis and (2) the risk factors for experiencing a diagnostic delay.

Design: A retrospective cohort study of patients with tuberculosis using longitudinal healthcare encounters prior to diagnosis.

Setting: Commercially insured enrollees represented in the Commercial Claims and Encounters or Medicare Supplemental IBM Marketscan Research Databases from 2001-2017.

Participants: All patients diagnosed with, and receiving treatment for, pulmonary tuberculosis, enrolled at least 365 days prior to diagnosis.

Primary and secondary outcome measures: We estimate the number of visits with tuberculosis-related symptoms prior to diagnosis that would be expected to occur in absence of delays and compare this estimate to the observed pattern. We compute the number of visits representing a delay and use a simulation-based approach to estimate the number of patients experiencing a delay, number of missed opportunities per patient and duration of delays (i.e., time between diagnosis and earliest missed opportunity). We also estimate risk factors for experiencing a missed opportunity.

Results: We identified 3,371 patients diagnosed and treated for active tuberculosis that could be followed for 1 year prior to diagnosis. We estimated 77.2% (CI: 75.6-78.7%) of these patients experienced at least one missed opportunity; of these patients, an average of 3.89 (CI: 3.65-4.14) visits represented a missed opportunity and the mean duration of delay was 31.66 days (CI: 28.51-35.11). Risk factors for delay included outpatient or ED setting, weekend visits, patient age, influenza season, chronic respiratory symptoms prior to infection and receipt of fluroquinolones.

Conclusions: Many patients with tuberculosis experience multiple missed diagnostic opportunities prior to diagnosis. Missed opportunities occur most commonly in outpatient settings and numerous patient-, environment- and setting-specific factors increase risk for delays.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Strengths and limitations of this study

- This study reviewed longitudinal healthcare records for a large population of insured enrollees (over 195-million represented) spanning an extensive time period (2001-2017) and covering a range of healthcare settings (inpatient, outpatient, and ED).
- A simulation-based analysis was conducted to identify visits most likely to represent a diagnostic delay, while excluding coincidental visits that may appear to be missed opportunities.
- This study relied on diagnostic codes (ICD-9/ICD-10) to identify index cases of tuberculosis, and such codes may lack specify for identifying active tuberculosis. Medication claims were used to help validate diagnosis codes by identifying patients receiving medications used to treat active tuberculosis.
- This study also relied on diagnostic codes to identify signs and symptoms of tuberculosis prior to diagnosis. Such records may not capture all visits where symptoms occurred (e.g., symptoms recorded in clinic notes). We conducted a sensitivity analysis to evaluate the potential sensitivity of our findings to visits without related symptom codes.
- Without more granular patient data, we cannot confirm that all patient visits we identify represent diagnostic opportunities.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BACKGROUND

The incidence of tuberculosis has been decreasing in the United States during the past several decades,[1,2] but in more recent years the rate of decrease has slowed.[1,3] To further reduce the incidence of tuberculosis, the rapid identification and treatment of new cases is essential.[3] However, as the incidence of tuberculosis decreases, so may familiarity with the disease among clinicians,[4] resulting in an increase in diagnostic delays.[5,6] Because these delays are, in part, a function of the familiarly and experience of clinicians with a particular disease,[5-8] as the disease becomes less common, diagnostic delays for tuberculosis may become more common.[5,7]

Diagnostic delays of tuberculosis are important to consider for several reasons. First, delays are common in the United States[6,7,9] and other lower-prevalence countries.[8,10-12] Second, delays may contribute to worse clinical outcomes,[13-15] and increased healthcare costs.[16] Third, diagnostic delays for tuberculosis are especially important because delays contribute to additional exposures and thus, additional cases of tuberculosis.[17,18] Substantial diagnostic delays contributing to increased transmission have occurred in both community[19-22] and healthcare settings.[10,23-25]

Traditional approaches to investigate diagnostic delays have focused on single centers, most commonly hospitals, or alternatively have depended on public health registries that rely on patient recall.[8,26] Although diagnostic errors occur in hospitals, opportunities to understand and reduce diagnostic delays may frequently occur in ambulatory settings where patients often first present with signs and symptoms of a disease. Multiple investigations focusing on emergency department visits have highlighted missed opportunities to diagnose tuberculosis.[6,27-29] Thus, to enable a more complete understanding of diagnostic delays requires consideration of sequential healthcare visits across outpatient clinic visits, emergency department visits and hospitalizations. Also, when diagnostic delays are detected, it may be difficult to learn about risk factors for diagnostic delays if patients present in multiple different settings before the diagnosis is made.

Before interventions to decrease diagnostic delays can be designed and implemented, a better understanding of the incidence of and risk factors for diagnostic delays is needed, especially in lower-incidence countries. Thus, the goal of this this study is to propose a population-based approach for estimating the incidence and duration of diagnostic delays associated with tuberculosis, and also to describe the risk factors associated with patients experiencing a diagnostic delay.

METHODS

Data Source: We used longitudinal insurance claims data from the Truven Marketscan Commercial Claims and Encounters and Medicare Supplemental databases from 2001 through 2017. These databases contain claims for over 195 million enrollees across the United States, representing over six-billion enrollment months. Claims from outpatient, emergency and inpatient visits are provided along with outpatient medications.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

This research used de-identified claims data, studies of this type are deemed non-human subjects research by the University of Iowa Institutional Review Board.

Study Population: We identified all patients diagnosed with primary, pulmonary, respiratory or miliary tuberculosis using the ICD-9-CM diagnosis codes 010.X, 011.X, 012.X and 018.X, and the ICD-10-CM codes A15.X and A19.X. Because non-pulmonary tuberculosis presents with different signs and symptoms, we did not include codes for tuberculosis of the central nervous system, intestines, peritoneum, mesenteric glands, bones, joints, genitourinary system or other organs. We required cases to be enrolled for at least one year prior to their initial tuberculosis diagnosis; this first diagnosis was labeled as the *index diagnosis*. Because diagnosis codes alone lack specificity for identifying active tuberculosis,[30] we restricted our analysis to patients with evidence of treatment for active tuberculosis near the index diagnosis using outpatient medication claims.[31] Specifically, we identified treatment with the following set of medications: *Isoniazid* and *Rifampicin/Rifampin*, *Pyrazinamide*, or *Ethambutol*. We considered patients whose treatment began within 1 year of the index diagnosis. (We performed a sensitivity analysis using cases where treatment occurred within 2 months of diagnosis.) If treatment began prior to the initial tuberculosis diagnosis, we used the treatment start date as the index diagnosis date. Only patients with non-missing enrollment information were included.

Statistical Analysis: We conducted two primary statistical analyses to address the following objectives: (1) to estimate the incidence and duration of diagnostic delays associated with tuberculosis, and (2) to estimate the risk factors for experiencing a diagnostic delay. We started by identifying *potential diagnostic delays* by looking for *symptomatically similar diagnoses* (SSDs) that occurred during healthcare visits prior to the index tuberculosis diagnosis. We define SSDs to be diagnoses that include, or share, similar symptoms to active pulmonary tuberculosis. SSDs may include diagnoses in one of four categories:

(1) ***General symptoms*** of active infection, such as cough, fever, weight loss, or hemoptysis;
(2) ***Symptomatically similar infections*** that share similar symptoms to tuberculosis, such as pneumonia, influenza or bronchitis;
(3) ***Symptomatically similar cardio-sino-pulmonary diseases or syndromes***, such as COPD, asthma or lung cancer;
(4) ***Testing, imaging or physical exam-based diagnoses***, such as anemia or swollen lymph nodes.

Supplementary Table 1 describes the individual diagnoses and ICD-9/10 codes used to identify the four types of SSD conditions. This list was developed based on a review of prior literature of diagnostic delays for tuberculosis.[6] We identified SSDs during visits in the time prior to the index diagnosis where diagnostic opportunities may plausibly occur, between 3 and τ days prior; we denote the period [3, τ] as the *diagnostic-opportunity window*. The value τ is the upper bound of the diagnostic-opportunity, reflecting the longest plausible diagnostic delay; this is estimated based on a change-point analysis

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

described below. We disregard visits within 3 days of the index diagnosis, to account for lags in diagnostic testing. Figure 1 depicts the process used to identify potential diagnostic opportunities. This type of "look-back" approach has been referred to as Symptom-Disease Pair Analysis of Diagnostic Error (SPADE), [32] and variations have been frequently used to identify diagnostic delays associated with numerous diseases.[6,33-36]

Estimating Incidence of Diagnostic Delays: Visits occurring prior to an index diagnosis of tuberculosis that contain an SSD may represent a missed diagnostic opportunity but may also represent a coincidental visit (e.g., unrelated respiratory infection). To account for visits representing the usual pattern of care or coincidental diseases, and not a missed opportunity, we compared the difference between *expected* and *observed* patterns of SSD visits prior to the index diagnosis. First, we estimated the *expected* number of SSD visits by analyzing the trend in the incidence of SSD visits in the time prior to the diagnostic-opportunity-window, where missed opportunities are unlikely to occur (e.g., τ-365 days prior to tuberculosis). We then computed the expected number of visits in the diagnostic-opportunity window (e.g., 3-τ days prior to tuberculosis diagnosis) by extrapolating the prior trend to the diagnostic-opportunity window. Second, we compared the *observed* pattern of SSD visits during the diagnostic-opportunity window to the expected number based on the extrapolated trend. Finally, the number of potential diagnostic opportunities was estimated by the *excess number* of SSD visits: the difference between the observed and expected number. This approach has been used in prior work to estimate the number of diagnostic opportunities associated with AMI, stroke, and other cardiovascular events.[33] To identify the point prior to the index diagnosis where diagnostic opportunities first begin to occur (i.e. the diagnostic-opportunity-window), we used a change-point analysis to detect the point where the trend between observed and expected number of SSD visits begins to deviate. We fit a piecewise regression model with a linear trend prior to the change-point τ and a cubic trend after the change-point, to account for the non-linear pattern in visit counts in the period just prior to diagnosis (see Figure 2 for a depiction of this trend). We use the Akaike Information Criterion (AIC) to select the optimal change-point.

To estimate the number of individuals that experienced a potential diagnostic delay, number of recurrent missed opportunities per patient and the typical duration of delays, we used a bootstrapping approach similar to that of Waxman et al.[33] Specifically, we randomly drew (with replacement) a sample of patients and re-estimated the observed and expected patterns of care. Next, at each period prior to the index tuberculosis diagnosis, we randomly labeled a portion of visits for the resampled patients as "diagnostic delays" based on the computed excess number of SSD visits at that time period. Finally, we computed the number of patients that experienced a diagnostic delay, the number of recurrent missed opportunities per patient and the durations of the diagnostic delays. We repeated this procedure 25,000 times to compute 95% bootstrap-based confidence intervals for the change-point τ, number of potential diagnostic opportunities, number of patients that experienced a diagnostic delay, number of recurrent missed opportunities per patient, and the durations of the diagnostic delays.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Sensitivity Analysis. Because diagnostic codes from administrative records may not capture all signs and symptoms present during a clinic visit (e.g., in clinic notes), SSD-related ICD-9/10 codes may undercount the true number of visits representing a diagnostic opportunity. As a sensitivity analysis, we repeated our estimates of the incidence of diagnostic delays by including all visits that occurred within the diagnostic opportunity window (regardless of the presence of an SSD code). Specifically, we repeat the change-point and bootstrapping analysis described above using all visits prior to the index tuberculosis diagnosis.

Estimating Risk Factors for Experiencing a Missed Diagnostic Opportunity: We analyze the potential risk factors for diagnostic delays by estimating the likelihood of a patient with symptoms experiencing a missed opportunity on a given day prior to diagnosis. We treat diagnostic opportunities as a binary outcome – where a patient who has tuberculosis can experience either a missed opportunity (i.e., SSD-related visit in the diagnostic-opportunity window [3, τ]) or a correct diagnosis (i.e., the index diagnosis). Because multiple visits occurring on a single day likely represent a linked episode of care, for each day during the diagnostic-opportunity window or the index diagnosis date, we aggregate all visits containing an SSD or index diagnosis. We create indicators on each day for the specific type of healthcare facility (e.g., inpatient, outpatient, ED). Days with an SSD-related visit during the diagnostic-opportunity window are assigned an outcome of 1 (i.e., missed opportunity) and days representing the index tuberculosis diagnosis are assigned an outcome value of 0 (i.e., not a missed opportunity). We then used logistic regression to estimate the likelihood of a visit representing a missed opportunity, while controlling for other risk factors for delay.

We considered a number of patient- and context-specific risk factors for diagnostic delay. Patient demographics include age, sex, and region (i.e., urban vs. rural). Environment and setting specific factors include the year and month of the SSD visit or the index diagnosis, whether visits during a given day involved inpatient, outpatient, or ED settings, or combinations of visits to multiple settings, and a term for tuberculosis incidence at patient location. Because many symptoms associated with pulmonary tuberculosis are similar to influenza like illness (ILI), we created an indicator for peak influenza season based on the national level of outpatient ILI as reported by the CDC [37]. ILI-based indicator values are provided in Supplementary Table 2. Finally, we considered a number of clinical factors: indicators for asthma and COPD prior to the diagnostic-opportunity window were included as markers for pre-existing pulmonary conditions. In addition, indicators for a chest X-ray or a chest CT scan prior to the diagnostic-opportunity window were included because imaging may also indicate pre-existing pulmonary conditions. We also included an indicator for receipt of a fluoroquinolone prior to the delay window. We performed variable selection using backward elimination, evaluating model performance at each stage of the procedure using the AIC. Standard errors were used to compute Wald-type 95% confidence intervals for the logistic regression analysis.

Patient and Public Involvement: No patient involved

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

RESULTS

From 2001 through 2017, a total of 5,681 individuals had a tuberculosis diagnosis accompanied by an outpatient prescription drug claim consistent with treatment for active tuberculosis. The final study sample included 3,371 enrollees that met eligibility criteria of having been enrolled for at least 1 year prior to the index tuberculosis diagnosis. Supplementary Figure 1 provides a flow diagram of inclusion criteria. Table 1 presents baseline criteria (age, sex, enrollment information, and region) for the final study cohort.

Figure 2 depicts the pattern of total visits and SSD visits that occurred in the 1-year period prior to the index tuberculosis patients. There is a dramatic increase in both the total number of visits and SSD-related visits that occur just prior to the index tuberculosis diagnosis. Supplementary Figure 2 depicts similar patterns broken down by different categories of individual SSD diagnoses. Across nearly all SSD visits, there is a consistent trend. The pattern of SSD visits appears fairly stable, with a very gradual increase from 1-year up to around 100-days prior to the index diagnosis. Starting around 100 days prior to the index diagnosis there is a dramatic spike in SSD visits.

Of the 3,371 case patients we identified, 3,306 (98.1%) patients had at least one healthcare visit in the year prior to their index tuberculosis diagnoses. Of these patients, 1,134 (34.3%) had at least one inpatient visit, 1,301 (39.4%) had at least one ED visit and 3,297 (99.7%) had at least one outpatient visit. Focusing on visits with SSDs, we found 3,084 (91.5%) patients had at least one SSD visit in the year prior to their index tuberculosis diagnosis. Over a third of all visits (37.2%) that occurred in the year prior to the index tuberculosis diagnosis involved one of the SSD conditions. The most common category of SSDs prior to the index tuberculosis diagnoses was alternative cardio sino-pulmonary-based diagnoses (2,322 [68.9%] patients among 15,332 [17.6%] visits), followed by symptom-based diagnoses (2,382 [70.7%] patients among 9,086 [10.5%] visits), testing imaging or physical exam-based diagnoses (2,123 [63.0%] patients among 8,373 [9.6%] visits), and alternative infectious disease-based diagnoses (2,129 [63.2%] patients among 7,921 [9.1%] visits).

Since not all SSD visits are likely to represent diagnostic opportunities, we used a bootstrapping/simulation approach to estimate the number of likely diagnostic opportunities based on the observed and expected number of SSD visits prior to the index tuberculosis diagnosis. Our change-point analysis detected a significant increase in the number of SSD visits occurring 127 days (95% confidence Interval (CI): 117-138 days) prior to the index diagnosis; this represents the start of the diagnostic-opportunity window (i.e., maximum duration of delay). Figure 3 summarizes the observed and expected trend lines estimated from our change-point analysis. Across all patients, 2,903 (86.1%) patients had at least one SSD during this diagnostic-opportunity window (between 127 and 3 days prior to their index diagnosis).

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

There was a total of 19,818 SSD visits that occurred during the diagnostic-opportunity window. Of these visits, based on our simulation analysis, we estimated that 10,118 (51.1%) represented a missed opportunity. We also estimated that approximately 528 missed opportunities occurred in inpatient settings, 9,001 in outpatient settings, and 589 in ED settings. Table 2 presents the estimated number of missed opportunities that each patient experienced. We estimate that 2,602 (CI: 2,549-2,652) or 77.2% (CI: 75.6-78.7%) of patients experienced at least one missed opportunity prior to diagnosis. Of the patients who experienced at least one missed opportunity, we estimated that, on average, they experienced 3.89 (CI: 3.65-4.14) visits representing missed opportunities, occurring in an estimated 3.46 (CI: 3.24-3.69) outpatient visits, 0.20 (CI: 0.19-0.22) inpatient visits, and 0.23 (CI: 0.21-0.24) ED visits.

Table 2 also presents a breakdown of the estimated duration of diagnostic delays among patients who experienced at least one missed opportunity. The mean and median duration of delays were 31.66 (CI: 28.51-35.11) days and 28.00 (CI: 25.00-31.00) days, respectively. On average, patients who experienced at least one missed opportunity had a delay between first SSD and diagnosis of 41.00 days (CI: 37.54-44.77) with 62.1% (CI: 58.4–65.5%) of these delays lasting 30 or more days.

As a sensitivity analysis, we re-estimated the incidence and duration of diagnostic delays using all visits during the diagnostic opportunity window. When using all visits prior to the index diagnosis, the estimated diagnostic-opportunity window began 136 days prior to diagnosis. Across all patients, 3,223 (95.6%) patients had a visit for any reason during this window. There was a total of 44,924 SSD visits that occurred during the diagnostic-opportunity window. We estimated that 14,371 (32.0%) of these visits represented a missed opportunity and 2,976 (CI: 2,923-3,027) patients had at least one missed opportunity. On average, patients experienced 4.83 (CI 4.42-5.34) missed opportunities and had a delay duration of 45.71 days (CI 40.23-52.27).

Table 3 presents results of the logistic regression model estimating the likelihood of experiencing a potential missed opportunity during a visit on a given day. A number of patient-level factors were associated with increased likelihood of being missed. The likelihood of a miss was greater among individuals age ≥ 65 with an odds ratio (OR) of 1.262 (CI: 1.156-1.377). Patients with a history of asthma (OR 1.331 [CI: 1.138-1.557]) or COPD (1.372 [CI: 1.230-1.531]) were more likely to be delayed. Patients who had received chest imaging in the year prior to diagnosis but before the diagnostic-opportunity window were more likely to experience a miss (OR of 1.149 [CI: 1.081-1.296] for chest CT and 1.231 [CI: 1.121-1.353] for chest X-ray). Patients who received a fluoroquinolone in the year prior to diagnosis but before the diagnostic-opportunity window were more likely to experience a miss (OR 1.578 [CI: 1.435-1.734]).

Context and healthcare-setting factors were also significantly associated with missed opportunities. Misses were more likely to occur during weekend visits (1.495 [CI: 1.272-1.758]) and less likely to occur among patients in metropolitan locations (0.874 [CI: 0.771-0.990]). Missed opportunities were more likely to occur in outpatient settings during periods of high influenza activity (1.259 [CI: 1.052-1.507]). Missed opportunities

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

were much less likely to occur in inpatient settings. Compared to outpatient settings alone, misses were less likely to occur on days involving only an inpatient visit (0.123 [CI: 0.106-0.142]), both an inpatient and outpatient visit (0.124 [CI: 0.105-0.145]), both an inpatient and ED visit (0.142 [CI: 0.110-0.184]), or all three setting types (0.128 [CI: 0.089-0.185]). Visits to the ED appeared to increase the risk of a miss. Compared to outpatient settings alone, misses were more common on days when patients visited ED settings only (2.340 [CI: 1.540-3.555]).

DISCUSSION

Our results show the majority of patients diagnosed with pulmonary tuberculosis have multiple interactions with the US healthcare system prior to receiving a diagnosis consistent with active tuberculosis. Many patients present on multiple occasions, each representing possible missed opportunities to diagnose tuberculosis. One-hundred-twenty days prior to the index diagnosis, we observed an increase in visits for either symptoms associated with tuberculosis or an increase in diseases that share symptoms with tuberculosis. At least 90% of patients have at least one visit with either a code recording a symptom of tuberculosis or a disease that shares similar symptoms. Common diagnoses included pneumonia, respiratory infections, and other pulmonary conditions. Diagnoses based on symptoms most frequently listed included fever, cough, hemoptysis and weight loss. A considerable proportion of patients experienced multiple visits representing missed opportunities to diagnose tuberculosis: in fact, 23.8% of patients had more than 5 possible missed opportunities.

We identified a number of risk factors for diagnostic delays. First, we found that delays are more common for patients who visited the ED, without an inpatient visit on the same day. Diagnostic errors may occur commonly in the ED setting: an estimated 12% of patients who revisit the emergency department do so because of an original misdiagnosis.[38] In the ED, physicians are often treating patients they see for the first time and may be unaware of medical histories. In addition, many patients have vague symptoms, and a range of severity.[39] Also, ED physicians frequently care for multiple different patients concurrently. In one study, ED physicians were caring for a median of 5 patients at one time, and they were interrupted an average of 30.9 times during a 180-minute study period.[40] Finally, when diagnostic errors do occur, ED physicians may not be able to learn from missed diagnostic opportunities because follow-up care occurs in other healthcare settings.

Additional risk factors that we identified included female sex and older age. Other studies have identified females as at higher risk for delays,[8,41,42] and there is a need to investigate the cultural, biological or epidemiological factors responsible for this finding. Also, similar to the findings of others, we found that older adults are at increased risk for diagnostic delays[8,11,41]. Older patients may be at greater risk because of more comorbidities or because they are less likely to exhibit some of the classic signs and symptoms of tuberculosis, perhaps due to the immunosenescence associated with aging. In addition to female sex and older age, several investigations also highlight the risk of fluoroquinolone use for increasing diagnostic delays.[43-45] Because

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

fluoroquinolones have some anti-tuberculosis activity, their inappropriate use prior to the diagnosis of tuberculosis (e.g., to empirically treat a misdiagnosed bacterial pneumonia) may transiently improve symptoms.

In addition to established risk factors, our results highlight two more novel risk factors for delay. First, we found that patients with a history of pulmonary diseases, specifically asthma or COPD, were more likely to experience a delayed tuberculosis diagnosis. Other groups have found that other comorbidities, especially pulmonary diseases, were associated with delays;[46] however, we also found that pulmonary imaging (prior to the risk window) was associated with delays. Prior history of pulmonary disorders is a risk factor because it creates a cognitive bias among clinicians. For patients with a history of asthma or COPD presenting with respiratory symptoms, it is less likely that tuberculosis may be considered as part of a differential diagnosis. While patients with a history of pulmonary imaging prior to the diagnostic window, presumably because of some long-standing pulmonary complaint, are more likely to experience a delay, delays are less common if patients received imaging during the diagnostic window because pulmonary imaging would help confirm a tuberculosis diagnosis. Our second novel finding is also related to cognitive bias. Interestingly, we found that if a patient presents during the influenza season, they are more likely to experience a delayed diagnosis for tuberculosis. Delays were also more common during periods of high ILI activity. This finding may reflect the fact that ILI symptoms and tuberculosis symptoms often overlap (e.g., fever, cough), and clinicians may be more likely to suspect influenza during a period of increased activity.

Our study has a number of limitations. First, we use diagnostic codes to identify tuberculosis cases. While such codes have poor sensitivity for identifying active tuberculosis[30], we used medications to validate our case definition, an approach previously used for identifying tuberculosis.[31] Second, we rely on claims data to determine the reason for visits prior to the tuberculosis diagnosis. Thus, our results may underestimate the number of visits that represent missed opportunities. Patients may have had a visit for hypertension, for example, and complained of a cough during that visit. However, cough or other respiratory symptoms may not be recorded in the health insurance claim. Indeed, in our sensitivity analysis the number and duration of diagnostic delays increased slightly when including all visits during the diagnostic-opportunity window, regardless of the presence of SSD-related diagnosis codes. Third, our data do not contain race or ethnicity. Tuberculosis is much more common among immigrants and family members of immigrants. In other studies of low-incidence countries, delays were more common among non-immigrant populations.[8,12,46,47] Fourth, our dataset is restricted to a privately insured population, with employer-sponsored health insurance and/or supplemental Medicare coverage. Thus, our findings may not be generalizable to an uninsured population or individuals with Medicaid coverage. However, vulnerable populations in inner cities or patients experiencing homelessness may be less likely to experience a delay.[7] Finally, our study excluded extra-pulmonary tuberculosis cases, and future work should focus on such cases given that they are at even greater risk for diagnostic delays.[8,12,46]

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Despite our limitations, our results highlight the number of missed opportunities to diagnose tuberculosis. Risk factors for diagnostic delays include older age, female sex, and living in a lower-incidence area. In addition, we identified new risk factors, including existing pulmonary conditions, previous pulmonary imaging, and circulating influenza. These novel risk factors are directly related to cognitive biases that will need to be overcome to improve the timely diagnosis of tuberculosis.

REFERENCES

1. Armstrong LR, Winston CA, Stewart B, Tsang CA, Langer AJ, Navin TR. Changes in tuberculosis epidemiology, United States, 1993-2017. *Int J Tuberc Lung Dis.* 2019;23(7):797-804.
2. Stewart RJ, Tsang CA, Pratt RH, Price SF, Langer AJ. Tuberculosis - United States, 2017. *MMWR Morb Mortal Wkly Rep.* 2018;67(11):317-323.
3. Salinas JL, Mindra G, Haddad MB, Pratt R, Price SF, Langer AJ. Leveling of Tuberculosis Incidence - United States, 2013-2015. *MMWR Morb Mortal Wkly Rep.* 2016;65(11):273-278.
4. Chida N, Brown C, Mathad J, et al. Internal Medicine Residents' Knowledge and Practice of Pulmonary Tuberculosis Diagnosis. *Open Forum Infect Dis.* 2018;5(7):ofy152.
5. Guderian LJ, Miller WC, Seña AC, Stout JE. Increased prevalence of advanced tuberculosis in rural low tuberculosis caseload counties in North Carolina. *Int J Tuberc Lung Dis.* 2011;15(11):1455-1460, i.
6. Miller AC, Polgreen LA, Cavanaugh JE, Hornick DB, Polgreen PM. Missed Opportunities to Diagnose Tuberculosis Are Common Among Hospitalized Patients and Patients Seen in Emergency Departments. *Open Forum Infect Dis.* 2015;2(4):ofv171.
7. Wallace RM, Kammerer JS, Iademarco MF, Althomsons SP, Winston CA, Navin TR. Increasing proportions of advanced pulmonary tuberculosis reported in the United States: are delays in diagnosis on the rise? *Am J Respir Crit Care Med.* 2009;180(10):1016-1022.
8. Loutet MG, Sinclair C, Whitehead N, Cosgrove C, Lalor MK, Thomas HL. Delay from symptom onset to treatment start among tuberculosis patients in England, 2012-2015. *Epidemiol Infect.* 2018;146(12):1511-1518.
9. Mindra G, Wortham JM, Haddad MB, Powell KM. Tuberculosis Outbreaks in the United States, 2009-2015. *Public health reports (Washington, DC : 1974).* 2017;132(2):157-163.
10. Jonsson J, Kan B, Berggren I, Bruchfeld J. Extensive nosocomial transmission of tuberculosis in a low-incidence country. *The Journal of hospital infection.* 2013;83(4):321-326.
11. Paynter S, Hayward A, Wilkinson P, Lozewicz S, Coker R. Patient and health service delays in initiating treatment for patients with pulmonary tuberculosis: retrospective cohort study. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 2004;8(2):180-185.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

12. Leutscher P, Madsen G, Erlandsen M, et al. Demographic and clinical characteristics in relation to patient and health system delays in a tuberculosis low-incidence country. *Scandinavian journal of infectious diseases.* 2012;44(1):29-36.
13. Lui G, Wong RY, Li F, et al. High mortality in adults hospitalized for active tuberculosis in a low HIV prevalence setting. *PloS one.* 2014;9(3):e92077.
14. Pablos-Méndez A, Sterling TR, Frieden TR. The relationship between delayed or incomplete treatment and all-cause mortality in patients with tuberculosis. *Jama.* 1996;276(15):1223-1228.
15. Enarson DA, Grzybowski S, Dorken E. Failure of diagnosis as a factor in tuberculosis mortality. *Canadian Medical Association journal.* 1978;118(12):1520-1522.
16. Kelly AM, D'Agostino JF, Andrada LV, Liu J, Larson E. Delayed tuberculosis diagnosis and costs of contact investigations for hospital exposure: New York City, 2010-2014. *American journal of infection control.* 2017;45(5):483-486.
17. Golub JE, Bur S, Cronin WA, et al. Delayed tuberculosis diagnosis and tuberculosis transmission. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 2006;10(1):24-30.
18. Sreeramareddy CT, Panduru KV, Menten J, Van den Ende J. Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature. *BMC infectious diseases.* 2009;9:91.
19. Castells Carrillo C, San José Rodríguez S, López Aranaga I, et al. Diagnostic delay as main contributing factor to a large outbreak of tuberculosis in a university. *Enfermedades infecciosas y microbiologia clinica.* 2019;37(8):496-501.
20. Raffalli J, Sepkowitz KA, Armstrong D. Community-based outbreaks of tuberculosis. *Archives of internal medicine.* 1996;156(10):1053-1060.
21. MacIntyre CR, Plant AJ, Hulls J, Streeton JA, Graham NM, Rouch GJ. High rate of transmission of tuberculosis in an office: impact of delayed diagnosis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 1995;21(5):1170-1174.
22. Calder L, Rivers J, Hayhurst M, et al. A school and community outbreak of tuberculosis in Palmerston North, New Zealand. *The New Zealand medical journal.* 2008;121(1278):50-61.
23. Rao VK, Iademarco EP, Fraser VJ, Kollef MH. Delays in the suspicion and treatment of tuberculosis among hospitalized patients. *Annals of internal medicine.* 1999;130(5):404-411.
24. Yilmaz A, Boğa S, Sulu E, et al. Delays in the diagnosis and treatment of hospitalized patients with smear-positive pulmonary tuberculosis. *Respiratory medicine.* 2001;95(10):802-805.
25. Harris TG, Sullivan Meissner J, Proops D. Delay in diagnosis leading to nosocomial transmission of tuberculosis at a New York City health care facility. *American journal of infection control.* 2013;41(2):155-160.
26. Evenden P, Roche A, Karo B, Balasegaram S, Anderson CS. Presentation and healthcare delays among people with tuberculosis in London, and the impact on treatment outcome. *BMJ open respiratory research.* 2019;6(1):e000468.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

27. Deponti GN, Silva DR, Coelho AC, Muller AM, Dalcin Pde T. Delayed diagnosis and associated factors among new pulmonary tuberculosis patients diagnosed at the emergency department of a tertiary care hospital in Porto Alegre, South Brazil: a prospective patient recruitment study. *BMC infectious diseases.* 2013;13:538.
28. Moran GJ, McCabe F, Morgan MT, Talan DA. Delayed recognition and infection control for tuberculosis patients in the emergency department. *Ann Emerg Med.* 1995;26(3):290-295.
29. Yen YL, Chen IC, Wu CH, Li WC, Wang CH, Tsai TC. Factors associated with delayed recognition of pulmonary tuberculosis in emergency departments in Taiwan. *Heart & lung : the journal of critical care.* 2015;44(4):353-359.
30. Ronald LA, Ling DI, FitzGerald JM, et al. Validated methods for identifying tuberculosis patients in health administrative databases: systematic review. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 2017;21(5):517-522.
31. Winthrop KL, Baxter R, Liu L, et al. The reliability of diagnostic coding and laboratory data to identify tuberculosis and nontuberculous mycobacterial disease among rheumatoid arthritis patients using anti-tumor necrosis factor therapy. *Pharmacoepidemiology and drug safety.* 2011;20(3):229-235.
32. Liberman AL, Newman-Toker DE. Symptom-Disease Pair Analysis of Diagnostic Error (SPADE): a conceptual framework and methodological approach for unearthing misdiagnosis-related harms using big data. *BMJ Qual Saf.* 2018.
33. Waxman DA, Kanzaria HK, Schriger DL. Unrecognized Cardiovascular Emergencies Among Medicare Patients. *JAMA Intern Med.* 2018.
34. Moy E, Barrett M, Coffey R, Hines AL, Newman-Toker DE. Missed diagnoses of acute myocardial infarction in the emergency department: variation by patient and facility characteristics. *Diagnosis.* 2015;2(1):29-40.
35. Newman-Toker DE, Moy E, Valente E, Coffey R, Hines AL. Missed diagnosis of stroke in the emergency department: a cross-sectional analysis of a large population-based sample. *Diagnosis (Berl).* 2014;1(2):155-166.
36. Schull MJ, Vermeulen MJ, Stukel TA. The risk of missed diagnosis of acute myocardial infarction associated with emergency department volume. *Ann Emerg Med.* 2006;48(6):647-655.
37. Centers for Disease Control and Prevention. CDC FluView Interactive: ILI Surviellence. https://www.cdc.gov/flu/weekly/fluviewinteractive.htm. Published 2020. Accessed May 14, 2020, 2020.
38. Verelst S, Pierloot S, Desruelles D, Gillet JB, Bergs J. Short-term unscheduled return visits of adult patients to the emergency department. *J Emerg Med.* 2014;47(2):131-139.
39. Medford-Davis LN, Singh H, Mahajan P. Diagnostic Decision-Making in the Emergency Department. *Pediatr Clin North Am.* 2018;65(6):1097-1105.
40. Chisholm CD, Collison EK, Nelson DR, Cordell WH. Emergency department workplace interruptions: are emergency physicians "interrupt-driven" and "multitasking"? *Acad Emerg Med.* 2000;7(11):1239-1243.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

41. Saldana L, Abid M, McCarthy N, Hunter N, Inglis R, Anders K. Factors affecting delay in initiation of treatment of tuberculosis in the Thames Valley, UK. *Public Health.* 2013;127(2):171-177.
42. Long NH, Johansson E, Lönnroth K, Eriksson B, Winkvist A, Diwan VK. Longer delays in tuberculosis diagnosis among women in Vietnam. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 1999;3(5):388-393.
43. Chen TC, Lu PL, Lin CY, Lin WR, Chen YH. Fluoroquinolones are associated with delayed treatment and resistance in tuberculosis: a systematic review and meta-analysis. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases.* 2011;15(3):e211-216.
44. Hogan CA, Puri L, Gore G, Pai M. Impact of fluoroquinolone treatment on delay of tuberculosis diagnosis: A systematic review and meta-analysis. *Journal of clinical tuberculosis and other mycobacterial diseases.* 2017;6:1-7.
45. Rush B, Wormsbecker A, Stenstrom R, Kassen B. Moxifloxacin Use and Its Association on the Diagnosis of Pulmonary Tuberculosis in An Inner City Emergency Department. *The Journal of emergency medicine.* 2016;50(3):371-375.
46. Zão I, Ribeiro AI, Apolinário D, Duarte R. Why does it take so long? The reasons behind tuberculosis treatment delay in Portugal. *Pulmonology.* 2019;25(4):215-222.
47. Farah MG, Rygh JH, Steen TW, Selmer R, Heldal E, Bjune G. Patient and health care system delays in the start of tuberculosis treatment in Norway. *BMC infectious diseases.* 2006;6(1):33.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Author Contributions

Aaron C. Miller – designed the study, developed the methodological approach, drafted and revised the final manuscript, and helped to obtained funding for the research.

Alan Arakkal – helped to conduct statistical analysis, helped draft the methods and results section, reviewed and revised the final manuscript.

Scott Koeneman – helped to conduct statistical analysis, helped draft the methods and results section, reviewed and revised the final manuscript.

Joseph E. Cavanaugh – helped in developing the methodological approach, provided guidance on the statistical analysis, reviewed and revised the final manuscript.

Alicia K. Gerke – provided clinical expertise and feedback, helped review and revise the final manuscript, and helped to obtain funding for the research.

Douglas B. Hornick – provided clinical expertise and feedback, helped review and revise the final manuscript.

Philip M Polgreen – helped conceive the study objective, provided clinical guidance, drafted and revised the final manuscript, and helped to obtain funding for the research

Data Sharing

The IBM Marketscan Research Databases can be obtained from *IBM Watson Health*. The code used for the simulation and statistical analysis is available on GitHub at https://github.com/aarmiller/tb_delay_2020.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Figure Headings

Figure 1 – Diagram for Identifying SSD Visits – SSD visits include symptoms, symptomatically-similar diagnoses and testing or exam-based diagnoses that suggest an active tuberculosis infection may be present in the patient. Potential diagnostic opportunities are defined as SSD-related visits that occur during the diagnostic opportunity window (i.e., the window prior to index diagnosis where delays are biologically plausible).

Figure 2 - **Upward Spike in Healthcare Visits Prior to Index Tuberculosis Diagnosis** – Counts of healthcare visits for both SSD-related diagnoses (blue) and any diagnosis (red) are depicted each week leading up to the index tuberculosis diagnosis. Before the index tuberculosis diagnosis there is a spike in SSD related healthcare visits in inpatient, ED and outpatient settings.

Figure 3 – Trends in the Observed and Expected Number of SSD-related Visits. The red line depicts the trend in expected SSD-related visits, which was estimated using data prior to the change-point. The blue line depicts the trend in the observed number of visits after the change-point. The area between the blue and red lines depicts the number of SSD-related visits that represent likely diagnostic opportunities.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Tables

Table 1 Baseline Study Population Characteristics

	Total Patients (% of patients)
Age at diagnosis	
<18	95 (2.8%)
18-35	436 (12.9%)
36-45	437 (13.0%)
46-55	600 (17.8%)
56-65	800 (23.7%)
>65	1003 (29.8%)
Sex	
Male	1613 (47.8%)
Female	1758 (52.2%)
Enrollment time prior to index (years)	
Mean	4.1
Median	3.1
Range	1.0 - 16.5
Count ≤ 1.5 years	533 (15.8%)
Count ≤ 2 years	981 (29.1%)
Count ≤ 3 years	1630 (48.4%)
Count > 3 years	1741 (51.6%)
Region	
Rural	355 (10.5%)
Urban	2998 (88.9%)
Missing	18 (0.5%)

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 2 – Estimated Number of Missed Opportunities and Duration of Diagnostic Delay Based on Simulation Model

Metric / Category	Count (Percentage of all patients) / Mean	CI (from bootstrapping)
Number of Missed opportunities		
0 Days	769 (22.8%)	719 - 822 (21.3 - 24.4%)
>= 1 Days	2602 (77.2%)	2549 - 2652 (75.6 - 78.7%)
>= 2 Days	2065 (61.2%)	1981 - 2148 (58.8 - 63.7%)
>= 3	1563 (46.4%)	1457 - 1667 (43.2 - 49.5%)
>= 4	1137 (33.7%)	1028 - 1248 (30.5 - 37.0%)
>= 5	803 (23.8%)	704 - 908 (20.9 - 26.9%)
Mean - Overall	3.89	3.65 - 4.14
Mean - Outpatient	3.46	3.24 - 3.69
Mean - Inpatient	0.20	0.19 - 0.22
Mean - ED	0.23	0.21 - 0.24
Duration of Delays		
>= 0 Days	2602 (100.0%)	2549 - 2652 (NA)
>= 10 Days	2354 (90.4%)	2284 - 2420 (89.3 - 91.5%)
>= 20 Days	1990 (76.5%)	1895 - 2080 (74.1 - 78.7%)
>= 30 Days	1615 (62.1%)	1495 - 1731 (58.4 - 65.5%)
>= 40 Days	1260 (48.4%)	1114 - 1401 (43.5 - 53.0%)
>= 50 Days	928 (35.6%)	769 - 1087 (30.0 - 41.1%)
>= 60 Days	635 (24.4%)	478 - 801 (18.7 - 30.3%)
>= 70 Days	388 (14.9%)	253 - 540 (9.9 - 20.4%)
>= 80 Days	204 (7.8%)	105 - 327 (4.1 - 12.4%)
>= 90 Days	86 (3.3%)	30 - 170 (1.2 - 6.4%)
>= 100 Days	25 (1.0%)	3 - 70 (0.1 - 2.7%)
>= 110 Days	4 (0.1%)	0 - 19 (0.0 - 0.7%)
Mean Among Delayed	41.00	37.54 - 44.77
Mean Everyone Included	31.66	28.51 - 35.11

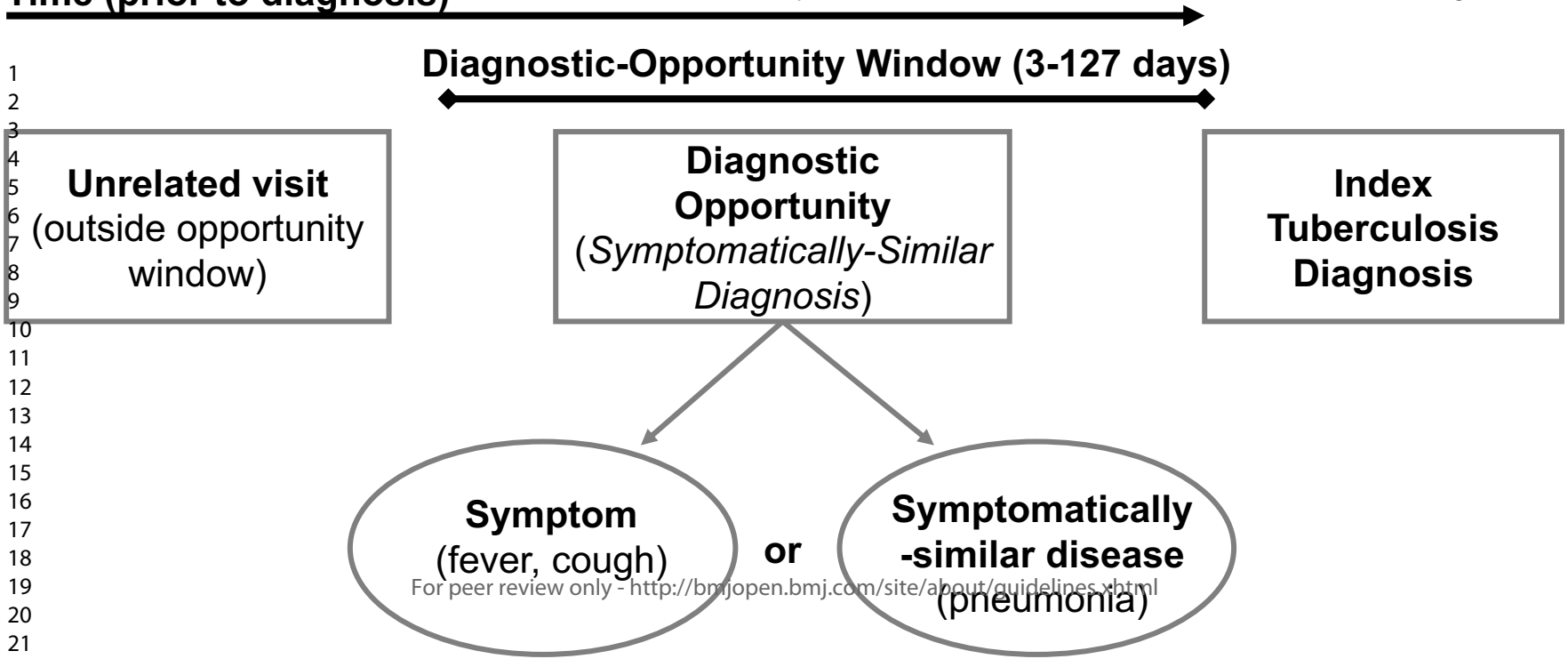
For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 3 – Regression Results for Likelihood of Experiencing a Missed Opportunity

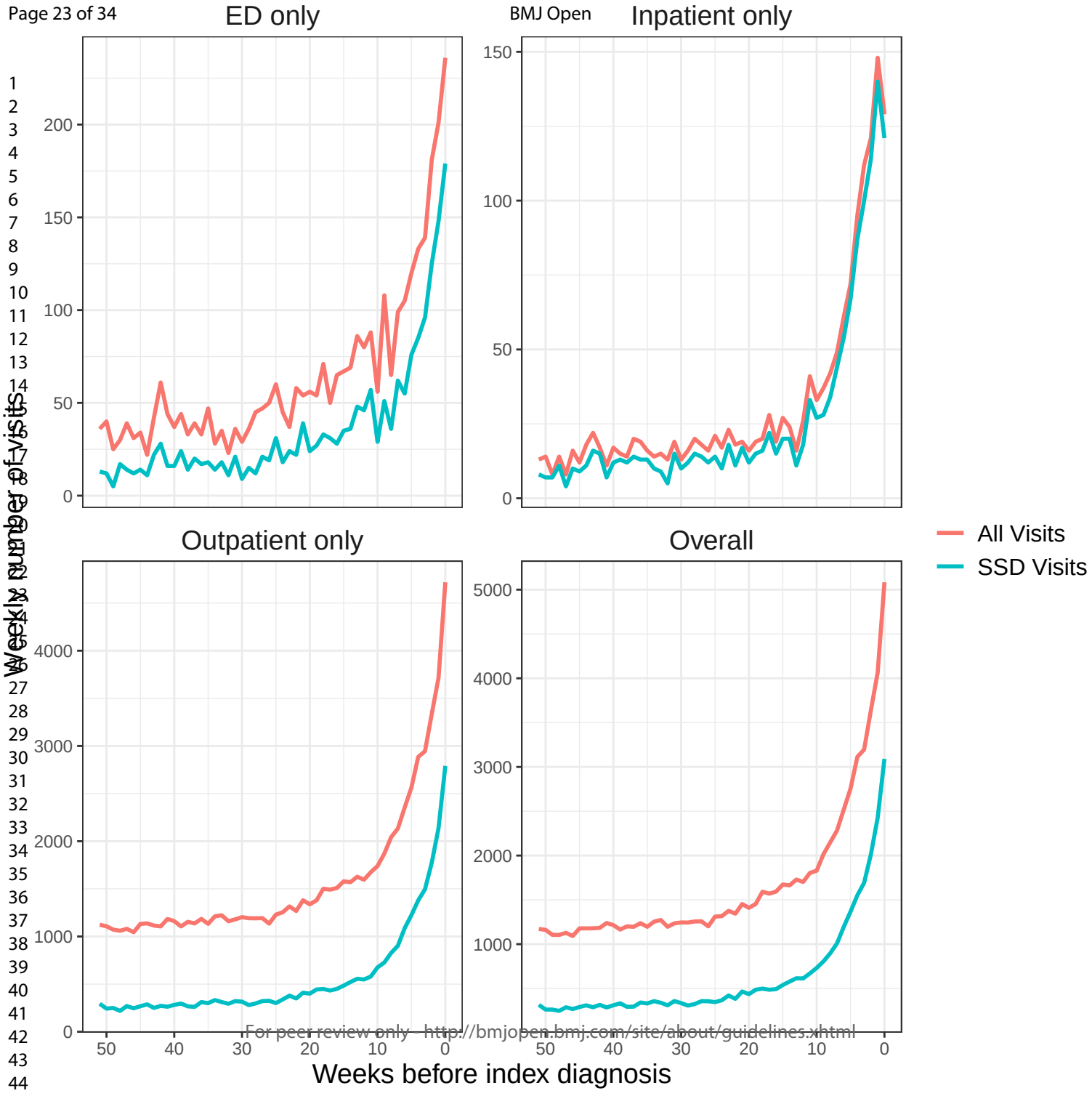
Coefficient	Effect Estimate	95% CI	P-value
Weekend (visits that occurred on a Saturday or Sunday)	1.495	1.272, 1.758	<0.001
Age > 65	1.262	1.156, 1.377	<0.001
Settings visited			
Outpatient only	Ref	Ref	Ref
All three (inpatient, outpatient, and ED)	0.128	0.089, 0.185	<0.001
ED only	2.340	1.540, 3.555	<0.001
Inpatient only	0.123	0.106, 0.142	<0.001
Inpatient and ED	0.142	0.110, 0.184	<0.001
Inpatient and outpatient	0.124	0.105, 0.145	<0.001
Outpatient and ED	1.324	0.968, 1.811	0.079
Urban vs. not urban	0.874	0.771, 0.990	0.034
ILI (≥ 3.8 %) * outpatient interaction	1.259	1.052, 1.507	0.012
Asthma prior to change point	1.331	1.138, 1.557	<0.001
COPD prior to change point	1.372	1.230, 1.531	<0.001
Chest CT prior to change point	1.149	1.018, 1.296	0.025
Chest X-Ray prior to change point	1.231	1.121, 1.353	<0.001
Fluoroquinolones between change point and 3 days prior to index	1.578	1.435, 1.734	<0.001

For peer review only

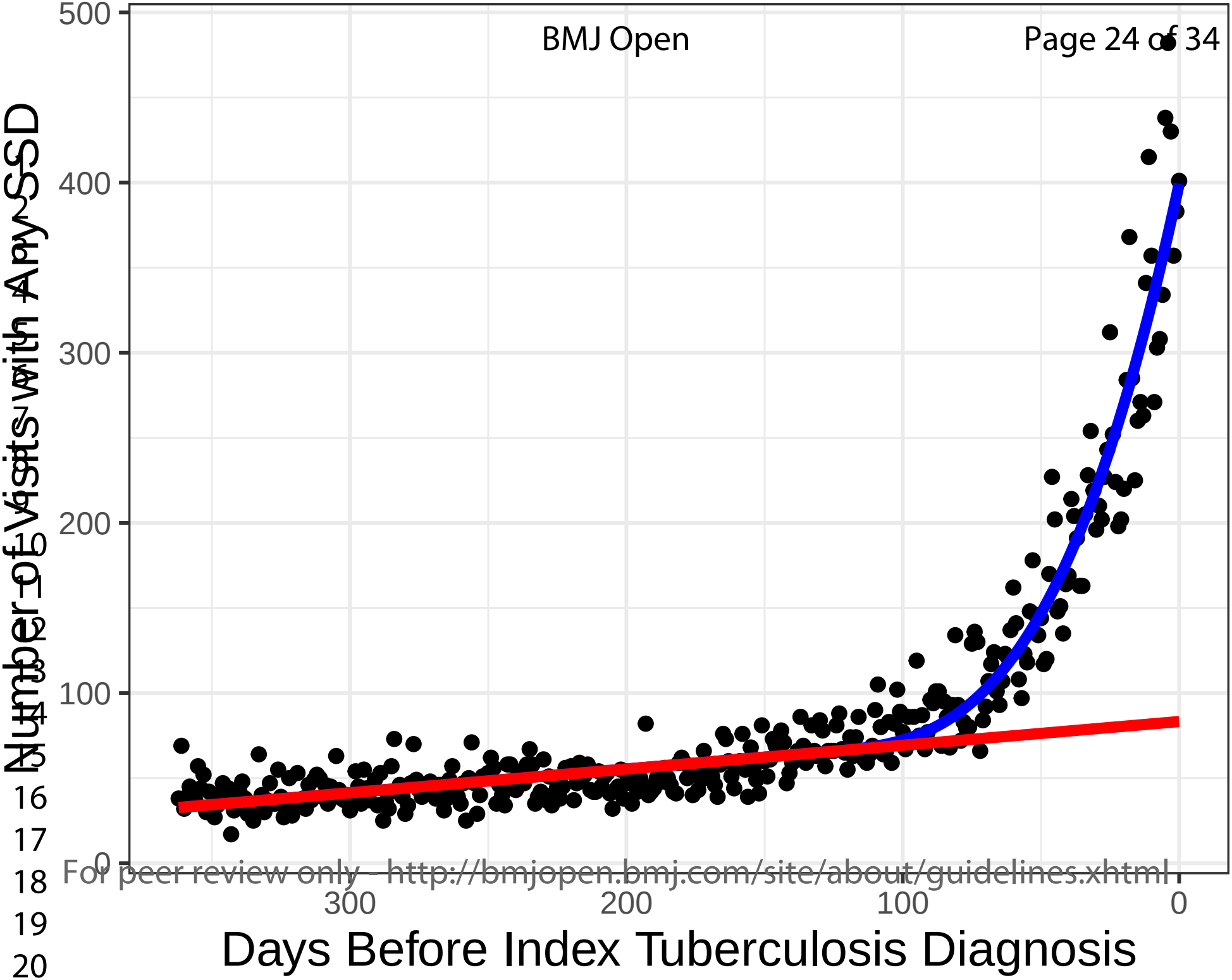
For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Supplementary Material

Supplementary Table 1 - List of SSDs used to identify potential diagnostic opportunities

Category	Symptomatically Similar Diagnosis (SSD)	ICD-9-CM	ICD-10-CM
Alternative-Cardio-Sino-Pulmonary-Based Diagnoses	Tonsillitis	474.12, 474.2, 474.8, 474.9	J35.1
	Respiratory Failure	517.3, 518.81, 518.82, 518.83, 518.84, 799.1	J80, J96.00, J96.01, J96.02, J96.10, J96.12, J96.20, J96.21, J96.22, J96.90, J96.92
	Respiratory Cancer	163.0, 163.1, 163.8, 163.9, 165.0, 165.8, 165.9, 231.1, 231.8, 231.9	C33, C38.4, C39.0, C39.9, C45.0, D02.1, D02.3, D02.4
	Pleurisy Pneumothorax	511.0, 511.1, 511.8, 511.89, 512.0, 512.8, 512.81, 512.82, 512.83, 512.84, 512.89, 518.1, 518.2	J86.9, J92.0, J92.9, J93.0, J93.12, J93.81, J93.82, J93.83, J94.0, J94.1, J94.2, J94.8, J94.9, J98.19, J98.2, J98.3, R09.1
	Other Upper Respiratory Disease	472.0, 477.0, 477.2, 477.8, 477.9, 478.1, 478.19, 478.20, 478.29, 478.30, 478.31, 478.32, 478.33, 478.34, 478.4, 478.5, 478.70, 478.74, 478.75, 478.79, 478.8, 519.1, 519.11, 519.19, 519.3, 784.40, 784.49, 784.7, 784.8	J30.0, J30.1, J30.2, J30.81, J30.89, J30.9, J31.0, J34.2, J34.89, J37.0, J37.1, J38.00, J38.01, J38.02, J38.1, J38.2, J38.3, J38.4, J38.5, J38.6, J38.7, J39.2, J39.3, J39.8, J39.9, J98.01, J98.51, R04.0, R04.1, R09.81
	Other Lower Respiratory Disease	514, 515, 516.0, 516.1, 516.2, 516.3, 516.30,	J81.0, J81.1, J82, J84.01, J84.02, J84.03, J84.09, J84.10, J84.111, J84.112,

For peer review only

		516.31, 516.32, 516.33, 516.34, 516.35, 516.36, 516.37, 516.4, 516.5, 516.8, 516.9, 517.2, 517.8, 518.3, 518.4, 518.89, 519.4, 519.8, 519.9, 786.00, 786.09, 786.39, 786.9, 793.11, 794.2	J84.113, J84.114, J84.115, J84.116, J84.117, J84.17, J84.2, J84.81, J84.82, J84.89, J84.9, J98.4, J98.6, J98.8, J99, R04.89, R04.9
	Lung Disease Due to External Agents	495.0, 495.1, 495.2, 495.3, 495.4, 495.5, 495.6, 495.7, 495.8, 495.9, 500, 501, 502, 503, 504, 505, 506.0, 506.1, 506.2, 506.3, 506.4, 506.9, 507.1, 507.8, 508.0, 508.1, 508.2, 508.8, 508.9	J60, J61, J62.0, J62.8, J63.0, J63.1, J63.2, J63.3, J63.4, J63.5, J63.6, J64, J66.0, J66.1, J66.2, J66.8, J67.0, J67.1, J67.2, J67.3, J67.4, J67.5, J67.6, J67.7, J67.8, J67.9, J68.0, J68.1, J68.2, J68.3, J68.4, J68.8, J68.9, J69.1, J69.8, J70.0, J70.1, J70.2, J70.3, J70.4, J70.5, J70.8, J70.9
	Lung Cancer	162.2, 162.3, 162.4, 162.5, 162.8, 162.9, 209.21	C34.00, C34.01, C34.02, C34.10, C34.11, C34.12, C34.2, C34.30, C34.31, C34.32, C34.80, C34.81, C34.82, C34.90, C34.91, C34.92, C7A.090, D02.20, D02.21, D02.22
	Hemoptysis	786.39	R04.8, R04.89, R04.9
	COPD	491.2, 491.20, 491.21, 491.22, 492.0, 492.8, 494, 494.0, 494.1, 496	J41.8, J43.0, J43.1, J43.2, J43.8, J43.9, J44.1, J44.9, J47.1, J47.9
	Asthma	493.00, 493.01, 493.02, 493.10, 493.11, 493.12, 493.20, 493.21, 493.22, 493.81, 493.82, 493.90, 493.92	J45.20, J45.21, J45.22, J45.30, J45.31, J45.32, J45.40, J45.41, J45.42, J45.50, J45.51, J45.52, J45.901, J45.902, J45.909, J45.990, J45.991, J45.998

For peer review only

	Aspiration Pneumonitis	507.0	J69.0
	Additional Codes	135, 197.0, 212.3, 235.7, 239.1, 289.1, 416.8, 423.9, 428.0, 446.4	I50.9, J85.0, N39.0
Alternative-Infectious-Disease-Based Diagnoses	Tonsillitis	463, 474.0, 474.00, 474.01, 474.02, 474.10, 474.11, 475	J03.80, J03.81, J03.90, J03.91, J35.01, J35.02, J35.03, J35.2, J35.3, J35.8, J35.9, J36
	Pneumonia	112.4, 114.0, 114.4, 115.05, 115.15, 115.95, 130.4, 136.3, 480.0, 480.1, 480.2, 480.8, 480.9, 481, 482.0, 482.1, 482.2, 482.3, 482.30, 482.31, 482.32, 482.39, 482.4, 482.40, 482.41, 482.42, 482.49, 482.8, 482.81, 482.83, 482.84, 482.89, 482.9, 483, 483.0, 483.1, 483.8, 484.1, 484.3, 484.6, 484.7, 484.8, 485, 486, 513.0	A31.0, A37.01, A37.11, A43.0, A48.1, B25.0, B37.1, B38.0, B38.1, B38.2, B39.0, B39.1, B39.2, B58.3, B59, B77.81, J12.0, J12.1, J12.2, J12.3, J12.89, J12.9, J13, J14, J15.0, J15.1, J15.20, J15.211, J15.212, J15.29, J15.3, J15.4, J15.5, J15.6, J15.7, J15.8, J15.9, J16.0, J16.8, J17, J18.0, J18.1, J18.8, J18.9, J85.1
	Pleurisy Pneumothorax	510.0	J86.0, J93.11
	Other Upper-Respiratory Infection	460, 461.0, 461.1, 461.2, 461.3, 461.9, 462, 464.0, 464.00, 464.01, 464.11, 464.20, 464.21, 464.30, 464.31, 464.4, 464.50, 464.51, 465.0, 465.8, 465.9, 473.0, 473.1, 473.2,	J00, J01.00, J01.01, J01.10, J01.11, J01.20, J01.21, J01.30, J01.31, J01.40, J01.41, J01.80, J01.81, J01.90, J01.91, J02.0, J02.8, J02.9, J03.00, J03.01, J04.0, J04.10, J04.11, J04.2, J04.30, J04.31, J05.0, J05.10, J05.11, J06.0, J06.9, J32.0, J32.1, J32.2,

For peer review only

		473.3, 473.8, 473.9	J32.3, J32.4, J32.8, J32.9, R09.82
	Other Upper Respiratory Disease	472.2, 476.0, 476.1, 478.21, 478.22, 478.24, 478.71, 478.9, 519.2	J31.1, J31.2, J39.0, J39.1, J98.09, J98.5, J98.59, R07.0
	Other Lower Respiratory Disease	513.1	J18.2, J22, J85.2, J85.3, J98.9, R06.6, R06.82
	Influenza	487.0, 487.1, 487.8, 488, 488.1, 488.11, 488.12, 488.19, 488.81, 488.82, 488.89	J09.X1, J09.X2, J09.X3, J09.X9, J10.00, J10.01, J10.08, J10.1, J10.2, J10.89, J11.00, J11.08, J11.1, J11.2, J11.81, J11.82, J11.83, J11.89
	COPD	490, 491.0, 491.1, 491.8, 491.9	J40, J41.0, J41.1, J42, J44.0, J47.0
	Bronchitis	466.0, 466.1, 466.11, 466.19	J20.0, J20.1, J20.2, J20.3, J20.4, J20.5, J20.6, J20.7, J20.8, J20.9, J21.0, J21.1, J21.8, J21.9
	Additional Codes	038.9, 079.99, 310, 340, 391, 599.0, 830, 995.91	A41.9, B34.9, D14.30, D38.1, D49.1, D86.0, D86.9
Symptom-Based Diagnoses	Other Upper Respiratory Disease	784.1, 784.41, 784.42, 784.9, 784.99	R49.0, R49.8, R49.9
	Other Lower Respiratory Disease	786.02, 786.05, 786.07, 786.2, 786.3, 786.30, 786.4, 786.52	R04.2, R05, R06.00, R06.01, R06.02, R06.03, R06.09, R06.2, R06.89, R06.9, R07.1, R07.81, R09.3
	Hemoptysis	786.3, 786.30	R04.2
	Fever	780.6, 780.60, 780.61	R50, R50.81, R50.9
	Cough	786.2	R05
	Additional Codes	780.79, 780.8, 783.21, 786.50, 786.51, 786.59	R07.2, R07.82, R07.89, R07.9, R53.1, R53.81, R53.83, R61, R63.4
Testing-Imaging-or Physical-Exam-Based Diagnoses	Pleurisy Pneumothorax	511.9, 518.0	J90, J91.8, J93.9, J98.11
	Other Lower Respiratory Disease	786.6, 786.7, 793.1, 793.19	R09.02, R91.1, R91.8

For peer review only

	Additional Codes	263.9, 276.1, 285.29, 285.9, 288.60, 289.3, 429.3, 782.2, 784.2, 785.0, 785.6, 799.02, 799.4	D64.9, D72.829, E871, I51.7, R00.0, R22.0, R22.1, R22.2, R59, R59.0, R59.1, R59.9
Procedure Codes		**CPT Code**	
CT – Chest		71260, 71250, 71270	
X-ray - Chest		71010, 71015, 71020, 71021, 71022, 71023, 71030, 71034, 71035, 71101, 71111, 71120,71045, 71046, 71047, 71048	

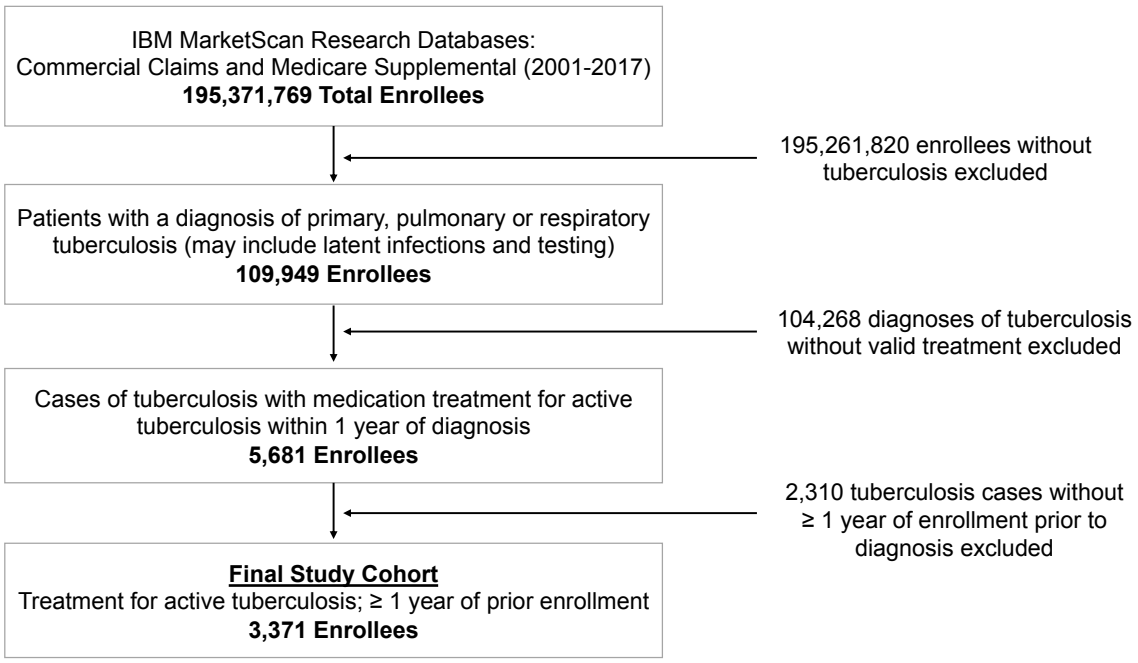
For peer review only

Supplementary Table 2 – ILI Indicators for the optimal ILI cutoff. We used AIC to select the optimal cutoff for defining peak ILI activity; this was determined to be an ILI level >3.8%. Appendix Figure 2 also depicts the trend in ILI across time along with the threshold used to define peak activity.

Period	ILI >= 3.8%
2001/01/01 - 2001/01/14	0
2001/01/15 - 2001/02/04	1
2001/02/05 - 2003/11/23	0
2003/11/24 - 2003/12/28	1
2003/12/29 - 2005/01/23	0
2005/01/24 - 2005/02/20	1
2005/02/21 - 2008/01/20	0
2008/01/21 - 2008/03/09	1
2008/03/10 - 2009/08/30	0
2009/08/31 - 2009/11/15	1
2009/11/16 - 2011/01/16	0
2011/01/17 - 2011/02/20	1
2011/02/21 - 2012/12/09	0
2012/12/10 - 2013/01/27	1
2013/01/28 - 2013/12/22	0
2013/12/23 - 2014/01/05	1
2014/01/06 - 2014/12/14	0
2014/12/15 - 2015/01/25	1
2015/01/26 - 2017/01/15	0
2017/01/16 - 2017/02/19	1
2017/02/20 - 2017/12/10	0
2017/12/11 - 2017/12/24	1

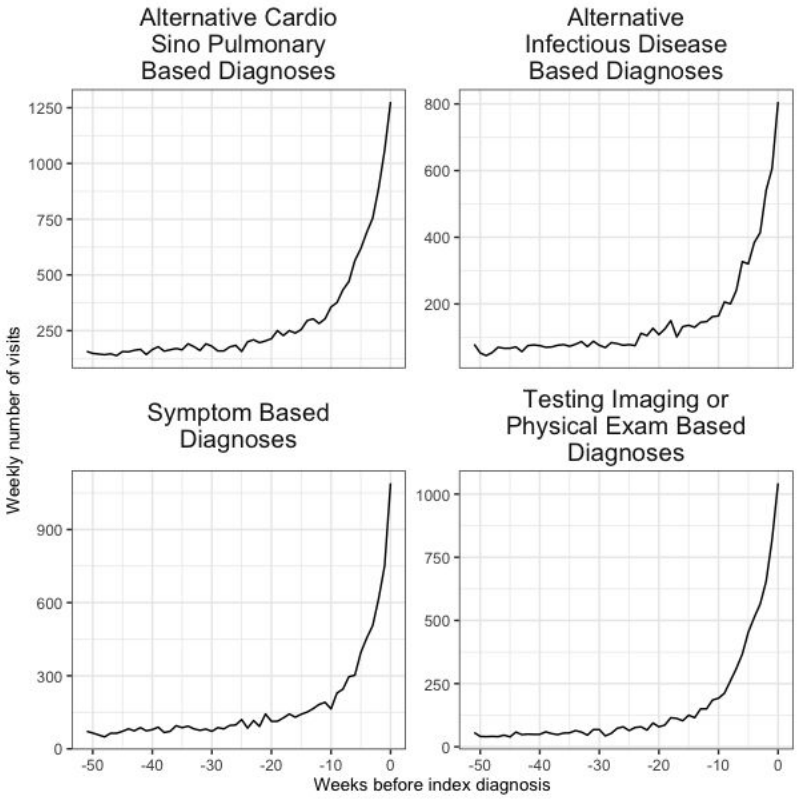
For peer review only

Supplementary Figure 1 – Flow diagram of patient inclusion and exclusion criteria



review only

Supplementary Figure 2 – Trend in SSD visits prior to diagnosis for the four SSD categories



view only

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

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	4-5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	5
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
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	*4*
Bias	9	Describe any efforts to address potential sources of bias	6-7
Study size	10	Explain how the study size was arrived at	4-5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6-7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5-7
		(*b*) Describe any methods used to examine subgroups and interactions	5-7
		(*c*) Explain how missing data were addressed	5
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	6-7

Continued on next page

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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	8
		(b) Give reasons for non-participation at each stage	8
		(c) Consider use of a flow diagram	See Supplement
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8 (Table 1)
		(b) Indicate number of participants with missing data for each variable of interest	N/A (all enrollees continuously enrolled)
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	N/A
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	*8-9*
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
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	8-9
		(*b*) Report category boundaries when continuous variables were categorized	8-9
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	9
Discussion			
Key results	18	Summarise key results with reference to study objectives	10-11
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	11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11
Generalisability	21	Discuss the generalisability (external validity) of the study results	11
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	1

*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.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

Incidence, duration and risk factors associated with delayed and missed diagnostic opportunities associated with Tuberculosis: A population-based longitudinal study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-045605.R1
Article Type:	Original research
Date Submitted by the Author:	30-Dec-2020
Complete List of Authors:	Miller, Aaron; University of Iowa, Epidemiology Arakkal, Alan; The University of Iowa, Epidemiology Koeneman, Scott; The University of Iowa, Biostatistics Cavanaugh, Joe; The University of Iowa, Biostatistics Gerke, Alicia; The University of Iowa, Internal Medicine Hornick, Douglas; University of Iowa, Internal Medicine Polgreen, Philip; University of Iowa Roy J and Lucille A Carver College of Medicine, Internal Medicine
Primary Subject Heading:	Infectious diseases
Secondary Subject Heading:	Epidemiology, Respiratory medicine
Keywords:	Tuberculosis < INFECTIOUS DISEASES, Epidemiology < INFECTIOUS DISEASES, GENERAL MEDICINE (see Internal Medicine), Respiratory infections < THORACIC 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.

For peer review only

Incidence, duration and risk factors associated with delayed and missed diagnostic opportunities associated with Tuberculosis: A population-based longitudinal study

Aaron C Miller[1]*, Alan Arakkal[1], Scott Koeneman[2], Joseph E Cavanaugh[2], Alicia K Gerke[3], Douglas B Hornick[3], Philip M Polgreen[3,1]

[1]Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, Iowa; [2]Department of Biostatistics, College of Public Health, University of Iowa, Iowa City, Iowa; [3]Department of Internal Medicine, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, Iowa.

*Corresponding Author: Aaron C Miller, Department of Epidemiology, University of Iowa College of Public Health, 145 N. Riverside Drive, Iowa City, IA 52242. Email: aaron-miller@uiowa.edu. Phone: (319) 384-1584.

Word Count: 299 (abstract), 3,990 (manuscript)

Funding: This work was supported by the *Agency for Healthcare Research and Quality* grant number 5R01HS027375, and PMP has received funding from the National Center For Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR002537.

Competing Interests: The authors have no competing interests to declare.

For peer review only

ABSTRACT

Objectives: Missed opportunities to diagnose tuberculosis are costly to patients and society. In this study we (1) estimate the frequency and duration of diagnostic delays among patients with active pulmonary tuberculosis and (2) determine the risk factors for experiencing a diagnostic delay.

Design: A retrospective cohort study of patients with tuberculosis using longitudinal healthcare encounters prior to diagnosis.

Setting: Commercially insured enrollees from the Commercial Claims and Encounters or Medicare Supplemental IBM Marketscan Research Databases, 2001-2017.

Participants: All patients diagnosed with, and receiving treatment for, pulmonary tuberculosis, enrolled at least 365 days prior to diagnosis.

Primary and secondary outcome measures: We estimated the number of visits with tuberculosis-related symptoms prior to diagnosis that would be expected to occur in absence of delays and compared this estimate to the observed pattern. We computed the number of visits representing a delay and used a simulation-based approach to estimate the number of patients experiencing a delay, number of missed opportunities per patient and duration of delays (i.e., time between diagnosis and earliest missed opportunity). We also explored risk factors for missed opportunities.

Results: We identified 3,371 patients diagnosed and treated for active tuberculosis that could be followed for 1 year prior to diagnosis. We estimated 77.2% (95% CI: 75.6-78.7%) of patients experienced at least one missed opportunity; of these patients, an average of 3.89 (95% CI: 3.65-4.14) visits represented a missed opportunity, and the mean duration of delay was 31.66 days (95% CI: 28.51-35.11). Risk factors for delays included outpatient or emergency department settings, weekend visits, patient age, influenza season presentation, history of chronic respiratory symptoms and prior fluroquinolone use.

Conclusions: Many patients with tuberculosis experience multiple missed diagnostic opportunities prior to diagnosis. Missed opportunities occur most commonly in outpatient settings and numerous patient-, environment- and setting-specific factors increase risk for delays.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Strengths and limitations of this study

- This study reviewed longitudinal healthcare records for a large population of insured enrollees (over 195-million represented) spanning an extensive time period (2001-2017) and covering a range of healthcare settings (inpatient, outpatient, and emergency department).
- A simulation-based analysis was conducted to identify visits most likely to represent a diagnostic delay, while excluding coincidental visits that may appear to be missed opportunities.
- This study relied on diagnostic codes (ICD-9/ICD-10) to identify index cases of tuberculosis, and such codes may lack specificity for identifying active tuberculosis. Medication claims were used to help validate diagnosis codes by identifying patients receiving medications used to treat active tuberculosis.
- This study also relied on diagnostic codes to identify signs and symptoms of tuberculosis prior to diagnosis. Such records may not capture all visits where symptoms occurred (e.g., symptoms recorded in clinic notes). We conducted a sensitivity analysis to evaluate the potential sensitivity of our findings to visits without related symptom codes.
- Without more granular patient data, we cannot confirm that all patient visits we identify represent diagnostic errors.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BACKGROUND

The incidence of tuberculosis has been decreasing in the United States during the past several decades,[1,2] but recently the rate of decrease has slowed.[1,3] To further reduce the incidence of tuberculosis, the rapid identification and treatment of new cases is essential.[3] However, as the incidence of tuberculosis decreases, so may familiarity with the disease among clinicians,[4] resulting in an increase in diagnostic delays.[5,6] Because these delays are, in part, a function of the familiarity and experience of clinicians with a particular disease,[5-8] as the disease becomes less common, diagnostic delays for tuberculosis may become more common.[5,7]

Diagnostic delays of tuberculosis are important to consider for several reasons. First, delays are common in the United States[6,7,9] and other lower-prevalence countries.[8,10-12] Second, delays may contribute to worse clinical outcomes,[13-15] and increased healthcare costs.[16] Third, diagnostic delays for tuberculosis are especially important because delays contribute to additional exposures and thus, additional cases of tuberculosis.[17,18] Substantial diagnostic delays contributing to increased transmission have occurred in both community[19-22] and healthcare settings.[10,23-25]

Traditional approaches to investigate diagnostic delays have focused on single centers, most commonly hospitals, or alternatively have depended on public health registries that rely on patient recall.[8,26] Although diagnostic errors occur in hospitals, opportunities to understand and reduce diagnostic delays may frequently occur in ambulatory settings where patients often first present with signs and symptoms of a disease. Multiple investigations focusing on emergency department visits have highlighted missed opportunities to diagnose tuberculosis.[6,27-29] Thus, to enable a more complete understanding of diagnostic delays requires consideration of sequential healthcare visits across outpatient clinic visits, emergency department visits and hospitalizations. Also, when diagnostic delays are detected, it may be difficult to learn about risk factors for diagnostic delays if patients present in multiple different settings before the diagnosis is made.

Before interventions to decrease diagnostic delays can be designed and implemented, a better understanding of the incidence of and risk factors for diagnostic delays is needed, especially in lower-incidence countries. Thus, the goal of this study is to propose a population-based approach for estimating the incidence and duration of diagnostic delays associated with tuberculosis, and also to describe the risk factors associated with patients experiencing a diagnostic delay.

METHODS

Data Source: We used longitudinal insurance claims data from the IBM Marketscan Commercial Claims and Encounters and Medicare Supplemental databases from 2001 through 2017. The Commercial Claims data contain information for individuals with employer-sponsored health plans (employees, retirees, dependents, and spouses) from participating large employers, health plans and government organizations. The

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Medicare supplemental databases contain information for Medicare-eligible individuals with employer sponsored Medicare Supplemental plans. Together, these databases contain claims for over 195 million enrollees across the United States, representing over six-billion enrollment months. Claims from outpatient, emergency and inpatient visits are provided along with outpatient medications.

Permission to use these data were granted to our research team from *IBM.* This research used de-identified claims data, studies of this type are deemed non-human subjects research by the University of Iowa Institutional Review Board.

Study Population: We identified all patients diagnosed with primary, pulmonary, respiratory or miliary tuberculosis using the ICD-9-CM diagnosis codes 010.X, 011.X, 012.X and 018.X, and the ICD-10-CM codes A15.X and A19.X. Because non-pulmonary tuberculosis presents with different signs and symptoms, we did not include codes for tuberculosis of the central nervous system, intestines, peritoneum, mesenteric glands, bones, joints, genitourinary system or other organs. We required cases to be enrolled for at least one year prior to their initial tuberculosis diagnosis; this first diagnosis was labeled as the *index diagnosis*. Because diagnosis codes alone lack specificity for identifying active tuberculosis,[30] we restricted our analysis to patients with evidence of treatment for active tuberculosis near the index diagnosis using outpatient medication claims.[31] Specifically, we identified treatment with the following set of medications: *Isoniazid* and *Rifampicin/Rifampin*, *Pyrazinamide*, or *Ethambutol*. We considered patients whose treatment began within 1 year of the index diagnosis. We performed a sensitivity analysis using cases where treatment occurred within 2 months of diagnosis. If treatment began prior to the initial tuberculosis diagnosis, we used the treatment start date as the index diagnosis date.

Statistical Analysis: We conducted two primary statistical analyses to address the following objectives: (1) to estimate the incidence and duration of diagnostic delays associated with tuberculosis, and (2) to estimate the risk factors for experiencing a diagnostic delay. We started by identifying *potential diagnostic delays* by looking for *symptomatically similar diagnoses* (SSDs) that occurred during healthcare visits prior to the index tuberculosis diagnosis. We defined SSDs to be diagnoses that include, or share, similar symptoms to active pulmonary tuberculosis. SSDs may include diagnoses in one of four categories:

(1) ***General symptoms*** of active infection, such as cough, fever, weight loss, or hemoptysis;
(2) ***Symptomatically similar infections*** that share similar symptoms to tuberculosis, such as pneumonia, influenza or bronchitis;
(3) ***Symptomatically similar cardio-sino-pulmonary diseases or syndromes***, such as COPD, asthma or lung cancer;
(4) ***Testing, imaging or physical exam-based diagnoses***, such as anemia or swollen lymph nodes.

Supplementary Table 1 describes the individual diagnoses and ICD-9/10 codes used to identify the four types of SSD conditions. This list was developed based on a review of

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

prior literature of diagnostic delays for tuberculosis.[6] We identified SSDs during visits in the time prior to the index diagnosis where diagnostic opportunities may plausibly occur, between 3 and τ days prior; we denoted the period [3, τ] as the *diagnostic-opportunity window*. The value τ is the upper bound of the diagnostic-opportunity, reflecting the longest plausible diagnostic delay; this is estimated based on a change-point analysis described below. We disregard visits within 3 days of the index diagnosis, to account for lags in diagnostic testing. Figure 1 depicts the process used to identify potential diagnostic opportunities. This type of "look-back" approach has been referred to as Symptom-Disease Pair Analysis of Diagnostic Error (SPADE),[32] which has been used to identify diagnostic delays associated with numerous diseases.[6,33-36]

Estimating Incidence of Diagnostic Delays: Visits occurring prior to an index diagnosis of tuberculosis that contain an SSD may represent a missed diagnostic opportunity but may also represent a coincidental visit (e.g., unrelated respiratory infection). To account for visits representing coincidental diseases, and not a missed opportunity, we compared the difference between *expected* and *observed* patterns of SSD visits prior to the index diagnosis. First, we estimated the *expected* number of SSD visits by analyzing the trend in the incidence of SSD visits in the time prior to the diagnostic-opportunity-window, where missed opportunities are unlikely to occur (e.g., τ-365 days prior to tuberculosis diagnosis). We then computed the expected number of visits in the diagnostic-opportunity window (e.g., 3-τ days prior to tuberculosis diagnosis) by extrapolating the prior trend to the diagnostic-opportunity window. Second, we compared the *observed* pattern of SSD visits during the diagnostic-opportunity window to the expected number based on the extrapolated trend. Finally, the number of potential diagnostic opportunities was estimated by the *excess number* of SSD visits: the difference between the observed and expected number. This approach has been used in prior work to estimate the number of diagnostic opportunities associated with AMI, stroke, and other cardiovascular events.[33] To identify the point prior to the index diagnosis where diagnostic opportunities first begin to occur (i.e. the diagnostic-opportunity-window), we used a change-point analysis to detect the point where the trend between observed and expected number of SSD visits begins to deviate. We fit a piecewise regression model with a linear trend prior to the change-point τ and a cubic trend after the change-point, to account for the non-linear pattern in visit counts in the period just prior to diagnosis (see Figure 2 for a depiction of this trend). We used the Akaike Information Criterion (AIC) to select the optimal change-point.

To estimate the number of individuals that experienced a potential diagnostic delay, number of recurrent missed opportunities per patient and the typical duration of delays, we used a bootstrapping approach similar to that of Waxman et al.[33] Specifically, we randomly drew (with replacement) a sample of patients and re-estimated the observed and expected patterns of care. Next, at each period prior to the index tuberculosis diagnosis, we randomly labeled a portion of visits for the resampled patients as "diagnostic delays" based on the computed excess number of SSD visits at that time period. Finally, we computed the number of patients that experienced a diagnostic delay, the number of recurrent missed opportunities per patient and the durations of the diagnostic delays. We repeated this procedure 25,000 times to compute 95% bootstrap-

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

based confidence intervals for the change-point τ, number of potential diagnostic opportunities, number of patients that experienced a diagnostic delay, number of recurrent missed opportunities per patient, and the durations of the diagnostic delays.

Sensitivity Analysis. Because diagnostic codes from administrative records may not capture all signs and symptoms present during a clinic visit (e.g., in clinic notes), SSD-related ICD-9/10 codes may undercount the true number of visits representing a diagnostic opportunity. As a sensitivity analysis, we repeated our estimates of the incidence of diagnostic delays by including all visits that occurred within the diagnostic opportunity window (regardless of the presence of an SSD code). Specifically, we repeated the change-point and bootstrapping analysis described above using all visits prior to the index tuberculosis diagnosis.

Estimating Risk Factors for Missed Diagnostic Opportunities: We analyzed the potential risk factors for diagnostic delays by estimating the likelihood of a patient experiencing a missed opportunity on a given day prior to diagnosis. We treated diagnostic opportunities as a binary outcome – where a patient who has tuberculosis can experience either a missed opportunity (i.e., SSD-related visit in the diagnostic-opportunity window [3, τ]) or a correct diagnosis (i.e., the index diagnosis). Because multiple visits occurring on a single day likely represent a linked episode of care, for each day during the diagnostic-opportunity window or the index diagnosis date, we aggregated all visits containing an SSD or index diagnosis. We created indicators on each day for the specific type of healthcare facility (e.g., inpatient, outpatient, ED). Days with an SSD-related visit during the diagnostic-opportunity window were assigned an outcome of 1 (i.e., missed opportunity) and days representing the index tuberculosis diagnosis are assigned an outcome value of 0 (i.e., correct diagnosis). We then used logistic regression to estimate the likelihood of a visit representing a missed opportunity, while controlling for other risk factors for delay.

We considered a number of patient- and context-specific risk factors for diagnostic delay. Patient demographics include age, sex, and region (i.e., urban vs. rural). Environment and setting specific factors include the year and month of the SSD visit or the index diagnosis, whether visits during a given day involved inpatient, outpatient, or ED settings, or combinations of visits to multiple settings, and a term for tuberculosis incidence at patient location. Because many symptoms associated with pulmonary tuberculosis are similar to influenza like illness (ILI), we created an indicator for peak influenza season based on the national level of outpatient ILI as reported by the CDC [37]. ILI-based indicator values are provided in Supplementary Table 2. Finally, we considered a number of clinical factors: indicators for asthma and COPD prior to the diagnostic-opportunity window were included as markers for pre-existing pulmonary conditions. In addition, indicators for a chest X-ray or a chest CT scan prior to the diagnostic-opportunity window were included because imaging may also indicate pre-existing pulmonary conditions. We also included an indicator for receipt of a fluoroquinolone prior to the delay window. We performed variable selection using backward elimination, evaluating model performance at each stage of the procedure

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

using the AIC. Standard errors were used to compute Wald-type 95% confidence intervals for the logistic regression analysis.

Patient and Public Involvement: No patients were involved.

RESULTS

From 2001 through 2017, a total of 5,681 individuals had a tuberculosis diagnosis and an outpatient prescription drug claim consistent with treatment for active tuberculosis. The final study sample included 3,371 enrollees that had been enrolled for at least 1 year prior to the index tuberculosis diagnosis. Figure 3 provides a flow diagram of inclusion criteria. Table 1 presents baseline criteria (age, sex, enrollment information, and region) for the final study cohort.

Figure 2A depicts the pattern of SSD visits that occurred in the 1-year period prior to the index tuberculosis diagnosis. Supplementary Figure 1 depicts similar patterns for all visits and SSD visits broken down by type of healthcare setting and Supplementary Figure 2 depicts trends for different categories of individual SSD diagnoses. Across nearly all settings and SSD visits, the pattern of SSD visits appears fairly stable, with a very gradual increase from 1-year up to around 100-days prior to the index diagnosis. Starting around 100 days prior to the index diagnosis there is a dramatic spike in SSD visits.

Of the 3,371 case patients we identified, 3,306 (98.1%) patients had at least one healthcare visit in the year prior to their index tuberculosis diagnosis. Of these patients, 1,134 (34.3%) had at least one inpatient visit, 1,301 (39.4%) had at least one ED visit and 3,297 (99.7%) had at least one outpatient visit. Focusing on visits with SSDs, we found 3,084 (91.5%) patients had at least one SSD visit in the year prior to their index tuberculosis diagnosis. Over a third of all visits (37.2%) that occurred in the year prior to the index tuberculosis diagnosis involved one of the SSD conditions. The most common category of SSDs prior to the index tuberculosis diagnoses was alternative cardio sino-pulmonary-based diagnoses (2,322 [68.9%] patients among 15,332 [17.6%] visits), followed by symptom-based diagnoses (2,382 [70.7%] patients among 9,086 [10.5%] visits), testing imaging or physical exam-based diagnoses (2,123 [63.0%] patients among 8,373 [9.6%] visits), and alternative infectious disease-based diagnoses (2,129 [63.2%] patients among 7,921 [9.1%] visits).

Since not all SSD visits represent diagnostic opportunities, we used a bootstrapping/simulation approach to estimate the number of likely diagnostic opportunities based on the observed and expected number of SSD visits prior to the index tuberculosis diagnosis. Our change-point analysis detected a significant increase in the number of SSD visits occurring 127 days (95% confidence Interval (CI): 117-138 days) prior to the index diagnosis; this represents the start of the diagnostic-opportunity window (i.e., maximum duration of delay). Figure 2B summarizes the observed and

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

expected trend lines estimated from our change-point analysis. Across all patients, 2,903 (86.1%) patients had at least one SSD during this diagnostic-opportunity window.

There was a total of 19,818 SSD visits that occurred during the diagnostic-opportunity window. Of these visits, based on our simulation analysis, we estimated that 10,118 (51.1%) represented a missed opportunity. We also estimated that approximately 528 (5.22%) missed opportunities occurred in inpatient settings, 9,001 (88.96%) in outpatient settings, and 589 (5.82%) in ED settings. Table 2 presents the estimated number of missed opportunities that each patient experienced. We estimate that 2,602 (CI: 2,549-2,652) or 77.2% (CI: 75.6-78.7%) of patients experienced at least one missed opportunity prior to diagnosis. Of the patients who experienced at least one missed opportunity, we estimated that, on average, they experienced 3.89 (CI: 3.65-4.14) visits representing missed opportunities, occurring in 3.46 (CI: 3.24-3.69) outpatient visits, 0.20 (CI: 0.19-0.22) inpatient visits, and 0.23 (CI: 0.21-0.24) ED visits.

Table 2 also presents a breakdown of the estimated duration of diagnostic delays among patients who experienced at least one missed opportunity. The mean and median duration of delays were 31.66 (CI: 28.51-35.11) days and 28.00 (CI: 25.00-31.00) days, respectively. On average, patients who experienced at least one missed opportunity had a delay between first SSD and diagnosis of 41.00 days (CI: 37.54-44.77) with 62.1% (CI: 58.4–65.5%) of these delays lasting 30 or more days.

As a sensitivity analysis, we re-estimated the incidence and duration of diagnostic delays using all visits during the diagnostic opportunity window. In this case, the estimated diagnostic-opportunity window began 136 days prior to diagnosis. Across all patients, 3,223 (95.6%) patients had a visit for any reason during this window. There was a total of 44,924 visits that occurred during the diagnostic-opportunity window. We estimated that 14,371 (32.0%) of these visits represented a missed opportunity and 2,976 (CI: 2,923-3,027) patients had at least one missed opportunity. On average, patients experienced 4.83 (CI 4.42-5.34) missed opportunities and had a delay duration of 45.71 days (CI 40.23-52.27).

Table 3 presents results of the logistic regression model estimating the likelihood of experiencing a potential missed opportunity during a visit on a given day. The likelihood of a miss was greater among individuals age ≥ 65 with an odds ratio (OR) of 1.262 (CI: 1.156-1.377). Patients with a history of asthma (OR 1.331 [CI: 1.138-1.557]) or COPD (1.372 [CI: 1.230-1.531]) were more likely to be delayed. Patients who had received chest imaging in the year prior to diagnosis but before the diagnostic-opportunity window were more likely to experience a miss (OR of 1.149 [CI: 1.081-1.296] for chest CT and 1.231 [CI: 1.121-1.353] for chest X-ray). Patients who received a fluoroquinolone in the year prior to diagnosis but before the diagnostic-opportunity window were more likely to experience a miss (OR 1.578 [CI: 1.435-1.734]).

Misses were more likely to occur during weekend visits (1.495 [CI: 1.272-1.758]) and less likely to occur among patients in metropolitan locations (0.874 [CI: 0.771-0.990]). Missed opportunities were more likely to occur in outpatient settings during periods of

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

high influenza activity (1.259 [CI: 1.052-1.507]). Missed opportunities were much less likely to occur in inpatient settings. Compared to outpatient settings alone, misses were less likely to occur on days involving only an inpatient visit (0.123 [CI: 0.106-0.142]), both an inpatient and outpatient visit (0.124 [CI: 0.105-0.145]), both an inpatient and ED visit (0.142 [CI: 0.110-0.184]), or all three setting types (0.128 [CI: 0.089-0.185]). Visits to the ED appeared to increase the odds of a miss. Compared to outpatient settings alone, misses were more likely on days when patients visited ED settings only (2.340 [CI: 1.540-3.555]).

DISCUSSION

Our results show the majority of patients diagnosed with pulmonary tuberculosis have multiple interactions with the US healthcare system prior to receiving a diagnosis consistent with active tuberculosis. Many patients present on multiple occasions, each representing possible missed opportunities to diagnose tuberculosis. Approximately 127 days prior to diagnosis, we observed an increase in visits for either symptoms associated with tuberculosis or diseases that share symptoms with tuberculosis. At least 90% of patients have at least one visit with either a code recording a symptom of tuberculosis or a disease that shares similar symptoms. Common diagnoses included pneumonia, respiratory infections, and other pulmonary conditions. Diagnoses based on symptoms most frequently listed included fever, cough, hemoptysis and weight loss. A considerable proportion of patients experienced multiple visits representing missed opportunities to diagnose tuberculosis: 23.8% of patients had more than 5 possible missed opportunities.

We identified a number of risk factors for diagnostic delays. First, we found that delays are more common for patients who visited the ED, without an inpatient visit on the same day. Diagnostic errors may occur commonly in the ED setting: an estimated 12% of patients who revisit the emergency department do so because of an original misdiagnosis.[38] In the ED, physicians are often treating patients they see for the first time and may be unaware of medical histories. In addition, many patients have vague symptoms, and a range of severity.[39] Also, ED physicians frequently care for multiple different patients concurrently. In one study, ED physicians were caring for a median of 5 patients at one time, and they were interrupted an average of 30.9 times during a 180-minute study period.[40] Finally, when diagnostic errors do occur, ED physicians may not be able to learn from missed diagnostic opportunities because follow-up care occurs in other healthcare settings.

Additional risk factors that we identified included female sex and older age. Other studies have identified females as at higher risk for delays,[8,41,42] and there is a need to investigate the cultural, biological or epidemiological factors responsible for this finding. Also, similar to the findings of others, we found that older adults are at increased risk for diagnostic delays[8,11,41]. Older patients may be at greater risk because of more comorbidities or because they are less likely to exhibit some of the classic signs and symptoms of tuberculosis, perhaps due to the immunosenescence associated with aging. In addition to female sex and older age, several investigations also highlight the

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

risk of fluoroquinolone use for increasing diagnostic delays.[43-45] Because fluoroquinolones have some anti-tuberculosis activity, their inappropriate use prior to the diagnosis of tuberculosis (e.g., to empirically treat a misdiagnosed bacterial pneumonia) may transiently improve symptoms.

In addition to established risk factors, our results highlight two novel risk factors for delay. First, we found that patients with a history of pulmonary diseases, specifically asthma or COPD, were more likely to experience a delayed tuberculosis diagnosis. Other groups have found that other comorbidities, especially pulmonary diseases, were associated with delays;[46] however, we also found that pulmonary imaging (prior to the risk window) was associated with delays. Prior history of pulmonary disorders is a risk factor because it creates a cognitive bias among clinicians. For patients with a history of asthma or COPD presenting with respiratory symptoms, it is less likely that tuberculosis may be considered as part of a differential diagnosis. While patients with a history of pulmonary imaging prior to the diagnostic window, presumably because of some long-standing pulmonary complaint, are more likely to experience a delay, delays are less common if patients received imaging during the diagnostic window because pulmonary imaging would help confirm a tuberculosis diagnosis. Our second novel finding is also related to cognitive bias. Interestingly, we found that if a patient presents during the influenza season, they are more likely to experience a delayed diagnosis for tuberculosis. Delays were also more common during periods of high ILI activity. This finding may reflect the fact that ILI symptoms and tuberculosis symptoms often overlap (e.g., fever, cough), and clinicians may be more likely to suspect influenza during a period of increased activity.

Our study has a number of limitations. First, we use diagnostic codes to identify tuberculosis cases. While such codes have poor sensitivity for identifying active tuberculosis,[30] we used medications to validate our case definition, an approach previously used for identifying tuberculosis.[31] Second, we rely on claims data to determine the reason for visits prior to the tuberculosis diagnosis. Not all symptoms present during a visit are recorded in the insurance claim (e.g., a patient visit for hypertension may also involve an unrecorded symptom of cough.) Indeed, in our sensitivity analysis the number and duration of diagnostic delays increased slightly when including all visits during the diagnostic-opportunity window, regardless of the presence of SSD-related diagnosis codes. In addition, some patients may have experienced diagnostic delays exceeding our detected opportunity window, who were not detected by our change-point algorithm because the volume of such visits is low. Thus, our results may underestimate the true number of visits that represent missed opportunities or the duration of longer individual delays. Third, our data do not contain race or ethnicity. Tuberculosis is much more common among immigrants and family members of immigrants. In other studies of low-incidence countries, delays were more common among non-immigrant populations.[8,12,46,47] Fourth, our dataset is restricted to a privately insured population, with employer-sponsored health insurance and/or supplemental Medicare coverage. Thus, our findings may not be generalizable to an uninsured population or individuals with Medicaid coverage. However, vulnerable populations in inner cities or patients experiencing homelessness may be less likely to

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

experience a delay.[7] Finally, our study excluded extra-pulmonary tuberculosis cases, and future work should focus on such cases given that they are at even greater risk for diagnostic delays.[8,12,46]

Despite our limitations, our results highlight the number of missed opportunities to diagnose tuberculosis. Risk factors for diagnostic delays include older age, female sex, and living in a lower-incidence area. In addition, we identified new risk factors, including existing pulmonary conditions, previous pulmonary imaging, and circulating influenza. These novel risk factors are directly related to cognitive biases that will need to be overcome to improve the timely diagnosis of tuberculosis.

REFERENCES

1. Armstrong LR, Winston CA, Stewart B, Tsang CA, Langer AJ, Navin TR. Changes in tuberculosis epidemiology, United States, 1993-2017. *Int J Tuberc Lung Dis.* 2019;23(7):797-804.
2. Stewart RJ, Tsang CA, Pratt RH, Price SF, Langer AJ. Tuberculosis - United States, 2017. *MMWR Morb Mortal Wkly Rep.* 2018;67(11):317-323.
3. Salinas JL, Mindra G, Haddad MB, Pratt R, Price SF, Langer AJ. Leveling of Tuberculosis Incidence - United States, 2013-2015. *MMWR Morb Mortal Wkly Rep.* 2016;65(11):273-278.
4. Chida N, Brown C, Mathad J, et al. Internal Medicine Residents' Knowledge and Practice of Pulmonary Tuberculosis Diagnosis. *Open Forum Infect Dis.* 2018;5(7):ofy152.
5. Guderian LJ, Miller WC, Seña AC, Stout JE. Increased prevalence of advanced tuberculosis in rural low tuberculosis caseload counties in North Carolina. *Int J Tuberc Lung Dis.* 2011;15(11):1455-1460, i.
6. Miller AC, Polgreen LA, Cavanaugh JE, Hornick DB, Polgreen PM. Missed Opportunities to Diagnose Tuberculosis Are Common Among Hospitalized Patients and Patients Seen in Emergency Departments. *Open Forum Infect Dis.* 2015;2(4):ofv171.
7. Wallace RM, Kammerer JS, Iademarco MF, Althomsons SP, Winston CA, Navin TR. Increasing proportions of advanced pulmonary tuberculosis reported in the United States: are delays in diagnosis on the rise? *Am J Respir Crit Care Med.* 2009;180(10):1016-1022.
8. Loutet MG, Sinclair C, Whitehead N, Cosgrove C, Lalor MK, Thomas HL. Delay from symptom onset to treatment start among tuberculosis patients in England, 2012-2015. *Epidemiol Infect.* 2018;146(12):1511-1518.
9. Mindra G, Wortham JM, Haddad MB, Powell KM. Tuberculosis Outbreaks in the United States, 2009-2015. *Public health reports (Washington, DC : 1974).* 2017;132(2):157-163.
10. Jonsson J, Kan B, Berggren I, Bruchfeld J. Extensive nosocomial transmission of tuberculosis in a low-incidence country. *The Journal of hospital infection.* 2013;83(4):321-326.
11. Paynter S, Hayward A, Wilkinson P, Lozewicz S, Coker R. Patient and health service delays in initiating treatment for patients with pulmonary tuberculosis: retrospective

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

cohort study. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 2004;8(2):180-185.
12. Leutscher P, Madsen G, Erlandsen M, et al. Demographic and clinical characteristics in relation to patient and health system delays in a tuberculosis low-incidence country. *Scandinavian journal of infectious diseases.* 2012;44(1):29-36.
13. Lui G, Wong RY, Li F, et al. High mortality in adults hospitalized for active tuberculosis in a low HIV prevalence setting. *PloS one.* 2014;9(3):e92077.
14. Pablos-Méndez A, Sterling TR, Frieden TR. The relationship between delayed or incomplete treatment and all-cause mortality in patients with tuberculosis. *Jama.* 1996;276(15):1223-1228.
15. Enarson DA, Grzybowski S, Dorken E. Failure of diagnosis as a factor in tuberculosis mortality. *Canadian Medical Association journal.* 1978;118(12):1520-1522.
16. Kelly AM, D'Agostino JF, Andrada LV, Liu J, Larson E. Delayed tuberculosis diagnosis and costs of contact investigations for hospital exposure: New York City, 2010-2014. *American journal of infection control.* 2017;45(5):483-486.
17. Golub JE, Bur S, Cronin WA, et al. Delayed tuberculosis diagnosis and tuberculosis transmission. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 2006;10(1):24-30.
18. Sreeramareddy CT, Panduru KV, Menten J, Van den Ende J. Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature. *BMC infectious diseases.* 2009;9:91.
19. Castells Carrillo C, San José Rodríguez S, López Aranaga I, et al. Diagnostic delay as main contributing factor to a large outbreak of tuberculosis in a university. *Enfermedades infecciosas y microbiologia clinica.* 2019;37(8):496-501.
20. Raffalli J, Sepkowitz KA, Armstrong D. Community-based outbreaks of tuberculosis. *Archives of internal medicine.* 1996;156(10):1053-1060.
21. MacIntyre CR, Plant AJ, Hulls J, Streeton JA, Graham NM, Rouch GJ. High rate of transmission of tuberculosis in an office: impact of delayed diagnosis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 1995;21(5):1170-1174.
22. Calder L, Rivers J, Hayhurst M, et al. A school and community outbreak of tuberculosis in Palmerston North, New Zealand. *The New Zealand medical journal.* 2008;121(1278):50-61.
23. Rao VK, Iademarco EP, Fraser VJ, Kollef MH. Delays in the suspicion and treatment of tuberculosis among hospitalized patients. *Annals of internal medicine.* 1999;130(5):404-411.
24. Yilmaz A, Boğa S, Sulu E, et al. Delays in the diagnosis and treatment of hospitalized patients with smear-positive pulmonary tuberculosis. *Respiratory medicine.* 2001;95(10):802-805.
25. Harris TG, Sullivan Meissner J, Proops D. Delay in diagnosis leading to nosocomial transmission of tuberculosis at a New York City health care facility. *American journal of infection control.* 2013;41(2):155-160.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

26. Evenden P, Roche A, Karo B, Balasegaram S, Anderson CS. Presentation and healthcare delays among people with tuberculosis in London, and the impact on treatment outcome. *BMJ open respiratory research.* 2019;6(1):e000468.
27. Deponti GN, Silva DR, Coelho AC, Muller AM, Dalcin Pde T. Delayed diagnosis and associated factors among new pulmonary tuberculosis patients diagnosed at the emergency department of a tertiary care hospital in Porto Alegre, South Brazil: a prospective patient recruitment study. *BMC infectious diseases.* 2013;13:538.
28. Moran GJ, McCabe F, Morgan MT, Talan DA. Delayed recognition and infection control for tuberculosis patients in the emergency department. *Ann Emerg Med.* 1995;26(3):290-295.
29. Yen YL, Chen IC, Wu CH, Li WC, Wang CH, Tsai TC. Factors associated with delayed recognition of pulmonary tuberculosis in emergency departments in Taiwan. *Heart & lung : the journal of critical care.* 2015;44(4):353-359.
30. Ronald LA, Ling DI, FitzGerald JM, et al. Validated methods for identifying tuberculosis patients in health administrative databases: systematic review. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 2017;21(5):517-522.
31. Winthrop KL, Baxter R, Liu L, et al. The reliability of diagnostic coding and laboratory data to identify tuberculosis and nontuberculous mycobacterial disease among rheumatoid arthritis patients using anti-tumor necrosis factor therapy. *Pharmacoepidemiology and drug safety.* 2011;20(3):229-235.
32. Liberman AL, Newman-Toker DE. Symptom-Disease Pair Analysis of Diagnostic Error (SPADE): a conceptual framework and methodological approach for unearthing misdiagnosis-related harms using big data. *BMJ Qual Saf.* 2018.
33. Waxman DA, Kanzaria HK, Schriger DL. Unrecognized Cardiovascular Emergencies Among Medicare Patients. *JAMA Intern Med.* 2018.
34. Moy E, Barrett M, Coffey R, Hines AL, Newman-Toker DE. Missed diagnoses of acute myocardial infarction in the emergency department: variation by patient and facility characteristics. *Diagnosis.* 2015;2(1):29-40.
35. Newman-Toker DE, Moy E, Valente E, Coffey R, Hines AL. Missed diagnosis of stroke in the emergency department: a cross-sectional analysis of a large population-based sample. *Diagnosis (Berl).* 2014;1(2):155-166.
36. Schull MJ, Vermeulen MJ, Stukel TA. The risk of missed diagnosis of acute myocardial infarction associated with emergency department volume. *Ann Emerg Med.* 2006;48(6):647-655.
37. Centers for Disease Control and Prevention. CDC FluView Interactive: ILI Surviellence. https://www.cdc.gov/flu/weekly/fluviewinteractive.htm. Published 2020. Accessed May 14, 2020, 2020.
38. Verelst S, Pierloot S, Desruelles D, Gillet JB, Bergs J. Short-term unscheduled return visits of adult patients to the emergency department. *J Emerg Med.* 2014;47(2):131-139.
39. Medford-Davis LN, Singh H, Mahajan P. Diagnostic Decision-Making in the Emergency Department. *Pediatr Clin North Am.* 2018;65(6):1097-1105.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

40. Chisholm CD, Collison EK, Nelson DR, Cordell WH. Emergency department workplace interruptions: are emergency physicians "interrupt-driven" and "multitasking"? *Acad Emerg Med.* 2000;7(11):1239-1243.
41. Saldana L, Abid M, McCarthy N, Hunter N, Inglis R, Anders K. Factors affecting delay in initiation of treatment of tuberculosis in the Thames Valley, UK. *Public Health.* 2013;127(2):171-177.
42. Long NH, Johansson E, Lönnroth K, Eriksson B, Winkvist A, Diwan VK. Longer delays in tuberculosis diagnosis among women in Vietnam. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease.* 1999;3(5):388-393.
43. Chen TC, Lu PL, Lin CY, Lin WR, Chen YH. Fluoroquinolones are associated with delayed treatment and resistance in tuberculosis: a systematic review and meta-analysis. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases.* 2011;15(3):e211-216.
44. Hogan CA, Puri L, Gore G, Pai M. Impact of fluoroquinolone treatment on delay of tuberculosis diagnosis: A systematic review and meta-analysis. *Journal of clinical tuberculosis and other mycobacterial diseases.* 2017;6:1-7.
45. Rush B, Wormsbecker A, Stenstrom R, Kassen B. Moxifloxacin Use and Its Association on the Diagnosis of Pulmonary Tuberculosis in An Inner City Emergency Department. *The Journal of emergency medicine.* 2016;50(3):371-375.
46. Zão I, Ribeiro AI, Apolinário D, Duarte R. Why does it take so long? The reasons behind tuberculosis treatment delay in Portugal. *Pulmonology.* 2019;25(4):215-222.
47. Farah MG, Rygh JH, Steen TW, Selmer R, Heldal E, Bjune G. Patient and health care system delays in the start of tuberculosis treatment in Norway. *BMC infectious diseases.* 2006;6(1):33.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Author Contributions

Aaron C. Miller – designed the study, developed the methodological approach, drafted and revised the final manuscript, and helped to obtain funding for the research.

Alan Arakkal – helped to conduct statistical analysis, helped draft the methods and results section, reviewed and revised the final manuscript.

Scott Koeneman – helped to conduct statistical analysis, helped draft the methods and results section, reviewed and revised the final manuscript.

Joseph E. Cavanaugh – helped in developing the methodological approach, provided guidance on the statistical analysis, reviewed and revised the final manuscript.

Alicia K. Gerke – provided clinical expertise and feedback, helped review and revise the final manuscript, and helped to obtain funding for the research.

Douglas B. Hornick – provided clinical expertise and feedback, helped review and revise the final manuscript.

Philip M Polgreen – helped conceive the study objective, provided clinical guidance, drafted and revised the final manuscript, and helped to obtain funding for the research

Data Sharing

The IBM Marketscan Research Databases can be obtained from *IBM Watson Health*. The code used for the simulation and statistical analysis is available on GitHub at https://github.com/aarmiller/tb_delay_2020.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Figure Headings

Figure 1 – Diagram for Identifying SSD Visits – SSD visits include symptoms, symptomatically-similar diagnoses and testing or exam-based diagnoses that suggest an active tuberculosis infection may be present in the patient. Potential diagnostic opportunities are defined as SSD-related visits that occur during the diagnostic opportunity window (i.e., the window prior to index diagnosis where delays are biologically plausible).

Figure 2 – Trend in SSD-related Healthcare Visits Prior to Index Tuberculosis Diagnosis. Figure (A), left, depicts the number of SSD-related visits each day prior to the index tuberculosis diagnosis summed across all patients and healthcare settings. Before the index tuberculosis diagnosis there is a large spike in SSD related healthcare visits. Supplementary Figures 1 and 2 provide similar counts of visits prior to the index diagnosis broken down by healthcare setting and type of SSD, respectively. Similar results are obtained for each healthcare setting and type of SSD. Figure (B), right, depicts the same counts but adds trend lines for *observed* and *expected* visits. The red line depicts the trend in expected SSD-related visits, which was estimated using data prior to the change-point. The blue line depicts the trend in the observed number of visits after the change-point. The area between the blue and red lines depicts the number of SSD-related visits that represent likely diagnostic opportunities.

Figure 3 – Flow diagram of patient inclusion and exclusion criteria – Counts of patients excluded and reasons for exclusion used to identify the final 3,371 index cases of tuberculosis.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Tables

Table 1 Baseline Study Population Characteristics

	Total Patients (% of patients)
Age at diagnosis	
<18	95 (2.8%)
18-35	436 (12.9%)
36-45	437 (13.0%)
46-55	600 (17.8%)
56-65	800 (23.7%)
>65	1003 (29.8%)
Sex	
Male	1613 (47.8%)
Female	1758 (52.2%)
Enrollment time prior to index diagnosis (years)	
Mean	4.1
Median	3.1
Range	1.0 - 16.5
Count ≥ 1.5 years	2846 (84.4%)
Count ≥ 2 years	2394 (71.0%)
Count ≥ 3 years	1744 (51.7%)
Region	
Rural	355 (10.5%)
Urban	2998 (88.9%)
Missing	18 (0.5%)

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 2 – Estimated Number of Missed Opportunities and Duration of Diagnostic Delay Based on Simulation Model

Metric / Category	Count (Percentage of all patients) / Mean	95% CI (from bootstrapping)
Number of Missed opportunities		
0 Days	769 (22.8%)	719 - 822 (21.3 - 24.4%)
>= 1 Day	2602 (77.2%)	2549 - 2652 (75.6 - 78.7%)
>= 2 Days	2065 (61.2%)	1981 - 2148 (58.8 - 63.7%)
>= 3 Days	1563 (46.4%)	1457 - 1667 (43.2 - 49.5%)
>= 4 Days	1137 (33.7%)	1028 - 1248 (30.5 - 37.0%)
>= 5 Days	803 (23.8%)	704 - 908 (20.9 - 26.9%)
Mean - Overall	3.89	3.65 - 4.14
Mean - Outpatient	3.46	3.24 - 3.69
Mean - Inpatient	0.20	0.19 - 0.22
Mean - ED	0.23	0.21 - 0.24
Duration of Delays		
>= 0 Days	2602 (100.0%)	2549 - 2652 (NA)
>= 10 Days	2354 (90.4%)	2284 - 2420 (89.3 - 91.5%)
>= 20 Days	1990 (76.5%)	1895 - 2080 (74.1 - 78.7%)
>= 30 Days	1615 (62.1%)	1495 - 1731 (58.4 - 65.5%)
>= 40 Days	1260 (48.4%)	1114 - 1401 (43.5 - 53.0%)
>= 50 Days	928 (35.6%)	769 - 1087 (30.0 - 41.1%)
>= 60 Days	635 (24.4%)	478 - 801 (18.7 - 30.3%)
>= 70 Days	388 (14.9%)	253 - 540 (9.9 - 20.4%)
>= 80 Days	204 (7.8%)	105 - 327 (4.1 - 12.4%)
>= 90 Days	86 (3.3%)	30 - 170 (1.2 - 6.4%)
>= 100 Days	25 (1.0%)	3 - 70 (0.1 - 2.7%)
>= 110 Days	4 (0.1%)	0 - 19 (0.0 - 0.7%)
Mean Among Delayed	41.00	37.54 - 44.77
Mean Everyone Included	31.66	28.51 - 35.11

For peer review only

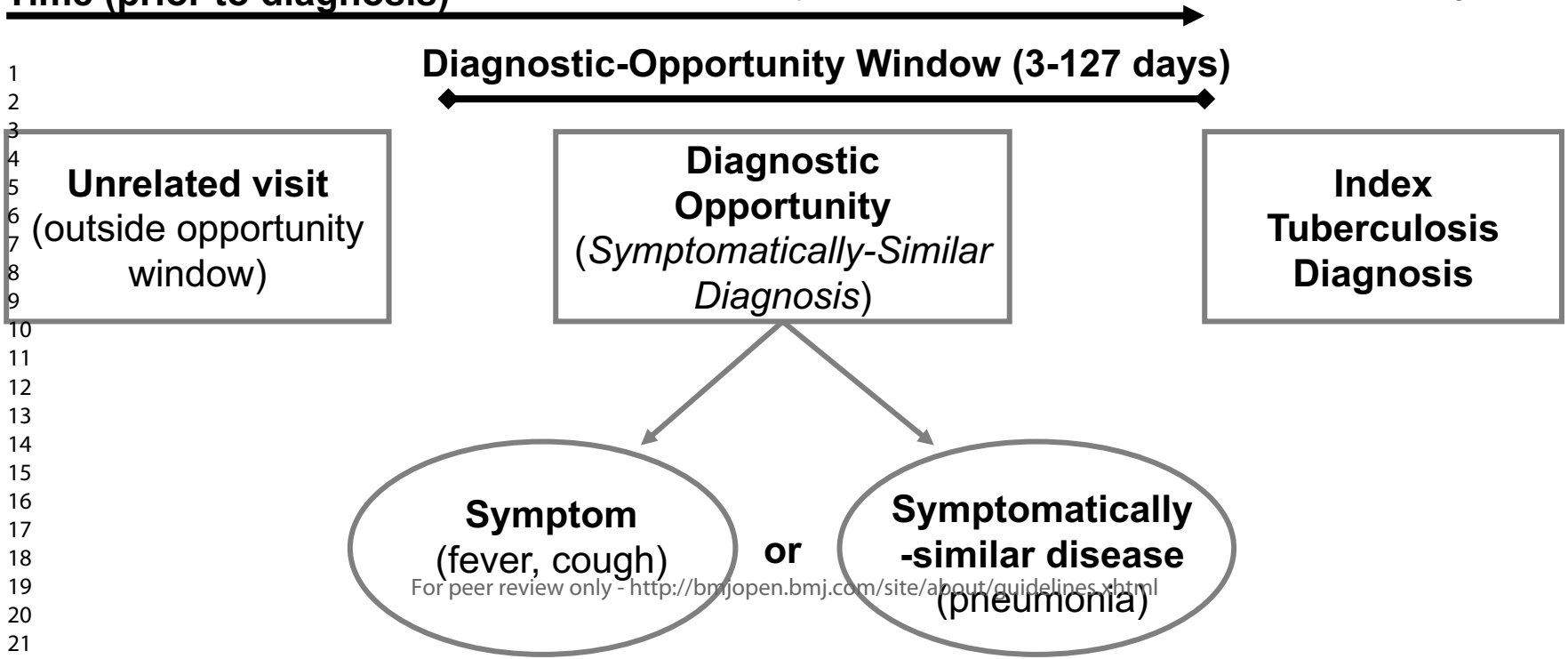
For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 3 – Regression Results for Likelihood of Experiencing a Missed Opportunity

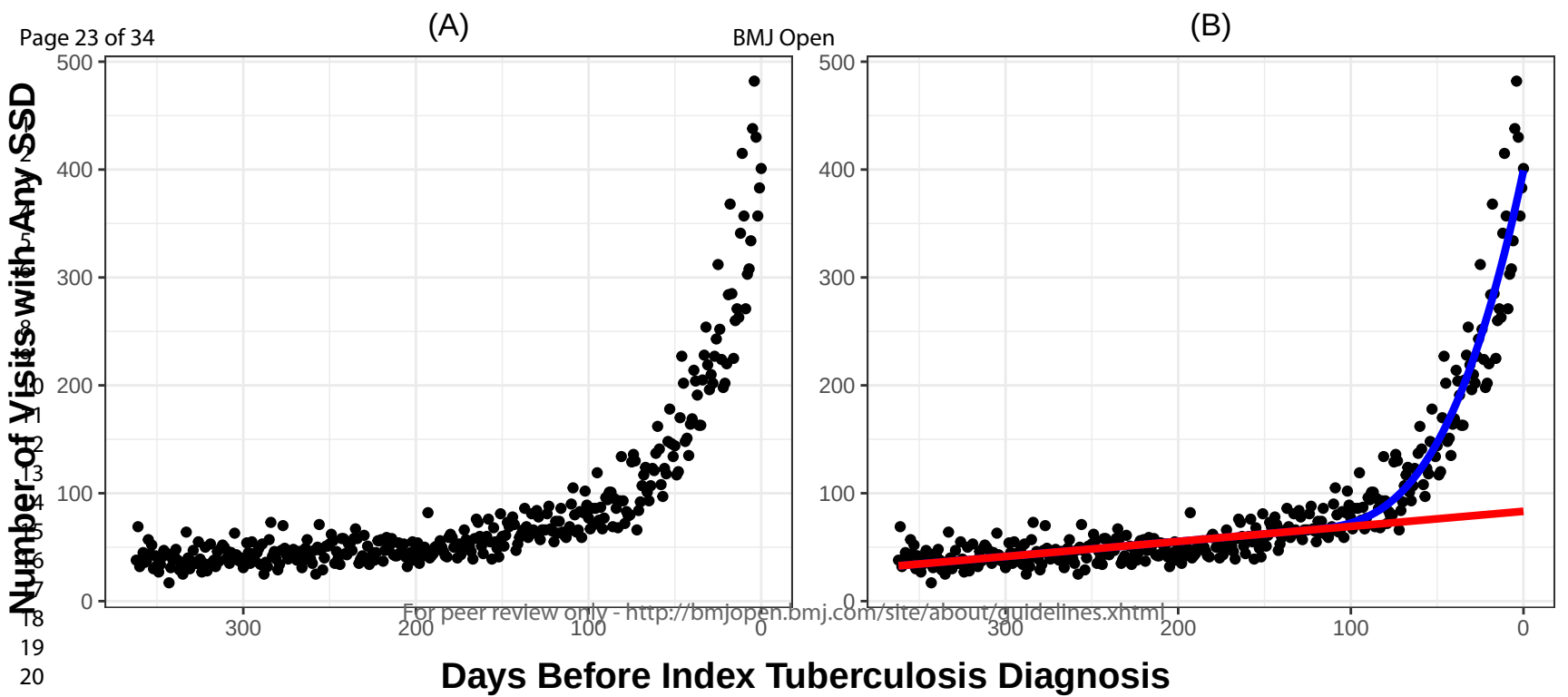
Coefficient	Effect Estimate	95% CI	P-value
Weekend (visits that occurred on a Saturday or Sunday)	1.495	1.272, 1.758	<0.001
Age > 65	1.262	1.156, 1.377	<0.001
Settings visited			
Outpatient only	Ref	Ref	Ref
All three (inpatient, outpatient, and ED)	0.128	0.089, 0.185	<0.001
ED only	2.340	1.540, 3.555	<0.001
Inpatient only	0.123	0.106, 0.142	<0.001
Inpatient and ED	0.142	0.110, 0.184	<0.001
Inpatient and outpatient	0.124	0.105, 0.145	<0.001
Outpatient and ED	1.324	0.968, 1.811	0.079
Urban vs. not urban	0.874	0.771, 0.990	0.034
ILI (≥ 3.8 %) * outpatient interaction	1.259	1.052, 1.507	0.012
Asthma prior to change point	1.331	1.138, 1.557	<0.001
COPD prior to change point	1.372	1.230, 1.531	<0.001
Chest CT prior to change point	1.149	1.018, 1.296	0.025
Chest X-Ray prior to change point	1.231	1.121, 1.353	<0.001
Fluoroquinolones between change point and 3 days prior to index	1.578	1.435, 1.734	<0.001

For peer review only

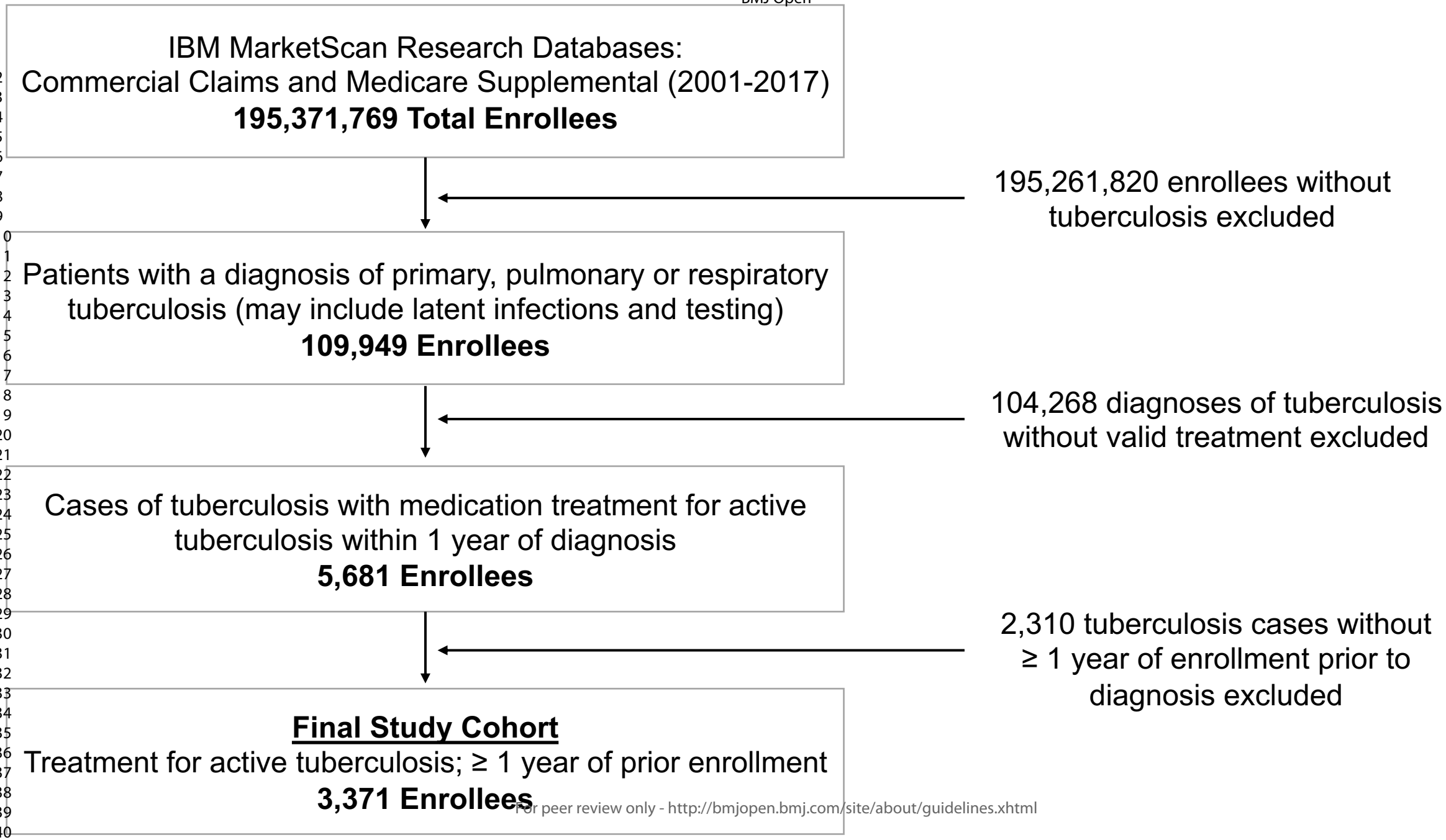
For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml


IBM MarketScan Research Databases:
Commercial Claims and Medicare Supplemental (2001-2017)
195,371,769 Total Enrollees
195,261,820 enrollees without tuberculosis excluded
Patients with a diagnosis of primary, pulmonary or respiratory tuberculosis (may include latent infections and testing)
109,949 Enrollees
104,268 diagnoses of tuberculosis without valid treatment excluded
Cases of tuberculosis with medication treatment for active tuberculosis within 1 year of diagnosis
5,681 Enrollees
2,310 tuberculosis cases without ≥ 1 year of enrollment prior to diagnosis excluded
Final Study Cohort
Treatment for active tuberculosis; ≥ 1 year of prior enrollment
3,371 Enrollees

Supplementary Material

Supplementary Table 1 - List of SSDs used to identify potential diagnostic opportunities.

Category	Symptomatically Similar Diagnosis (SSD)	ICD-9-CM	ICD-10-CM
Alternative-Cardio-Sino-Pulmonary-Based Diagnoses	Tonsillitis	474.12, 474.2, 474.8, 474.9	J35.1
	Respiratory Failure	517.3, 518.81, 518.82, 518.83, 518.84, 799.1	J80, J96.00, J96.01, J96.02, J96.10, J96.12, J96.20, J96.21, J96.22, J96.90, J96.92
	Respiratory Cancer	163.0, 163.1, 163.8, 163.9, 165.0, 165.8, 165.9, 231.1, 231.8, 231.9	C33, C38.4, C39.0, C39.9, C45.0, D02.1, D02.3, D02.4
	Pleurisy Pneumothorax	511.0, 511.1, 511.8, 511.89, 512.0, 512.8, 512.81, 512.82, 512.83, 512.84, 512.89, 518.1, 518.2	J86.9, J92.0, J92.9, J93.0, J93.12, J93.81, J93.82, J93.83, J94.0, J94.1, J94.2, J94.8, J94.9, J98.19, J98.2, J98.3, R09.1
	Other Upper Respiratory Disease	472.0, 477.0, 477.2, 477.8, 477.9, 478.1, 478.19, 478.20, 478.29, 478.30, 478.31, 478.32, 478.33, 478.34, 478.4, 478.5, 478.70, 478.74, 478.75, 478.79, 478.8, 519.1, 519.11, 519.19, 519.3, 784.40, 784.49, 784.7, 784.8	J30.0, J30.1, J30.2, J30.81, J30.89, J30.9, J31.0, J34.2, J34.89, J37.0, J37.1, J38.00, J38.01, J38.02, J38.1, J38.2, J38.3, J38.4, J38.5, J38.6, J38.7, J39.2, J39.3, J39.8, J39.9, J98.01, J98.51, R04.0, R04.1, R09.81
	Other Lower Respiratory Disease	514, 515, 516.0, 516.1, 516.2, 516.3, 516.30, 516.31, 516.32, 516.33, 516.34,	J81.0, J81.1, J82, J84.01, J84.02, J84.03, J84.09, J84.10, J84.111, J84.112, J84.113, J84.114, J84.115, J84.116,

For peer review only

		516.35, 516.36, 516.37, 516.4, 516.5, 516.8, 516.9, 517.2, 517.8, 518.3, 518.4, 518.89, 519.4, 519.8, 519.9, 786.00, 786.09, 786.39, 786.9, 793.11, 794.2	J84.117, J84.17, J84.2, J84.81, J84.82, J84.89, J84.9, J98.4, J98.6, J98.8, J99, R04.89, R04.9
	Lung Disease Due to External Agents	495.0, 495.1, 495.2, 495.3, 495.4, 495.5, 495.6, 495.7, 495.8, 495.9, 500, 501, 502, 503, 504, 505, 506.0, 506.1, 506.2, 506.3, 506.4, 506.9, 507.1, 507.8, 508.0, 508.1, 508.2, 508.8, 508.9	J60, J61, J62.0, J62.8, J63.0, J63.1, J63.2, J63.3, J63.4, J63.5, J63.6, J64, J66.0, J66.1, J66.2, J66.8, J67.0, J67.1, J67.2, J67.3, J67.4, J67.5, J67.6, J67.7, J67.8, J67.9, J68.0, J68.1, J68.2, J68.3, J68.4, J68.8, J68.9, J69.1, J69.8, J70.0, J70.1, J70.2, J70.3, J70.4, J70.5, J70.8, J70.9
	Lung Cancer	162.2, 162.3, 162.4, 162.5, 162.8, 162.9, 209.21	C34.00, C34.01, C34.02, C34.10, C34.11, C34.12, C34.2, C34.30, C34.31, C34.32, C34.80, C34.81, C34.82, C34.90, C34.91, C34.92, C7A.090, D02.20, D02.21, D02.22
	Hemoptysis	786.39	R04.8, R04.89, R04.9
	COPD	491.2, 491.20, 491.21, 491.22, 492.0, 492.8, 494, 494.0, 494.1, 496	J41.8, J43.0, J43.1, J43.2, J43.8, J43.9, J44.1, J44.9, J47.1, J47.9
	Asthma	493.00, 493.01, 493.02, 493.10, 493.11, 493.12, 493.20, 493.21, 493.22, 493.81, 493.82, 493.90, 493.92	J45.20, J45.21, J45.22, J45.30, J45.31, J45.32, J45.40, J45.41, J45.42, J45.50, J45.51, J45.52, J45.901, J45.902, J45.909, J45.990, J45.991, J45.998
	Aspiration Pneumonitis	507.0	J69.0

For peer review only

	Additional Codes	135, 197.0, 212.3, 235.7, 239.1, 289.1, 416.8, 423.9, 428.0, 446.4	I50.9, J85.0, N39.0
Alternative-Infectious-Disease-Based Diagnoses	Tonsillitis	463, 474.0, 474.00, 474.01, 474.02, 474.10, 474.11, 475	J03.80, J03.81, J03.90, J03.91, J35.01, J35.02, J35.03, J35.2, J35.3, J35.8, J35.9, J36
	Pneumonia	112.4, 114.0, 114.4, 115.05, 115.15, 115.95, 130.4, 136.3, 480.0, 480.1, 480.2, 480.8, 480.9, 481, 482.0, 482.1, 482.2, 482.3, 482.30, 482.31, 482.32, 482.39, 482.4, 482.40, 482.41, 482.42, 482.49, 482.8, 482.81, 482.83, 482.84, 482.89, 482.9, 483, 483.0, 483.1, 483.8, 484.1, 484.3, 484.6, 484.7, 484.8, 485, 486, 513.0	A31.0, A37.01, A37.11, A43.0, A48.1, B25.0, B37.1, B38.0, B38.1, B38.2, B39.0, B39.1, B39.2, B58.3, B59, B77.81, J12.0, J12.1, J12.2, J12.3, J12.89, J12.9, J13, J14, J15.0, J15.1, J15.20, J15.211, J15.212, J15.29, J15.3, J15.4, J15.5, J15.6, J15.7, J15.8, J15.9, J16.0, J16.8, J17, J18.0, J18.1, J18.8, J18.9, J85.1
	Pleurisy Pneumothorax	510.0	J86.0, J93.11
	Other Upper-Respiratory Infection	460, 461.0, 461.1, 461.2, 461.3, 461.9, 462, 464.0, 464.00, 464.01, 464.11, 464.20, 464.21, 464.30, 464.31, 464.4, 464.50, 464.51, 465.0, 465.8, 465.9, 473.0, 473.1, 473.2, 473.3, 473.8, 473.9	J00, J01.00, J01.01, J01.10, J01.11, J01.20, J01.21, J01.30, J01.31, J01.40, J01.41, J01.80, J01.81, J01.90, J01.91, J02.0, J02.8, J02.9, J03.00, J03.01, J04.0, J04.10, J04.11, J04.2, J04.30, J04.31, J05.0, J05.10, J05.11, J06.0, J06.9, J32.0, J32.1, J32.2, J32.3, J32.4, J32.8, J32.9, R09.82

For peer review only

	Other Upper Respiratory Disease	472.2, 476.0, 476.1, 478.21, 478.22, 478.24, 478.71, 478.9, 519.2	J31.1, J31.2, J39.0, J39.1, J98.09, J98.5, J98.59, R07.0
	Other Lower Respiratory Disease	513.1	J18.2, J22, J85.2, J85.3, J98.9, R06.6, R06.82
	Influenza	487.0, 487.1, 487.8, 488, 488.1, 488.11, 488.12, 488.19, 488.81, 488.82, 488.89	J09.X1, J09.X2, J09.X3, J09.X9, J10.00, J10.01, J10.08, J10.1, J10.2, J10.89, J11.00, J11.08, J11.1, J11.2, J11.81, J11.82, J11.83, J11.89
	COPD	490, 491.0, 491.1, 491.8, 491.9	J40, J41.0, J41.1, J42, J44.0, J47.0
	Bronchitis	466.0, 466.1, 466.11, 466.19	J20.0, J20.1, J20.2, J20.3, J20.4, J20.5, J20.6, J20.7, J20.8, J20.9, J21.0, J21.1, J21.8, J21.9
	Additional Codes	038.9, 079.99, 310, 340, 391, 599.0, 830, 995.91	A41.9, B34.9, D14.30, D38.1, D49.1, D86.0, D86.9
Symptom-Based Diagnoses	Other Upper Respiratory Disease	784.1, 784.41, 784.42, 784.9, 784.99	R49.0, R49.8, R49.9
	Other Lower Respiratory Disease	786.02, 786.05, 786.07, 786.2, 786.3, 786.30, 786.4, 786.52	R04.2, R05, R06.00, R06.01, R06.02, R06.03, R06.09, R06.2, R06.89, R06.9, R07.1, R07.81, R09.3
	Hemoptysis	786.3, 786.30	R04.2
	Fever	780.6, 780.60, 780.61	R50, R50.81, R50.9
	Cough	786.2	R05
	Additional Codes	780.79, 780.8, 783.21, 786.50, 786.51, 786.59	R07.2, R07.82, R07.89, R07.9, R53.1, R53.81, R53.83, R61, R63.4
Testing-Imaging-or Physical-Exam-Based Diagnoses	Pleurisy Pneumothorax	511.9, 518.0	J90, J91.8, J93.9, J98.11
	Other Lower Respiratory Disease	786.6, 786.7, 793.1, 793.19	R09.02, R91.1, R91.8
	Additional Codes	263.9, 276.1, 285.29, 285.9, 288.60, 289.3,	D64.9, D72.829, E871, I51.7, R00.0, R22.0,

For peer review only

		429.3, 782.2, 784.2, 785.0, 785.6, 799.02, 799.4	R22.1, R22.2, R59, R59.0, R59.1, R59.9
Procedure Codes		**CPT Code**	
CT – Chest		71260, 71250, 71270	
X-ray - Chest		71010, 71015, 71020, 71021, 71022, 71023, 71030, 71034, 71035, 71101, 71111, 71120,71045, 71046, 71047, 71048	

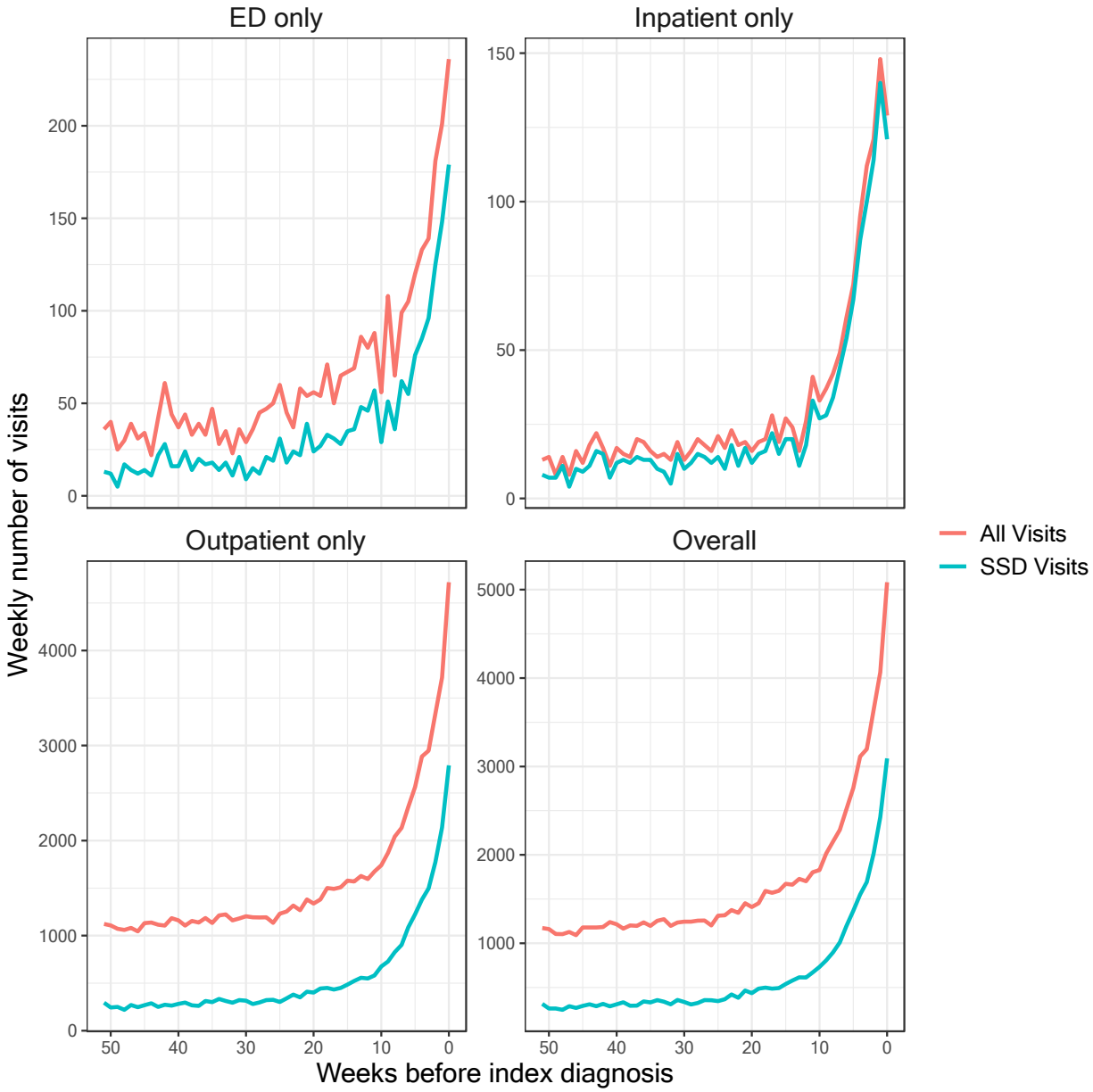
For peer review only

Supplementary Table 2 – ILI Indicators for the optimal ILI cutoff. We used AIC to select the optimal cutoff for defining peak ILI activity; this was determined to be an ILI level >3.8%.

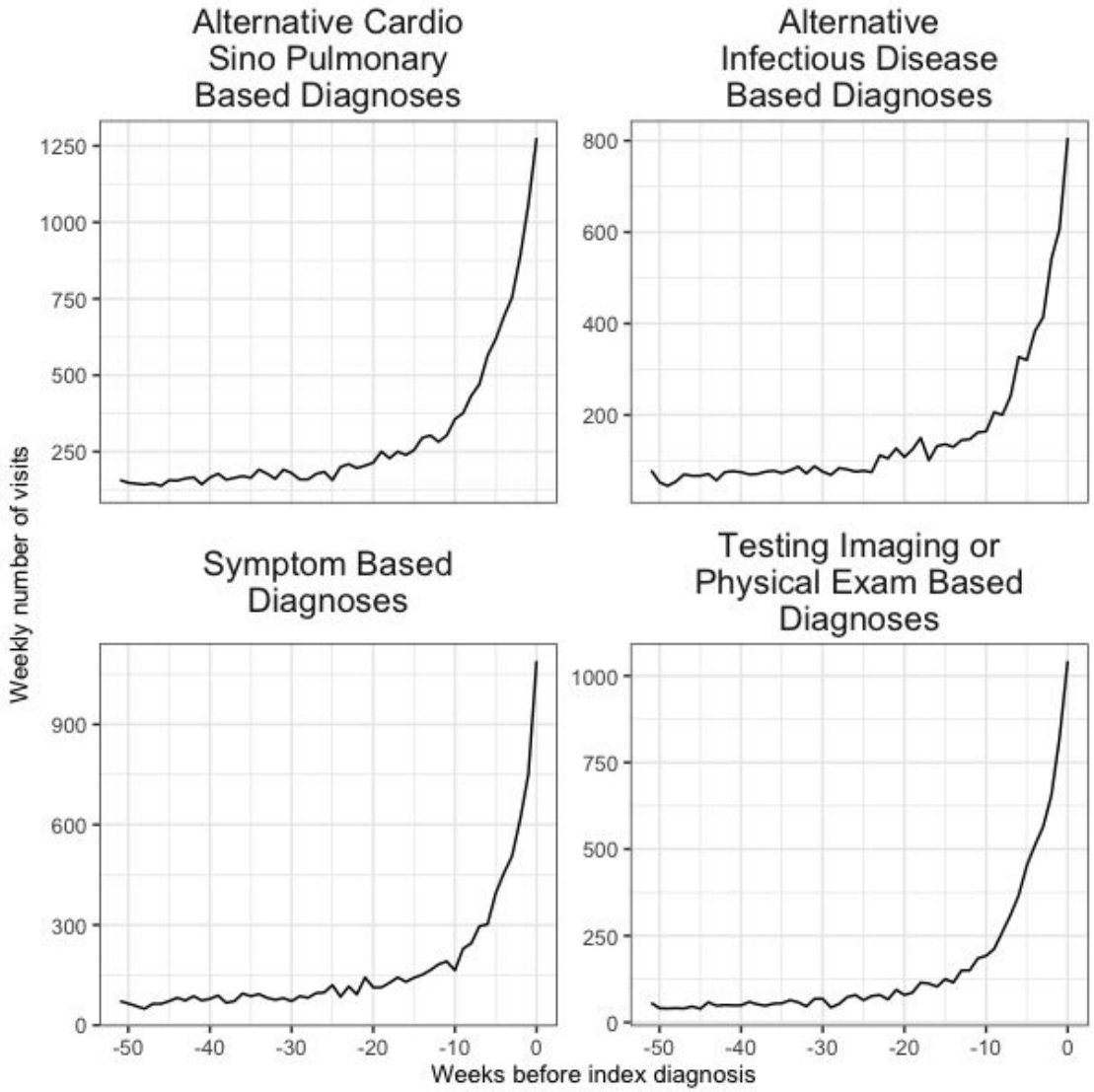
Period	ILI >= 3.8%
2001/01/01 - 2001/01/14	0
2001/01/15 - 2001/02/04	1
2001/02/05 - 2003/11/23	0
2003/11/24 - 2003/12/28	1
2003/12/29 - 2005/01/23	0
2005/01/24 - 2005/02/20	1
2005/02/21 - 2008/01/20	0
2008/01/21 - 2008/03/09	1
2008/03/10 - 2009/08/30	0
2009/08/31 - 2009/11/15	1
2009/11/16 - 2011/01/16	0
2011/01/17 - 2011/02/20	1
2011/02/21 - 2012/12/09	0
2012/12/10 - 2013/01/27	1
2013/01/28 - 2013/12/22	0
2013/12/23 - 2014/01/05	1
2014/01/06 - 2014/12/14	0
2014/12/15 - 2015/01/25	1
2015/01/26 - 2017/01/15	0
2017/01/16 - 2017/02/19	1
2017/02/20 - 2017/12/10	0
2017/12/11 - 2017/12/24	1

For peer review only

Supplementary Figure 1 – Trend in SSD visits and all visits prior to the index tuberculosis diagnoses broken down by type of healthcare setting. The red lines depict all visits, and the blue line depicts visits with SSD-related conditions.



Supplementary Figure 2 – Trend in SSD visits prior to diagnosis for the four SSD categories.



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

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	4-5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	5
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
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	*4*
Bias	9	Describe any efforts to address potential sources of bias	6-7
Study size	10	Explain how the study size was arrived at	4-5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6-7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5-7
		(*b*) Describe any methods used to examine subgroups and interactions	5-7
		(*c*) Explain how missing data were addressed	5
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	6-7

Continued on next page

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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	8
		(b) Give reasons for non-participation at each stage	8
		(c) Consider use of a flow diagram	See Supplement
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8 (Table 1)
		(b) Indicate number of participants with missing data for each variable of interest	N/A (all enrollees continuously enrolled)
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	N/A
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	*8-9*
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
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	8-9
		(*b*) Report category boundaries when continuous variables were categorized	8-9
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	9
Discussion			
Key results	18	Summarise key results with reference to study objectives	10-11
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	11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11
Generalisability	21	Discuss the generalisability (external validity) of the study results	11
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	1

*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.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml